Lefteris Tsoulfidis

Competing Schools of Economic Thought

Springer

Competing Schools of Economic Thought

Lefteris Tsoulfidis

Competing Schools of Economic Thought

Springer

Professor Lefteris Tsoulfidis
University of Macedonia
Department of Economics
156 Egnatia Street
540 06 Thessaloniki, Greece
Lnt@uom.gr

ISBN 978-3-540-92692-4 e-ISBN 978-3-540-92693-1
DOI 10.1007/978-3-540-92693-1
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010923615

Cover design: WMXDesign GmbH, Heidelberg, Germany

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Introduction

1.1 Introduction

This book was born out of our reaction to the way in which the usual texts cover the subject of the history of economic thought. In most of these texts, there is a tendency to emphasize the similarities and differences between all the important economists and form a repository of encyclopedic knowledge where one can study the seemingly important economic ideas. In this book, we argue that it is much more fruitful to focus on the essential ideas of each and every school of economic thought and relate them to present-day problems, than to engage into a sterile discussion of the ideas and the lives of the great economists of the past.

Thus, although this book deals with the history of economic thought, it does not necessarily follow a historic (in the sense of the order of presentation) approach, but rather a logical one, that is to say it deals with the social conditions associated with the emergence of a school of economic thought, its evolution, and its contemporary influence. One cannot write a book on the history of economic thought without writing separate chapters on the major economists of the past, that is, Adam Smith, David Ricardo, Karl Marx, and J.M. Keynes. Of course these economists formed schools of economic thought, that is, the classical and the Keynesian. As for the neoclassical school of economic thought, the ideas of its founders, that is, Stanley Jevons, Karl Menger, Léon Walras, and Alfred Marshal are put together into a single chapter and school of economic thought. The book also studies the evolution of current mainly macroeconomic approaches, that is, monetarism, new classical economics, real business cycles, and new Keynesian economics. Furthermore, we include separate chapters such as the microeconomic revolution of the 1930s, the upshot of Sraffa's-based critique of the neoclassical theory of the firm, the neoclassical synthesis, capital theory, and a final chapter that summarizes and critically evaluates the major schools of economic thought.

The main goal of the book is to present those theories that survived over time and that can inform us about current developments and economic policies. Thus, although the book includes the major economists, such as Adam Smith and John M. Keynes,

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_1, © Springer-Verlag Berlin Heidelberg 2010

the objective is not an encyclopedic narrative of their lives and works but rather to use their theories to the understanding of the underlying mechanisms that govern the current economic system and address contemporary problems and issues. In this sense, although we recognize that Malthus is one of the major economists of the past, nevertheless we do not examine him in a separate chapter, as is usual in the history of economic thought texts. In fact, we do justice to his views by referring to the mechanics of his population law in the chapters of Smith and Ricardo in connection to their theories of economic growth as well as to policy issues such as taxation. Furthermore, Malthus's underconsumption theory is discussed in connection to the classical conception of effective demand and the possibility of generalized gluts and Keynes's theory of effective demand. In this sense, we discuss Marx's theory of money, not in order to present just another theory among many, but rather because this specific theory may be used to explain current phenomena of inflation and exchange rate fluctuations. The same is true with the labor theory of value, which can shed new light on the variations in actual prices of contemporary economies. Similarly, Marx's theory of the tendential fall in the rate of profit is discussed in order to explain the actual state of the economy (growing or stagnating).

Another salient feature of our approach is the detailed study of the conditions within which modern schools of economic thought have developed and unlike other texts does not leave such a discussion to advanced macroeconomic courses. In these chapters we argue that only in microeconomic theory there is a consensus among, at least, neoclassical economists. In fact, we have known that microeconomic texts are used for many years and when they change, the change is about the use of new techniques and not about the development of new theories. In microeconomics, we do not really have new theoretical approaches as opposed to macroeconomics, which since the late 1960s is a deeply divided into rival approaches and so macroeconomics has been in a stage of flux making the distinction of each of its strands an increasingly more difficult task. If there is a consensus in macroeconomics this is the need for the provision of solid theoretical microeconomic foundations, that is, the need to assign optimizing behavior to all economic agents. The provision of microfoundation is an expression that in microeconomics there is agreement and the disagreement is in the macroeconomic level. In our opinion, this is also true in the recently celebrated New Consensus Macroeconomics where the word consensus is used rather as a euphemism for the current state of macroeconomics and its inability to address the problems of the severe recession that started for most economists in the end of year 2007 and continues up until of this writing.

As the efforts for achieving consensus among macroeconomists intensify in the years ahead the deeply recessionary years of 2008 and 2009 may contribute to the emergence of new and even radical ideas. The usual approach in macroeconomics treats the various schools of macroeconomic thought as rather random events, where graph follows graph and equation follows equation; as a consequence, the knowledge that one obtains cannot be placed in the appropriate theoretical context and historical conditions. Pedagogically speaking it is much more interesting, and at the same time a deeper understanding of macroeconomics is achieved much more

effectively, when one follows the logical–historical approach. This means that each approach is presented in the historical context of its emergence and is evaluated according to logical and not necessarily ideological criteria.

By way of an example, it would be a mistake to present and, at the same time, discard monetarism on the basis of ones opposition to the ideology of neoliberalism, which is associated with monetarism. According to our approach, the presentation must be based on the texts of monetarists and their evaluation must be carried out on the basis of primarily logical criteria and only secondarily on empirical and ideological criteria. The same should be applied also to the other schools of economic thought and respective economists. For example, Ricardo's theory of value should be based on Ricardo's text and not on the secondary literature. As we argue in the relevant chapter the presentation of Ricardo's theory of value (*e.g.*, in Marshall or Marx) differ (often in substantial ways) from that which is derived from Ricardo's text.

In this book, we advance the claim that new economic theories might be the result of three nonmutually exclusive conditions: first, a new theory might be the product of the elaboration of an existing theory; second, it might be the outcome of systematic failures of the dominant theory to account for phenomena that it was designed to explain; third, it may lead to economic policy conclusions that are more relevant to current situations than the policies proposed by the existing and prevailing theory. When at least one of these conditions is met then various adjustment processes might be activated that could lead to the advancement of an altogether new theory. In order to establish our claim we combine economic history and the history of economic thought. The intuitive idea is that economic history constitutes the testing terrain for economic theory, thus enabling us to understand the past and present in a fuller and more precise manner, while also enabling us to confront the reality of the future in a more prepared way.

While each chapter is self-contained and can be read in isolation and according to one's interests or needs, the structure of the book is oriented to contemporary economies and in this sense it compares and critically evaluates the core ideas of the major schools of economic thought and places them into proper historical perspective. An additional feature of the book is that it explores the dynamics of shifting from one school of economic thought to another. The above not only stimulate one's interest to explore further the subject and its historical development, but also provides the reader with the necessary background to deal with more advanced current developments in macroeconomics. This is something that differentiates our approach from others because it also enables the discussion of the modern schools of macroeconomic approaches, which are usually left out of analysis in the usual books of the history of economic thought.

The manuscript came out from the course of history of economic theories that I teach for many years in the University of Macedonia in Thessaloniki. I also taught parts of the manuscript in the graduate courses of macroeconomics and political economy. During this long gestation period, I am particularly indebted to Persefoni Tsaliki (Aristotle University) with whom I discussed the structure of the book and benefited from her comments on each and every of its chapters. Our joint works

have certainly influenced the writing of the book in many ways. At various stages of the work I have also benefited from the comments of Theodore Mariolis (Panteion University), Michalis Psalidopoulos (University of Athens). Special thanks also go to Aris Papageorgiou, who read the whole manuscript, and his remarks gave solutions to many problems. Of course, I have the full responsibility for the ideas expressed in this book.

Chapter 2
The Origins of Economic Theory

Independent of men's will [. . .] far from being an arbitrary value or a value which is established by agreement between the contracting parties.

(Quesnay cited in Meek 1962, p. 90)

The fundamental price of commodities is determined by the expenses or costs which have to be incurred in their production or preparation. If they are sold for less than they have cost, their price sinks to a level at which a loss is made. If they are sold at a price which is high enough to yield a gain sufficient to encourage people to maintain or increase their production, they are at their proper price.

(Quesnay cited in Meek 1962, p. 93)

2.1 Introduction

Economics as a scientific discipline originated with the emergence of capitalism, that is, the system that solves the problems of organisation of production and distribution through the generalised functioning of markets. Markets in capitalism do not work accidentally, occasionally or in the fringes of economic life, but at its epicentre, and they tend to encompass the operation of the entire economy. The salient feature of capitalism is the presence and the systematic operation of markets for the factors of production and in particular the market for the services of labour activity. Capitalism was established through a long and, at the same time, slow process that started in Western Europe in the sixteenth century and gained momentum with the advent of the industrial revolution in the mid to late eighteenth century. The publication of Adam Smith's book the *Wealth of Nations* in 1776 marks the period when capitalism dominated (at least in Great Britain) and at the same time the beginning of economics as a scientific discipline. This is not to say that in the period before the industrial revolution, there was no economic thinking; on the contrary, the philosophers, theologians and social thinkers in general were constantly making policy recommendations to the authorities to deal with specific

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_2, © Springer-Verlag Berlin Heidelberg 2010

economic problems.[1] These policy recommendations, however – whether they were correct or not for the period that they were applied to – were not characterised by generality. The reason is that the circumstances under which these policies were applied did not normally (re)appear and, therefore, there were no hidden economic forces that were working beneath the surface and could be theorised. For example, the slaves in ancient Greece or Rome worked according to the dictates of their owners. The product of their labour was appropriated by the owners, who also determined the standard of living of their slaves. Similarly, in feudalism, the serfs worked according to the dictates of tradition. The product of their labour was distributed according to the customs and also the power of the lord. Under these circumstances, non-economic forces such as political power or tradition clarified the rules of who produced what, how it was produced and who received the fruits of production. Consequently, in these societies, everything was simple and plain, and there was not much that could be theorised.

Social thinkers before the industrial revolution therefore dealt with economic questions without the existence of markets that operate systematically and determine the economic outcomes. Consequently, their analyses were based on the lack of competition. We know, however, that in the case of insufficient competition – for example, as with bilateral monopoly and isolated exchanges, in general – the equilibrium price and quantity are usually uncertain. The idea is that non-economic forces, such as the social status and the bargaining power of the traders, affect the price–quantity combination that finally prevails. This is the reason why scholastic writers of the past introduced the notion of just price (*justum pretium*). The idea was to protect the weak side of the market from the possible discretionary policies of those possessing political power. The just price is in essence a normative and at the same time administrative way of price determination of products in non-competitive conditions. With the establishment of just price, the benefits from trade are secured even for those merchants or producers with a weak bargaining position. Only with the dominance of the market as a mechanism for the arrangement of the questions of production and distribution of the social product is it possible to study the appearance of economic forces independently. Under these conditions, the development of economic thought with scientific character becomes possible, since only under these circumstances do economic phenomena make a systematic appearance and become subject to the operation of laws that govern their appearance. The evolution of economic thought reflects, to a great extent, the evolution of the capitalist economy.

In this chapter, we provide a broad outline of the essential aspects of two schools of economic thought, the mercantilist and the physiocratic. The common characteristic of these two schools is that they were essentially developed in an

[1]Hence, it is fair to say that there were many economists in the eighteenth century or even before whose work had anticipated, to a great extent, Adam Smith's ideas. However, it is also fair to say that in the wealth of Nations we have for the first time a systematic exposition of economic ideas that are applicable to capitalism.

environment, the backdrop of which was the just emerging systematic operation of the market mechanism.

2.2 Mercantilism

Mercantilism is a label that was invented by Adam Smith, who defined 'the system of commerce' or 'mercantile system' as the set of economic ideas that prevailed in Europe during the period that started around 1500 A.D. and continued up until the end of the eighteenth century. Smith, by virtue of his personality, imposed his viewpoint on the historians of economic thought. For Smith, mercantilists claim that the wealth of a society is estimated by the availability of precious metals. A country secures wealth, especially gold, if it exports more goods than it imports. On the basis of this, it is evident that the policies of a country in general must promote exports and discourage imports. The trade balance, therefore, should not be left to be determined by the free operation of market forces; the government must assume an active role so as to achieve, and at the same time maintain, a growing trade surplus. In this framework, military power and state intervention constituted the *sine qua non* for the establishment and maintenance of trade surplus.

The views of mercantilists have been developed in various countries and during different time periods. In England, for example, we have Thomas Mun (1571–1641) and William Petty (1623–1687), while in France, Jean-Babtiste Colbert (1619–1683), Louis XIV's powerful economic adviser, dominated. The texts that refer to mercantilists are relatively few and mostly do not do justice to the ideas of these thinkers. For example, some authors of late mercantilism such as Petty developed ideas that are closer to those of classical economists. The same can be said of James Steuart (1712–1780) and David Hume (1711–1776) to refer to just two from a long list of names. As a first step, it is important to point out that there have never been economists who would call themselves Mercantilists. Authors who are usually classified in the area of mercantilism are characterised by such heterogeneity that they cannot constitute a school of economic thought as, for example, is the case of the Physiocrats that we examine below.

The policies that mercantilists suggested included the maintenance of low prices for agricultural products as a precondition for low wages[2] and cheap industrial products. Mercantilists encouraged the export of industrial products, while their imports were discouraged; hence, we have the case of 'fear of goods'. Exactly the opposite policies were followed for agricultural products. The idea is that if, for example, cheap agricultural products were exported, then it might be possible for rival countries to produce their own industrial products at lower cost and so expand their market share in the foreign markets at the expense of the home industries.

[2]High wages, according to mercantilists, tend to reduce workers effort and by decreasing their productivity raise the cost of production.

A characteristic example of mercantilist policy would be the famous navigation acts that were introduced by the British Parliament in 1651 and 1660. These laws allowed the import of goods into Great Britain if, and only if, they were carried on British ships or on ships from the country of origin of the goods. Some colonial goods had to be sold exclusively in England, while some other goods had to be exported, firstly to England and then to other countries. Colonies were not allowed to import goods from countries other than England. In addition, they were not allowed to produce industrial goods and they were restricted to exporting their raw material to England and importing industrial goods. With respect to domestic trade, mercantilists developed policies encouraging domestic trade by lifting many restrictions such as the payment of tolls going from one region to another. Mercantilists, however, did not want trade to be carried out under conditions of absolute freedom. This is the reason why they were in favour of the granting of monopoly rights and exclusive privileges in the trading of goods.

In France, during the period of Colbert's ministry from 1661 to 1683, there were specific policy proposals for the encouragement of manufacturing. Among them were included the institution of state factories, the provision of incentives for the growth of population, the prohibition of migration of skilled workers, *etc*. Moreover, there were instituted specifications for the production of various commodities that guilds had to observe, while some of the impediments to international trade were gradually lifted. Colbert's influence was so strong that it continued even after his death and his name became synonymous with French mercantilism.

The German version of mercantilism is known as 'cameralism' from the German word 'camera', which means state vault. The cameralists were state employees, whose main concern was the application of effective economic policies. Their ideology was nationalism and their policy prescriptions had as a goal the strengthening of the exporting capacity of a nation, the growth of population, *etc*. One of the reasons for the development of cameralism in Germany can be traced to the slow growth of its commercial class relative to that of England or France. So in Germany the nationalistic concerns of the state bureaucrats made them assume the role of providing the necessary requirements to the commercial class so as to grow stronger and be able to compete internationally.

Smith's simplistic description of mercantilism and his emphasis on cohesiveness do not do justice to the contributions of a host of economists of the seventeenth and eighteenth centuries. It is important to stress that mercantilism is not characterised by a set of common principles that are shared by the economists that are usually classified as mercantilists. We know that the views that are attributed to mercantilists differ between countries and also differ in the same country in various time periods. Consequently, the ideas of these authors are characterised by such heterogeneity that they cannot constitute a school of economic thought in the strict sense of the term.

A rather more balanced view on mercantilism has been expressed by Marx in his *Theories* I, where he distinguishes between two basic components of the mercantile system. First, with regard to economic policy, which is oriented towards the accumulation of precious metals and second, with regard to the theory, where mercantilism seeks to discover economic laws in the sphere of circulation of

goods (in the domestic and mostly international trade). The underlying principle of the mercantilist system is 'buying cheap and selling dear'; the difference between the two prices is called 'profit upon alienation', which is realised in the form of precious metals. The fact that surplus is created in the sphere of production is something that, although pointed out, by many mercantilists, (*e.g.*, William Petty, James Hume, *inter alia*) did not receive the attention it deserved. Of course, there is nothing wrong with the idea that the difference in price between purchase and sale gives rise to profit. The trouble, however, is that by no means can this constitute a general explanation for the creation of wealth, since the profit of one party is necessarily the loss of the other.

Keynes praised the 'practical wisdom' of mercantilists. In the *General Theory* (1936), he devoted an entire chapter to mercantilism, where he develops its essential points. However, it is important to point out that Keynes treats them as if they were a unified school of economic thought. In mercantilism, Keynes finds support for his views of government intervention and the encouragement of demand in an effort to correct the weaknesses of the market system. For example, he mentions a doctrine that the classical school has repudiated as childish, but that deserves rehabilitation and honour (Keynes 1936, p. 351).

Some of the ideas of mercantilist authors that are praised by Keynes are as follows: the stock of money must be at such a level that the rate of interest is maintained at a low level so as to encourage investment; the idea that creeping inflation may exert a positive effect on the level of economic activity; and the lack of adequate amounts of money is what might be responsible for unemployment, and therefore, the expansion of credit and government deficit spending can reduce unemployment. Mercantilist authors had no illusions about the nationalistic character of their policies and their tendency to promote wars (Keynes 1936, p.346). However, according to Keynes, protectionism and nationalism are policies that contribute to full employment in a given country.

Keynes's version of mercantilism is based, to a large extent, on Heckscher's (1931) book on Mercantilism. In the second edition of his book, Heckshser (1955), however, criticised Keynes, for treating the economic categories of modern economies as if they were similar to those of the period of mercantilism. In particular, unemployment that is caused from withholding of investment is no doubt a modern phenomenon that was unknown or its meaning was quite different in the seventeenth century. In fact, for mercantilists unemployment was related to the latent labour force in agriculture and to its widespread migration to towns. Consequently, unemployment of this type does not have much in common with Keynes's unemployment due to the lack of effective demand.

2.3 Physiocracy

Physiocrats constitute the first school of economic thought that was developed, almost exclusively, in France during a relatively short period of time (from approximately 1750 to 1780). The founder of this school is François Quesnay

(1694–1774), the personal doctor of Madame de Pompadour, the powerful mistress of Louis XV. At the age of 54, Quesnay decided to deal with economic problems in an effort to contribute to the economic development and social reforms of France and to prevent the social unrest and forthcoming revolution. Quesnay's personality was strong and exerted a lasting influence on his followers. The writings of the other major Physiocrats such as Du Pont de Nemours (1739–1817) and Lemercier de la Rivière (1720–1793) were mostly interpretations and further elaboration of the ideas of their teacher. However, Anne Robert Jacques Turgot's (1727–1781) contributions are characterised by originality and extend many aspects of Quesnay's work. Such aspects include the theory of price and the determination of profit as a separate category from rent. However, the highest scientific achievement of Quesnay and of the entire Physiocratic School is the *Tableau Economique*, which was published in 1758.[3]

2.3.1 Fundamental Physiocratic Principles

Physiocrats transferred the inquiry of the sources of the wealth of a society from the sphere of circulation to the sphere of production and in particular of agricultural production.[4] Agriculture, Physiocrats argued, is endowed with the capacity to produce more output than that required as input. The difference between the output produced and the inputs that went into its production is called surplus, or in Physiocratic terminology, *produit net* (net product), which is appropriated by the landlords. Consequently, the form of surplus is that of rent, which is divided into a series of other incomes such as profits, interest and taxes. The increase in surplus is what makes a country rich or poor. The immediate question is why did physiocrats think that surplus is created only in agriculture? The answer is that in agriculture the production and appropriation of surplus is much more transparent than in other sectors of the economy. This idea stems from the simple observation that when the farmer sows a given quantity of seeds, at the end of the production period, he receives a multiple quantity of output that allows him not only to recoup the required inputs but also to secure a surplus quantity, which during the next period can be either consumed or invested. Consequently, for Physiocrats agriculture is the only sector with the capacity to create surplus. The other sectors of the economy such as industry simply transform the inputs without really adding any new value to them. Physiocrats viewed the creation of surplus as a 'gift of

[3]Marquis de Mirabeau (1715–1789), Quesnay's student and the father of Honoré Gabriel Mirabeau one of the protagonists of the French revolution, puts the *Tableau* on equal footing with the inventions of writing and money! This claim reveals after all the powerful influence that Quesnay exerted on his students.

[4]This is the reason why Marx characterises the Physiocrats "the true fathers of modern political economy" (Marx, *Theories* I, p. 44).

nature'. This is the reason why this school of economic thought has been called Physiocratic.[5]

Physiocrats believed that economies are governed by natural laws that operate 'independent of people's will'. The task of economists is, on the one hand, to lay bare the operation of these laws and, on the other hand, to propose policies that accord with them. Thus, they suggested policy measures that encourage agricultural production, the only sector of the economy with surplus-yielding capacity. Physiocrats wanted to liberate market forces from the constraints of state regulations. For example, they opposed Colbert's policies according to which there were restrictions in the exports of agricultural products. They thought that if free trade works, then the price of agricultural products will rise, whereas, the price of industrial products will fall. Consequently, agricultural production will expand and with that the surplus-yielding capacity and wealth of the nation. The price of agricultural products in such a case covers the cost of their production and, at the same time, provides the possibility for surplus. Physiocrats called this kind of price *bon prix*, which is a concept similar to that of classical economists' natural prices, or Marshall's long-run prices, which we will study in the next chapters. Physiocrats are mostly known for their policy of a single tax, which we study after we discuss their concept of *Tableau Economique* and its functions.

2.3.2 The *Tableau Economique*

In the following section, our attention is focused on the operation of *Tableau Economique*, which describes a rather ideal society with an advanced agricultural sector. The economy of Great Britain, perhaps, was close to this ideal society that the *Tableau* described and France had to emulate it as soon as possible. The central question that Physiocrats posed and tried to provide an answer to through their *Tableau* was to illustrate how society's gross and net product created in agriculture circulated in its natural and monetary form between the three social classes. The important idea that the *Tableau* conveys is the way in which the process of reproduction and realisation takes place, when certain proportions are strictly observed. For simplicity's sake, let us assume (with the Physiocrats) that we are at the end of the production period and the farmers have already selected their production, which is evaluated to 5 milliards – a 'milliard' being the Physiocratic monetary unit of account. For the production of this output, farmers used the following:

- Circulating capital (wages, raw materials, *etc*.) worth 2 milliards.
- Capital advanced (tools, animals, *etc*.) whose depreciation is equal to 1 milliard (10% of the 10 milliards of capital advanced)

[5]The name has been coined by Pierre S. Dupont de Nemours (1758), whose book is titled *Physiocracy*, which literally means "rule of nature".

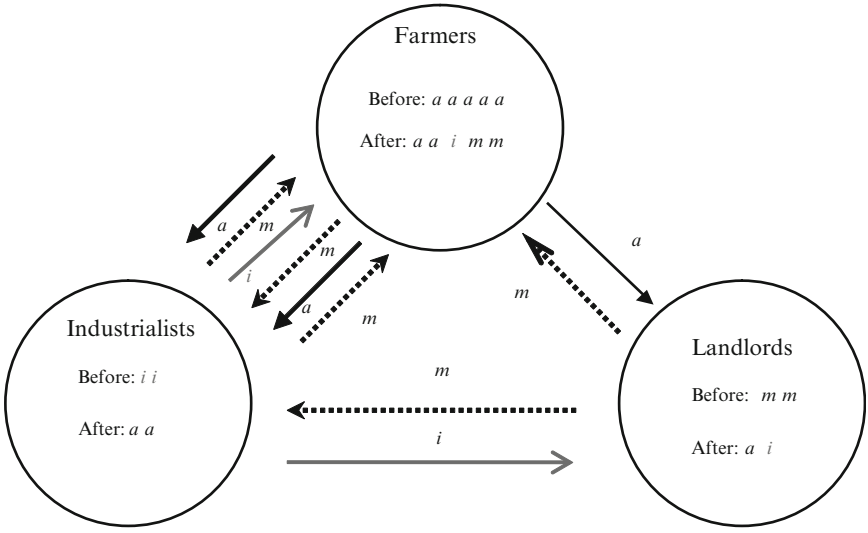

Fig. 2.1 Distribution of social product to social classes

Consequently, farmers used 3 milliards as input and produced output worth 5 milliards. It follows that the surplus equals two milliards and takes the form of rent. Farmers have already paid 2 milliards in rent in the beginning of the production period for their right to use the land of landlords, who now posses 2 milliards to spend on either agricultural or industrial products. Finally, industrialists, or the sterile class, start the new production period with a stock of industrial products worth 2 milliards that they 'produced' the previous year. Thus, at the beginning of the production period, the different classes possess, either in money or in kind, the following:

- *Farmers*: Agricultural products worth 5 milliards, of which 4 milliards are in food and 1 milliard in raw materials that can be used by the industrialists.
- *Landlords*: They already have received rent from the farmers worth 2 milliards.[6]
- *Industrialists*: They have goods worth 2 milliards.

The situation is illustrated in Fig. 2.1: where α = agricultural products worth a milliard, i industrial products worth a milliard and m is money (in milliards). The broken lines show money flow, while the flow of goods is shown with solid lines. The distribution of the social product takes place in five stages that are shown in Fig. 2.1. More specifically, we have the following transactions:

1. Landlords exchange their first 1 milliard with the farmers to purchase agricultural products of equal value.

[6]It is worth pointing out that rent is the only monetary income and that the total quantity of money is equal to 2 milliards and it is much smaller from the value of goods that are exchanged.

2. Landlords exchange their other milliard with the industrialists to acquire luxury goods of equal value.
3. Industrialists (artisans) exchange the 1 milliard with the farmers for the provision of raw materials of equal value.
4. Farmers exchange their one milliard with the industrialists for the purchase of industrial products (raw materials) of equal value.
5. The industrialists purchase with their one milliard agricultural products (food) of equal value.

After the distribution, we end up with a situation in which the farmers have agricultural products worth 2 milliards at their disposition and moreover posses industrial products worth one milliard and also 2 milliards in money so that they can rent the land to start the next production period. Industrialists possess agricultural products worth 2 milliards (one for food and the other for raw materials), which in the next period will be transformed into industrial products of equal value. Finally, landlords possess agricultural and industrial products worth 1 milliard each, which they consume, while they expect the rent payments from the farmers. The new production period starts from the moment that farmers pay their rent worth two milliards to the landlords, who in turn give the right to cultivate the land to the farmers. The final picture is displayed in Fig. 2.1.

It is important to point out that in the aforementioned description, the role of money is purely to mediate exchange. The circulation of goods is what sets the money in motion and not the other way around. Moreover, the quantity of money that is required for the circulation of goods is much less than the value of goods that actually circulate. Moreover, some of the exchanges cancel each other out and there is no need for the physical presence of money, except for its function as a standard of value of goods. These are two important corollaries that are derived from the functioning of the *Tableau Economique* and can be proved extremely useful in the formulation of a theory of money.

2.3.3 The *Tableau Economique* as an Input–Output Model

A modern way of presenting the *Tableau Economique* is an input–output table. We know that the *Tableau* had inspired Wassili Leontief (1906–1999) to create his own input–output tables (Leontief 1939). An input–output table is a double-entry bookkeeping procedure that records the transactions between sectors and the final demand. More specifically, the rows of an input–output table indicate the distribution of the total product (output) of a sector to itself and others as well as to final demand. In short, the rows of an input–output table show the allocation of total output to intermediate and final demand.[7] However, the columns of an input–output

[7]In the final demand we include the consumption expenditures of landlords only. However, the final demand includes investment and government expenditures as well as the foreign sector of the economy.

Table 2.1 The *Tableau Economique* as an input–output table

Inputs	Outputs			Total
	Agriculture	Industry	Final demand	
Agriculture	2	2	1	5
Industry	1	0	1	2
Rent (Net product)	2	0	(2)	
Total	5	2		7

table indicate the input of a sector to itself and from the other sectors. Included in the inputs are the factors of production and their respective payments. In short, the columns of an input–output table show the cost structure of an industry. On this basis, the exchange between the three sectors of the *Tableau* can be cast in terms of an input–output table as above.

The advantage of this presentation as opposed to others (*e.g.*, Maital 1972; Phillips 1955) lies in the fact that it makes clear that only the agricultural sector produces output over and above what it uses as input (Tsoulfidis 1989). The value of total agricultural output is greater than that of agricultural input by an amount of 2 milliards, and this is shown on the first column of Table 2.1. In contrast, in the industry the value of input is equal to that of output; consequently, there is no net product in this sector. In addition, landlords are not considered as a separate sector – as is the case with either Phillips's or Maital's presentation – but rather as the recipients of rent, which is identified with the net product. The landlords' consumption appears in the final demand column of our input–output representation of the economy.

The matrix of technological coefficients (A) is the matrix whose elements are derived by dividing each of the inputs of a sector by its total output produced and is given below:

$$A = \begin{bmatrix} 2/5 = 0.4 & 2/2 = 1 \\ 1/5 = 0.2 & 0/2 = 0 \end{bmatrix}$$

Every element of matrix A indicates the quantity of input needed in a sector to produce a unit of its output. In this case matrix A includes the materials, depreciation and the wages that are required for the production of a unit of output. By $p = [p_a\, p_i]$ we denote the row vector of prices with p_a the price of agricultural products and p_i the price of industrial products. Lastly, $<\rho>$ is the diagonal matrix of the rent rates. We can, therefore, write:

$$p = pA + pA <\rho> \; with <\rho> = \begin{bmatrix} \rho_a & 0 \\ 0 & \rho_i \end{bmatrix}$$

where p is the row vector of relative prices and pA denotes the cost of intermediate input; $pA <\rho>$ denotes the rent that accrues for the landlords with ρ_a the rent rate of the agricultural sector, and ρ_i the rent rate in manufacture, which is equal to zero. This system solves for the relative prices of the two goods and the rent rate $\rho = \rho_a$ of the agricultural sector:

$$p[I - A] = PA <\rho> \text{ and } p = pA <\rho> (\mathrm{I} - \mathrm{A})^{-1}$$

This is a system of two equations with three variables, that is, the two prices and the rent rate. After substitution, we get

$$[p_a \quad p_i] = [p_a \quad p_i] \begin{bmatrix} 0.4 & 1 \\ 0.2 & 0 \end{bmatrix} \begin{bmatrix} \rho_a & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 2.5 & 2.5 \\ 0.5 & 1.5 \end{bmatrix}$$

which solves for:

$$[p_a \quad p_i] = [2.5(4p_a + 0.2p_i)\rho_a \quad 2.5(4p_a + 0.2p_i)\rho_a]$$

and

$$p_a = p_i = 1 \text{ and } \rho_\alpha = 0.66$$

It is important to point out that in this transformation we can raise comparative static questions such as how do relative prices and the rent rate change in the case of the introduction of taxation? It is true that Physiocrats addressed these types of questions within the framework of their *Tableau Economique*. The same questions can also be addressed once the Tableau has been cast in terms of an input–output table.

For Physiocrats, the level and the composition of output constitutes an essential factor that affects the manner in which the total product changes. The level and composition of landlords' expenditures is of extreme importance for the scale of society's reproduction and its characterisation as expansionary, stationary or contractionary.[8] If, for example, the landlords undercut their expenditures, it follows that the fall in demand will lead to a fall in the net product. From this, we can conclude that it is the interest of the landlords to spend their rent on agricultural products, since this is how agricultural production expands and with it the net product. It is worth noting that Quesnay assumes that the landlords spend the entire net product they receive. If some of it is hoarded, then Quesnay argued that the 'failure of the annual net product to return to the productive expenditure class' leads to an economic decline (in Kuczynski and Meek 1972, p. xij). Similarly, if the state spends the tax proceeds on the maintenance of mercenary armies or as a subsidy of a foreign state, then the tax will have adverse effects on the growth of the Economy (Kuczynski and Meek 1972, p. 3). Consequently, for Physiocrats the level and composition of final demand determines the level of wealth in a society, a fact that

[8]It seems that the main concern of Physiocrats was the simple, or stationary, reproduction of society, that is, the case where net investment is zero. The idea is that the profit motive was not well established in France of the eighteenth century, a mainly rural country with the landlords consuming and not investing their rent income.

they took into account in their policy proposals. Meanwhile, in France Physiocrats had to deal with a series of problems, such as:

1. Taxes were collected mainly from the poor farmers.
2. There were many obstacles to domestic trade (tolls and different units of measurement in various areas of France).
3. Agricultural production was on a small scale and the methods of production too old.
4. There were monopoly privileges in the production and circulation of goods.

As a result of these, the tax revenues that the government could raise were not enough for its expenditures, and the government was forced to borrow and thus to increase its debt. These conditions of course were not conducive at all to the increase in the demand for agricultural products and led to the worsening of the economic position of the peasants.

Physiocrats felt that the market forces in and of themselves were not strong enough to sidestep the various legal and political obstacles that their society inherited from feudalism. This is the reason why they suggested the removal of all barriers to free competition through government intervention. In principle, Physiocrats were fervent supporters of the *laissez–fair*, *laissez–passer* policies. However, with regard to the particular circumstances of France in their day, they thought that government intervention was necessary until market forces fully dominated in the economy. Physiocrats are also known for the idea of a single tax that replaces all forms of taxation and must be paid exclusively by the landlords. Naturally, such a proposition faced the opposition of the landlords; meanwhile, the social unrests that erupted throughout France and were hiated in the 1789 revolution fully justified the Physiocratic concerns.

2.3.4 The Single-Tax Scheme

The idea of replacing the entire tax system with a single-tax system, which must be collected from the landlords, seems strange at first glance. Nevertheless, within the framework of the Physiocratic model, it appears to be the best solution for all the social classes including the landlords! The logic of the proposition is as follows: Physiocrats thought that workers have no capacity to pay taxes since their income simply allows their reproduction at subsistence. If, however, they are taxed, they shift the tax on higher wages, and the tax in the final instance is paid out of surplus, and, therefore, of rent. If the government turns to the farmers, as the source of its tax revenue, then it is certain that in their case (as with the workers), since they live in conditions of subsistence, their tax will in the final analysis be collected from rent, and, therefore, will be, once again, paid by the landlords. The mechanism of tax shifting is as follows: the tax is nominally paid by the farmers who in turn, in order to continue their activities, reduce their agricultural product that was intended for

investment, which means that the net product decreases and thus the tax burden shifts once again to the landlords.

Finally, if taxes are imposed on the sterile class, then the idea of subsistence income comes back again. As a result, the industrial products become more expensive, which means that the tax will be collected in the final analysis from the rent of landlords and the farmers, who in turn reduce their investible product and so forth. Such a tax, Physiocrats argued, leads to a reduction in demand for agricultural products, and, therefore, diminishes the net product. Physiocrats had reasons to believe that the landlords and society at large are better off with a single tax on rent incomes. The idea of a single tax was proposed later on by a number of social reformers such as Henry George (1879). The proposals for tax simplification and the replacement of all taxes with only a few containing the same revenue-raising capacity reflected after all the Physiocratic idea that a single tax is much easier to impose since its implementation is simple, predictable and at the same time has the lowest administrative cost.

2.4 Summary and Conclusions

Mercantilism is a period of time that lasted approximately three centuries during which the dissolution of feudalism occurred as did the emergence of market economy. Because of the long period of time and the divergent economic conditions from one era to another and from one place to another, the economic views that were developed are characterised by a high degree of heterogeneity. Nevertheless, there is an overriding concern that perhaps unifies all the authors during this period of time and that is the growth of the wealth of their nations. For this purpose, the policy recommendations of these authors included the abolition of internal trade barriers and the encouragement of exports especially of industrial products. Imports were allowed only to the extent to which they enhanced the exporting capacity of a nation. Clearly, the strategic goal of these authors was national prosperity and strength; while the trade surplus and the influx of precious metals were the crucial mechanism for the achievement of their strategic goal, full employment (the meaning of the term, as noted earlier, was different from today) was their major tactical goal. Hence, we observe a sort of mercantilist virtuous cycle, where the trade surplus gives rise to national wealth and power, which in turn maintain if not increase the trade surplus and so forth.

Physiocrats, by contrast, criticised the mercantilist view that surplus is created in trade and turned the focus of economic analysis from the sphere of circulation of commodities to the sphere of production and in particular agricultural production, where the creation of surplus was incomparably easier to discern than the other sectors of the economy. Physiocrats described the production and circulation of commodities and money as well as the distribution of income to social classes through their *Tableau Economique*, their major intellectual achievement. There is no doubt that Marx's schemes of reproduction, Walras' general equilibrium view

and Leontief's input output economics have been inspired by the *Tableau*. We showed that the *Tableau* can be reformulated to an input–output table and be used to address various important questions. However the idea that surplus is created only in agriculture limited, to a great extent, Physiocrats' capacity to understand the operation of a fully developed market economy and not just an underdeveloped economy based on agriculture.

Questions for Thought and Discussion

1. Discuss the rationale behind major mercantilist economic policy recommendations.
2. Was Keynes right in his evaluation of mercantilism? Discuss.
3. Discuss the Physiocrats' major ideas. What purpose(s) did the *Tableau Economique* serve?
4. Write out the *Tablaeau Economique* as an input–output model and calculate the relative prices and the rent rate. Then assume a change (for example due to taxation) in the elements of matrix A and re-estimate the relative prices and the rent rate.
5. Discuss the logic behind the Physiocratic proposal for a single tax on incomes from rent.
6. Does economics become more of a science in the mercantilist or in the physiocratic school of economic thought? Explain.

Notes for Further Reading

Rubin's book constitutes perhaps the most comprehensive presentation of the ideas of the so-called mercantilists. Rubin (1979) puts mercantilism in historical context and he argues that economic forces although were working nevertheless were proved weaker than the legal and political obstacles. Consequently, during the mercantile period it is impossible to study the action and reaction of economic forces independently of the legal and political environment within which they operated. In the same spirit Heilbroner (1981, 1985) gives a detailed historical account of the necessary and sufficient economic conditions, required for economics to obtain a scientific character (of course, with the restrictions that usually apply and differentiate social from positive sciences). The reader who is interested in a strict neoclassical view of economics is encouraged to read the book by Hirchleifer (1981) or North (1981) where it is claimed that economic theory is applicable to all periods of human history.

For additional literature on mercantilism the classic book by Heckscher (1931) is frequently cited. However it has been criticised by Blaug (1985, p. 16) who asserts with regard to Heckscher that 'his analysis of mercantilism displays almost absurd irritation with anything that smacks of economic determinism', while Niehans (1994, p. 18) describes the book as a 'massive though analytically disappointing treatise'. Viner's (1968) extensive summary

(continued)

of the mercantilists' views is worth reading; the same is true with the entry in the New Palgrave dictionary by Allen (1989).

Most of the Physiocrats' writings are in French; however, Meek (1963) has collected perhaps the most important essays and has published them in English. The same is true with the various versions of the *Tableau Economique* (Kuczynski and Meek 1972). Once again Rubin's (1979) treatment of physiocracy is a must before the interested reader delve into the other literature. It is important to note that Joseph Schumpeter (as we learn through his student Samuelson) considers Quesnay as one of the four greatest economists of all times, the other two are Walras and Cournot. He does not mention the fourth. Despite the fact that Schumpeter (1954) does not devote many pages to the study of the physiocracy nevertheless it seems that he praised the general equilibrium approach contained in the *Tableau* that bears a lot of similarities to his own beloved circular flow (Samuelson 1982).

There are various efforts to reformulate the *Tableau* in an input–output table. The story starts with a group of students who followed Leontief, when he was teaching at Harvard. Firstly, George Malanos (1946) wrote his dissertation under Leontief's supervision and then the effort continued with Almarin Phillips (1955), Shlomo Maital (1972) and Paul Samuelson (1982), perhaps the last of Leontief's students that dealt with the *Tableau*. However, the reformulation of the *Tableau* in a way such as to become operational is discussed in Tsoulfidis (1989). Pressman (1992) criticised such a reformulation, on the grounds that the *Tableau* refers to conditions of distribution and not to production that the input–output tables refer to. However, it is not difficult to see that the *Tableau* shows the technique of production at the same time (matrix A, which also includes the given real wage) and so one can determine relative prices. In the recent decade there has been renewed interest in the ideas of Physiocrats and a number of papers and books have been published (Steilze 2000; Giacomin 2001, *inter alia*). The major characteristic of these efforts is the restatement of the *Tableau* in the form of an input-output model similar to the one presented above. The Physiocratic theory of taxation is discussed in Einaudi (1967), Bilginsoy (1994) and Eltis (2000), while a rigorous formulation of taxation in an input–output framework is discussed in Tsoulfidis (1989).

Chapter 3
Adam Smith's *Wealth of Nations*

> *The value which the workmen add to the materials, therefore, resolves itself in this case (in which land is excluded) into two parts, of which one pays their wages, the other the profits of their employer upon the whole stock of materials and wages which he advanced. He could have no interest to employ them, unless he expected from the sale of their work something more than what was sufficient to replace his stock to him; and he could have no interest to employ a great stock rather than a small one, unless his profits were to bear some proportion to the extent of his stock.*
>
> (*Wealth of Nations*, p. 48)

3.1 Introduction

In this chapter, our objective is to present the main contributions of Adam Smith (1723–1790) starting from his first book *The Theory of Moral Sentiments* (1764) and continuing with his main book, *An Inquiry Into the Nature and Causes of the Wealth of Nations,* which established him as the founding father of economics. In fact, the *Wealth of Nations* (henceforth *WN*) was published in 1776, the same year of the Declaration of Independence of the American Colonies. Heilbroner (1981) notes that there is a close relation between the two events, since the declaration of independence describes the requirements of a capitalist society, while Smith's book lays bare the mechanisms that govern the operation of such a society. In what follows, we deal with the first effort of Smith to create what we call a theory-generating concept and then with his theory of value, his growth model and his theory of the falling rate of profit and the stationary state. A discussion on taxation and public debt follows, and finally, a summary and some remarks conclude the chapter.

3.2 The Theory of Moral Sentiments

The fundamental question that Smith pursues in his book *The Theory of the Moral Sentiments* is what kind of forces hold society together? Smith's question and answer are certain to have been inspired by Newton's model of self-regulated

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_3, © Springer-Verlag Berlin Heidelberg 2010

universe, where the law of gravity determines the movement of planets. Turning to the social universe, Smith argued that society is held together by the force of morality. The notion of morality in Smith is quite different from the Christian one based either on the fear of punishment or the reward in after-life. In his book, Smith claimed that moral standards are determined from the social system in power and not by innate invariable moral standards that hold true always and everywhere. Thus, Smith replaced the Christian conception of innate moral sense with a theory based on 'sympathy'. Smith argued that all sentiments and actions in relation to other people emanate from the human ability to see things from someone else's point of view and feel the sentiments of the other, that is, to sympathise with him.

This idea of sympathy does not spring from people's instincts but from their social education. Human beings reach a level of perfection, when they can distance from themselves and become 'impartial spectators' and evaluate the situation and either approve or disapprove of it. The immediate consequence of the sympathy principle is to provide us with an explanation of moral perceptions. In a society where freedom prevails, that is, the customs and traditions as well as the power of a central agency (such as the feudal lord) no longer play any significant role, people are constrained by a perception of morality that has been implanted in them right from their young age. A child, under normal conditions, is taught to behave in such a way as to obtain the recognition and approval of others. This training begins with the family, continues in the neighbourhood and at school and finally, extends to the society at large. In this process, people develop a conscience that enables them to evaluate situations and also to be able to pass judgement on them.

Smith further pointed out that it is one thing to assume the role of the impartial spectator and another thing to sympathise with someone and act accordingly. For that, it takes both parties involved to give their best; he further suggested that in order to attain the sympathy of the spectator, the person under judgement must be humble and fulfil, at least in part, the spectator's expectations.

Thus, to the question what prevents a society, where everyone is pursuing his self-interest, from collapsing to chaos – where there is no central agency (such as, the feudal lord) to ensure order and discipline – Smith's answer is that the principle of sympathy and the associated notion of benevolence generate the forces that bind people together in a society with cohesion. Smith's theory about morality, though it was extremely important and popular for the time that was developed – the book went through six editions[1] – it is doubtful that this book could make Smith known until our days. Smith is known because of his second important book on which we turn our focus.

[1]In these different editions, there were many changes indicating that Smith was not fully satisfied with the text.

3.3 The Market Mechanism

The *Wealth of Nations* is the book that made Smith famous and established him as the Father of Economics. The whole work is divided into five books; the first two are theoretical, while the other three books refer mainly to questions of economic policy and also include a lot of historical material. The content of the book is truly revolutionary if we take into account that it was written during a time period in which feudalism was still widespread and also quite strong, and that the part of the economy that was organised according to the market mechanism was extremely small, albeit rapidly growing.[2] Smith argued that wealth consists of all goods in the disposition of people in a society. Unlike mercantilists for Smith, neither gold nor precious metals are viewed as the exclusive constituent components of wealth. And unlike Physiocrats, Smith did not confer any special status either to the (foreign) trade activity (as the mercantilists) or to agricultural production, as the Physiocrats did. Smith described a society in continuous motion, where the flow of goods and services that are consumed constituted the goal of economic life.

The first great question that Smith pursued was to reveal the mechanism through which a social system of the type outlined above would function over time and remain stable. At first sight, it comes as a surprise for a social system that is based primarily on self-interest to develop internal mechanisms that lead to social cohesion, where the needs of individuals are not only satisfied, but are compatible with the needs of society at large. In addition, this system is based primarily on decentralised decisions of numerous individuals without the mediation of any central co-ordinating agency and without the customs and traditions to play any significant role; nevertheless, such an individualistic society is capable of reproduction not only on the same scale but also on an expanding scale.

Observations of this sort led Smith to the discovery of the laws that govern modern society. To borrow Quesnay's expression, these laws work 'independent of men's will' and guide self-interest and other human passions to a direction, which is co-ordinated with the interests of the entire society. Smith described the spontaneous operation of objective laws that guide human action in a way that accords to the interests of society as a whole, the 'invisible hand'.

> As every individual, therefore, endeavours as much as he can both to employ his capital in the support of domestic industry, and so to direct that industry that its produce may be of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither indents to promote the public interest, nor he knows how much he is promoting it. By preferring to support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an *invisible hand* to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it.

[2]The ideas that are developed in the next pages are based to a great extent on Heilbroner (1981).

> By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. (*WN*, p. 423)

In spite of the fact that the notion of the 'invisible hand' is one of the most successful academic metaphors of all times, in Smith's works, there are only two occurrences of the concept. One is in the *Wealth of Nations* cited earlier, and the second is in *The Theory of Moral Sentiments* that we cite below:

> (The business owners) are led by an *invisible hand* to make nearly the same distribution of the necessaries of life which would have been made had the earth been divided into equal portions among all its inhabitants; and thus, without intending it, without knowing it, advance the interest of the society, and afford the means to the multiplication of the species [. . .]. (*Moral Sentiments*, p. 386)

By introducing the notion of economic laws and their operation, Smith placed economics on a scientific foundation. This is the reason why he is called the founding father of modern economics. Smith showed that a society of egoistically motivated individuals does not collapse to chaos as one would expect but actually manages to solve the fundamental problems of production and distribution more effectively than traditional (feudalistic) society. Furthermore, the emerging society is capable not only of holding together all centrifugal forces giving rise to cohesion but also of growing, a feature that was absent in traditional society.

The mechanism through which this ostensibly paradox result takes place is based on self-interest, which, as a separate and, at the same time, an independent force, guides people to activities that are socially desirable.

> It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantages. Nobody but a beggar chooses to depend chiefly upon the benevolence of his fellow-citizens. (*WN*, p. 14)

Self-interest, nevertheless, constitutes only one part of the whole story, because it only explains human actions. While, one would expect that self-interest would lead to a chaotic society, Smith instead argues that society does not collapse because of competition. When each and every member of the society seeks to serve his or her own interest, without regard for social costs, his or her interest confronts with those of other similarly motivated individuals. Every individual is ready to take advantage of the unscrupulous behaviour of his or her competitors when the latter deviate from what is considered as the socially accepted norm. If, for example, a producer charges too high a price, then it is expected that the buyers will turn to other producers. If a producer underpays his workers, well-qualified workers will seek employment elsewhere. We observe that as in the case of *The Moral Sentiments*, self-interest (a socially unacceptable motive) gives rise (via the complex social interrelations) to the most surprising result which is social cohesion.[3]

[3]We purposely avoid the word harmony because we do not want to rule out social conflict, which is inherent in Smith's work.

The self-interest of each individual producer would dictate a price that would give rise to excessive profits. Producers, however, would be very careful not to charge such a price because apprehension that they might lose their market share to competitors. If now all producers of the same commodity decided to raise their price, this would be possible, other things constant, only for a limited period of time. The excess profits would soon attract producers from other industries, and the expansion of supply would lower the price to the point that would restore the normal rate of profit. Meanwhile, there is no certainty that all producers would stick to their agreement.

Thus, market forces dictate to producers the prices that they charge and producers supply the quantities the market wants. For example, if the price of a product is too high, this is an indication that society wants larger quantities of the product in question. High prices give rise to excess profits, which lure investment in this profitable activity; the ensuing accumulation of capital expands the supply of the product, something that society wanted in the first place and at lower prices. The converse process takes place in the case where the price is below normal. Consequently, society, through the market mechanism, solves the problem of production by supplying the amounts of goods that are socially desired, at prices that incorporate the normal rate of profit, and what is really remarkable is that all the above take place without the mediation of any central co-ordinating agency dictating what is to be produced and at what quantities and what prices must be charged so as the producers plans to align with those of the consumers. The problem of distribution is implicitly solved, since the prices of the products incorporate the normal rewards of the factors of production, that is, normal wages, profits and rents.

Until now, the analysis is simple without any complications, and the results are plausible. On further investigation, we discover that the analysis based on the entwining of the notions of self-interest and competition leads to the following results:

1. Producers sell their commodity at approximately uniform price.
2. Market prices of all commodities fluctuate around the normal (natural) prices.
3. Social wants are satisfied through the operation of spontaneous market forces.

In short, Adam Smith discovered in the market mechanism a self-regulated system, whose operation does not lead to chaos and to eventual breakdown, but rather to a system that manages to give rise to social cohesion and to hold together all centrifugal forces.

3.4 The Theory of Value

The theory of value is the effort to connect the surface phenomena of economic life to some inner law. The idea is that if we are looking at the fundamental economic problem of the production of wealth of a nation as well as of its distribution, then we are unavoidably confronted with the problem of price determination. There is no

doubt that economic agents in a market society follow price signals in order to decide upon questions of production and distribution. At first glance, it seems that prices are dependent on the forces of supply and demand, but on further examination, one discovers that supply and demand are themselves governed by something more fundamental, which is (for Smith and the classical economists in general) the 'principle of equal profitability', and that there is another set of more fundamental prices (natural prices) that regulate the fluctuations of actual prices. The theory of value deals with the connection of market prices to something more essential and enduring, that is, the normal prices, which, unlike market prices, can be subjected to theorisation. The theory of value addresses the question of the determination of relative and absolute prices of commodities, and thus, it makes possible the estimation of the current production and distribution of wealth. The absolute prices enable inter-temporal comparisons of production and wealth in a country.

Smith is characterised by inconsistency with regard to the theory of value, since he supported at least three different views. Smith initially assumes a primitive society in which there is neither capital nor wage labour. In such a society, everyone produces his goods using his own labour and the surplus of goods that he may produce exchanges with other goods of similarly situated producers.[4] Under these conditions, Smith argues that the relative prices of goods must be proportional to the quantities of labour that are spent on their production. Hence, we refer to the often-cited deer and beaver parable.

> In that early and rude state of society which precedes both the accumulation of stock and the appropriation of land, the proportion between the quantities of labour necessary for acquiring different objects seems to be the only circumstance which can afford any rule for exchanging them for one another. If among a nation of hunters, for example, it usually costs twice the labour to kill a beaver which it does to kill a deer, one beaver should naturally exchange for or be worth two deer. It is natural that what is usually the produce of two days or two hours labour, should be worth double of what is usually the produce of one day's or one hour's labour. (*WN*, p. 49)

The aforementioned example is formally stated as follows:

$$\frac{P_D}{P_B} = \frac{L_D}{L_B}$$

where P_D and P_B are the prices of the deer (D) and beaver (B), and L_D and L_B are the respective quantities of labour time that are required. If, for example, it takes one labour hour to kill a deer and two labour hours to kill a beaver, then it follows that two deer are exchanged for one beaver,

$$\frac{P_D}{P_B} = \frac{L_D}{L_B} = \frac{1 \text{ hour}}{2 \text{ hours}} \text{ and } 2D = 1B$$

This is Smith's first view and it is known as the labour theory of value.

[4]Hence, the production of surplus and its exchange with other products is not systematic but accidental.

In Smith's example, it is natural that the proportion $2D = 1B$ is the only equilibrium one, where the hunter spends 2 h of labour and catches either two deer or one beaver. Any other exchange ratio would be a disequilibrium and therefore, ephemeral. Let us suppose that the proportion is $1D = 1B$. In such a proportion, the hunting of beavers is discouraged, since by hunting for an hour, one can catch a deer and then exchange it for a beaver, which requires 2 h of hunting. Such proportions cannot last long; the oversupply of deer would diminish their price, whereas the shortage of beavers would raise their price until the point that the exchange rate of the two goods is restored to $2D = 1B$. In this exchange ratio, the hunters have no interest in the exclusive hunting of deer or beaver. This conclusion is based on the assumption of free competition, that is, a process of free mobility of resources from the production of one to the production of another good. Of course, such mobility is dictated by the self-interest of individuals.

Smith, however, introduced a second theory of value according to which the value of commodities is determined by the quantity of labour that a given commodity can purchase. Smith notes:

> Every man is rich or poor according to the degree in which he can afford to enjoy the necessaries, conveniences and amusements of human life. But division of labour has once thoroughly taken place, it is but a very small part of these with which a man's own labour can supply him. The far greater part of them he must derive from the labour of other people, and he must be either rich or poor according to the quantity of that labour which he can command, or which he can afford to purchase. The value of any commodity, therefore, to the person who possesses it, and who means not to use or consume it himself, but to exchange it for other commodities, is equal to the quantity of labour which enables him to purchase or command. Labour, therefore, is the real measure of the exchangeable value of all commodities. (WN, p. 30)

This second view has been established in the economic literature as labour commanded the theory of value. Clearly, in conditions of simple commodity production, where there is no wage labour and the producers of commodities work with their own means of production, the two definitions of value coincide. By way of an example, let us suppose a hat maker who exchanges a hat for a pair of shoes. We say that *ceteris paribus* the work of the hat maker is equated with the work of the shoemaker or that a hat is equivalent to the work of the shoemaker who produces a pair of shoes. Consequently, in primitive societies or in conditions of simple commodity production, the same relative prices are the result of the labour theory of value or the labour-commanded theory of value.

When Smith tried to apply these two definitions to modern society, where there is accumulation of capital, wage labour and landed property, a series of questions arose for which the two theories of value could not provide the same answers. In order to show the difficulties that Smith encountered in his application of the two definitions of value to modern societies, let us suppose that the shoemaker in the aforementioned example now becomes a worker in a shoe-producing factory. The wage that he receives is lower than the total value of the shoes that he produces. This means that the entrepreneur who hires the worker, in the final analysis, receives more value for his commodity than that which he paid to hire the worker.

In other words, the labour time spent on the production of shoes is no longer equal to the labour that the produced commodity purchases (or commands). Consequently, in modern societies, where there is capital and wage labour, the relative prices of commodities may be affected by non-labour factors. The production of shoes, for example, in modern societies, where there is employment of capital and ownership of land, and, therefore, there are profits and rent, which must be subtracted from the value of the commodity. The labour commanded by the commodity is different from the labour that it contains.[5]

These difficulties led Smith to abandon the previous two theories and to accept a third theory of value in an effort to account for the incomes of capital and landed property. This theory is known as the adding-up theory of value,[6] according to which the value of a commodity is determined by the sum of three natural incomes, that is, natural wages, natural profit and natural rents. Smith notes:

> When the price of any commodity is neither more nor less than what is sufficient to pay the rent of the land, the wages of the labour, and the profits of the stock employed in raising, preparing, and bringing it to market, according to their natural rates, the commodity is then sold for what may be called its natural price. (*WN*, p. 55)

Hence, Smith does not simply argue that the constituent components of the price of a commodity include the wage, the profit and rent (something that is always true), but he argues about something that is quite stronger, that is, the above independently determined incomes jointly determine the natural price of the commodity.

The adding-up theory of value has two disadvantages. The first relates to the fact that the three constituent components of price are themselves prices which need to be determined. If in order to determine the natural price of a commodity we are required to know about other prices, then the question that comes to the fore is how these other prices are determined in the first place? If the other prices are also determined by another set of prices, as is the case with Smith's theory of value, then we end up in a vicious cycle, where the determination of the natural price of commodities requires the prior determination of another set of prices and so forth.

The second disadvantage of this theory is that it comes up with conclusions with respect to the price level and not to relative prices. More specifically, if one of the constituent components of price changes, then the price of the commodity changes in the same direction. This view, however, leads to problems of logical consistency in Smith's theory. David Ricardo was opposed to the adding-up theory of value, since, in his theory, an increase, for example, in wages implies an equivalent decrease in profits which leads to changes in relative prices without necessarily affecting the general price level.

---

[5]For a formal statement of the labour-commanded theory of value, see the Appendix to this chapter.

[6]The term is attributed to Sraffa (1951, pp. xxxv) see also Dobb (1973, p. 46) and Kurz and Salvadori (1995, p. 7).

This third theory of value, in fact, constituted a theory of price determination based on the cost of production, which were to play an important role in our days. This notion of value became the starting point for the development of theories of value that place labour, capital and land as equivalent factors of production and creators of value. This interpretation was supported by many economists of the nineteenth century, who include J.B. Say and J.S. Mill. Ricardo, as we will examine in the next chapter, adopted Smith's first view of value known as the labour theory of value and made an effort to apply it to modern society and solved some of the riddles that made Smith to abandon his first (and correct according to Ricardo) view.

3.5 The Making of Economic Science

If we examine the prices of commodities, we observe that they are subject to continuous fluctuations for which there cannot be any single explanation. At first sight, one could be led to the conclusion that these fluctuations in market prices are completely arbitrary. A more careful examination, however, would reveal that the fluctuations in market prices are around a set of more fundamental prices, which are called natural prices, and operate as centres of gravity for market prices.

> The actual price at which any commodity is commonly sold is called its market price. It may either be above, or below, or exactly the same with its natural price. (WN, p. 56)

Adam Smith was the first economist who made such a fundamental distinction between the natural price and the market price.[7] Market prices are just a descriptive category, they can go up or down for a variety of reasons which are not liable to a theoretical analysis. The natural prices, however, reflect persistent economic forces for the explanation of which we can set up a theory. More specifically, the term 'natural', as Rubin (1979, p. 174) notes, in Smith has two dimensions: First, 'spontaneity' and second 'law-determined regularity'. With respect to the first dimension, the word 'natural' signifies a result that emerges from the operation of free competition and it is found to be in conflict with self-interest. More specifically, the word 'natural' is used as opposed to the words 'just price' or 'legally set price' fixed by custom, the state or some other authority, and second, the conflict of individual interests in the sense that it is not a monopoly price. With respect to the second dimension, the word 'natural' indicates that not all prices qualify as natural prices, but only a single one can be characterised as a central price towards which the market prices are continuously gravitating. This is because only for a single price, demand and supply are equal to each other. Only this central price is equal to the sum of the natural payments of labour, capital and land.

[7]It is true that Physiocrats and some forerunners of the classical economists (e.g., William Petty and James Steuart) made such a clear distinction, but it is only after Smith's analysis that the distinction became widely known and accepted.

This discovery of 'natural prices' by Smith constituted the object of economic theory and since then, economists have the same object of analysis although they use different theories to explain it. For example, the natural price in Marx is called the price of production, and the 'transformation problem' (whose details are discussed in Chap. 4) refers to the complex mediations between the natural and market prices. This distinction can also be found in neoclassical economics, and in particular, in the work of Alfred Marshall, whose natural price is called the long-run price on which the market prices converge. Consequently, both the classical and neoclassical schools (whose structure is discussed in Chap. 6) share a common object of analysis which is the determination of the natural prices of commodities.[8]

Adam Smith defined natural price as the price that is adequate to pay workers, capitalists and landlords their natural wages, profits and rents, respectively. When a commodity is sold at its natural price, there will be a quantity of that commodity that will be in demand. Smith called this kind of demand 'effectual demand'. More specifically, Smith suggested:

> The market price of every particular commodity is regulated by the proportion between the quantity which is actually brought to market, and the demand of those who are willing to pay the natural price of the commodity, or the whole value of the rent, labour, and profit, which must be paid in order to bring it thither. Such people may be called the effectual demanders, and their demand the effectual demand; since it may be sufficient to effectuate the bringing of the commodity to market. It is different from the absolute demand. A very poor man may be said in some sense to have a demand for a coach and six; he might like to have it; but his demand is not an effectual demand, as the commodity can never be brought to market in order to satisfy it. (*WN*, p. 56)

It is interesting to note that the natural price in Smith constitutes a necessary prerequisite for the determination of demand (Garegnani 1997, p. 417) and not the other way around. This is a characteristically different approach from the neoclassical one (see Chap. 6), where the demand (curve) is determined first and through it the supply or offer (curve) is then derived. The two curves then jointly determine the equilibrium price and quantity at the same time.

The role of effectual demand in Smith is to interpret the attraction of the market price of commodities to its natural price and not to determine the natural price.

Consequently, if we want to make a graphical depiction, then, the effective demand would not constitute a curve, but rather a single point on the price-quantity diagram. In terms of Fig. 3.1, point e stands for the effectual demand, where the quantity x^* is being demanded at the natural price p^*.

Smith argues that when the quantity supplied is smaller than the effectual demand, then the market price is greater than the natural price, and thus, there are excess profits in the production of this commodity. Consequently, there is encouragement of investment which leads to an increase in the produced quantity (see the north-western quadrant in Fig. 3.1). On the other hand, if the quantity produced is

[8]Since the 1960s, the object of analysis of neoclassical economics, at least in the so-called intertemporal equilibrium, strand is no longer natural prices. On this question, we return in Chaps. 7 and 8).

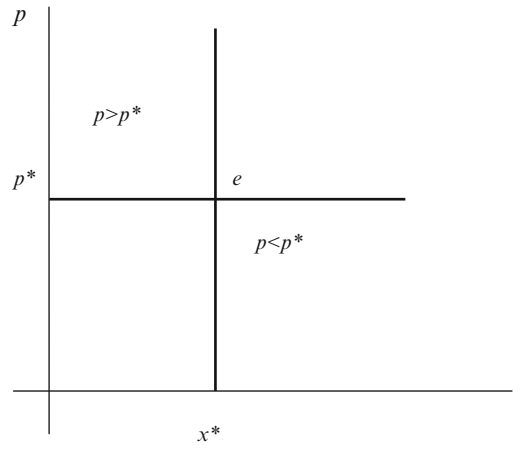

Fig. 3.1 Natural price and effectual demand

greater than x^*, then the market price falls below the natural price. As a result, producers make profits smaller than the normal amounts (see the south-eastern quadrant in Fig. 3.1). Consequently, investment slows down and thus, reduces the quantity produced. A characteristic of this analysis is the ever-present disequilibrium situation (*i.e.*, $x \neq x^*$). According to Smith, only accidentally, and, therefore, temporarily is it possible to reach equilibrium, (*i.e.*, $x = x^*$ and $p = p^*$), and the economy is continuously gravitating towards the equilibrium point which, nevertheless, does not rest on it in any single period of time, but only on an average and after the passage of long time (years) of continuous fluctuations.

This notion of natural price and its connection to competition led to the formation of economic theory as a scientific discipline. This is Smith's great contribution. Before Smith, economics was mainly a descriptive discourse or an appendage to Philosophy. With Adam Smith, however, economic theory becomes a systematic and coherently expounded independent scientific inquiry. This became possible with the appearance of an abstract category, the natural price that made economic theorisation possible. Smith's analysis did not explain how precisely the natural price of a commodity is determined, but that competition tends to maintain the market price close to the natural price.

This analysis, at first sight, leads to the temptation of drawing ordinary demand and supply curves as one finds in most history of economic thought texts. The temptation is so strong that even established economists give in. A characteristic example is that of Blaug (1983) who, on the one hand, admits that there are no such curves in Smith, and, on the other hand, claims that this is because Smith is not familiar with such curves. Specifically, Blaug notes: 'Smith thinks of demand and supply as referring to people's willingness to buy or sell at a particular price rather than at all possible prices; the former is expressed in actual amounts desired or offered, the latter in a schedule of amounts, each corresponding to different price. Still, the whole of the passage given above (see the above quotation as well as the *WN*, pp. 56–57) has no real meaning unless demand at any rate is interpreted in the

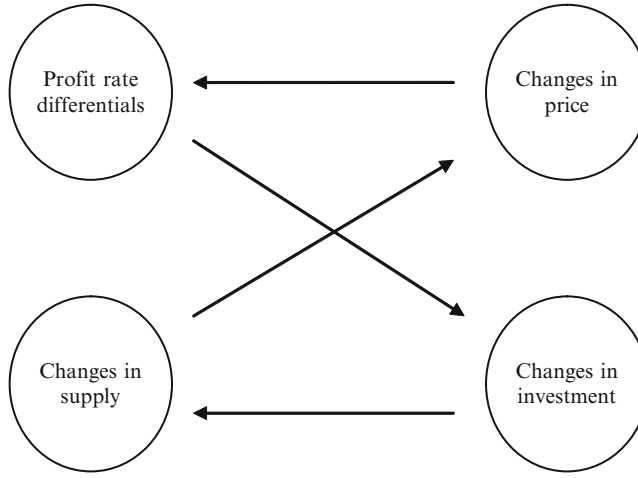

Fig. 3.2 Smith's dynamic equilibrium process

sense of a schedule, and a negatively inclined schedule at that. Here and elsewhere, Smith intuitively gropes his way toward the right answer' (Blaug, p. 44). On closer examination, however, we discover that such curves have no place in Smith's analysis. The reason is that every point on a demand curve represents a possible equilibrium, which is equivalent to saying that each point on an ordinary demand curve corresponds to a natural price. The price that finally prevails in the market is the one that is determined by the intersection of demand and supply curves. Market prices in Smith, however, do not bear any relation with equilibrium situations, since they are by definition points away from equilibrium, where profits, wages and rent are not on their natural magnitudes.[9]

The mechanism through which market prices orbit around natural prices can be depicted in terms of the following illustration:

As it is shown in Fig. 3.2, changes in the rate of profit away from the normal (natural) rate of profit of an industry elicit changes in investment, which lead to changes in the quantity supplied, which in turn leads to changes in prices and the rate of profit, which comes closer to its natural level along with the other variables.

Adam Smith was the first economist who understood this process and posed it in a way so as to be able to make generalisations. Smith argued that the process of competition tends to establish normal and systematic relations between prices. This systematic behaviour that Smith studied relates to the tendential equalisation

[9]Marshall who had an inclination to explain almost everything in neoclassical terms, attributed to Smith the wish to have a demand curve, where each point on the curve was a possible equilibrium point or natural price (a notion of demand which is different from that of the effectual demand of Smith). After this, it was easy for Marshall to argue that the difference between the market and the natural price was a difference of degree related with the time interval during which the process of equilibrium was taking place (see Garegnani 1983, pp. 309–313).

of the price of similar goods and services. Smith described this organisation as the principle of competition, by referring to the relation between natural price (*i.e.*, the equilibrium price) and market price that is the actual price at which exchange takes place. Smith described this relationship as the law of supply and demand,[10] and it is characteristically different from the theory of supply and demand of the neoclassical economists.

The law of supply and demand does not refer to the determination of the natural (normal) price, but only to the proposition that market prices will orbit continuously around their normal prices. Smith's market prices refer to prices which bear no relevance to the neoclassical demand and supply curves and do not correspond to uniform profit rates. In other words, these are disequilibrium prices, while the supply and demand curves are merely the locus of equilibrium points. Eatwell (1977, pp. 61–68) argues that Smith thought that the market prices are determined by supply and demand but not by the intersection of demand and supply curves. Market prices depend on a variety of forces which include uncertainty, bad crops, monopoly, *etc*. Consequently, they cannot be determined in terms of systematic forces, as is the case with natural prices.

3.6 Smith's Model of Economic Growth

It is important to bear in mind that Smith was describing an economy in transition from feudalism to the market system. Smith had to show that, on the one hand, the emerging new social organism is viable without any central authority to dictate what is to be produced and how the fruits of production are to be distributed among the members of such a society. Smith showed that although the market society functions in a completely decentralised way, where each individual seeks to satisfy his own self-interest, yet this society creates internal mechanisms that bind people together in a cohesive totality. On the other hand, Smith had to demonstrate that such a society is capable of economic growth which, however, is not out of control but rather contained by internally generated mechanisms. So speaking of Smith's description of economic growth, we must stress the fact that he was describing a society in transition, where customs and traditions were still important. For example, production was taking place by small firms, and in different professions, there were still specifications that were followed in the production of each good, behavioural codes that ought to be observed, *etc*. In addition, the State was granting various kinds of monopoly privileges; for example, the commerce with the East India was granted exclusively to the East India Company.[11] We know that the

[10]Though Smith never used this expression (see also Löwe 1975, p. 417).

[11]Although Smith was writing during the industrial evolution, there is no evidence that he was aware of the tremendous changes that were taking place and which, in a short period of time, would transform economic life completely.

Physiocrats believed that market forces were weak and that state intervention was necessary to make the economy work according to natural law. By contrast, Smith believed that the market mechanism is powerful enough to cope with any obstacles that were caused by feudalistic relics.

As a consequence, the natural tendency of the market economy is the growth of wealth that is created from the division of labour and the ensuing increase in productivity and reduction in cost. Smith showed the advantages of the division of labour by observing the production process in a pin factory, where each worker out of a group of ten could construct at most a few pins by accomplishing all stages of the pinmaking process himself. If, however, the labour process was subdivided into many different tasks designed for each worker, then the total production of pins would be in the range of thousands of pins a day! Hence, it is worth quoting this famous example of Adam Smith:

> I have seen a small manufactory of this kind where ten men only were employed, and where some of them consequently performed two or three distinct operations. But though they were very poor, and, therefore, but indifferently accommodated with the necessary machinery, they could, when they exerted themselves, make among them about twelve pounds of pins in a day. There are in a pound upwards of four thousand pins of a middling size. Those ten persons, therefore, could make among them upwards of 48,000 pins in a day. Each person, therefore, making a tenth part of 48,000 pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made 20, perhaps not one pin in a day; that is, certainly, not the two hundred and fortieth, perhaps not the four thousand eight hundredth part of what they are at present capable of performing, in consequence of a proper division and combination of their different operations. (*WN*, p. 5)

Smith refers to three reasons that the division of labour leads to this remarkable increase in productivity:

> first to the increase of dexterity in every particular workman; secondly, to the saving of the time which is commonly lost in passing from one species of work to another; and lastly, to the invention of a great number of machines which facilitate and bridge labour, and enable one man to do the work of many. (*WN*, p. 7)

It is important to point out that while Smith recognised the advantages of the division of labour on the productivity and growth of a society, he did not disregard the disadvantages, and pointed out the adverse effects on the mental development of workers. In fact, according to Smith, the material wealth resulting from the introduction of machines and the further subdivision and routinisation of the labour process are associated with the physical and mental degradation of workers. The following passage from the *Wealth of Nations* is revealing of the antagonistic relation between material wealth and human degradation, and the means for the possible prevention of the latter.

> In the progress of the division of labour, the employment of the far greater part of those who live by labour [. . .] comes to be confined to a few very simple operations; frequently to one or two. But the understanding of the greater part of men are necessarily formed by their employments. The man whose life is spent in performing a few simple operations, of which

> the effects, too, are always the same, or very nearly the same, has no occasion to exert his understanding. [. . .] He naturally loses, therefore, the habit of such exertion, and generally becomes as stupid and ignorant as it is possible for a human creature to become. [. . .] In every improved and civilised society this is the state into which the labouring poor, that is, the great body of the people, must necessarily fall, unless government takes some pains to prevent it. (*WN*, p. 734)

The division of labour does not increase indefinitely but is conditioned by the 'extent of the market' and the accumulation of capital. The larger the size of the market, the greater the demand for goods, and consequently, the greater the likelihood for further division of labour. We say that the division of labour is constrained by the accumulation of capital, since the production process requires time that intervenes between the beginning of the production process and the disposition of the products. This is how Smith justified his view:

> As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or, in other words, by the extent of the market. When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment, for want of the power to exchange all that surplus part of the produce of his own labour, which is over and above his own consumption, for such parts of the produce of other men's labour as he has occasion for. (*WN*, p. 17)

In conditions of simple commodity production, the problem between the time of production and the time of the disposition of goods is almost non-existent, since a few resources are needed for maintenance (food, clothing, shelter, *etc*.) during the production process. To the extent that the division of labour increases, workers produce even fewer goods for self-consumption and thus, the need to purchase a greater amount of goods from the market arises. This required stock of money for starting the production process comes from savings, and Smith calls it capital. Smith suggested that one of the basic functions of a capitalist is to take care of the required means for the survival of workers, for the time that the latter are engaged in the production process and before he receives his revenues from sales. Consequently, the division of labour depends directly on the size of the so-called 'wage fund'.

The division of labour unquestionably leads to a rapid increase in productivity, which in turn leads to the growth of the economy. Hence, two interrelated questions are posed: First, what is the motive that compels society to the continuous division of labour, and, therefore, to economic growth? Second, how do we know that this growth tendency will continue and will not be interrupted? With respect to the first question, Smith suggested that people in all societies are characterised by the insatiable desire for bettering,

> But the principle which prompt to save [=invest], is the *desire of bettering our condition*, a desire which, though generally calm and dispassionate, comes with us from the womb, and never leaves us till we go to the grave. In the whole interval which separates those two moments, there is scarce perhaps a single instant in which any man is so perfectly and completely satisfied with his situation, as to be without any wish of alteration or improvement of any kind. An augmentation of fortune is the means by which a greater part of men

> propose and wish to better their condition. It is the means the most vulgar and the most obvious; and the most likely way of augmenting their fortune is to *save and accumulate* some part of what they acquire. (*WN*, pp. 324–325)

While the desire for bettering is (perhaps) true in all societies, in the modern one, it takes on a very specific form, which Smith describes in the following way:

> The consideration of his own private profit, is the sole motive which determines the owner of any capital to employ it either in agriculture, in manufactures, or in some particular branch of the retail trade. (*WN*, p. 355)

Let us now hypothesise a typical entrepreneur whose objective is to increase his profits. Having made profit the goal of his entrepreneurial activity, it follows that this goal is only served if the entrepreneur invests in new plant and equipment, which allows the even-finer division of labour, the subsequent increase in productivity and the reduction of unit cost which expands demand and leads to further investment, and, therefore, economic growth.[12]

Until now, we have an answer to the first question; that is, the causes that lead to the growth of a business and by extension of the economy as a whole. In fact, for the first time, we have the operation of the laws of motion of the economy that work 'independently of people's will'. It is not that the entrepreneur is ordered by someone to invest his profits, but rather competition compels him to invest (particularly in new technologies) in order to increase productivity, which is the only effective way to reduce cost, and to undercut prices in order to remain competitive. If the entrepreneur decides otherwise, it is certain that sooner or later he will be undersold by his competitors.[13]

The next question is whether this growth of the economy will continue forever.[14] Smith argues that investment increases the demand for labour, wages increase and profits fall, and, therefore, the initial motive for investment is lessened but not for long. According to Smith, the population law is the control mechanism through which profits are restored and the growth process starts once again. In particular, Smith argued that as in the case with the price of every good, the price of labour depends on its demand (*WN*, p. 80). The higher demand for labour in a growing society means higher wages which gradually lower profits and slow down the growth process. Here is where the population law enters the picture; higher wages make the living conditions of workers much more comfortable, which is equivalent to saying, in Smith's time, that the child mortality rate was falling, and, therefore,

[12]Hence, it is assumed that the introduction of machinery does not replace workers, so in Smith, employment can only increase (see Löwe 1975, p. 419 and Heilbroner 1975, p. 526).

[13]There are many references in the *Wealth of Nations*, where Smith describes competition as a race or contest and not as the static situation of perfect competition envisioned in neoclassical economics.

[14]Nowhere does Smith describe any model of economic growth in the current sense of the term. Simply, many authors – among them are included: Adolph Löwe (1975), Robert Heilbroner (1975, 1981) and Walter Eltis (1975, 2000) – by reading Smith's text, managed to put together the necessary elements and create what today is called a model of economic growth.

labour supply was rising, which would result in lower wages and higher profits.[15] The opposite would be true in the case where the wage is lower than the subsistence level, which implies bad living conditions, and, therefore, increase in the child mortality rate and decrease in the supply of labour and in higher wages. Only if the wage is at the subsistence level is there, a tendency for the population and the labour supply to remain constant.

3.7 The Falling Rate of Profit and the Stationary Economy[16]

Until now, we have established that the goal of the entrepreneur is to invest his profits in order to make more profits. As a consequence, the investment in machinery leads to a further division of labour, productivity increases, cost, and, therefore, price decline and demand increases, and this virtuous cycle is ready to repeat itself. One stays with the impression that this growth process may go on forever. It is true that investment increases the demand for labour and wages while profits decrease, but only temporarily, since the higher wages set in motion the population law, and because as labour supply expands, wages fall once again, profits increase and so does investment, and the economy expands and so forth. As Heilbroner (1981) notes, we do not have here any kind of business-cycle theory but rather a process of continuous improvement, and naturally, this raises the question whether there is an end to this spiral-like way with which society evolves. Smith argued that the growth of the economy depends principally on the profit rate, which displays a long-run falling tendency.[17] Smith notes:

> The increase of stock, which raises wages, tends to lower profit. When the stocks of many rich merchants are turned into the same trade, their mutual competition naturally tends to lower its profit and when there is a like increase of stock in all the different trades carried on in the same society, competition must produce the same effect in them all. (WN, p. 87)

The aforementioned often-cited quotation is interpreted to mean that competition tends to reduce the rate of profit in two ways: first in the product market where each producer is compelled to 'sell what he deals in somewhat cheaper' and second, in the labour market, where the competition of producers raises the wages paid to the labourers, and, therefore, profits are diminished (WN, 89).

This interpretation of the falling rate of profit as a result of the intensification of competition has been formulated by Ricardo. For example, he notes that Smith 'uniformly ascribes the fall of profits to accumulation of capital, and to the

[15]Malthus explicitly drew upon Smith's suggestion when he formulated the population principle.

[16]This section is based on Paitaridis and Tsoulfidis (2009).

[17]The profit rate expresses the total profit as a percentage of invested capital in the production process. With respect to this issue, Smith notes: The entrepreneur would have "no interest to employ a great stock rather than a small one, unless his profits were to bear some proportion to the extent of this stock" (*WN*, p. 48).

competition which will result from it' (Ricardo 1951, p. 289). Thus, Ricardo could level his criticisms charging Smith for inconsistencies because competition can only bring about an equalisation of profit rates across industries to a general rate of profit. As a consequence, the generally accepted view of the falling rate of profit remained an open question on which Ricardo could advance his own interpretation.

Ricardo's interpretation of Smith was so convincing and also convenient that even Marx attributed the falling rate of profit in Smith to excessive competition (see for example, *Grundrisse*, p. 751 and also his *Theories of Surplus Value* p. 438). Thus, Ricardo and Marx have essentially established the view that in Smith, excessive competition is the root cause for the falling rate of profit, and a majority of historians of economic thought and all the widely used textbooks subscribe to this view. Of course, there are a few exceptions, which include Hollander (1973), Eltis (1984, 1987) and Verdera (1992).

From the three authors, Eltis presents a pretty sophisticated argument, which seems to 'square the circle', for he combines the increase in stock as the cause for the falling rate of profit, together with the intensification in competition. More specifically, according to Eltis, (1987) as the capitalisation of production in the agricultural sector increases, productivity, however, remains more or less the same, and this happens so because in agriculture, the division of labour is much more difficult to apply (*WN*, 234).[18] As capital accumulates in agriculture and the demand for labour increases, leading to rapidly growing wages (which cannot increase any more), it follows that the resulting surplus will be split between profits and rent, and if there is a division that is approximately equal and constant, then it follows that the profit share to income remains approximately constant. Given the rising trend in the capital–output ratio (because of improvements in soil, fertilisation and the number of oxen), it follows that the rate of profit falls below that of industry and trade. Hence, we have the intensification of competition inasmuch as the falling profitability in agriculture, leading to an acceleration of accumulation to the more profitable industry and trade. The result of this acceleration of capital accumulation is the inter-sectoral profit rate equalisation towards the agricultural lower profit rate. In Eltis's analysis, one wonders why should capitalists in the agricultural sector of the economy increase their stock of capital if they are going to get a profit rate lower than that of the rest of the economy and, furthermore, why should they invest in agriculture if their profit picture is bleak.

Smith's argument of the falling rate of profit is couched on the idea that the profit motive as an end in itself is what stands behind the actions of capitalists. This insatiable desire for profit leads to the division of labour as the means to increase the productivity of labour, which in turn reduces the unit cost of production and the selling price. The expansion of demand that follows necessitates even further division of labour, which becomes possible through the introduction of fixed capital. The presence of machines facilitates the further division of labour which in turn increases

[18]With the exceptions of vegetables by products of improved methods of production and hides, as when the demand for cattle increases the price of hides falls.

productivity and reduces the unit cost and prices, thereby leading to another repetition of this 'virtuous cycle' that we described. It is important to point out that the invention of various machines that form the newly introduced fixed capital is a direct consequence of the (sub)division of labour (*WN*, p. 260). As a result, we expect an increase in the total capital in a country that is on a growth trajectory (*WN*, p. 326).

The increase in total capital has two dimensions (a) fixed capital and (b) circulating capital. Although we know that the growth in fixed capital was relatively low in the eighteenth century, in the *Wealth of Nations*, this growth is very important for an increase in labour productivity through the division of labour (*WN*, p. 277). On the other hand, the circulating capital naturally tends to increase following the rising tendency of the fixed capital. As we know, by circulating capital, Smith and the classical economists (*e.g.*, Ricardo) mean raw materials and wages (*WN*, p. 267).

Having identified the role of fixed capital, Smith points out that though technical change may be, at first instance, labour saving, it by no means displaces labour; rather the introduction of fixed capital manages to 'facilitate and abridge labour'. Thus, by means of fixed capital, 'an equal circulating capital (mainly wages) can afford a much greater revenue to its employer' (*WN*, pp. 265–266).

It is important to stress at this point that for Smith, profits are not created by fixed capital and so, we dispense with any idea of attributing to Smith a marginal productivity theory of distribution. Clearly, for Smith, productive labour is the source of profits. He writes:

> [T]he labourers and labouring cattle, therefore, employed in agriculture, not only occasion, like the workmen in manufactures, the reproduction of a value equal to their own consumption, and of the capital which employs them, together with its owner profits [. . .] (*WN*, p. 344)

The continuous increase in the stock of capital will require an increasing share of output to be devoted in the replacement of capital. This share is directly related (in both absolute and relative terms) to the wealth of a nation. Writes Smith:

> That part of the annual produce, which, [...] is destined for replacing a capital, is not only much greater in rich than in poor countries, but bears a much greater proportion to that which is immediately destined for constituting a revenue either as rent or as profit. (*WN*, pp. 318–319)

As a consequence, for Adam Smith, the increase in the capital–output ratio will lead to a profit squeeze. In the limit, the rate of profit will vanish and with this, the growing state of the economy will give its place to the stationary one, where net investment is zero or what amounts to the same thing there is only replacement investment.

3.8 The Stationary Economy

Equilibrium as an outcome is more likely to be realised in a stationary state, where all accumulation and growth come to a halt. It is in such a situation that it becomes more realistic to talk about the perfect equality of market and natural

prices or the rates of return in different employments to be perfectly equal. The act of accumulation occurs to take advantage of supernormal profits, and these profits are not likely to be the same in every employment of capital. For Smith, this end-state was still far away, as nations were yet to acquire the full complement of riches that was possible. In the chapter on 'the natural progress of opulence', Smith talks about the various employments of capital. In accordance with the natural course of things, the correct sequence for the employment of capital of a country is first agriculture, then manufacturers and finally, foreign commerce. Smith reasons:

> As capitals increase in any country, the profits which can be made by employing them necessarily diminish. It becomes gradually more and more difficult to find within the country a profitable method of employing any new capital. (*WN*, p. 375, Vol. I)

This is equivalent to saying that any new investment is unlikely to secure positive profits. Thus, we reach a point at which there are no new investments and as a result, output and employment do not change. Thus, it is only the fall of profits, and at the same time, the lack of opportunities to direct capitals to more profitable activities that actually stimulate competition between capitals, rather than the other way around. Smith notes:

> There arises in consequence a competition between different capitals, the owner of one endeavouring to get possession of that employment which is occupied by another. But upon most occasions he can hope to justle that other out of this employment, by no other means but by dealing upon more reasonable terms. (*WN*, p. 375, Vol. I)

Consequently, the rising capital stock leads to a rising capital–output ratio which is the principal cause for the falling rate of profit. The intensification of competition will be the result of a situation where capitals cannot be used in a profitable manner anymore. In other words, the stagnation of profits will be the result of overaccumulation, a point of view very close to Marx's (see Chap. 5).

In the stationary state, one would expect that the low rate of profit is accompanied by high wages, but this is not the case, because on Smith's claims, both the rate of profit and wage are low:

> In a country which had acquired that full complement of riches which the nature of its soil and climate, and its situation with respect to other countries, allowed it to acquire; which could, therefore, advance no further, and which was not going backwards, both the wages of labour and the profits of stock would probably be very low. (*WN*, p. 94)

However, there are two countertendencies which may prevent the fall in the rate of profit, the discovery of new territories and new trades (WN, pp. 93–94). Smith, however, argued that these countertendencies are only ephemeral interruptions of the falling tendency. One such example is the navigation laws of 1651 and 1660 that gave the monopoly to England of trade with her colonies. Smith pointed out that the stationary economy is not for the present circumstances and that it will take quite a long time until society reaches the stationary level. In fact, such a state is not attainable in the foreseeable future (*WN*, p. 95).

What happens in reality with regard to the evolution of the profit rate? In a period of time, with no national income account data, Smith argued that the rate of profit is subject to a lot of fluctuations which make it difficult for someone to estimate it over time. Nevertheless, he claimed that a satisfactory approximation of the rate of profit would be the rate of interest. Smith based his view on the following argument:

> But though it may be impossible to determine with any degree of precision, what are or were the average profits of stock, either in the present, or in ancient times, some notion may be formed of them from the interest of money. It may be laid down as a maxim that wherever a great deal can be made of the use of money, a great deal will commonly given for the use of it; and that wherever little can be made by it, less will commonly be given for it. According, therefore, as the usual market rate of interest varies in any country, we may be assured that the ordinary profits of stock must vary with it, must sink as it sinks, and rise as it rises. The progress of interest, therefore, may lead us to form some notion of the progress of profit. (*WN*, p. 88)

Consequently, the rate of interest and the profit rate are connected to each other and the long-term movement of the one is indicative of the movement of the other. According to Smith, the rate of profit (approximately) must be twice as high as the interest rate, which is charged on a low-risk loan (*WN*, p. 97). Thus, when Smith was interested in examining the extent to which his theory of the falling tendency of the rate of profit holds true, he looked at the data on the rate of interest whose evolution verified (according to Smith) the falling tendency in the profit rate. In Fig. 3.1, we portray the evolution of the interest rate (or yield) of the console bonds of the UK, which clearly displays a falling tendency, a pattern which fully agrees with Smith's conjecture for the long-term movement of the profit rate of the whole economy.[19]

In an overall evaluation of Smith's theory for the long-term tendency of the rate of profit, some attention is required. It is true that in Smith's analysis, the rate of profit displays a long-run falling tendency which eventually reaches zero for the reasons that we have outlined earlier. But even in this situation, the picture is not so bleak for society. This rather overall optimistic view is based on the following considerations: First, the wage stabilises to its natural level, however, for a longer period of time, it was above the subsistence level. In fact, Smith claims that there is a long-term increase in wages over and above the subsistence level so long as the growth of wages does not exceed the labour supply required by the growth of capital. This idea in the prospect of long-term improvement in the workers' welfare which was based on the fact that the wages of English workers had actually risen from the seventeenth to the mid-eighteenth centuries, distinguished Smith's optimistic world view from the pessimistic views of the economists after him; for instance, Ricardo Malthus and J.S. Mill. Figure 3.3(b) portrays the evolution of the real daily wage of the English workers during the seventeenth and eighteenth centuries. Clearly, the real wage follows an upward trend which fully accords with Smith's conjecture. Second, the landlord no doubt improves his position

[19]Smith argues that in the beginning of the eighteenth century the rate of interest in England and elsewhere was somewhat higher than 10%, and at the time of the writing of the *Wealth of Nations*, did not exceed 6%, Fig. 3(a) above is consistent with Smith's view.

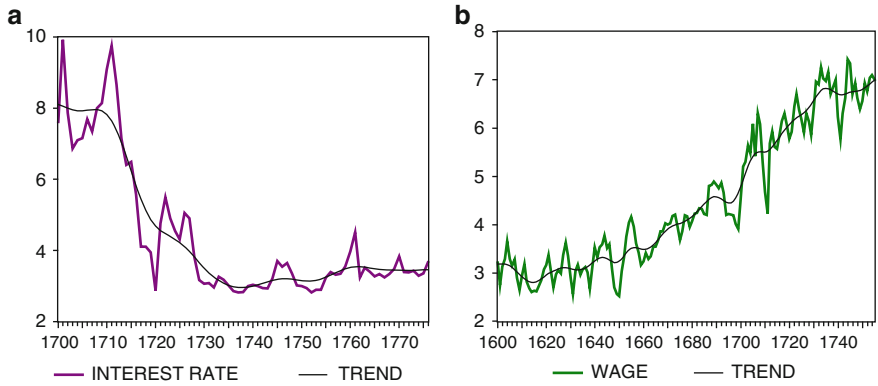

Fig. 3.3 Interest rate on consoles and daily real wages of English workers
Source: the data on the UK consol bonds yield are from the Global Financial Data and start from the year 1700, whereas the daily real wage rate data are from Makridakis, Wheelwright and Hyndman (1998) and are available from the thirteenth century, we just arranged the data to fit better Smith's argument and time period

since he possesses resources whose scarcity raises their prices. Consequently, Smith points out that the progress of society will see an increase in the real rent directly or indirectly. This is not due to the efforts of the landlords who, according to Smith, 'love to reap where they never sowed' (*WN*, p. 21) and they live on the work of the other social classes. Third, the capitalist class is certainly with the greater losses, since the profit rate tends to zero. On closer examination, however, we discover that capitalists do not make any excess profits, and, therefore, do not live on luxury. In fact, the capitalist, argues Smith, secures for himself a kind of wage, somewhat higher than that of workers, as a reward for his administrative and, in general, his entrepreneurial abilities. Consequently, even the capitalists are forced to work in order to secure for themselves a kind of wage somewhat higher than that of their workers, what is important, however, is that they do work in order to survive (Heilbroner 1975, p. 529).

Smith pointed out that the stationary economy is not for the present circumstances, and it will take quite a long time for society to reach the stationary level; meanwhile, society must be subject to substantial reforms in order for the market forces to work without feudalistic impediments. Thus, Smith refers more to his contemporary circumstances which were not what they ought to be, and, therefore, substantial changes were needed.

3.8.1 Productive and Non-Productive Labour

There are conflicting views on the importance of the distinction of labour activity in productive and unproductive. Some economists view this distinction as critically

important for the growth of the economy, and for some other economists, this distinction is simply superfluous. In what follows, we examine Smith's position and, at the same time, we make some more general comments on this issue which remains unsettled even in our days. To begin with, the distinction initially was referring to whether or not a type of labour was morally accepted. Subsequently, the distinction was restricted on whether or not a kind of labour was creating surplus, and, therefore, wealth in a society. Consequently, according to the dominant theory of the generation of surplus, labour activity was characterised as either productive or non- productive.

More specifically, in the mercantilistic literature, surplus is created from the difference of the price that a merchant buys a commodity and the selling price of the commodity. We know that Mercantilists surplus is created in (international) trade. Consequently, it was natural for them to think that surplus is created in (international) trade, materialising itself in the flows of precious metals. This conceptualisation of surplus led to the conclusion that labour engaged in export trade, such as the labour of seamen, merchants and the like, was surplus-yielding, and, therefore, productive. As a consequence, the expansion of exports and the labour employed in these activities was regarded conducive to economic prosperity and to national strength.

Physiocrats, as we know, refuted the mercantilist doctrine according to which exchange, in general, and international trade, in particular, were the sources of surplus. They argued that although exchange can become beneficial for one of the parties involved at the expense of another, the establishment of free trade and therefore, the abolition of any restrictions and monopoly rights necessarily leads to the exchange of equivalents; thus, 'profit upon alienation' cannot be made in any systematic way. For Quesnay and Physiocrats, surplus was created in production, and, therefore, preceded exchange, which was normally among equivalent values. Thus, they transferred the inquiry of surplus creation from exchange to production. For Physiocrats, surplus arises exclusively in agriculture and it was the natural outcome of land's exclusive physical capacity to create more output than input going into production. Hence, for Physiocrats, the term productive labour referred exclusively to that engaged in agriculture, the only kind of activity endowed with this surplus-yielding capacity.

Smith agreed with Physiocrats' critique of the mercantilist productive nature of international trade, since he also conceived the latter as a source of mutual benefit rather than differential gain, leading to a larger and cheaper aggregate output. In fact, Smith had a great admiration for many aspects of physiocratic thinking and considered their ideas as 'the nearest approximation to the truth that has yet been published upon the subject of political economy'. Nevertheless, he could not accept their narrow idea of surplus and its creation because of land's physical productivity. He remarked:

> The capital error of this system, however, seems to lie in its representing the class of artificers, manufacturers, and merchants as altogether barren and unproductive. (*WN*, p. 176)

Smith's distinction of productive labour became broader due to the development of a notion of surplus based on production in general and not just agricultural

production. For Smith, all wage labour engaged either in agriculture or manufacturing was productive, insofar as it was being paid directly from capital and gave rise to a profit for the owners of capital. By contrast, the wage labour paid out of revenue, that is, profit, rent or even wage, was considered unproductive. In his own words:

> There is one sort of labour, which adds to the value of the subject upon which it is bestowed: there is another which has no such effect. The former, as it produces a value, may be called productive; the latter, unproductive labour. Thus the labour of a manufacturer adds, generally, to the value of the materials which he works upon, that of his own maintenance and of his master's profit. The labour of a menial servant, on the contrary, adds to the value of nothing. Though the manufacturer has his wages advanced to him by his master, he, in reality, costs him no expense, the value of those wages being generally restored, together with a profit, in the improved value of the subject upon which his labour is bestowed. But the maintenance of a menial servant never is restored. A man grows rich by employing a multitude of manufacturers: he grows poor, by maintaining a multitude of menial servants. (*WN*, p. 314)

In the aforementioned quotation, Smith made a distinction between productive and unproductive labour according to social relations. This first distinction, however, was followed by a second, which, like that of Physiocrats, was restricted to the character of the product. Thus, continuing the long aforementioned quotation on productive and unproductive labour, Smith pointed out:

> But the labour of the manufacturer fixes and realises itself in some particular subject or vendible commodity, which lasts for some time at least after the labour is past [...] The labour of the menial servant, on the contrary, does not fix or realise itself in any particular subject or vendible commodity, which endures after that labour is past, and for which an equal quantity of labour could afterwards be procured. The sovereign, for example, with all the officers both of justice and war who serve under him, the whole army and navy, are unproductive labourers. They are the servants of the public, and are maintained by a part of the annual produce of the industry of other people. Their service, how honourable, how useful, or how necessary soever, produces nothing for which an equal quantity of service can afterwards be procured. The protection, security, and defence of the commonwealth, the effect of their labour this year will not purchase its protection, security, and defence for the year to come. In the same class must be ranked, some both of the gravest and most important, and some of the most frivolous professions: churchmen, lawyers, physicians, men of letters of all kinds; players, buffoons, musicians, opera-singers, opera-dancers, *etc*. The labour of the meanest of these has a certain value, regulated by the very same principles which regulate that of every other sort of labour; and that of the noblest and most useful, produces nothing which could afterwards purchase or procure an equal quantity of labour. Like the declamation of the actor, the harangue of the orator, or the tune of the musician, the work of all of them perishes in the very instant of its production. (*WN*, pp. 314–315)

Hence, the two above-mentioned distinctions are put together, and are treated as if they were identical; thus, services were not considered production because they were not materialised in a tangible product and, by extension, labourers employed in services were viewed unproductive. Smith's assumption was perhaps a good approximation for his time, since services were mainly provided by the government and were not as extensive as they are today. But with the progress of time, the amount of non-tangible products expanded and so if we were to follow Smith's

definition, we would fall into a series of riddles. These riddles were addressed systematically in the work of Marx who continued this tradition by praising Smith's penetrating approach to the classification of labour activity and to the eloquent discussion on the importance of this distinction to the growth of wealth of a capitalist society.

The distinction between productive and unproductive labour was gradually abandoned by the neoclassical economists in the last quarter of the nineteenth century. Alfred Marshall (1890) is the first neo-classical economist who attempted in a systematic way to discredit this distinction that prevailed in the history of economic thought for two centuries. Marshall notes:

> we may define labour as any exertion of mind or body undergone partly or wholly with a view to some good other than pleasure derived directly from the work. And if we had to make a fresh start it would be best to regard all labour as productive except that which failed to promote the aim towards which it was directed, and so produced no utility. (Marshall 1890, p. 54)

Marshall's rationale became the foundation of today's orthodox economics according to which 'productive labour' is any kind of labour insofar as it is considered useful. The criterion for defining usefulness is that someone is willing to pay for this kind of labour. Thus, with the turn of the century, the demarcation line drawn between productive and unproductive labour in orthodox economics was rendered synonymous with marketed versus non-marketed goods and services.

Schumpeter (1954) has exemplified the view held by neoclassical economists by criticising the dichotomy between productive and non-productive labour in the following harsh way:

> The only reason why this dusty museum piece interests us at all is that it affords an excellent example of the manner in which the discussion of meaningful ideas may lose sight of their meanings and slip off into futility [...] And so meaningless discussion became a standard item of nineteenth-century textbooks in spite of the increasing awareness of its futility, which eventually killed it. (Schumpeter 1954, p. 631)

To our view, the distinction between productive and unproductive labour continues to occupy the attention of economists from different theoretical perspectives dealing with questions, such as the slow down in productivity and growth performance of the economy, and the explanations they offer are based on the distinction of productive and unproductive labour. The construction of national income accounts is also based on a notion of productive–unproductive labour and the dissatisfaction with these accounts led several times until now to their redefinitions. Bearing these considerations in mind, it becomes particularly important to have a clear understanding of this dichotomy, which dominated economic theory for at least two centuries.

3.8.2 General Statements on Taxation

Smith's views on taxation and in general on public finance, are included in Book V of the *Wealth of Nations* titled: *Of the Revenue of the Sovereign or Commonwealth*,

which is more than one-fourth of the *Wealth of Nations*. Historians of economic thought either disregard or just give a cursory consideration to these chapters. The reason is that they usually think of these chapters as inferior to the rest of the book; after all, the student of economics can find anything he needs on taxes and government expenditures in the books of Public Finance. In our view, there are many important lessons to be learnt from Book V of the *Wealth of Nations* and hence, solely for reasons of space, we restrict ourselves to a summary version of Smith's statements on the question of taxation, encouraging the reader to go through the relevant chapters and literature cited in the notes for further reading.

In what follows, we begin with Smith's famous four maxims of taxation that must characterise a socially just taxation system:

1. The subjects of every state ought to contribute towards the support of the government, as nearly as possible, in proportion to their respective abilities; that is, in proportion to the revenue which they respectively enjoy under the protection of the state.
2. The tax which each individual is bound to pay ought to be certain, and not arbitrary. The time of payment, the manner of payment and the quantity to be paid, all ought to be clear and plain to the contributor, and to every other person.
3. Every tax ought to be levied at the time, or in the manner, in which it is most likely to be convenient for the contributor to pay it.
4. Every tax ought to be so contrived as both to take out and to keep out of the pockets of the people as little as possible over and above what it brings into the public treasury of the state. (*WN*, pp. 350–351)

The first maxim emphasises equity, however, there is ambiguity whether it supports the ability-to-pay or the benefits received criterion according to modern parlance in the philosophy behind the imposition of certain taxes. The other three maxims are administrative and refer to certainty, convenience of payments, and economy in collection, respectively. It is important to point out that Smith's emphasis was placed on the last three canons, which he felt that once society follows, the first canon, which is about equality and equity, is much better served.

Smith's analysis of the effects of various taxes is intrinsically connected to his theory of value and distribution. We know that Smith did not have a single theory of value and distribution; this becomes particularly pronounced in his treatment of rent, where sometimes he gives the impression that rental income is determined by price and at other times that rent determines the price (see, for instance *WN*, p. 85). This ambiguity is natural to (re) appear in his theory of taxation, whose major statements are summarised below. We begin with taxes on income flows (or direct taxes) and then we proceed with taxes on commodity flows (indirect taxes).

The first income flow that Smith considers is land rent and he, as the Physiocrats before him, argues that a tax on land rent is paid by the landowner and it cannot be shifted to other incomes, since the farmer must make the natural rate of profit. A tax on house rent, however, Smith argued has different implications whose examination requires first of all the distinction of house rent into building and ground rent. Starting from the ground rent which is paid for some real or supposed advantage of the location

of the house, we observe that the effects of taxation are similar to those of land rent. The incidence of this tax falls exclusively on the landowner (WN, p. 795).

As for the building rent, Smith says that it is essentially profits of the capital expended in building the house. This part of tax will fall on the resident (occupier), the idea is that the builder must receive his normal rate of profit and competition ensures the receipt of such payment (*WN*, p. 792). Thus, a tax on the building rent will be paid by the resident, and, therefore, it may fall on any of 'the different sources of revenue' (*WN*, p. 794).

The analysis of taxation of profits begins with the two-part division of profit income: the interest part and the surplus part which we could call net (of interest) profit. A tax on the net profits of the farmer is shifted to landlords, argued Smith. The idea is that,the farmer, to make a normal rate of profit, deducts the tax from the rent that he pays and so, the burden of this tax falls on the landlords. If the profits of the other activities are taxed, the effect will be higher prices (which usually compensate more than fully for the amount of the tax) that are paid by (rich) consumers. A general profit tax, Smith argued, is passed (usually with an overcharge) on higher prices and it is ultimately paid by the (rich) consumers. As for the tax on interest, Smith argued that it is not shifted and falls wholly on the lenders exactly as in the case with the land rent. The difference between these two types of income is that the interest on money is not that easy to ascertain from net profits and so, by 'taxing' interest income, it might be possible to tax net profits with completely different effects. According to Smith, the same is not true with land rent, since land is a tangible asset whose value can be easily ascertained. By contrast, capital is an altogether different asset whose value cannot be easily identified.

A tax on wage income is always shifted, in other words, workers pass this tax on higher money wages which are paid by the owners of capital who finally pass it on higher prices and rich consumers once again bear the burden of this taxation. This is due to the fact that wages in Smith are regulated by the demand for labour and by the prices of goods that constitute the basket of goods that workers consume (*WN*, p. 816). If the real wage remains the same – in fact, this is how the real wage is treated not only in Smith but also in Ricardo and the other classical economists – then a direct tax on wages simply raises the money wage in order for workers to be able to purchase the same bundle of goods as before the tax. The shifting mechanism is different in manufacturing and agriculture. Specifically, a wage tax in manufacturing raises wages which producers pass on higher prices and the tax finally is being paid by the (rich) consumers. In agriculture, however, there is an increase in wages as a result of tax, and farmers deduct this increase not from their profits but from the rent. In Smith's analysis, there is the idea that the effect of the tax on agricultural wages on rent incomes is much worse for the landlords when it is paid indirectly (through a tax on wages) and perhaps landlords would rather pay such a tax directly out of their rent incomes (*WN*, p. 815).

Smith's analysis of taxation of commodity flows has differential effects depending on the categories of commodities. He divided commodities into necessities and luxuries. A tax on necessities will increase their prices which will make more expensive the bundle of goods that workers normally consume and so, the money

wage will increase. The increase in money wage will lead to even higher prices, and the tax in the final analysis will be paid by the (rich) consumers. A tax on luxury goods, however, will increase the prices of these goods and will be paid wholly by the rich consumers. Money wages are not affected since workers are not supposed to consume luxury goods.

From the above, it follows that in Smith's analysis, the taxation of net profits, wages and necessary goods is always shifted to (rich) consumers, whereas, the taxation on rent and luxury goods cannot be shifted. In modern parlance, the nominal and the real incidence of taxation on luxury goods and rent are one and the same because landlords and rich consumers bear the burden of these taxes. In similar fashion, the taxation of interest income rests on the lenders of capital. Undoubtedly, there are certain questions with regard to the logical consistency of Smith's theory, which for reasons of space, we do not pursue here. Such a discussion is postponed until the next chapter, where we present Ricardo's theory of taxation which, in many respects, starts as a critique of Smith's theory.

3.9 On Public Debt

In the last chapter of the *Wealth of Nations*, Smith analyses the economic effects of public debt. In the 41 pages of this chapter, Smith argues that governments should not run budget deficits, because the accumulation of debt is considered 'pernicious' for the nation even if all of it is owed to domestic investors. In fact, Smith attacks the mercantilist notion according to which the payment of interest on public debt is like 'the right hand which pays the left'. For Smith, this is an 'apology founded altogether on the sophistry of the mercantile system' (Smith 1937, p. 879). The reason is that soon the need to redeem the debt will lead to increased taxation, causing the flight of domestic capital and the devaluation of the currency with negative effects for the remaining domestic producers (*WN*, pp. 927–929). The debt, according to Smith, severely retards the 'natural progress of a nation towards wealth and prosperity' (*WN,* p. 674) since resources that could be used productively from the private sector of the economy are diverted by the state in order to finance its unproductive activities. Consequently, Smith proposed balanced budgets, where all government expenditures are financed by taxation. Budget deficits can be justified only in emergencies, as those that arise during wars or natural disasters. In such circumstances, Smith argues that the method of financing public expenditures (*i.e.*, via taxation or issue of public bonds) is crucial for capital accumulation (*WN*, p. 878).

The aforementioned implies that taxation decreases mainly the households' expenditures and savings only to a limited extent. Taxation, therefore, does not have the same effects with those of public borrowing which encroaches mainly on savings. Consequently, the amount of money raised by government through borrowing crowds out an equal amount of private investment. Hence, the underlying

idea is that for Smith and the classical economists, in general, savings are identified with investment. Consequently, taxation interferes with new investment and thus, with the accumulation of new capital, leaving the existing productive capacity intact; the same is not true, however, with public borrowing which undermines the existing productive capacity by displacing savings from the 'maintenance of productive labour' to unproductive and wasteful usages. In general, the two methods of financing of government expenditures are not equivalent, and taxation is preferred to borrowing since the latter diminishes savings, that is, the investible product and hence the accumulation capacity of the nation.

3.10 Summary and Conclusions

Adam Smith's first task was to show that the market system emerging out of feudalism was not only viable but also capable of economic growth. This was not an easy task given that he was arguing for a society where there was no central co-ordinating agency such as the feudal lord, and traditions and customs were important. In fact, it was not easy to prove that a society based on the activities of merchants and small manufacturers (two social classes that were not in high social esteem in Smith's time) would be capable of creating a viable and prosperous society.

Smith, like most people of the Enlightenment, was inspired by Isaac Newton's (1642–1727) model of the universe, developed in 1687. In this model, the heavenly bodies display trajectories which at first sight seem chaotic. A more thorough examination, however, reveals that their trajectories are explained by the law of gravity. Smith applied this magnificent idea of a self-regulating universe to the social universe and tried to create a theory-generating concept that will explain the laws of motion of society. The first systematic effort was in his book the *Theory of Moral Sentiments*, where he developed the principle of sympathy and benevolence which binds selfish individuals together in a society with cohesion. Smith, however, soon realised that the moral sentiments, although undeniably strong in a civilised society, were not strong enough to be used as an acceptable theory-generating concept. After all, the impartiality of the 'impartial spectator' is questionable, when large vested interests are at stake. As a result, Smith, in his more mature work, the *Wealth of Nations*, downplayed moral forces and replaced them with economic forces and especially, competition (the self-interest of many market participants) which guides the self-interest of individuals like an 'invisible hand' in a way so as to establish social cohesion, exactly as the law of gravity holds together the planets in their trajectories.

Smith's opening chapter in the *Wealth of Nations* is on the division of labour indicating that, like the Physiocrats before him, the economic inquiry is placed in the sphere of production. In the division of labour, Smith discovered the key to rising productivity. The next step for him was to put together the division of labour

with the pursuance of self-interest as the motivation that characterises human action. The result was that, contrary to commonsense, society did not collapse to chaos by the unleashing of so many centrifugal forces, rather through the operation of free competition, social order was being established, which was also accompanied by economic growth that benefited all social classes. He anticipated the stationary economy, a central feature of all major economists after Smith, and he thought that the situation was not as gloomy as was envisaged by Malthus or Ricardo. For Smith, the worse that happens in the stationary state is that capitalists are forced to work in order to survive.

Smith, as other economists before him, argued for the need of a theory of value and distribution, since this is how one can measure the surplus and the wealth of a society in every particular period of time and also during different time periods. Inter-temporal comparisons are needed in order to ascertain whether wealth (and thus, social welfare) has increased or not. Smith tried to solve the riddle of value using three different theories, but none of these is without problems as we discussed in Sect. 3.4 earlier. Smith, however, had a grand plan to complete and the limitations of his theory of value could not prevent him from giving the whole picture of the way in which the economy works.

Smith's social philosophy emanates from his belief in the capacity of the market mechanism that when it works without government intervention, it leads the economy to the best possible performance. It is certain that Smith's views were in full support of the emerging market system and that he opposed feudalism and the absolutist state. Consequently, merchants and industrialists found the necessary theoretical justification that they needed to oppose government interventions and, at the same time, to create the necessary institutional framework for their own growth. This did not mean that Smith consciously defended the interests of the merchants or industrialists. There are many parts in the *Wealth of Nations*, where Smith was critical of them. For example, he pointed out:

> People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance, to raise price. (*WN*, p. 128)

It is interesting to note that while Smith is perhaps the most authentic representative and, at the same time, interpreter of the emerging capitalism, in reality, however, as Heilbroner (1981) notes, Smith was 'a slave to his system'. He argued in favour of private property, because without it, competition has no meaning, and competition is extremely important for the accumulation of capital, which in turn is responsible for the level of employment and the level of wages.

Governments, according to Smith, have not been characterised by responsibility with regard to their finances, as they tend to increase their expenditures by limiting the economy's investible product and undermining economic growth. In spite of all these, Smith does not oppose all government activities. For instance, he argues for the provision of education and health services by the government because the private sector motivated by profit may not provide these goods to the required extent.

Smith, in general, argues that the role of the government is to promote competition. He thought, however, that there are three major areas of exclusive government responsibility. These are:

(1) Provision of national defence
(2) Provision of justice
(3) Provision of the necessary infrastructure

Smith pointed out that the government should also institute laws that enforce the performance of contracts. Money supply is an area that must not be left to private banking and the government should apply some control. He also suggested that under certain circumstances, some regulation of the interest rate is needed, and favoured laws that provide copyrights and patents in an effort to encourage innovative activity.

State's activities, in Smith's analysis, are necessary, nevertheless they continue to be unproductive and this is the reason why he thought that they should be kept to a minimum. Smith attributed special importance to the distinction of productive (or surplus yielding) and unproductive (surplus-consuming) labour activity. Smith simply wanted to limit the State's activities because their expansion interfered with the productive activities of the private sector.

Although a lot has been written in books of public finance about Smith's cannons of taxation, we know, however, that the cannons are not so original. They can be found; for example, in the writings of William Petty and also of James Stewart. What we think as original is Smith's analysis of taxation and especially, his treatment of the question of public debt. It seems that public finance theorists may find many interesting, albeit neglected, ideas that perhaps will help to shed new light on some of the current debates and issues. For example, in Smith's treatment of public debt, we discover that taxes and public loans are not equivalent ways of financing government expenditures, a difference that emanates from Smith's distinction of labour activity in productive and unproductive. The idea is that taxes diminish primarily households' consumption expenditure and secondarily potential investment, whereas government loans diminish savings (=investment). According to Smith, although various government activities are absolutely necessary; nevertheless, they diminish the economy's investible product, and, therefore, society's growth potential.

Questions for Thought and Discussion

1. Discuss Smith's theory of morality as it is presented in the *Theory of Moral Sentiments*.
2. Discuss the notion of order and stability of the market system according to the *Theory of Moral Sentiments* and the *Wealth of Nations*.
3. How is selfishness defeated in *The Theory of Moral Sentiments* and in the *Wealth of Nations*? Explain.
4. Present Smith's theory of value and distribution as well as its limitations.
5. Why is the notion of natural price so important for the development of economics as a scientific discipline?
6. Discuss the difference between the Smithian law of supply and demand and the neoclassical supply and demand theory.

7. Is there a growth model in the *Wealth of Nations?* Discuss.
8. Why does the rate of profit fall? What is the usual explanation? What is the explanation offered in the text?
9. What is productive and unproductive labour in Adam Smith? Is this distinction useful? Explain.
10. Present Smith's statements of taxation. In what ways (if any) do these statements relate to the theoretical parts of his book?
11. Are taxation and public borrowing equivalent forms of financing of government expenditures? Explain.

Notes for Further Reading

The quotations of the text are from the Modern Library edition of the *Wealth of Nations* forwarded by Cannan. This, however, is not the only edition of the *Wealth of Nations*; there is also an edition by Penguin which, however, contains only Books I and II, from the five books of the *Wealth of Nations*. This edition is recommended only for its introduction by Skinner. Finally, there is *The Essential Adam Smith* (1986) which contains carefully selected parts of all the works of Adam Smith starting from his early writings (*i.e.*, *The History of Astronomy*; and the *Lectures on Jurisprudence*) to the *Theory of Moral Sentiments* and finally *The Wealth of Nations*. Each of the books is forwarded by Robert Heilbroner. For the reader, who is interested in the full texts, there is the *Glasgow Edition of the Works and Correspondence of Adam Smith*. Oxford: Oxford University Press.

One issue that comes to the fore is the originality of Smith's writings, where there are conflicting views that vary from 'complete independence' to 'he simply expressed ideas that were around for years.' For example, Schumpeter (1954, p. 184) notes that the '*Wealth of Nations* contained no really novel ideas and [. . .] it cannot rank with Newton's *Principia* or Darwin's *Origin* as an intellectual achievement'. In fact, many of the central ideas of the *Wealth of Nations* can be traced to authors before Smith. For example, the labour theory of value can be found, at least in part, in William Petty, who also wrote on the effects of the division of labour; the maxims of taxation can be found in James Steuart, and so forth. Marx, on the other hand, seems to provide us with a rather fair account on this issue which is worth quoting: 'The Scottish proverb that if one has gained a little it is often easy to gain much, but the difficulty is to gain little, has been applied by Adam Smith to intellectual wealth as well, and with meticulous care, he accordingly keeps the sources secret to which he turns indeed into much' (Marx 1970, pp. 167–168).

On Smith's theory of value, the interested reader may start with Heilbroner's (1983) article and then continue with the relevant chapters in Rubin ([1929], 1979), Dobb (1973), but also Marx's *Theories of Surplus Value* ([1861], 1963). If the reader is interested in a challenging reinterpretation of Smith's theory of

(continued)

value, Hollander's (1973) book is a must. The author presents Smith's theory of value in a neoclassical framework and argues that in Smith, we find a general equilibrium perspective of the economy. No one would deny that in Smith's analysis, there exists general interdependence of the various economic sectors, an idea that permeates Smith's work and also that of the Physiocrats. The trouble, however, starts when one tries to attribute to Smith a utilitarian theory of value. Hollander's contribution is that he further elaborates the earlier efforts by Marshall to put initially Smith, and later Ricardo, in a neoclassical 'dress'.

On the issue of natural and market price, the interpretation of Garegnani (1983, 1998) and Eatwell (1977, 1983) are extremely important. The interested reader perhaps could contrast this view of the effectual demand as being a single point on the price–quantity plain with those that one finds in the usual books on the history of economic thought (Blaug 1983; Ladreth and Collander 1992; Ekelund and Hebert 1983, *inter alia*). No matter how one would represent demand in Smith (as a point or as a curve), the essential thing is that the distinction between market and natural price constitutes the scientific foundation of modern economic analysis.

The discussion on productive and unproductive labour is usually not well presented in the history of economic texts; if someone wants a good treatment, then he must look in books such as the classic one by Coontz (1966). However, this issue is also discussed adequately by Rubin (1979), who presumably reflects on Marx's views as presented in the *Theories of Surplus Value* (1963). The issue of productive–unproductive labour appears in a somewhat disguised form; for instance, in Baumol's (1967) distinction of goods vs. services, in Bacon and Eltis (1976) marketed vs. non-marketed activities and more recently, with the notion of 'rent seeking' (see Bhagwati 1982). The natural extension of this discussion is in the national income accounts. The current dissatisfaction with the national income accounts and the frequent redefinition of the variables are a reflection of issues involved in a very old tradition in which there was no debate.

On the issue of growth, we must stress once again that in Smith, there is no proper growth model. There are various remarks which some authors, starting with Löwe (1975) and continuing with Heilbroner (1982, 1975) and Eltis (1975, 1984), tried to make a model out of these. Other efforts include a mathematical presentation by Samuelson (1977); also, a summary of various views along with a connection of Smith's ideas with current developments in the theory of economic growth can be found in Reid (1989). On the issue of the falling rate of profit and the stationary state, we follow an interpretation of Smith which makes him a progenitor of Marx's theory, something that Marx did not notice. It seems that Ricardo's interpretation, in combination with Smith's confusing way of writing, established the idea of excessive competition as the reason for the falling rate of profit in Smith.

(continued)

On the question of taxation and public debt, an issue of great significance in Smith after all this is the second focus of political economy '[. . .] first, to provide a plentiful revenue or subsistence for the people [. . .] and secondly, to supply the state or commonwealth with a revenue sufficient for the public services' (*WN*, p. 397?). It seems that taxation, as an integral part of the theory of value and distribution, has been downplayed by historians of economic thought. According to Schumpeter, 'The fifth and longest (book) – taking 28.6 percent of total space – is nearly self-contained treatise on Public Finance and was to become and to remain the basis of all the nineteenth-century treatises on the subject [. . .]. The length of the book is due to the masses of material it contains: its treatment of public expenditure, revenue, and debts is primarily historical. The theory is inadequate, and does not reach much below the surface' (Schumpeter 1954, pp. 184–186). Blaug (1978, p. 61) has expressed a similar view, whereas, Rubin (1979, p. 166) claimed that only the first two books of the *Wealth of Nations* are important, from a theoretical perspective, the rest of the books are only of historical interest. On closer examination, however, it seems that Smith has much to offer in this area and this is why we presented his views in length in Sects. 3.9 and 3.10. For the reader who is interested in this relatively unexplored area, we recommend the book by Seligman (1969) and also the articles by Lynn (1977) and Tullio (1987), among others.

Appendix: The Labour Commanded Theory of Value

Smith's labour commanded theory of value can be presented in terms of a linear model of production, where for reasons of simplicity of presentation, we assume away the landed property.[20] The value of commodities can be written:

$$\lambda = a_o + \lambda \mathrm{A} \text{ or } \lambda = a_o (I - A)^{-1}$$

where λ the row vector of labour values, a_o is also the row vector of quantities of labour required to produce of unit of good and $\lambda\varsigma$ is the value of inputs used, where A is the well-known square matrix of technological coefficients (see Chaps. 2 and 8). It is important to point out that Smith was interested in a numeraire commodity and on the basis of which to express the value of all commodities. We know that in a

[20]The inclusion of landed property in a linear system of production is difficult and the difficulty increases when we come to Smith who does not really have a single theory of rent. Since in one instance he argues that "High or low wages and profit, are the causes of high or low price; high or low rent is the effect of it. It is because high or low wages and profit must be paid, in order to bring a particular commodity to market, that its price is high or low" (*WN*, p. 146). How does this view reconcile with the adding up theory of value where it is assumed that absolute rent is a question for which the researcher cannot find definitive answers in Smith's text.

given moment in time any given commodity could be used as a numeraire. However, if we need inter-temporal comparisons of wealth, then a numeraire of this type must be of constant value. Smith thought that the labour time concentrates this property of constancy better than any other commodity and this is the reason why he expressed the value of commodities in terms of wages or labour commanded. More specifically, if by p we symbolise the row vector of equilibrium prices of commodities, we will have:

$$p = wa_o + pA + rpA$$

where r is the uniform profit rate and w is the uniform wage rate and in terms of labour command, we will have

$$\frac{p}{w} = a_o + \frac{p}{w}A + r\frac{p}{w}A \text{ and } \frac{p}{w} = \underbrace{a_o(I-A)^{-1}}_{\lambda=\text{labour embodied}} + \underbrace{r\frac{P}{w}A(I-A)^{-1}}_{\text{profits}}$$

From the above equation, it follows that the labour commanded is greater than the labour embodied, $\lambda = a_o(I-A)^{-1}$) precisely because of the presence of profit income $r\frac{p}{w}A(I-A)^{-1} \geqslant 0$. Finally, the idea of labour commanded is also in Keynes with his idea of money-wage rigidity in the downward direction and the idea of expressing the nominal values of the variables in terms of wage units.

Chapter 4
David Ricardo's *Principles of Political Economy*

I know that I shall soon be stopped by the word price.

(*Works*, vol. VI, p. 348)

Political Economy you think is an enquiry into the nature and causes of wealth – I think it should rather be called an enquiry into the laws which determine the division of the produce of industry among the classes who concur in its formulation. No law can be laid down respecting quantity, but a tolerably correct one can be laid down respecting proportions. Every day I am more satisfied that the former enquiry is vain and delusive, and the latter only the true objects of the science.

(Ricardo letter to Malthus, 9 October 1820, The Works and Correspondence of David Ricardo, *Works*, vol. VIII, pp. 278–279)

4.1 Introduction

David Ricardo (1772–1823) was trained, from the age of 14, by his father (a Jew stockbroker, who had migrated from Amsterdam to London) in the 'secrets' of stockbrokerage. Although Ricardo did not have a formal university education, he had a keen interest in mathematics, chemistry, mineralogy, geology and, late in his life, political economy on which he focused his attention when he decided to retire. Ricardo managed to earn a lot of money through his activities in the stock market and, as Heilbroner (1981) notes, became 'the richest of all the economists'. His early publications included The High Price of Bullion (1810) and, about five years later, An Essay on the Influence of a Low Price of Corn on the Profits of Stock, which contained his, still, controversial corn model. In the same year, he set out to write a book on issues of political economy, and in 1816, he had completed a draft of the first seven theoretical chapters. Ricardo was hesitant to publish in general, so he sent his manuscript to his friend James Mill (1773–1836) and asked for his comments and advice, and J. Mill encouraged him to proceed with the publication. The rest of the book was completed in a short period of time and was published as *On the Principles of Political Economy and Taxation* (henceforth, *Principles*) in 1817.

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_4, © Springer-Verlag Berlin Heidelberg 2010

A second edition was out in 1818, and finally, the third revised edition was published in 1821, the major change in the third edition being the publication of the chapter on 'the effects of machinery on employment.' From the 32 chapters of the book, the first six chapters refer to the theory of value and distribution, while the seventh chapter deals with international trade. The remainder of the book is literally an elaboration and further extension of these seven chapters, the first one of which, concerning the theory of value, is the longest and the one that has attracted the attention of most researchers. All Ricardo's works have been collected and published in eleven volumes titled *The Works and Correspondence of David Ricardo* (henceforth *Works*) and edited by Pierro Sraffa, with the collaboration of Maurice Dobb. The hallmark of Ricardo is his analytical abilities and capacity to reduce difficult economic problems to their fundamental components and to treat them in terms of numerical models on the basis of which sweeping generalisations are drawn. In this sense, Ricardo can be literally considered as the first model builder.

In what follows, we focus mainly on Ricardo's views on the theory of value, the long-run prospects of the profit rate and the associated stationary state, the theory of international trade, the question on machinery and, finally, his theory of taxation and public debt.

4.2 The Theory of (Exchange) Value

Ricardo in his *Principles* states in a straightforward way that his purpose is '[t]o determine the laws which regulate [...] distribution [. . .] the principle problem of political economy' (*Works*, vol. I, p. 5). However, the understanding of the 'laws of distribution' requires the prior understanding of the theory of relative prices:

> Before my readers can understand the proof I mean to offer, they must understand the theory of currency and of price [. . .]. If I could overcome the obstacles in the way of giving a clear insight into the origin and the law of relative or exchangeable value I should have gained half the battle. (*Works*, vol. I, pp. xiv–xv)

Ricardo undertook the Herculean task of understanding the determinants of the relative prices of commodities (the theory of value) until the very end of his life but without complete success. While he advanced the theory of value more than any other economist of his time, it was, however, impossible to supersede a series of riddles that continue even today to occupy the focus of the economists' attention.

Ricardo knew well that market prices are, prima facie, determined by the forces of supply and demand, but he observed that, over time, these market prices fluctuate ceaselessly, with these fluctuations being regulated by something more fundamental than supply and demand, that is, the 'principle of equal profitability'. Thus, if in an industry, the rate of profit is above the economy's average, the inflow of capital would increase the supply of this industry relative to its demand, thereby reducing the market price to a level that would restore the average profitability. By contrast, if the rate of profit of an industry is below the economy's average, then the outflow

of capital will decrease the supply relative to demand, so the price increases to a point that restores the economy's average profit rate. The resulting prices are called 'natural prices', and Ricardo's major contribution to the theory of value is his insistence on the idea that the labour time is the central determining factor of the movement of natural prices.

The first chapter of the *Principles* headed 'On Value' begins with a discussion with regard to Smith's distinction between use value and exchange value. Ricardo argues that the use value of a commodity is a prerequisite for exchange, yet the exchange value of a commodity cannot be assessed in terms of its usefulness:

> Possessing utility, commodities derive their exchangeable value from two sources: from their scarcity, and from the quantity of labour required to obtain them. (*Works*, vol. I, p. 6)

For Ricardo, scarcity is important only in the case of non-reproducible goods, such as works of art, rare coins, rare books, *etc*. The value of these goods

> is wholly independent of the quantity of labour originally necessary to produce them, and varies with the varying wealth and inclinations of those who are desirous to possess them. (*Works*, vol. I, p. 12)

These goods whose relative prices are determined by subjective forces are just a trifling percentage of the totality of goods. The large majority of goods are reproducible, and, with regard to them, Ricardo accepts Smith's labour theory of value, according to which the relative prices of goods are approximately proportional to the relative quantities of labour that were spent for their production.

> The value of a commodity, or the quantity of any other commodity for which it will exchange, depends on the relative quantity of labour which is necessary for its production, and not on the greater or less compensation which is paid for that labour. (*Works*, vol. I, p. 11)

Formally, this can be written as $P_i/P_j = f(L_i/L_j)$, where P_i/P_j are the relative prices of goods i and j, whereas L_i and L_j are the respective quantities of labour that were spent for the production of these goods. Ricardo then argues that if, for some reason, the ratio P_i/P_j changes, then it is not enough to know that this is due to the change in the ratio of L_i/L_j, because a change of the latter can come from various movements. For example, supposing that the ratio P_i/P_j rises, this might be due to a rise in L_i while L_j remained constant, or due to a fall in L_j while L_i remained constant or because L_i increased at a faster rate than L_j and so forth.

In such aforementioned situations, it would be very helpful if we could find a commodity whose production would always require the same quantity of labour and whose price would not change with every change in the income distribution. If we discover such a commodity, such an 'invariable measure of value', we can use it as a numéraire. Then, we would be able to identify the source of each change in the relative prices in the quantity of labour contained in the production of the numéraire commodity.[1] Ricardo devoted the rest of his life to the discovery of such a

[1]The invariable measure of value could also be used as an index for the intertemporal estimation of the wealth, and, therefore, the welfare of a society.

commodity (practical or analytical), however, without success.[2] He tried various possibilities, but none of these gave an acceptable solution, because the production of each of the commodities with which he was experimenting required a different quantity of labour over time. Although unsuccessful in his efforts, Ricardo supposed that gold could at least partially fulfil such a role.

For Ricardo, the value of gold was determined in exactly the same manner as that of the other commodities, that is, from the labour time required for their production. Using gold as an approximate invariable measure of value, he also provided himself with a means to connect the theory of value with the theory of money. Thus, while the labour time involved in the production of gold was supposed to be approximately constant, it did not prevent Ricardo from engaging in discussions concerning the effects of the supply of money (that is, gold) on the general price level. This separation between the real and monetary effects is of critical importance to the interpretation of Ricardo's theory of value, distribution, international trade and taxation.

In spite of the lack of a perfect invariable measure of value, Ricardo accepted the fact that the principle according to which the exchange ratios of products are regulated by relative labour times expended in their production is correct not only in 'the rude and early stage of society' but also in capitalism. The difference is that in capitalism, the previously correctly stated labour theory of relative prices needs several qualifications and modifications. These modifications relate to the presence of factors such as capital–labour ratios and changes in income distribution. Ricardo's fundamental thesis, however, is that the relative prices of products, subject to reproduction, are determined by the relative quantity of labour expended on their production. The quality of the expended labour is independent of the subjective evaluations of individual producers; this is an issue that is resolved through the market mechanism. Moreover, Ricardo argued that the value of a product is determined not only by the direct labour expended on its production but also by the indirect labour expended on its means of production. He notes:

> Not only the labour applied immediately to commodities affect their value, but also the labour which is bestowed on the implements, tools, and buildings, with such labour is assisted (*Works*, vol. I, p. 23).

In other words, the exchange ratios of commodities are determined by their respective labour times with fixed capital only transferring its exchange value gradually and through its depreciation.[3]

---

[2]Marx called these efforts "blind alley" precisely because they could not reach any satisfactory answer

[3]"Suppose [Ricardo notes] that the same quantity of labour was necessary to make both weapons, but they were of very unequal durability; of the durable implement only a small portion of its value would be transferred to the commodity" (*Works*, vol. I, p. 23). However, in the numerical examples that follow, Ricardo assumes, for simplicity purposes, that fixed capital does not depreciate.

4.2.1 Modifications due to Unequal Capital–Labour Ratios

Let us now focus on Ricardo's thesis when it is applied to full-fledged capitalism where: 'The principle that the quantity of labour bestowed on the production of commodities regulates their relative value, considerably modified by the employment of machinery and other fixed capital' (*Works*, vol. I, p. 30). Ricardo argues that capital is transferred wholly or partly (through depreciation) into the final product, and the owner of capital gets a profit. Under these circumstances, the presence of capital and of the rate of profit affect the exchange ratio of commodities only marginally. For this purpose, he constructs a numerical example of two industries ('trades'), the first producing cotton and the second corn; each of the trades employs the same quantity of labour, a hundred men (working the same number of hours) at a wage rate of £50 per year (see Table 4.1). He further supposes that the cotton trade uses a machine worth of £5,500 while the corn trade uses only direct labour. The rate of profit (r) is assumed at 10%. Thus, we can construct Table 4.1 below.

Where K is the value of the machine (indirect labour),[4] w the wage rate which multiplied by the employment L gives the wage bill $W = wL$, Π is the amount of profit, P_{ct} and P_{cn} are the prices of cotton and corn trades, respectively. From the above Table, we get that the relative prices or exchange values of commodities $P_{ct}/P_{cn} =$ £6,050/£5,500 = 1.10 are close to relative direct labour times $L_{ct}/L_{cn} = 100/100 = 1$. The deviation stems from the differences in the K/W ratios and since the wage rate is uniform, we can say that the dependence is on the capital–labour ratio.

The following price equations can be formed on the basis of the above numerical example: $P_{ct} = W + r\,(W + K)$ and $P_{cn} = W + rW$ and the relative prices will be:

$$\frac{P_{ct}}{P_{cn}} = \frac{wL_{ct} + r(wL_{ct} + K_{ct})}{wL_{cn} + rwL_{cn}} = \frac{L_{ct}}{L_{cn}} \frac{1 + r(1 + K_{ct}/wL_{ct})}{(1 + r)}$$

Clearly, the relative prices of production are affected by the presence of capital and rate of profit, but only in a limited way. Ricardo's example is reasonable, since he demonstrates his thesis under unfavourable circumstances. For example, instead of taking the two trades as having K/W ratios close to each other, he demonstrates that even where there are large differences between capital/wage (or labour) ratios, the deviations between relative prices and labour times are still relatively small

Table 4.1 Ricardo's numerical example

	K	$W = wL$	K/W	$\Pi = r(K + wL)$	$P = W + \Pi$
Cotton industry:	£5,500	£5,000	1.1	£1,050	£6,050
Corn industry:	0	£5,000	0	£500	£5,500

[4]Fixed capital in Ricardo's numerical examples is assumed that does not depreciate and also there are no materials used up; in short, there is no "constant capital" in Marx's sense (see chapter 4).

when K/W differences are accounted for. He examines a rather extreme case where the cotton trade uses accumulated labour, yielding a positive capital–wage bill ratio, $K_{ct}/W_{ct} = 1.1$ or $K/L_{ct} = 5.500/100 = 55$£ per labourer, and the corn trade uses no accumulated labour, yielding a zero capital–wage bill (or capital labour) ratio, and he demonstrates that the effect on relative prices is not only limited but also predictable. More specifically, the effect on relative prices is proportional to the differences in the capital–wage bill ratios of the two trades. It is important to point out that the rate of profit exerts an influence on relative prices but this diminishes with the passage of time because of the falling tendency of the rate of profit.[5]

Ricardo's example can be straightforwardly generalised. Supposing two different industries i and j with different amounts of accumulated capital (K_i and K_j) and wages (wL_i and wL_j), we get:

$$\frac{P_i}{P_j} = \frac{wL_i + r(wL_i + K_i)}{wL_j + r(wL_j + K_j)} = \left[\frac{L_i}{L_j}\right]\left[\frac{1 + r(1 + K_i/wL_i)}{1 + r\left(1 + K_j/wL_j\right)}\right]$$

Clearly, if there are any differences between relative prices and relative labour times, these differences depend on the capital–wage bill ratios as follows:

$$sign\left(\frac{P_i}{P_j} - \frac{L_i}{L_j}\right) = sign\left(\frac{K_i}{wL_i} - \frac{K_j}{wL_j}\right)$$

4.2.2 Modifications due to Changes in Distribution

Ricardo continues his analysis by examining the effects of changes in the distributive variables on relative prices. In so doing, Ricardo shows first that his theory of value remains valid; and second, he discredits Smith's cost of production theory of value. Supposing that wages rise across sectors and further supposing a uniform capital–labour ratio across the economy, it follows that profits fall proportionally across sectors in the economy and there is no reason for relative prices to change. However, in the usual case where capital intensities differ between sectors, an increase in wage diminishes profits disproportionally, that is, labour-intensive (capital-intensive) industries will experience (all else equal) a proportionally higher (lower) reduction in their rate of profit. The resulting unequal rates of profit between the two sectors are only ephemeral, since the acceleration of accumulation in the one sector and the deceleration of accumulation in the other sector will restore a new, albeit lower, equilibrium rate of profit through changes in relative prices. In the labour-intensive industries, prices have to rise so as to raise the rate of profit to a point that is equalised to the new lower uniform rate of profit. In the capital-intensive

[5]Thus, the effect of different capital-labour ratios on relative prices is secondary, whereas the labour expended is the principal term exerting most of the influence.

industries, prices have to fall to the point where their rate of profit becomes equal to the (lower) average one. Thus, individual prices twist due to changes in distribution in a systematic way, but for the average price level, there is no reason to change. This is a remarkable result to support in a period where Smith's cost of production theory of value was dominant.

In terms of the numerical example of Table 4.1, Ricardo argues that a rising wage results in a fall in the usual rate of profit. Hence, P_{cn} will remain the same, since we only have a redistribution between wages and profit, and their sum must be the same. In the case of P_{ct}, the manufacturer estimates a profit of his machine equal to 9% instead of 10%. The final price, therefore, will be $P_{ct} = £5{,}995$, and $P_{ct}/P_{cn} = 1.09$. We observe, firstly that a fall in the rate of profit by only 1% made the relative prices of production to come even closer to their respective labour times. Secondly, a significant reduction in profits by £155 (=10% × £1,550 total profits) leads to a change in relative prices by only 1%. Ricardo, after doing a kind of sensitivity analysis, concludes that 'The greater effects which could be produced on the relative prices of these goods from a rise of wages, could not exceed 6 or 7%; for profits could not, probably, under any circumstances, admit of a greater general and permanent depression than to that amount' (*Works*, vol. I, p. 36). If we differentiate the relative prices with respect to the rate of profit and also take into account that $w = w(r)$, $\mathrm{d}w/\mathrm{d}r < 0$ and $\mathrm{d}K/\mathrm{d}r = 0$, we get:

$$\frac{d}{dr}\left(\frac{P_i}{P_j}\right) = \frac{d}{dr}\left[\frac{wL_i + r(wL_i + K_i)}{wL_j + r(wL_j + K_j)}\right] = \left(r\frac{dw}{dr} - w + r^2\frac{dw}{dr}\right)\frac{L_iK_j - K_iL_j}{(wL_j + rwL_j + rK_j)^2}$$

Since the first term is always negative and the denominator of the second term is always positive, it follows that the sign of the aforementioned derivative depends on the term $L_iK_j - K_iL_j$, or in symbolic terms; we have

$$\operatorname{sign}\left[\frac{d}{dr}\left(\frac{P_i}{P_j}\right)\right] = \operatorname{sign}\left(L_iK_j - L_jK_i\right) = \operatorname{sign}\left(\frac{K_j}{L_j} - \frac{K_i}{L_i}\right)L_iL_j.$$

4.2.3 Modifications due to Unequal Turnover Times

Ricardo continues with the modifications of his theory of value by introducing in the analysis the problem of differences in the time that elapses between the beginning and the completion of the production process when the product is available for sale. Ricardo sets up an example of two industries i and j with a uniform rate of profit ($r = 10\%$). In industry i, we have an investment of £2,000 (40 workers × £50 annual wage) where £1,000 is invested in the first year and the rest in the second year. At the end of the second year, the exchange value of the commodity must be £2,310. This figure is estimated as follows:

$$P_i = \underbrace{£1{,}000^*(1 + 0.10)}_{\text{First year}} + \underbrace{£1{,}000^*(1 + 0.10)^2}_{\text{Second year}} = £2{,}310$$

The idea is that the estimation of the price in the second year must account for the fact that the producer should have estimated a 10% profit rate for his capital in the first year that he did not invest and 10% for the investment in the current year. By contrast, in industry *j*, all the money is invested in the first year, so the exchange value of the commodities will be

$$P_j = £2,000 \times (1 + 0.10) = £2,200$$

We observe that the same quantities of labour give rise to two different prices. The deviation arises precisely because of the assumption of a given rate of profit and also of different turnover times, which make relative prices differ from relative labour times.[6]

> This case appears to differ from the last, but is, in fact the same. In both cases the superior price of one commodity is owing to the greater length of time which must elapse before it can be brought to market. [. . .] one commodity is more valuable than the other, although no more labour was employed on its production. The difference in value arises [. . .] from the profits being accumulated as capital, and is only a just compensation for the time that the profits were withheld. (*Works*, vol. I, p. 37)

It follows that the inclusion of turnover time modifies the initial principle. We observe, once again, that deviations between relative prices from relative labour times, despite the significant differences in turnover times, are negligible. In terms of Ricardo's numerical example, the deviation will be £2,300/£2,200 = 1.05, that is to say, a 5% difference (Tsoulfidis 2005).

Overall, one wonders whether Ricardo was right or wrong in his claims. To our view, it seems that Ricardo's intuition was on the right direction, however, it was impossible to subject his propositions to any empirical test. Nowadays, as economists not only have more data but also can use econometric techniques and input–output models, the empirical evidence is in favour of Ricardo's intuition. Long-run prices are closely related to labour time contained in commodities. The available econometric studies on prices lend overwhelming support to Ricardo's theory.[7] Similarly, the input–output studies show that the relative prices of production and the relative labour times are surprisingly close to each other and that both are close to market prices (Ochoa 1989; Shaikh 1998; Tsoulfidis and Maniatis 2002; Tsoulfidis 2008a; Mariolis and Tsoulfidis 2009, *inter alia*).

[6]This assumption led Marshall to the conclusion that Ricardo like neoclassical economists accepts a cost of production theory of value, which includes 'time or waiting as well as labour' (Marshall, 1920, p. 672)

[7]Most of the econometric studies that test the explanatory power of relative labour times on relative prices at the industry level have been conveniently summarised in Semmler (1984, ch. 3).

4.3 The Rate of Profit in the Long Run and the Stationary State

Ricardo, in his analysis of the evolution of profitability, accepts Smith's view that '[t]he natural tendency of profits is to fall' (*Works*, vol. I, p. 120). He disagreed, however, with an argument that he attributed to Smith, according to which competition leads to higher prices of resources, which in turn increase cost, and, therefore, in the long run, it leads to a reduction in the profit rate. For Ricardo, competition can only establish the tendential equalisation of interindustry profit rates to a general one, but it cannot explain the long-run tendency of the general rate of profit. The tendential equalisation of the interindustry profit rates is independent of the falling tendency of the economy's general rate of profit.

Ricardo, like Smith, starts his analysis by considering the three social classes, that is, the workers, the capitalists and the landlords. Workers get their socially determined subsistence wage, which must be understood as a fixed bundle of goods required for the normal reproduction of themselves and their families according to given social standards. If the current wage exceeds the subsistence wage, then the control mechanism, which is called population law – the details of its operation have been discussed in the chapter on Adam Smith – is initiated and restores the current wage to its subsistence level.

Capitalists, according to Ricardo, are driven by the insatiable desire to save their profits in order to invest and expand their production basis hiring more workers, and so forth. Capitalists, in other words, make profits, which they spent mainly on the expansion of their business. Excess profits that might come, for example, through an innovation of any sort are gradually eliminated through competition. At the same time, profits can be depressed from the workers' demands, because profits depend on wages; hence, wages are assumed to be at least at their subsistence level.

Until now, we see that Ricardo's analysis does not differ essentially from the analysis of Adam Smith. The essential differences appear when the role of landlords and rents is taken into account. Rent for Ricardo does not constitute an ordinary reward for the services of land as, for example, is the case of interest as a reward for the services of capital, or of wage as a reward for the services of labour. Rent is viewed as a sort of income that comes from the fact that all the parcels of land do not have the same fertility, so differential fertility (cost) of land gives rise to differential rent and there is no absolute rent. This is an idea that Ricardo borrowed from his close friend Thomas Malthus (1815); however, the idea of zero absolute rent had been already developed in an embryonic form by James Anderson (1739–1808).

Let us suppose two adjacent parcels of land of the same size. The first parcel of land happens to be more fertile than the second. Suppose that in the first parcel, ten workers with the necessary equipment fertiliser, *etc*. are employed and that the output that they produce amounts to 200 tons of wheat. Suppose also that in the second, less fertile parcel of land, the same production technique is employed (same number of workers, equipment, fertiliser, *etc*.), but the output that the ten workers produce is only 150 tonnes of wheat. It follows, therefore, that the unit cost of the first parcel of land would be much lower than the unit cost of production of the second parcel of land.

Ricardo argues that differential rent springs from this cost difference. The reason is that if the demand for wheat is increased so that its selling price is so high such as to encourage the cultivation of the second parcel of land, then the first parcel of land gives rent. If the normal price per tonne of wheat, that is, the price that includes normal profits is say £10, then the revenues of the first parcel of land will be (200) × (£10) = £2,000, while the revenue from the second parcel of land is only (150) × (£10) = £1,500. With the total cost (normal profits are included) equal to £1,500, it follows that the first parcel of land secures the normal profit together with some rental income, while the second parcel of land with the price of £10 simply secures normal profits for the cultivator (capitalist) and no rent for the landlord. The second parcel of land is considered marginal in the sense that it is brought to cultivation last and as such does not pay rent. It is important to note that in Ricardo there is no absolute rent. Consequently, as the demand for food increases and less fertile parcels of land (or alternatively, lands of higher unit cost) are brought successively into cultivation.[8] Since the price of agricultural products is determined by the marginal land,[9] it follows that the production cost of agricultural products increases over time.[10] If we integrate the theory of rent in Ricardo's overall analysis, then all its repercussions on the future of the evolution of profitability and the capitalist system as a whole are revealed.

We start with the capitalists, given their insatiable desire to save and, at the same time, to invest their revenues from their enterprise. As capitalists buy additional machinery and expand output, they set into motion a chain of economic consequences: the first is the diminishing returns to land, and the second is the rising wages. More specifically, Ricardo pointed out that the increase in population increases the demand for consumer goods and thus, less and less fertile parcels of land are brought into cultivation in order to satisfy the rising demand for food. The prices of consumer goods increase because of the law of diminishing returns to land and, given the basket of goods that a worker and his family normally consume (*i.e.*, the subsistence wage), the wage bill for capitalists increases, and, therefore, profits and the rate of profit fall. Ricardo notes:

> however abundant capital may become, there is no other adequate reason for a fall in profit but a rise of wages, and further it may be added, that the only adequate and permanent cause for the rise of wages is the increasing difficulty of providing food and necessaries for the increasing number of workmen. (*Works*, vol. I, p. 296)

[8]We say higher cost, because a more fertile land might be located at a distant place and so the transportation expenses increases the unit cost of production.

[9]That is to say, "last land taken into cultivation" (*Works*, II, pp. 276–277).

[10]Alternatively, we consider that the total cost of production (that contains normal profits) is £1,500, consequently, the second parcel of land for the price wheat at £10 simply generates revenues that cover the cost of production which includes also the normal profits of the cultivators and so there is no money left to pay any rent. After all, rent is a residual income. If the demand for agricultural products increases and so the price increases to £11, it becomes possible to cultivate a third parcel of less fertile land, which now becomes the new marginal land, while the second parcel of land now pays rent, albeit lower than the first parcel of land, and so forth.

The wage increase sets in motion the control mechanism which is called population law, according to which the increase in population and in labour supply lead wages back to their subsistence level. The increase in population leads to an increase in the demand for food, so less fertile (productive) parcels of land are brought into cultivation. The result would be that the price of food would rise and with it the total rent and wages. The increase in wages is attributed to the increase in the price of the goods that comprise the workers' subsistence wage. As a result, if workers want to purchase the same bundle of goods, their wage must rise. It is important to stress at this point that rising wages do not necessarily imply improvement of the standard of living of workers, simply because workers need more money to be able to purchase the same (but more expensive) basket of goods that constitute their (socially determined) subsistence wage. In reality, however, it seems that workers' standard of living deteriorates as the profit rate tends to zero. Ricardo notes:

> The fate of the labourer will be less happy; he will receive more money wages, it is true, but his corn wages will be reduced. (*Works*, vol. I, p. 102)

Hence, Ricardo describes a very interesting phenomenon where real wages (or the cost of labour) increase from the point of view of the capitalists, who in order to set in motion the same number of workers must pay higher wages; whereas, from the point of view of the workers, who find that although their money wage has increased, their standard of living remains, at best, approximately the same. Meanwhile, the only increase observed concerns the rent of the landlords, who while sitting idle, they paradoxically enough see their wealth only to increase.

In this analysis, the capitalist is regarded as the tragic person, since he saves his profits, often living in difficult conditions, to be able to carry out his accumulation function, and what he achieves at the end is to see his wages increase so much as to eat up his profits, while the landlord, who was passive throughout this period, manages to increase his rents year after year. And the capitalist who is responsible, after all, for the activation of this chain of events, nevertheless finds his profits diminished from both the rising wages of workers and the rising rents of landlords. All these are the result of diminishing returns to agriculture, the given real wage and the capitalists drive to accumulate as an end in itself.

> The natural tendency of profits then is to fall; for in the progress of society and wealth, the additional quantity of food required is obtained by the sacrifice of more and more labour. (*Works*, vol. I, p. 111)

We observe that Ricardo's conclusions are different from those of Smith with respect to the progress of society. In Smith, we saw that all the social classes continuously improve their position with the passage of time.[11] In Ricardo, only the landlord's position really improves. The worker at best secures a (socially determined) subsistence level. If the wage happens to be above the subsistence level, this can only be temporary since the worker is bound to increase the size of his family,

[11]Hence, we do not account for Smith's "stationary state" which in any case is relevant only in the distant future.

the supply of labour increases and restores the current wage to its subsistence level, that is, the wage rate that ensures equilibrium in the labour market. The capitalist whose main function is to save his profits and, at the same time, to invest them, soon discovers that in the final analysis all his efforts were in vain, since the wages that he pays increase at the expense of his profits.

> [T]he very low rate of profit will have arrested all accumulation and almost the whole produce of the country, after paying the labourers, will be the property of the owners of land. [. . .] The farmer and manufacturer can no more live without profit than the labourer without wages. Their motive for accumulation will diminish with every diminution of profit. (*Works*, vol. I, pp. 120–122)

In conclusion, the long-term fall in the rate of profit discourages accumulation of capital. The growth of demand for labour will slow down, a development which workers cannot understand from the start, but after the passage of time. As a result, there is a lag in the adjustment of the growth of population, and, therefore, the growth of labour supply will exceed the growth of the demand for labour and the real wage will fall.

> Instead [. . .] of the money wages of labour falling, they would rise; but they would not rise sufficiently to enable the labourer to purchase as many comforts and necessaries as he did before the rise in the price of those commodities. [. . .] The condition of the labourer will generally decline. And that of the landlord will always be improved. (*Works*, vol. I, pp. 101–103)

Hence, two interrelated questions come to the fore, first is this process of the economy towards the stationary state reversible? And second, are there any counteracting forces? Clearly, if Ricardo's assumptions hold true, then the answer to the first question is negative, and the stationary state is inevitable. With regard to the second question, Ricardo pointed out that the stationary state could be postponed via the increase in productivity caused by technological progress. Another 'solution' is the reduction in wages via the reduction in the price of agricultural products. This is the reason why Ricardo opposed the famous 'Corn laws'. These laws instituted high tariffs for imported agricultural products. Ricardo argued that if the imported agricultural products were not taxed, they would become cheaper, and, therefore, would lead to lower wages without affecting the standard of living of workers.

From these predictions what is certain is that the prices of agricultural products do not increase over time, rather they fall relative to industrial products. What is certain, however, is that the rate of profit over a long period of time displays a falling tendency, though it is doubtful whether this has anything to do with Ricardo's analysis. In Fig. 4.1, we display the price index of wheat and the real daily wages of British workers.[12] We observe that the price index of wheat (WINDX) increased sharply during the Napoleonic wars, but then its tendency was falling, whereas the real wage (WR) on an average rises for the same time period. Ricardo's claim for the rising prices of agricultural products due to diminishing returns is really hard to sustain.

[12]The data come from Makridakis, Wheelwright and Hyndman (1998).

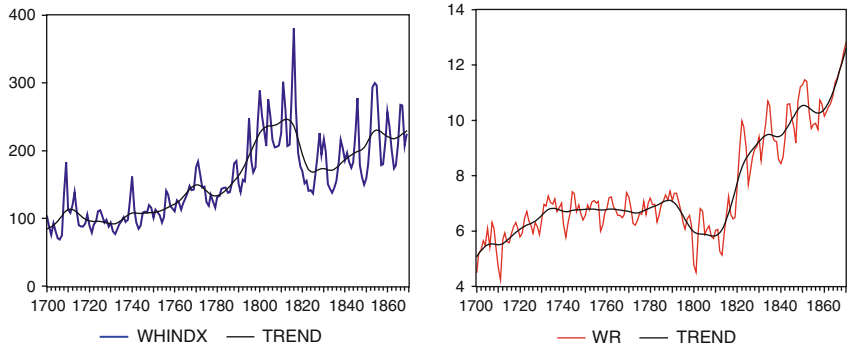

Fig. 4.1 The price index of wheat and the real daily wages of British workers

In retrospect, however, we find that some authors advocated the view that the stagnation of the 1970s and early 1980s occurred as a result of the high prices of oil. In fact, the production of oil, and the associated falling profitability, is a very good example where Ricardo's theory of differential rent could find a nice application.[13] The argument is that the growth of demand for oil led to the exploitation of less productive, and, therefore, higher cost oil wells (*e.g.*, USA, Mexico and the North Sea). The extraction of oil became an increasingly high cost activity as less and less productive oil wells were brought to production which, in combination with the presence of strong unions in the labour markets, preserved if not increased the real wage, and so, profits were gradually eroded, thereby leading to a fall in the rate of profit and the stagnation of the late 1960s that lasted until the early 1980s. Aspects of this scenario were also played out in the recent slowdown of the late 2000s, which may be explained along Ricardian lines where the rising prices of raw materials and food, in combination with inflexible real wages, led to a slowdown in economic activity, which resembles that of the 1970s and to some extent that of the 1930s.

4.4 The Principle of Comparative Advantage

Ricardo showed particular interest in questions of international trade. He thought that his labour theory of (exchange) value in the first chapter of the *Principles* could be extended to include not only domestic but also international markets. Ricardo, as

[13]It is interesting to note that according to Ricardo, the slowdown of the level of economic activity is internally generated and so is the price of agricultural and mineral products. By contrast, in the usual macroeconomic explanations, we find that the slowdown is attributed to external shocks such as the price of oil decided by the OPEC! Only in recent years and only for still relatively few macroeconomists, is has been recognised that the prices of agricultural and mineral products are determined endogenously by the system and not by monopolies such as the OPEC in the case of the price of oil.

is his wont, starts by reviewing and at the same time criticising Adam Smith's views on international trade in which he tries to detect weaknesses and by correcting these weaknesses to advance the correct (in his view) approach. Smith argues that free trade promotes the welfare of trading nations. If, for example, Smith argues, Scotland produces both wine and wheat but wine production in Scotland requires more labour than that required in Portugal, while the exact opposite is true for the production of wheat, then it is to the advantage of the two nations to specialise in the commodity for which they have an absolute cost advantage; that is, Scotland in wheat production and Portugal in wine. The needs of Scotland for wine would be met by the export of wheat to Portugal; the converse would be true for Portugal. The welfare of the two countries increases through international trade and specialisation on the commodity that these two countries have an absolute (cost) advantage.

Ricardo shares Smith's view that international trade is beneficial to the countries that engage in it, provided the countries have an absolute advantage on the goods they export. The question that comes to the fore is what happens in the case where one of the two countries has an absolute advantage in the production of both goods? In this case, Smith would say that there is no basis for trade, since the country that possesses such a cost advantage will naturally export both goods to the country with the cost disadvantage, which perhaps can import these goods to the extent that it can borrow money (gold), but soon trade is expected to be disrupted as the deficit country will accumulate an international debt. Ricardo, contrary to Smith, argues that even in this case – that Smith thought unsustainable – international trade is not only possible but also beneficial to the countries involved. The condition, however, is that each country has an absolute advantage in one of the two traded commodities. Ricardo's approach is called the 'theory of comparative cost', and it is intrinsically connected to his quantity theory of money, whose basic elements are worth stressing before proceeding to the main analysis.

Ricardo (*Works*, vol. I, pp. 85–86) argues that changes in prices lead to changes in the quantity of money. This conclusion is based on his labour theory of value. Thus, if the labour content of gold declines, for example, because of the discovery of more productive goldmines, or the introduction of new technologies in gold mining, it follows that the value of gold would fall and so, the general price level would increase, since gold is used as the numéraire commodity. In other words, if by λ_i we denote the quantity of labour required in the production of good i and by Lg the quantity of labour required for the production of a unit of gold, then the relative (to gold) price of good i will be $P_{ig} = L_i/L_g$. If the value of gold falls (rises) for some reason, then it follows that the price of good i rises (L_i for the sake of simplicity is assumed constant). The converse will be true if L_i is constant and L_g rises.

As a result, variations in the quantity of money affect the value of money and thus, the price level. If we suppose an increase in the output produced, it follows that the quantity of money needed in circulation increases (money becomes scarce). Consequently, as it happens with other goods, the price of money increases (its current price is higher than its normal) and the general price level falls. Suppose, now, that gold (*i.e.*, the monetary commodity) is produced inside the country; it follows that the excess profits in the production of money (gold) increase the output

gold and reduce its price, and simultaneously increase the quantity of money and cause a rise in the price level. If gold is imported from another country, which is the usual case, it follows that the demand for gold as well as its price increase, the excess profits in the gold production accelerate its accumulation and the expansion of supply lead to a lower price of gold. The converse will be true if there is a deficient demand for gold. Then, the price of gold will fall and the price level will rise. Ricardo argued that in the short run, there is going to be an increase in the price level as a consequence of the increase in the quantity of money, while the quantity of money (gold) adjusts to the needs of production only in the long run. Consequently, according to Ricardo, in the short run, the quantity of money determines the price level. In the long run, the natural price of gold determines the quantity of money. In Ricardo, the quantity of money is determined by the price level and output. The level of output is determined by Say's law, the details of which we examine in Chap. 6.

With these points in mind, we return to the original example assuming two goods, cloth and wine that are produced in England and Portugal. On the basis of Ricardo's numerical example, we construct the Table 4.2

We observe that Portugal produces both goods cheaper than England. The production of a unit of cloth requires 90 labour hours, while in England, it requires 100 labour hours. Similarly, the production of a unit of wine requires 100 labour hours in Portugal, while in England, it requires 120 labour hours.

In the beginning, producers in Portugal have an absolute advantage on both goods. Hence, Smith would argue that there is no basis for international trade. The two countries cannot trade with each other since each country pays for its imports through its exports; consequently, if a country, in this case England, does not export goods, then it cannot import goods either. It is at this point the strength of Ricardo's analysis lies, since he claims that in the beginning it is true that the Portuguese entrepreneurs would be able to export to England both goods. England, therefore, would run a trade deficit with Portugal, but this deficit would be covered by the export of gold to Portugal.

In what follows, the decisive role is played by the quantity theory of money of the classical economists and Ricardo, according to which the export of gold from England will reduce the supply of money there and will increase the supply of money in Portugal. Consequently, the price level in England falls while it increases in Portugal. Portugal continues to produce both goods at cheaper rates; however, it is expected that with the passage of time and with the continual increase in prices in this country and the continual decrease in prices in England, England will produce one of the two goods at a cheaper rate. Ricardo notes that the good that is produced cheaper in England cannot be wine, since the initial cost difference in both

Table 4.2 The principle of comparative cost

Goods	Countries	
	England	Portugal
Cloth	100 labour hours	90 labour hours
Wine	120 labour hours	100 labour hours

countries is bigger than that of cloth.[14] Thus, the product that England would specialise is textiles, while Portugal would specialise in wine. This specialisation is supposed to be beneficial for both countries.[15]

The aforementioned analysis shows that Ricardo believed that the labour theory of value holds within each country, but may not hold between countries, presumably because capital and labour are not free to move from one country to another to bring about an equilibrium. However, with the specialisation of each country to the good with a comparative advantage, the validity of the labour theory of (exchange) value is restored.

4.5 On the Question of Machinery

Technological progress has a twofold effect, it creates jobs on the one hand and destroys them on the other. This dual nature of technological progress may create totally different expectations with regard to its effects.[16] Ricardo was the first economist who posed the problem of technological progress with precision and always in the context of his overall analysis. According to Pierro Sraffa (the editor of the collected works of Ricardo), '[t]he most revolutionary change in edition 3 is the new chapter on Machinery' (*Works*, vol. I, p. lvii). In this chapter, Ricardo revises his initial view of the effects of mechanisation on the production process. Originally, Ricardo, as Smith before him, advanced the view that the mechanisation of the production process benefits, in general, all social classes.[17] The justification of this view is that as productivity increases, the cost of production falls, and in conditions of competition, the prices of commodities fall. Consequently, society at large benefits from lower prices (*Works*, vol. I, pp. 386–387).

More specifically, Ricardo's position in the first edition of the *Principles* in that technological progress in one sector reduces the quantity of labour that is required for the production of a good. If the demand for a good does not increase proportionally with the increase in labour productivity, then some workers may be unemployed. Hence, there is no doubt that Ricardo accepts Say's law (its detailed

[14]The principle of comparative cost would not hold if the initial cost differences were the same

[15]Historically, we know that the type of product that a country specialises in is of critical importance for its economic development. In terms of Ricardo's examples, the specialisation of England in textiles (the leading sector of that time) contributed to her growth. By contrast, Portugal that specialised in a traditional product could not follow England's growth performance.

[16]In the first decades of the nineteenth century, British workers opposed the use of labour-saving machines. Due to the prospect of being unemployed, they were destroying the machines. This is the famous movement of Luddites. The name comes from the leader of this movement, Ned Ludd.

[17]For example, in his *Essay on Profits*, Ricardo argued that the introduction of machinery in the production process increases the real wages. Ricardo expressed similar views in his speeches in the parliament and in his correspondence to McCulloch.

presentation is postponed until Chap. 5) as this can be conferred from the following quotation:

> there is no amount of capital, which may not be employed in a country, because demand is only limited by production. (*Works*, vol. I, p. 290)

Consequently, unemployment that is caused by technological change cannot but be only temporary and the operation of Say's law ensures the full employment of labour.[18]

In the third edition, Ricardo revised his initial optimistic view and argued that the mechanisation of the production process – although it continues to benefit the landlords and the capitalists – has negative consequences for the labouring class. More specifically, Ricardo admits:

> My mistake arose from the supposition, that whenever the net income [profits and rent] of a society increased, its gross income [net income plus wages] would also increase; I now, however, see reason to be satisfied that the one fund, from which landlords and capitalists derive their revenue, may increase, while the other, that upon which the labouring class mainly depend, may diminish, and, therefore, it follows, if I am right, that the same cause which may increase the net revenue of the country, may at the same time render the population redundant, and deteriorate the condition of the labourer. (*Works*, vol. I, p. 388)

Ricardo constructs an arithmetical example, where he assumes a capitalist who invests £20,000 for the production of consumer goods with proportions £7,000 in fixed capital (C^f) and £13,000 in circulating capital (C^c). Given this and assuming that the rate of profit (r) is at 10%, the value of the produced output will be equal to the sum of circulating capital and profits (Π), that is to say $C^c + \Pi$ (Table 4.3).

Consequently, when the capitalist sells his product for £15,000, he covers his own consumption expenditures of £2,000 and has £13,000 for the maintenance of labour the next year. If, however, the following year the capitalist employs half his workers in the production of consumer goods and the other half in the production of a machine, then we have (Table 4.4).

Since the value of the machine is £7,500, it follows that this sum is not available for the employment of labour. Thus, the capitalist has in his disposition £13,000 – £7,500 = £5,500 for the employment of labour, while he possesses a total fixed capital of £14,500 that consists of the £7,000 that he had initially and the £7,500 that are added with the production of the machine. The total capital remains at £20,000; only its composition has changed to C^f = £14,500 and C^c = £5,500.

Table 4.3 Production of consumer goods

K	C^f	C^c	$\Pi = rK$	Total value $C^c + \Pi$
£20,000	£7,000	£13,000	£2,000	£15,000

[18]Ricardo notes that the problem of additional persistent unemployment cannot arise "because the capitalist would have the power of demanding and employing the same quantity of labour as before, although he might be under the necessity of employing it in the production of a new or at any rate of different commodity" (*Works*, I, p. 387).

Table 4.4 Production of consumer goods and a machine

	K	C^f	C^c	$\Pi = rK$	Total Value $C^c + \Pi$
I	10,000	3,500	6,500	1,000	7,500 Consumer goods
II	10,000	3,500	6,500	1,000	7,500 Machine
Total	20,000	7,000	13,000	2,000	15,000

Ricardo accepts that the labour that is employed from the £5,500 circulating capital can set in motion the machines and tools of £14,500 of fixed capital. The reduction of the circulating capital from £13,000 to £5,500 means that labour that was initially employed and was worth £7,500 becomes redundant in the future production process, whereby fewer workers with more machines and tools produce consumer goods of value equal to £7,500, while the entrepreneur receives the same profit on the invested capital, that is, £2,000 = (0.10) × (£20,000). Consequently, the introduction of machines will be 'injurious' to the labouring class when this is accompanied by a reduction in the wage fund, because the unemployed cannot find employment elsewhere:

> That the opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests, is not founded on prejudice and error, but is conformable to the correct principles of political economy. (*Works*, vol. I, p. 392)

Ricardo relaxes somewhat his negative conclusions by arguing that if the preferences of the entrepreneurs remain the same, then the fall in prices will increase savings, and, therefore, investment and employment. Notwithstanding, Ricardo does not say that unemployment will be eliminated. Consequently, if the expenditures for the construction of machines are financed from savings and not from the transformation of circulating to fixed capital, then the mechanisation of production will have beneficial results for the working class and society in general.

Ricardo leaves no doubt that the introduction of machines and the concomitant technological progress in general must be encouraged regardless of the injurious side effects for the workers. Ricardo notes:

> The statements which I have made will not, I hope, lead to the inference that machinery should not be encouraged. (*Works*, vol. I, p. 395)

Ricardo argues that whatever be the results of the mechanisation of the production process, the government should not discourage the introduction of machines. The idea is that if the government prohibits a capitalist to acquire the maximum possible profit from the introduction and employment of machines in his own country, then it is likely that he would be forced to invest his capital in another country. It is true, argues Ricardo that the introduction of machines reduces the domestic demand for labour, but the domestic demand for labour becomes zero if capital is invested in a foreign country. Moreover, the mechanisation of the production process in a country reduces the production cost with the effect that this country remains competitive internationally.

One of the critiques to Ricardo's argument was that he did not account for feedback effects that are caused by unemployment. For example, the increase in unemployment because of the introduction of machines drives wages down, raises profits and investment, and as a result, employment increases. In addition, it has been argued that the technological progress that results from the introduction of machines leads to an increase in productivity, which in turn reduces the cost of production, causes prices to fall, and boosts demand. The subsequent expansion of output is likely to lead to a (substantial) reduction in unemployment. Feedback effects are usually left out in Ricardo's analysis. It seems that he considers that feedback effects perhaps modify but do not alter the fundamental results of the theoretical analysis.

4.6 Ricardo's Theory of Taxation

The formulation of the theory of value was absolutely necessary for Ricardo to develop his theory of taxation. He starts with Adam Smith's views, and after a short presentation, he either approves them and continues with the next topic or proceeds with the correction of Smith's 'original errors regarding value' as a necessary step in advancing a 'true doctrine of distribution' and a logically consistent theory of taxation. The corrections come mainly from the strict application of the labour theory of value and distribution.

Ricardo devoted 14 chapters of his *Principles* to the question of taxation; nevertheless, this part of the book was easy to complete in a relatively short period of time, precisely because the chapters on taxation were an exercise and further elaboration of the theory of value. It is interesting to note that in Ricardo's analysis, we discover the technique that we call today comparative statics. That is, he compared one equilibrium situation with the one that the economy would gravitate towards, when some of its elements changed.

Ricardo and Smith, in their analysis of taxation, regarded the real wage as a fundamental socio-economic variable that cannot be taxed under normal conditions (Shoup 1960; Eagly 1983, *inter alia*). Any effort to tax real wages leads to an increase in the money wage which reduces profits at least by the amount of the tax. Consequently, the money wage is virtually the means through which the tax is finally collected from surplus incomes, that is, profits or rents. The idea is that a tax on real wage, which is considered to be at the socially determined subsistence level, triggers an increase in the infant mortality rate and thus, the supply of labour decreases, thereby increasing the money wage and restoring the real wage to its pretax level (*Wealth of Nations*, pp. 79–80). In the general case, a similar idea was articulated by Ricardo with the difference that the reduction in labour supply was caused primarily by the deferment of marriages to old age, rather than by the increase in the infant mortality rate. This does not mean that Ricardo ruled out the case of increasing infant mortality rate, but that he considered it to be of secondary importance as far as the increase in the real wage was concerned (*Principles*, p. 93).

For the questions of taxation, however, Ricardo argued that the population law (in any version) would be inadequate, because it would require a long time to work itself out. For this reason and for the short term, Ricardo thought about the operation of an alternative mechanism based on the idea that the income received by the government is spent on the employment of workers in the non-productive activities of the state which competes with the private sector over a given supply of labour. As a result, the overall demand for labour increases, since the public demand for labour is added to the private, which remains the same, because the income of the owners of capital has not been taxed.

> the owners of capital who would have nothing to pay towards such a tax, would have the same funds for employing labour [. . .]. (*Works*, vol. I, p. 220)

In other words, capitalists have the same income to spend in order to produce the same level of output with the given state of technology. These circumstances are particularly favourable for workers, who, by increasing their money wage, manage to pass their entire tax burden on profits (*Works*, vol. VIII, pp. 169–170)

It is interesting to note that in Ricardo's analysis, the change in the relative prices of each sector depends on its relative capital–labour ratio, and also that the general rate of profit of the economy falls, and, therefore, the growth capacity of the economy diminishes. This is an issue that has been stressed by classical economists and Ricardo, who notes: 'there are no taxes which have not a tendency to lessen the power to accumulate' (*Principles*, p. 152). Ricardo discusses circumstances under which wages are taxed, and thus, the purchasing power of workers diminishes even below their customary standard of living, for example, when the government uses the income from taxation as foreign aid or to finance a war or, finally, to spend it on the output produced by the private sector instead of on the employment of labour. In all these cases, the mechanism that restores the money wage to its subsistence level does not work itself out.

> [If taxes are given] as a subsidy to a foreign state, and if therefore these funds were devoted to the maintenance of foreign, and not of English labourers, [...] then indeed, there would be a diminished demand for labour, and wages might not increase, although they were taxed. (*Works*, vol. I, p. 221)

Regarding the taxation of profits, Ricardo's analysis is based on the principle of equal profitability. If, for example, we have a partial tax on the profits of capitalist farmers, then they will increase their prices by the amount of the tax in their effort to secure for themselves the economy's average rate of profit. In this case, according to Ricardo, landlords benefit since they see that their rental incomes increase as a result of higher agricultural prices. If the government, on the other hand, imposes tax on the profits of the industrial capitalists, then these capitalists will increase their prices to a point that they can obtain the economy's average rate of profit, while landlords' rental incomes decrease as a result of lower relative agricultural prices. If, finally, the tax on profits is general, then there is no reason for any change

in the relative prices because the post-tax average rate of profit, though lower, continues to be the same across industries.[19]

In the analysis of the taxation of rent, one expects different conclusions and therefore, different policy proposals. We know that Smith accepts the case of absolute rent, which means that a tax on rent may be passed on higher prices. In contrast, Ricardo argues that the regulating price of agricultural products is formed on the marginal land on which there is no rent. As a consequence, a tax on rental incomes cannot be passed on higher prices, and, therefore, land rent becomes an ideal source of taxation. Nevertheless, Ricardo did not propose the taxation of incomes from rent for the same reasons that have been invoked by Smith. Included among the reasons are:

1. Real rent is practically very difficult to be separated from profits.
2. A tax that is imposed on only one social class is 'unfair' (*Principles*, p. 204).
3. Such a tax feeds speculative revaluations of the price of land which interfere with 'sober trade'.

Finally, indirect taxation falls mainly on profits since the real wage is fixed and therefore, it cannot be really taxed. It is important to reiterate that the rigidity of the real wage is not assumed axiomatically, but because of a specific economic process that is attributed either to some form of population law (Smith) or to the demand effects emanating from government expenditures (Ricardo and also J.S. Mill). The idea is that the private sector, when faced with a given technology and a given level of output, competes with the public sector of the economy over the same supply of labour; as a result, the money wage rises by the amount of the tax, so the end result is the restoration of the real wage to its pretax level and the reduction of profits by the amount of the tax.

This argument does not hold true in the case of indirect taxation on luxury goods, which rests exclusively on the capitalists and other rich consumers. The idea is that a tax on luxuries merely increases the price of these goods, and given that luxuries are neither inputs to production nor part of workers' basket of goods that constitute their real wage, it follows that their higher price affects only the income of the rich consumers, mainly capitalists and landlords, without further consequences to the economy.

From the aforementioned points, it follows that the rigidity of real wage is the idea *sine qua non* for the classical propositions of taxation according to which all taxes, in the final analysis, fall on surplus incomes (*i.e.*, profits or rents). A majority of modern economists are uncomfortable with the classical economists' idea for the flexibility of money wage that leads to a real wage fixed at the customary standard of living of workers. However, on closer examination, this idea becomes stronger than is usually imagined. The following considerations can be used in favour of

[19]For a review of the classical analysis of taxation and especially the views of Ricardo, see Shoup (1960); Eagly (1983) Tsoulfidis (1993).

a theorisation of a constant real wage in the classical economists' sense.[20] First, employers are willing to offer a real wage that enables workers to acquire the customary standard of living and also perform well in their jobs (the efficiency wage hypothesis). Second, unions, in their collective bargaining with employers, demand wage increases that secure the customary standard of living. Third, it is recognised as a stylised fact of modern societies that the money wage and the price level move in tandem so that they can be used interchangeably; for example, in the case of a Phillips curve. This increase in money wages approximately at the same (or somewhat higher) rate with the price level indicates that the real wage is characterised by relative rigidity at least in the downward direction and in any case is not subjected to wide fluctuations. Fourth, the mobility of labour, together with the rapid flow of information about the standard of living which workers would be willing to emulate, increase the pressure for higher money wages in order to bring them in line with the new consumption patterns and standard of living.

4.7 Ricardo on Public Debt

In the analysis of public debt, Ricardo shares Smith's views on the unproductive character of state expenditures and on the notion that their financing via public borrowing decreases the investible product, and, therefore, it becomes detrimental to society's capacity to accumulate wealth. Nevertheless, many modern economists attribute to him the idea of equivalence of the two forms of financing (taxes and public debt) in the so-called Ricardian Equivalence Theorem (Barro 1974), which is particularly popular in the literature of public finance, as well as in modern macroeconomics. This theorem ascribes to Ricardo the view that taxation and public borrowing constitute essentially equivalent forms of financing public expenditures. The truth, however, is that Ricardo categorically rejects the notion of equivalence of the two ways of financing government expenditures:

> [The system of borrowing] is a system which tends to make us less thrifty – to blind us to our real situation. If the expense of a war be 40 millions per annum, and the share which a man would have to contribute towards that annual expense were 100 l., he would endeavour, on being at once called upon for this portion, to save speedily the 100 l. from his income. By the system of loans, he is called upon to pay only the interest of this 100 l., or 5 l. per annum, and considers that he does enough by saving this 5 l. from his expenditure, and then deludes himself with the belief, that he is as rich as before. (*Works*, vol. I, p. 247)

In other words, the public would not perceive the debt as a tax of an equal amount, and, therefore, people would tend to save less than in the case of taxation and so, capital accumulation would slow down. As a consequence, income and tax

[20]The new Keynesian literature provides much more theoretical justification and empirical evidence in favour of the rigidity of real wages.

revenues would fall, and the government would raise the tax rates in an effort to raise the same tax revenues, slowing down capital accumulation further and eventually leading to national bankruptcy.

Similar arguments were developed three years later in his article on the Funding System (1820) (*Works*, vol. IV), where Ricardo evaluates three ways of financing a war of an annual cost of 20,000,000 l. First, through direct taxation of 20,000,000 l; second, through a loan, where the government will pay annually a specified sum in perpetuity and in agreed interest rate; if, for example, the interest rate is 5% and remains constant, then the annual sum (or the sinking fund) will be 1,000,000 l; third, with a loan which will be paid in a specific time interval. If, for example, the interest rate is 5% and remains constant, then the annual payments of taxes will be counterbalanced with 1,000,000 l plus a sum, for example, with 200,000 l for the settlement of loan, whose maturity date is calculated to 45 years. From these three ways of financing, Ricardo argues that the first is preferred over the others and the reason is

> [t]he burthens of the war are undoubtedly great during its continuance, but at its termination they cease altogether. When the pressure of the war is felt at once, without mitigation, we shall be less disposed wantonly to engage in an expensive contest, and if engaged in it, we shall be sooner disposed to get out of it [. . .]. (*Works*, vol. IV, p. 186)

He further draws attention to the idea that from a purely economical point of view, there is no real difference among the various ways of financing, because in the end, the same sum is being paid with the same collection cost. More specifically, he notes:

> In point of economy, there is no real difference in either of the modes; for twenty millions in one payment, one million per annum for ever, or 12,000,000 l for 45 years, are precisely of the same value; (*Works*, vol. IV, p. 186)

Hence, Ricardo gives the impression that he is in favour of equivalence between taxation and various forms of borrowing. However, a more careful reading of his text reveals that Ricardo claims that only 'in point of economy' the alternative ways of financing of war are equivalent. Ricardo explains:

> But the people who pay the taxes never so estimate them, and, therefore, do not manage their private affairs accordingly. We are too apt to think, that the war is burdensome only in proportion to what we are at the moment called to pay for it in taxes, without reflecting on the probable duration of such taxes. It would be difficult to convince a man possessed of 20,000 l., or any other sum, that a perpetual payment of 50 l. per annum was equally burdensome with a single tax of 1,000 l. He would have some vague notion that 50 l. per annum would be paid by posterity, and would not be paid by him; (*Works*, vol. IV, p. 186)

One criticism to the above argument is that individuals have a limited lifetime, and, therefore, they do not care very much about the tax, since what they will pay at the end of their lives will be less than what they are called to pay once and for all. As a result of tax reductions, individuals engage themselves more in consumption spending than in saving. The counterargument here is that the bequests must also be accounted for. Ricardo further argued that the economic unit of his analysis should not be limited to the individual, but it must be expanded to include the household.

In Ricardo's analysis, the household becomes an institution with infinite lifetime (since the income and the assets in general are transferred as bequests) and thus, his initial assumption about the difference in the method of financing continues to hold unaltered. More specifically, he remarks:

> But if he leaves his fortune to his son, and leaves it charged with this perpetual tax, where is the difference whether he leaves him 20,000 l., with the tax, or 19,000 l. without it? This argument of charging posterity with the interest of our debt, or of relieving them from a portion of such interest, is often used by otherwise well informed people, but we confess we see no weight in it. It may, indeed, be said, that the wealth of the country may increase; and as a portion of the increased wealth will have to contribute to the taxes, the portion falling on the present amount of wealth will be less, and thus posterity will contribute to our present expenditure. That this may be so true; but it may also be otherwise – the wealth of the country may diminish – individuals may withdraw from a country heavily taxed; and therefore the property retained in the country may pay more than the just equivalent, which would at the present time be received from it. (*Works*, vol. IV, pp. 186–187)

Hence, Ricardo claims that the methods of financing government expenditures are not equivalent because loans deceive individuals that their income remains intact. Ricardo further argues that in the exceptional case where public deficit is financed through taxation, it may produce, in the short-run, the same results with those that would be caused from public borrowing. But in the long run, the ruinous results of public borrowing in society's capacity to accumulate are even worse than those that are caused by taxation, since borrowing drains savings that are ready to be invested productively, while taxation falls on current incomes for which we do not really know whether they were to be invested or consumed. For example, in a letter to McCullough (3/29/1820), Ricardo points out:

> But when we are carrying on an expensive war and I is necessary to raise large funds within the year, either by loan, or by taxes equal in amount to such loan the former will I think be most injurious to the labourer, because it will more materially affect the accumulation of capital. (*Works*, vol. IX, p. 170)

The reason is that so long as capitalists or rich consumers are forced to pay an unusually high amount of tax, they will try to increase their income to the pretax level by limiting their consumption and other unnecessary expenses. Thus, it is possible for capitalists in one year to decrease the amount of their savings; however, in the next few years, they will make every possible effort to replenish their savings so as to maintain their capital on the pretax scale.

Ricardo's analysis refers to the usual case where government expenditures are not productive (as for example in the case of a war). If individuals anticipate future taxes, it follows that the present value of their future incomes decreases, which entails a *pari pasu* reduction in their consumption expenditures. However, since the income level remains the same, it follows that the increase in the absolute and not the relative amount of savings will function as anticipation for future tax payments. The interest rates are expected to increase, thereby, crowding out private investment expenditures. Consequently, if the government spends its money on wars or on consumption, then according to Ricardo, taxation is preferred to borrowing. There are cases, however, where there is equivalency in the modes of financing of

government expenditures. For example, if the deficit is caused by public investments in the provision of infrastructures in general, it follows that the results may not be different if these expenditures are financed through taxation or borrowing.

4.8 Concluding Remarks

Until now we have shown that the unequal composition of capital and changes in the income distribution affect only marginally, and in a predictable way, the relative (natural) prices. In Ricardo, the labour time is the primary regulator of natural prices, which in turn are the centre of gravity of market prices. For example, Ricardo notes:

> all the great variations which take place in the relative value of commodities to be produced by the greater or less quantity of labour which may be required from time to time to produce them. (*Works*, vol. I, pp. 36–37)

The next step for Ricardo was to incorporate in his analysis the theory of rent and Malthus's population law and treat the real wage as a datum to arrive at a falling rate of profit argument, which, however, is very weak, if we think that is based on the rising prices of agricultural products. This idea does not really hold true in general, and historical evidence shows that it was also not true in Ricardo's time. International trade was also an area that Ricardo investigated, and by using the quantity theory of money, he derived the principle of comparative cost, an idea that has survived (albeit modified) until our days and can be found in all books of international trade. Technological progress is another area that Ricardo offered us with an analysis rich in insights that shed light on structural unemployment.

Turning to Ricardo's chapters on taxation, we observe that his analysis is in accordance with his overall theory developed in the first seven chapters of his book. His conclusions on the effects of taxation are derived after a strict application of his theory of value and distribution. Ricardo's analysis is not confined to the microeconomic effects of the tax but also includes the macroeconomic consequences emanating from the government's activities. This is an aspect that distinguishes Ricardo from other economists of his time, who neglected the effects that arise from the government's demand for labour. This negligence is still present in the modern theories of tax incidence, since the standard treatments of both differential and balanced budget incidence are ways to evade the difficulties associated with the question of demand emanating from the state's expenditures.

Questions for Thought and Discussion

1. Discuss Ricardo's theory of value and its modifications. To what extent do these modifications change Ricardo's original principle?
2. What was the invariable measure of value and what are its functions? What was Ricardo's position with regard to the properties of the invariable measure of value? Is there a commodity (practical or analytical) that can play this role?

3. Let us suppose that for some reason (taxation, price of agricultural products, *etc*.), wage increases. What is going to be the effect of this increase, according to Ricardo, on relative and absolute prices? Would Smith agree with such a view? Discuss.
4. Contrast Ricardo's theory of international trade with that of Adam Smith.
5. Discuss Ricardo's views on the introduction of machines. Does technological change increase unemployment? Explain.
6. Discuss Ricardo's views on taxation on wages. To what extent does Ricardo differentiate his views from those of Adam Smith?
7. Discuss Ricardo's views on taxation on profits; does his principle of non-neutrality of money in the case of general tax on profits hold if the monetary commodity is imported?
8. Does the tax of rent change prices according to Ricardo? Does Ricardo reach similar policy conclusions to those of the Physiocrats with regard to taxation on rent? Explain.
9. Discuss the incidence of indirect taxation in Ricardo. To what extent does an indirect tax differ from a tax on profits?
10. Critically evaluate 'Ricardo's Equivalency Theorem'.

Notes for Further Reading

There is a voluminous literature on Ricardo that focuses on the theory of value. Despite the large number of articles, to our view, only a few really accord with Ricardo's text. This must be due to the effort of major authors to read in Ricardo their own theory of value. For example, Marshall's (1920) reading of Ricardo gives the impression that Ricardo was a forerunner of neoclassical economics. By contrast, Sraffa (*Works*, vol. I, introduction) argued that Ricardo's theory embodies a clear break from Adam Smith and is incommensurate with modern neoclassical economics. S. Hollander's (1985) reading of Ricardo revived Marshall's interpretation, with the difference that now the emphasis is placed on the general equilibrium perspective of the (neoclassical) theory of value. The case is different when we come to Marx (1968) who generally differentiates his theory of value from that of Ricardo's. However, he often gives the impression that the usage of the word value is similar to his own and then he charges Ricardo with various failures (see Tsoulfidis 1998). This is not the case when we come to Marxist authors, who charge Ricardo for not differentiating between different kinds of labour, for not explaining profits as a result of exploitation of labour and so forth. These points do not really affect Ricardo's theory, as it was shown so eloquently by Steedman (1982) in his comparison of Ricardo and Marx, where he finds that Marx and Marxists misused Ricardo's concepts. While this is true, at least in part, Steedman's conclusion that Ricardo held a cost of production theory of value and that there is no need for a labour theory of value are some of the issues that are debated until today. Stigler (1958) gives

(continued)

a very good account of the source of misinterpretations of Ricardo, but when he comes to give his own interpretation, he claims that Ricardo held an empirical 93% labour theory of labour value, but not an analytical (theoretical) one, since he reverted to a cost of production theory of value similar to Smith's with the difference that in Ricardo's theory, rent is not a component of price. Peach's (1998) book on Ricardo presents, to our view, a fair account of his theory of value.

On the long-run tendency of the rate of profit, the reader can find a very simple but accurate exposition in Heilbroner (1981) and also in Rubin (1979). For the mathematically oriented reader, the articles by Baumol (1962), Pasinetti (1974, 1977) and Samuelson (1978) are the best starting points. In this regard, we also recommend Harris (1988) who discovers the case of chaotic dynamics in the usual Ricardian model. The reader is also encouraged to see S. Hollander's (1985) so-called 'new view' according to which the rate of profit falls as a result of falling and not constant real wages. As we explained, this is an accurate description of Ricardo's view; that is, the rate of profit falls because of higher wages that capitalists have to pay, and at the same time, the workers' standard of living may deteriorate. This view can also be found in Rubin (1979, pp. 284–285).

On the theory of international trade, the article by Shaikh (1980a) is perhaps the most representative of Ricardo's Chapter 7 of the *Principles*. In this article, Shaikh argues that the quantity theory of money is the cornerstone for deriving Ricardo's principle of comparative cost; moreover, Ricardo's conclusions change if an alternative theory of money is used. The usual presentations of Ricardo's principle of comparative cost in international trade books are very far from what Ricardo really said and the serious reader should take this into account. For the reader interested in Ricardo's version of the quantity theory of money, the article by Green (1977) is strongly recommended.

On the question of machinery, one finds very good and interesting accounts by Hollander (1971), Eltis (1985), Meacci (1985), *inter alia*. Having known Ricardo's chapter "On Machinery" one wonders how it is possible for so many authors even historians of economic thought to claim that Ricardo, and the classical economists in general, assumed full employment of labour!

On the question of taxation, we recommend the book by Shoup (1960). Since then, however, some relevant articles that deal with the various aspects of Ricardo's tax chapters have been published. In this sense, the articles by Carr and Ahiakpor (1982), Eagly (1984) and Tsoulfidis (1993, 2005) are being included. The mathematically oriented reader is encouraged to read the articles by Semmler (1983), Dome (1992, 2000) and Erreygers (1995). On the question of debt, we recommend the articles by Tullio (1987), Eagly (1984) and Tsoulfidis (2007). The best source, however, on this issue is

(continued)

Ricardo's views published in Encyclopaedia Britannica in 1820 under the title 'Funding System' and has been included in the supplement of *Works*, IV.

For an overall presentation of Ricardo, we recommend the book by Clair (1965). In this respect, classical are by now the articles by Pasinetti (1974), which constitute a stimulating presentation of Ricardo's theory of value and distribution. Also, the entry on Ricardo by de Vivo (1987) in the New Palgrave is an excellent summary. For advanced and challenging reading, we recommend Morishima's (1989) mathematical treatment of Ricardo's *Principles*, and for a critique of Morishima's book, see Kurz and Salvadori (1998).

Chapter 5
Karl Marx's *Das Kapital*

Every child knows that a nation which ceased to work, I will not say for a year, but even for a few weeks, would perish. Every child knows, too, that the volume of products corresponding to the different needs require different and qualitatively determined amounts of the total labour of society. That this necessity of the distribution of social labour in definite proportions cannot possibly be done away with by a particular form of social production but can only change the mode of its appearance, is self evident. Natural laws cannot be abolished at all. What can change in historically different circumstances is only the form in which these laws assert themselves. And the form in which this proportional distribution of labour asserts itself, in a social system where the interconnection of social labour manifests itself through the private exchange of individual products of labour, is precisely the exchange value of the products. Science consists precisely in demonstrating how the law of value asserts itself. So that if one wanted at the very beginning to "explain" all the phenomena which seemingly contradict the law, one would have to present the science before science.

(Marx, letter to Kugelmann, 11 July 1868)

As long as capital is weak, it still itself relies on the crutches of past modes of production, or of those which will pass with its rise. As soon as it feels strong, it throws away the crutches, and. moves in accordance with its own laws. As soon as it begins to sense itself and become conscious of itself as a barrier to development, it seeks refuge in forms which, by restricting free competition, seem to make the rule of capital more perfect, but are at the same time the heralds of its dissolution and of the dissolution of the mode of production resting on it.

(Marx, Grundrisse, p. 651)

5.1 Introduction

Karl Heinrich Marx (1818–1883) was born in Trier, Germany, in a Jewish family that converted to Protestantism during his childhood. He studied philosophy in the Universities of Bonn, Berlin and Jena from where he earned his doctorate in philosophy at the age of 23. As a student, he was involved in circles of young philosophers known as the Young-Hegelians. He worked as a journalist and editor for the influential newspaper *Rheinische Zeitung* of Cologne. The radical perspective of the newspaper led the Prussian authorities initially to censorship and later to the closing of the newspaper and to the exile of Marx. He took refuge in France and settled in Paris, where he had the opportunity to study French utopian socialism and

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_5, © Springer-Verlag Berlin Heidelberg 2010

English Political Economy, while at the same time, he was involved in the socialist movement in Prussia. He negotiated the publication, in Paris, of a short-lived magazine called *Deutsch-Französische Jahrbücher*. Here the idea was the integration of the ideas of French Utopian socialism with the philosophical ideas of the Young-Hegelians. These activities, however, were not appreciated by both the French and the Prussian governments. After the advice of the Prussian authorities, the French government deported Marx, who went to Brussels in 1845 to return to Paris (1848) and then again to Germany, where he was involved in German politics, this time through another newspaper the *New Rheinische Zeitung*. Not after long, this paper was banned, and Marx had to emigrate to England in May 1849 and settled in London, where he lived for the rest of his life.

Marx published the first volume of *Capital* in 1867, while the other two volumes of this work were published posthumously by Friedrich Engels (1820–1895) and Karl Kautsky (1854–1938), one of the founders of the Second International, who also assumed the difficult task of publishing some other manuscripts of Marx in a three-volume work with the title *Theories of Surplus Value*.[1]

Friedrich Engels, the close friend, collaborator and financial supporter of Marx, was the son of a German industrialist. On the one hand, Engels occupied himself with the business activities of his family, assuming the responsibility of the operation of a factory in Manchester, England, and, on the other hand, being an intellectual, he published significant work and was involved in the socialist movement. He met with Marx in Paris in 1844, while he was visiting this city for a short trip. Thus, the beginning of a lifelong friendship and joint work was marked.[2] The collaboration between Marx and Engels led to the writing of many works including *The Holly Family* (1845), *The German Ideology* (1845–1846) and *The Manifesto of the Communist Party* (1848).

There are different interpretations of Marx, and this often gives the impression that his work is vague. A careful examination of the history of the socialist movement, the origins of which were inspired to a great extent by Marx himself, shows, however, that those engaged in this movement often ignored Marx's major work, that is, the three volumes of *Capital*. More specifically, during the period of the First International (1864–1876), Marxists paid particular attention to the political or philosophical writings of Marx and Engels (*e.g.*, the *Communist Manifesto* (1848), the *German Ideology* (1845–1846/1932), etc.). Volume I of *Capital* (1867) was not read as much as one would expect, except for the "historical" chapters that refer to the exploitation

[1]There exist still unpublished manuscripts by Marx's pen that are prepared for publication. One such publication has been the book *Grundrisse* published in 1973, 893 pages long. A book similar in size in the form of hand-written notes refers to the theory of money. The total of Marx's works (together with Engels) published in English (Marx/Engels Collected Works, MECW) is 50 volumes, and it took 30 years for its completion (1975–2005). There is another project under way called the Marx–Engels–Gesamtausgabe (MEGA) that purports to publish all the works of Marx and Engels, a total of 120 plus volumes!

[2]Engels met with Marx for the first time in the Rheinische Zeitung offices in Cologne on his way to England in 1842, however, their essential friendship and collaboration started in 1844.

of workers and their struggles for the reduction of the length of the working day. The Marxists of the Second International (1889–1916) began using volume I of *Capital* and to some extent volume II (published in 1885 and dealing with the circulation of capital and the mechanisms of reproduction). Their discussions focused on whether Marx's purpose was to demonstrate the possibility of a balanced growth for capitalism or to reveal capitalism's instability and predict its inevitable collapse, unless the system expanded to incorporate the non-capitalist economies as an additional source of cheap raw materials and as a market to dispose products. Hence, the foundation for an economic theory of imperialism was developed by a number of radicals. Volume III of *Capital* was published in 1894 but was considered "too scientific", as Rosa Luxemburg (1870–1919) once remarked, and there were only a few Marxists during the 1930s, notably Henryk Grossmann (1881–1950) and Maurice Dobb (1900–1976), who read it attentively. The subsequent Keynesian revolution in economics led many Marxists to "keynesify" the economic theory of *Capital* and "marxify" the economic policy conclusions of Keynesian economics. Many Marxists of the time abandoned partly or completely the economic analysis of *Capital* adopting the Keynesian analysis instead, from which they tried to derive radical policy conclusions with regard to the treatment of monopolies and income distribution. More specifically, Paul Sweezy (1910–2004) and also Paul Baran (1910–1964) claimed that Marx's analysis was more appropriate for the conditions of nineteenth-century capitalism, where there were neither monopolies nor a powerful state, than for the current economic situation in which powerful monopolies dominate and, together with the state, influence economic outcomes. Naturally, many radical economists claimed that once the workers' party seized power, it could use the state for its own purposes. In this context, it has been argued that the (Marxian) law of value no longer holds in conditions of monopoly capitalism and also that Marx's laws of motion should be revised because the economy is under the control of monopolies and the state. Notable exceptions to this stream of thought were Paul Mattick (1904–1981) and Ernest Mandel (1923–1995), who, during the 1970s and 1980s, were for Marxian economic thought what Grossmann and Dobb were in the 1920s and 1930s. That is to say, they were among the very few Marxists who consistently used and expanded the analysis of *Capital* to incorporate the conditions of their time.

In what follows, we focus on some important aspects of Marx's work, starting from the theory of value and money, continuing to the transformation problem, the theory of competition and the law of the falling tendency of the rate of profit and its association with economic crises.

5.2 Commodity Production and Value

The goal of Marx's economic studies was to "lay bare the laws of motion of modern society" (*Capital*, vol. I, p. 10), that is to say, to discover social regularities described mainly as long-run tendencies. Marx observes that capitalism is a historically specific system characterised by generalised commodity exchange, so, naturally, the starting point of his inquiry is the analysis of commodities, the most

elementary form of the wealth of a capitalist society. Commodities appear to be the only logical starting point for the analysis of capitalism, since any other starting point leads back to the notion of commodities. For example, starting from production or income, we soon discover that we are dealing, essentially, with the exchange of commodities, and, therefore, the commodity becomes the starting point for the analysis of capitalism, that is, the generalised system of exchange of commodities.

A commodity has a dual property: it can be used to satisfy needs and it can also be exchanged. An individual purchases a commodity to satisfy his/her own manifold needs, real or fictitious. The seller of commodities, on the other hand, aims, through the exchange, at making profits. Consequently, adopting the established terminology in Marx's time, a commodity is at the same time a use value and a value. In order to understand the notion of value, we need to first understand the relation between use value and exchange value, which constitute the opposite poles of a commodity. We say opposite poles since the seller of the commodity is interested in its exchange value, while the buyer in its use value. As a use value, a commodity is socially useful, and, therefore, it can be exchanged against another.

Turning now to the exchange value, we observe that in exchange commodities are compared to each other. For example, when we say that a commodity contains value, then we recognise essentially that x quantity of commodity A is equal to y quantity of commodity B or z quantity of good gold, and so forth. It follows that commodities must have something in common; otherwise, there is no basis for their comparison and exchange. Hence, the property that gives to commodities exchange value and thus makes them comparable must be distinguished from the measurement of their value. If, for example, we measure commodity A in terms of commodity B, we will get a different result from what we would receive had we measured commodity A in terms of commodity C or of commodity G (gold = money) and so forth. The measurement of values of commodities reminds one of the weighting of goods with weight units (kilos). Thus, a definite quantity of metal is taken as a standard for the measurement of the weight of the other goods. It is not the weight units that make goods to have weight but their mass. Similarly, it is not gold (or money) that gives rise to the worthiness of commodities. Consequently, when we refer to the price of good A, we essentially pose two questions: first, what is the cause of exchangeability of commodities? And second, how exactly is this worthiness of commodities manifested and measured in exchange?

5.3 Concrete and Abstract Labour

We repeat that, in commodity production, each product is useful and, at the same time, worth of something. The usefulness or use value and value aspects of a commodity stem from the dual nature of labour, which, in its concrete form (the labour of the carpenter, painter, designer, *etc*.), gives rise to the useful properties of a commodity to satisfy human needs (real or imaginary), and the value aspect of the commodity comes from the abstract character of labour. In fact, by experimenting

with different properties of commodities, Marx ends up with the idea that the only economically meaningful property characterising all commodities is that they are all products of human labour. This common property allows the comparison of commodities, according to the quantity of labour that they contain. Consequently, labour that is employed in the production of commodities gives them their worthiness (value).[3] This characteristic is common to all commodities, and it means that all types of labour must be equalised qualitatively, since different types of labour differ with respect to the quantity of value that they create. As in the case of different commodities which when expressed to a common denominator become comparable to each other, the quantity of labour that is contained in different commodities must be quantitatively comparable. Every commodity, therefore, incorporates human labour that makes comparisons with other commodities possible on the basis of the quantity of labour that is required for their production.

Marx calls abstract labour a kind of labour whose particular characteristics are abstracted already in the production process through a real (not a mental) process of abstraction. The secondary concrete differences of commodities are subjugated to their common property, that is, the amount of abstract labour that has been expended on their production. In other words, the labour time becomes abstract, and, therefore, creates value from the beginning in the sphere of production on the basis for the purpose that it was assumed, that is, the production of commodities. As a consequence, the concept of abstract labour is not a mental generalisation that we somehow choose to make, but rather the reflection in thought of a real social process. This in turn means that abstract labour and hence value are also real (Shaikh 1982, p. 273).[4] It is important to point out that only the labour in a capitalist society can be viewed as abstract, in the sense that only in capitalism, we have the dominance of the market in all aspects of economic life, so only this kind of labour regulates the exchange value of commodities. Moreover, from a social point of view, the total labour time required for the production of each commodity consists of the direct and the indirect – that is, the materialised labour time in the non-labour inputs – labour time. Marx calls the total quantity of abstract labour time, that is, incorporated in a commodity the immanent measure of value of the commodity (*Theories*, I, p. 403).

5.4 Socially Necessary Labour Time

The value of a commodity is equal to the quantity of the abstract labour time that is socially necessary for the production of the commodity in question. Hence the notion of socially necessary abstract labour time is different from the homogenised labour time of Smith and Ricardo. According to Marx

[3]Another common characteristic of commodities is the usefulness (or utility) that they possess. Marx, however, excluded such a possibility on the basis of the subjective character of utility.

[4]Sweezy (1942), among others, views abstract labour as the result of a mental abstraction from the particular characteristics of labour.

> the labour-time socially necessary is that required to produce an article under the normal conditions of production, and with the average degree of skill and intensity prevalent at the time. (*Capital*, vol. I, p. 39)

This means that the unit value of a commodity is equal to the quantity of the socially necessary labour time that is required for its production, and, therefore, is inversely related to the productivity of labour. It is important to point out that socially necessary labour time is subject to changes that are independent of concrete labour. For example, Marx (*Capital*, vol. I, p. 39) refers to a characteristic example that was observed in England when the introduction of power looms reduced the socially necessary labour time for the production of cloth by about fifty percent. The traditional producers who continued working with hand looms found out that the value of their commodity was slashed by one-half not because of the reduction of their own labour time but because of the reduction of the socially necessary labour time.

The notion of socially necessary labour time is used in a dual sense (Shaikh 1978, 1982, 1984). According to the first, which is also the usual one, the socially necessary labour time for the production of a specific commodity is the total labour time expended on the production of all commodities of this type divided by the number of such commodities (*Capital*, vol. I, p. 39). The second sense of socially necessary labour time refers to the labour time required for the production of a type of a commodity in a quantity, which is compatible with the social demand. If the quantity of labour that was expended for the production of a type of commodity is smaller (larger) than that actually demanded, it follows that the market price of the commodity will be higher (lower) than the average price (*Capital*, vol. III, p. 635). Although the notion of average socially necessary time (first sense) is an extremely good first approximation, when the analysis becomes more concrete as is in volume III of *Capital*, the notion of socially necessary time is expanded to account for the specific conditions of each industry. For example, in agriculture the value of agricultural commodities depends on the conditions of demand relative to supply. If demand and supply grow at the same rate then the average value is derived from the average industry, if, however, supply is not easily expandable as is in the case of agriculture or mining then the value of commodities is determined by the marginal land, that is, the only type of land that production can be expanded.

Thus, in production, we have to determine both total and individual value of the commodities that constitute the basis for the formation of the regulated price (direct price, price of production). The regulated price, in turn through demand, which is expressed according to its level, determines the socially necessary labour time in the second sense, which expresses the total need of the economy for this product (and, therefore, the total quantity of labour). There is always the possibility for some quantity of value and surplus value that has been created in the production sphere not to be realised in the sphere of circulation because of the lack of adequate demand. In other words, we have the phenomenon of a deviation between the two senses of socially necessary labour time.

Finally, the magnitude of value of a commodity depends not only on the quantity but also on the quality of labour. On this basis, we distinguish between simple and complex (skilled) labour. Simple labour is labour with no prior training. By contrast,

skilled labour presupposes a preparatory stage during which the labourer obtains the necessary educational training for the performance of his work. Consequently, an hour of skilled labour will create a multiple of the value created by unskilled labour.[5]

5.5 The Law of Value in Marx

A more careful examination of the issue at hand shows that even though we suppose that commodities embody different quantities of labour, these are not automatically comparable. The labour as in the case of the commodity has a dual property. On the one hand, it is useful labour, that is, expenditure of human labour power in a concrete form and for a specific purpose and with this property creates use values. On the other hand, it is abstract labour, that is, expenditure of human labour power without specific characteristics and creates the value of the commodity in which it is embodied. This abstraction allows the realisation of the exchange. The same act of exchange ascertains the reality of abstract labour. A specific quantity of labour of one producer is equated with a specific quantity of labour of another producer. Marx argues that in the process of exchange, the concrete labour (*i.e.*, the labour that creates use values) is transformed to abstract labour, and he calls the labour that creates exchange value abstract labour.

Hence, the basic contradiction in the normal operation of capitalism appears. We know that every individual labour process is assumed privately, as if it were independent from the rest and with profit as the principle goal. At the same time, every labour process presupposes that the other labour processes will take place on the right time and in the right proportions. This is equivalent to saying that not only that there will be buyers for the commodities produced, but also that there will be sellers of the means of production through which commodities will be produced at the time that are demanded and all these will take place repeatedly, so as to meet the necessary and sufficient conditions for the continuous reproduction of society.

Consequently, every private independent labour presupposes the social division of labour. Moreover, to realise this presupposition in praxis, the private and anarchic labours must in someway co-ordinate and become part of the social division of labour. This co-ordination takes place in the sphere of exchange, where every individual producer ascertains (through prices and profits) whether the results of the labour process that he has been engaged in are satisfactorily rewarded. If not, then he will modify his behaviour accordingly (Shaikh 1981). In short, Marx argued that labour time also regulates the surface phenomena of

[5]The wage differentials, as a rule, are taken as approximate indications of the skill-differences. Higher wages correspond to more skilled labour, and so forth. Of course, one may attribute, at least partly, the wage differentials to other factors such as un-ionisation, or monopoly (Botwinik 1994).

capitalism, such as the price of commodities. Hence, the law of value – according to which the socially necessary labour time is directly and indirectly embodied in a commodity – is the regulator of the movement of market prices. Prices are the means through which capitalists realise their profits and losses and regulate their behaviour accordingly. Consequently, a dual relation follows:

(1) Prices and profits are the immediate regulators of reproduction.
(2) Socially necessary labour time constitutes the regulator of prices and profit and, therefore, of social reproduction.

The operation of this dual relation is what Marx calls the law of value whose role is analogous to Adam Smith's "invisible hand", for it provides an explanation of how the capitalist society reproduces itself and the various scales of its reproduction. What Adam Smith calls "invisible hand" is called by Marx the "law of value", a law that holds only in conditions of capitalism, where the majority of goods making up the material basis of the social reproduction are produced without any direct connection to social needs. By contrast, goods are produced by independently undertaken private labour processes and each one of these is motivated by profit.

It follows that both abstract labour and value are real, that is, measurable quantities. In capitalist production, use values are produced as commodities, and the whole production process is characterised by the fact that the exchange value of a commodity constitutes the dominant side of the entire procedure. Consequently, use values are considered as commodities from the time of production (conception) and the labour is, at the same time, concrete and abstract from the outset of the production process. It is important to stress at this point that the labour that is employed in the production of commodities produces value, while the exchange simply realises this value in money form. This point is worth reiterating because it has been the source of much discussion in the post-Marxian literature, where it has been argued that the exchange process is what validates the labour expended in production and so value is realised in the process of exchange. This view point has been supported by many Marxist economists, starting from Rubin (1928) and continuing to more recent authors (Mohun 1994).[6]

5.6 Money and Price

The term value in Marx takes on a specific meaning according to which the commodities reflect in exchange the existence of abstract labour time expended on their production and this abstract labour time is the regulator of their prices. The price of a commodity in other words reflects the quantity of abstract labour time

[6]Usually, those who consider circulation (or exchange) as the determinant of abstract labour time claim that this view originated in Rubin (1928). However, this is not exactly right as Rubin (1928, p. 148, 155) also argues that abstract labour is determined in production and is modified in exchange.

required for its production.[7] The price of a commodity, that is to say, the form of value in the sphere of exchange, is always monetary. At first sight, it is obvious that there are so many ways to express the exchange value of a commodity as the number of commodities against which it is compared. Yet such a comparison procedure is complex on the one hand and presents difficulties in the operation of exchange on the other. Consequently, it becomes necessary to find a particular commodity in terms of which the other commodities could express their value. This commodity is called the *general equivalent* or the *universal commodity*. Historically, the precious metals such as gold and silver have played this role. Gold can function as the universal commodity, since it contains value, and, therefore, exchange value. Consequently, the money price of a commodity constitutes the external measure of its exchange value, that is, the form that value takes in exchange.

Thus, the value of a commodity and its price are two strictly related but at the same time distinct entities. The value of a commodity is measured by the quantity of necessary labour time and its price by the quantity of money that the commodity commands in its exchange. Thus, if, for example, gold plays the role of money, then the price of a commodity equals a specific quantity of gold. For example,

1 h of labour = €1 = 1/2 ounce of gold.

It follows that the value of a commodity and its price are altogether different entities, for if they were not, then it would be meaningless to investigate their relationship. The transformation of values (direct prices) into prices of production that we study in the next sections refers essentially to the way in which the value and its monetary expression (the direct price) of a commodity set the limits within which the price of production varies.[8]

5.7 Surplus Value and Profit

Marx made the distinction between labour and labour power, a distinction that is absolutely crucial for the understanding of the source of profit.

> By labour-power or capacity for labour is to be understood the aggregate of those mental and physical capabilities existing in a human being, which he exercises whenever he produces a use-value of any description. (*Capital*, vol. I, p. 186)

In other words, labour is the utilisation of labour power, that is, of the useful labour that a labourer performs in a specific period of time. As we know, the value of a commodity is equal to the socially necessary quantity of abstract labour time (direct and indirect) which is contained in it. The value of labour power, accordingly, is equal to the socially necessary labour time required for the production of

[7]The magnitude of value as Marx points out is determined by the current cost of reproduction in terms of abstract labour and not from the historical cost of reproduction (*Capital*, vol. I, p. 39).

[8]The theory of money and the theory of competition refer precisely to this relation.

commodities that a worker purchases with his money wage to reproduce himself and his family. But the labour time embodied in the commodities normally purchased by the worker for the reproduction of himself and his family in a day is less than the labour time that a worker actually offers to the owner of capital during the same time period. The result is that for any given time period, the worker produces more value than the wage equivalent which is paid by the owner of capital for the use of the labour power. This difference, Marx calls "unpaid labour" and "surplus labour".

Labour time in Marx's work constitutes the cornerstone of the reproduction process, since it is labour that produces and reproduces not only the social wealth but also the social relations in the context of which this wealth is created. All societies, to reproduce themselves, require labour that must be used to produce the material requirements of reproduction. In all class societies, a particular social class performs more work than what is required for its own reproduction with the excess labour being appropriated by the dominant classes through property relations, traditions, the legal system and also force. Such exploitative relations are quite transparent in pre-capitalist modes of production (*e.g.*, slavery and feudalism), whereas in capitalism they are embedded in monetary transactions that give the impression of equal, and, therefore, fair exchanges. Marx was the first to argue that in capitalism workers are exploited not because they are not paid their full wage, but because even when they are paid their full wage they can only buy the basket of goods required for the reproduction of their capacity to work (their labour power), which is acquired through what is only a portion of the total labour time that they expend during a workday. The difference between total labour time and that required to reproduce the workers' capacity to work is called *surplus labour time* and its monetary expression, the *surplus value*, is appropriated by the propertied classes (capitalists and landlords) and the state. The wealth accumulated in a society is directly related to the amount of surplus labour time, which is inversely related to the necessary labour time.

Marx, to prove the aforementioned proposition, supposes initially that the money price of every commodity is proportional to the total abstract labour time, that is, socially necessary for its production. Regarding the category of wage, this means that wages are proportional to the labour time (v) that workers must spend to produce the necessary means for the reproduction of themselves and their families. Labourers are considered as selling their labour time, that is, their capacity to work. The production of a commodity requires two types of abstract labour time: dead labour time (c) and living labour time (l). Dead labour refers to the inputs used in the production process (*i.e.*, raw materials as well as depreciated capital). Marx calls these inputs constant capital, since their value is recovered in its entirety when the commodity is sold. In other words, the value of the new commodity already includes the value of constant capital. Living labour, on the other hand, refers to the direct flow of labour, for example 8 h a day times the number of workers, which the capitalist appropriates from the workers, and which creates new value. If the price of a commodity is proportional to its value (λ), then we may write:

$$\lambda = c + l$$

The corresponding money cost of production is proportional to $c + v$. From here it follows that profits exist if and only if

$$c + l > c + v \text{ and } l > v, \text{ consequently } s = l - v > 0$$

The labour time that exceeds the amount that is required for the reproduction of the labouring class is called surplus value. In such a case, where commodities are sold at their values, exploitation is proved not in the easy case, that is, in the case where a worker is underpaid, but rather in the case where the worker is paid the normal wage. Marx calls that part of total capital that pays wages variable capital. The name is not accidental but it indicates that labour power enters in the process of production as the price of labour power (*i.e.*, wage) and generates more value than the value required for of its own reproduction.

5.8 Marx's Theory of Money

Marx's theory of money is extensive, and in a book on the history of economic thought, we are bound to present only the absolutely essential aspects of his theory, which may provide the solid theoretical foundation one needs to address even current monetary issues. Marx's writings on money are scattered throughout his work and an extensive discussion would require a rather specialised book. This is the reason why our attention focuses on the functions of money and then on Marx's critique on the quantity theory of money; in other words, our analysis refers basically to *Capital* Chaps. 1–3 and the *Critique of Political Economy*, while issues of credit and of other more advanced material are necessarily left out of our analysis.

Money in the analysis of Marx is a commodity that has been selected in the process of exchange for a range of convenient and desired properties (*e.g.*, it is easily divided, transferred, recognised, stored, etc.) and that serves as the general equivalent or universal commodity. It is obvious that the universal commodity has value itself, as is the case with the other goods against which it is being exchanged. The functions of money are:

(1) *Measure of value or unit of account and standard of price.*

This function does not require the physical presence of money. Hence, money may have a virtual presence; that is, it can be used mentally to assign monetary prices (exchange ratios) to commodities. The function of money as a measure of value is very important and occurs prior to the actual exchange of commodities. An example that can elucidate this function concerns periods of galloping inflation in one country, where all commodity prices and assets may be quoted in a foreign currency (*e.g.*, dollars or euros) because domestic prices rise quite frequently even during the same day. In such situations the prices of commodities are quoted in terms of a stable currency (say US dollars), which at the time of actual transaction are converted to domestic currency (the official money) according to its exchange rate against the dollar.

(2) *Medium of circulation.*

Here, we discern two sub-functions (a) medium of exchange and (b) means of payment. The function of money as an instrument of payment introduces the operation of credit.

(3) *Money as money.*

The previous two functions do not necessarily require the physical presence of money but this is not the case with the function of money as money. Marx does not claim that the universal commodity must be gold, but rather argues that gold, during a long historical process, eventually dominated over other possible "money" candidates because of a number of desirable properties that no other commodity possesses.

On the basis of the aforementioned functions of money, we can form the Marxian quantity theory of money, starting from the usual equation of exchange (see the appendix to Chap. 3). According to this, if the output of an economy, $\bar{Q}$, is given, assuming full utilisation of the productive capacity of the economy[9] and if the velocity of circulation, $\bar{V}$, is also given because of institutional factors, then we know that the usual quantity theory of money posits that the arrow of causality runs from left to right, that is, we have:

$$M\bar{V} \rightarrow P\bar{Q}$$

Here, changes in the quantity of money (M) are translated into changes in the price level (P). The proportional relationship between M and P is true regardless of the form of money. In neoclassical economics in general and in the quantity theory of money in particular, the function of money as a *means of account*, which is in the realm of the virtual, coincides with the function of money as a means of circulation, in the sphere of the actual, and from this alleged coincidence emanates the notion of money as a veil that covers the real economy. While most macroeconomists certainly argue for the non-influence of money on the real economy in the long run, there are also quite a few others, such as the New Classical Economists (see Chap. 11), who would argue that money plays no role even in the short run. In neoclassical economics, by and large, the behaviour of the monetary economy is put on par with the barter system, where equilibrium is the rule rather than the exception and the presence of money serves only to fix the price level.

For Marx on the other hand, the function of money as a measure of value need not necessarily coincide with its function as a medium of exchange.[10] Money as a means of circulation virtually guarantees its function as a means of calculating prices and that the quantity of money in circulation (hence are included all forms of

[9]It is important to note that the full utilisation of capital does not imply the full employment of labor.

[10]For example, in conditions of galloping inflation prices of goods are valued against third stable currencies (*e.g.*, dollar or euro), while actual transactions are carried out in terms of local (official) currency.

money) should adapt to the needs of the circulation of goods whose quantity varies over time. Therefore, the amount of money should be variable so as to correspond to the needs of circulation of commodities. One corollary of this is that the commonly used term velocity of money is devoid of any meaning, because in reality it is the velocity of circulation of commodities that gives the impression of the circulation of money. The reason for this is that as commodities are entering or leaving circulation, money still remains there, in order to mediate other transactions, and so on.

Thus, it is the circulation of commodities that determines the quantity of money and not *vice versa*, something that Marx repeats several times in *Capital* I (Chap. I, Sect. 3). In addition, money and goods move in opposite directions. Goods are brought by the seller to the buyer and money from the buyer to the seller. This, coupled with the fact that the quantity of money circulating is not identical with the quantity of goods in circulation, creates the impression that money is the key factor that determines the circulation of commodities.

In what follows we focus on Marx's quantity theory of money. We know that at the level of analysis in the first volume of *Capital*, the regulators of the market prices of commodities are the direct prices, that is, the prices which are proportional to values (*e.g.*, Shaikh 1977, 1980a). If by p_0 we denote the direct price of a commodity, then we have:

$$p_0 = \frac{\lambda}{\lambda_g} = \frac{\text{value of commodity}}{\text{value of gold}}$$

Assuming that the unit value of the commodity is equal to 10 h, while a unit of gold (oz) embodies 2 h of labour per oz, then we have:

$$p_0 = \frac{\lambda}{\lambda_g} = \frac{10\,\text{hours/unit output}}{2\ \text{hours/oz}} = 5\,\text{oz/unit of output}$$

Let us further suppose that the currency in circulation is the euro and that its exchange rate against gold, denoted by R, is assumed to be at the level where 1 oz gold = 2£.[11] Therefore, the direct price of the commodity expressed in euros will be:

$$p = \frac{\lambda}{\lambda_g} R = p_0 R = (5)\ \text{x}\ (2) = 10\,€/\text{unit of output.}$$

We observe that the price level may change for a number of reasons, that is, a rise in the price $p^{€}$ may be due, for example, to a reduction of λ_g with everything

[11]The discussion is hypothetical in the sense that there is no such exchange rate and official currencies are not converted to gold; this does not mean that gold is irrelevant to modern monetary issues. It was not long ago that the dollar was supposed to be exchanged against gold at the ratio \$35 = 1 oz gold, according to the Bretton Woods agreements signed in 1944.

else (λ, R) constant or due to a decrease in λ_g with λ and R decreasing even more, just to mention two cases. The amount of money required for the movement of goods (M_{circ}) thus equals:

$$\mathrm{M}_{\text{circ}} = \frac{PQ}{k} = \left(\frac{\lambda}{\lambda_g}\right)(R)\left(\frac{1}{k}\right)Q$$

where k is the velocity of circulation and PQ is the product of the general price level times the output produced.[12]

As we noted, the quantity of money in circulation is a variable magnitude depending on the needs of the circulation of commodities. The quantity of money in circulation should not be confused with the stock of money in the economy, $\bar{\mathrm{M}}$, which equals, $\bar{\mathrm{M}} = \mathrm{M}_{\text{circ}} + \mathrm{M}_h$where M_h stands for the amount of money stored not in the usual sense but rather in the sense of reserves that one needs to possess to deal with varying situations. In other words, we are speaking of planned or active, and therefore not passive, reserves. In Marx's analysis, it is not the quantity of money that determines the general price level, but rather the general price level is determined by the value ratio (λ/λ_g), which in turn (given the level of output, Q) determines the quantity of money in circulation, a result which is quite opposite to the quantity theory of money in its classical (see Chap. 3) or modern (see Chap. 13) version. Marx by referring to Tooke's empirical studies about the quantity theory of money made the following remark:

> But continued investigation of the history of prices compelled Tooke to recognise that the direct correlation between prices and the quantity of currency presupposed by this theory is purely imaginary, that increases or decreases in the amount of currency when the value of the precious metals remains constant are always the consequence, never the cause of price variations, that altogether the circulation of money is merely a secondary movement. (Marx 1859 [1981], p. 186)

5.8.1 Convertible Paper Money

Let us suppose now the case of convertible (to gold) symbols of money (convertible tokens).[13] In this case, instead of having a direct circulation of gold coins, we have a circulation of paper notes that are covered by the relevant quantity of gold.[14] The government (or generally the issuing authority, *i.e.*, the central bank and in the past even private banks) guarantees the convertibility to gold of the symbols of money that are put into circulation according to the amount specified in each symbol of money.

[12]This analysis is based on Shaikh's (1980a).

[13]We can assume any type of symbols of money, *e.g.*, bronze coins or other symbols, which do not have a greater intrinsic value than that they symbolise. To facilitate the presentation, we assume notes converted into gold by the issuing authority.

[14]Historically, gold coins were circulated upon which an indication of their gold content together with the «head» of the king or emperor was inscribed.

What would happen if, for example, the State issued an excessive quantity of money convertible into gold? The answer depends on the level of productive capacity of the economy. Assuming that the economy's productive capacity is underutilised, the excess quantity of money would lead, *ceteris paribus*, to an increase in output and employment. But if the economy were using its full production capacity, the excess quantity of money would be withdrawn from circulation in hoarding (M_h). If, however, we assume that the amount of money that was already hoarded was at the normal level, then the excess quantity of money would appear as an additional demand for gold which would be an indication that the government over-issued money. Alternatively, this mechanism works as follows: the market price of gold is governed by the current currency exchange rate to gold (R), which is the official exchange rate of gold. If the market price of gold is greater than R, then the public will buy more gold and it is expected that it will seek to convert its money into gold at the official rate. If the government seeks to stabilise the exchange rate, then it should buy all the surplus notes with its gold to absorb the excess quantity of money. It is likely, however, that the government may reduce the exchange rate R at the market price, so that any monetary unit that it has issued will now be worthless. In other words, money is losing part of its value, and, therefore, we have inflation. If, on the other hand, the government issues less money than the amount needed in circulation, then we would expect to observe exactly the opposite results. As a consequence, the State should suspend the issuing of currency to comply with the requirements of circulation or even to interrupt the issue of money because of the lack of an adequate quantity of gold to back up the circulation of convertible money. In this case, we have:

$$p^{€} = \frac{\lambda}{\lambda_g} R^{€}$$

where $R^{€}$ is related to gold in a looser manner than in the case of golden coins and this rather looser relationship allows the state to proceed to currency devaluations; in other words, the state may reduce the coverage of € with gold.

Once again we observe that the price level, $p^{€}$, depends on the influence of the following set of variables λ, λg, and $R^{€}$. As a consequence, $p^{€}$ is not related to the quantity theory of money according to which changes in the amount of money translate into changes in prices. In our case, changes in prices depend on a triad of forces (λ, λg, and $R^{€}$), which do not exist in the quantity theory of money. Of course, in both theories the increase in the money supply may lead to an increase in the price level. Marx's analysis, however, is based on an entirely different transmission mechanism and predictions are expected to differ quantitatively not qualitatively.[15] It would be totally absurd if the major proposition of the quantity theory

[15]In our analysis of monetarism and the associated notion of the quantity theory of money, we also discuss analogous transmission mechanism, applicable to both fixed and variable exchange rates.

of money (in any version) predicted wrongly the effects of the change in the supply of money on the price level. The issue is not only the rise in the price level resulting from the increase in the quantity of money but also the manner and extent of the increase. Also, the answers that the Marxian theory of money and the quantity theory of money are not the same, although both predict, other things being constant, the rise in the price level.

5.8.2 Non-convertible Paper Money

Non-convertible currency symbols are notes the convertibility of which, into a certain quantity of gold, is not guaranteed by the State (the issuing authority in general). The State simply issues money and ensures its function as a means of circulation but not its convertibility to a specified amount of gold or other precious metals. In other words, the State issues notes to meet the needs of circulation and, at the same time, we assume that there is a market for gold and that although the domestic currency is not immediately converted to gold; nevertheless, it is attached to a currency that is convertible into gold and that this convertibility is guaranteed by a government or an international organisation. In fact, this was an economic reality until around 1971; that is, the US dollar was convertible into gold at a fixed rate and the other countries pegged the exchange rate of their currency to the dollar. Therefore, currencies were not convertible into gold domestically but internationally. Now, let us suppose for example that the government whose currency is converted to gold (hence the US government from 1945 to 1971 that had set the parity rate at $35 per ounce of gold) issued an excess quantity of (paper) money. What would happen then? The excess amount of notes (hence dollars) will appear in the gold market, which means that the price of gold will tend to increase. It is obvious then that the parity between gold and dollar should be changed and be stabilised at a higher level and that the price level in terms of the dollar will increase as it will happen also with the price level internationally. If however, the US government wanted to keep the parity rate at $35 per ounce of gold, she should buy up the excess amount of dollars by selling its gold reserves. Let us now suppose a country that pegged its currency to the dollar and whose government happens to over-issue paper money. This excess would appear in the exchange rate market. The domestic currency would tend to depreciate in terms of the international money (dollar) unless the government had enough foreign (dollar) reserves to buy up the money that it issued in the effort to preserve the fixed exchange ratio. Therefore, non-convertible notes involve flexibility in exchange rates.

In the situation where there is no convertibility to an international currency, it is again expected that gold and the foreign exchange markets should enter the picture. The truth is that even the Bretton Woods system did not work as it was designed inasmuch as the convertibility of dollar to gold was not taking place as freely as one would expect. In reality, no county would like to supply large quantities of its gold reserves to maintain a specified parity ratio. This was the situation with the US,

when pressure was applied to her by some counties (France and W. Germany) that demanded to sell their excess dollar reserves for gold at the specified parity rate; ultimately, the US unilaterally brought the Bretton Woods agreements to an end. It is important to note that one major difference of the current situation from that which existed prior to 1971 is that the convertibility of the international currency (dollar) to gold is no longer guaranteed at a fixed parity rate. In other words, the gold market has been "liberated" as is the case with most markets. This however should not lead to the conclusion that governments (central banks) buy and sell their gold reserves freely. Gold reserves continue to be an important asset among those possessed by central banks. In Switzerland for example the backing up of its currency by gold reserves makes the Swiss franc one of the major currencies internationally.

5.9 The Transformation Problem

The publication of the first volume of *Capital* (1867) did not attract the attention of academics. The other two volumes of *Capital* published posthumously, Volume II in 1885 and volume III in 1894 attracted the attention of some mainstream economists. Böhm-Bawerk (1852–1914) was from the first orthodox economists that criticised Marx's overall work. Specifically, Böhm-Bawerk argued that there is a logical inconsistency between the first two volumes of *Capital*, in which the analysis is based on labour values (embodied labour time), and the third volume of *Capital*, in which Marx's analysis is in terms of prices of production, that is, prices that are sufficient to pay for the inputs and at the same time secure a normal (average) rate of profit on the invested capital, hence, the so-called "Great Contradiction" between the first two and the third volumes of *Capital*, since in a single market we cannot have two price systems, one in terms of values and the other in terms of prices of production. Marxist authors reacted to Böhm-Bawerk's critique by downplaying its importance. For example, the usual way to circumvent this critique was to claim that the determination of relative prices was not Marx's purpose and that Marx was interested in qualitative and not in quantitative relations, and so forth. But as we argued earlier, Marx's labour theory of value has a quantitative aspect along with the qualitative one.

Starting from the formula of value $c + l = c + v + s = \lambda$, Marx forms three ratios whose evolution reflects the state of the economy. Let us suppose, for the sake of simplicity and clarity of presentation, an economy with circulating capital only; in other words, an economy with no fixed capital. We form the following "great ratios" in Marxian economics:

The ratio of constant to variable capital (c/v), which is called value composition of capital and expresses the relation of dead to living labour, or of the constant capital to the value-creating variable capital. We remind that the variable capital not only reproduces its own value but also creates additional value, that is, surplus value.

The ratio of surplus value to variable capital (s/v), which is called the rate of surplus value. If we suppose that the means of workers' subsistence (*i.e.*, the real wage) require 4 h of labour and if the worker works 8 h, the extra 4 h of labour constitute the unpaid part of labour time and the rate of surplus value is 100%.

The ratio of total surplus value to the sum of constant and variable capital $r = s/(c + v)$, which Marx calls the rate of profit. It is important to note that this is an elementary definition of the profit rate and it is based on the assumption that there is no fixed capital.

We have shown that the direct price is a form of value that is manifested in the sphere of circulation; value as Marx argues takes on the form of the production price. The movement from one kind of prices to another more complex kind of prices has come to be known as the transformation problem.

Let us suppose that the regulator of the market price is the direct price, that is, the price proportional to value. We discover that even in this seemingly simple case the relation between labour values and direct prices are already complex enough. Let us for example symbolise by p_0^x the direct price of commodity x, which is equal to the ratio of the value of commodity x to the value of commodity gold, that is

$$p_0^x = \frac{\lambda_x}{\lambda_g} = \frac{\text{value of commodity } x}{\text{value of commodity gold}}$$

From the definition of the direct price of a commodity, it follows that it is possible to increase it if the value of the commodity rises with the value of gold remaining constant or if the value of the commodity decreases with the value of gold decreasing even more, and so forth. Clearly, even in this simple case the determination of the direct price of a commodity is much more complex from the mere determination of its value, as Marx's analysis in volumes I and II of *Capital* has shown.

The idea that the direct price determines the market price does not mean that the two prices are identical. In general, the two prices differ from each other ($p_o \neq p_m$), because of the permanent imbalance of supply and demand. If, for example, demand exceeds supply in a particular industry it follows that $p_m > p_o$. This implies that supply in the future will be rising at a faster rate than demand. In the case where $p_m < p_o$, we will have a contraction of the accumulation of capital, supply will fall short of demand, the market price will increase and so forth. Consequently, Marx's approach (in the first two volumes of *Capital*) can be cast as follows:

$$\frac{\lambda}{\lambda_g} = p_o \rightarrow p_m \,\text{simple transformation},$$

which is the simplest case, where market price is determined from the direct price, which in turn is determined by the value of the commodity in question relative to the value of gold.

In the third volume of *Capital*, there develops a much more complex discussion according to which the direct prices-market prices relation is mediated by a third kind of prices, called prices of production, which we symbolise by p^*. Consequently, we have:

$$\frac{\lambda}{\lambda_g} = p_o \rightarrow p* \rightarrow p_m \text{ complex transformation,}$$

The latter more complex case refers to the transformation problem of direct prices to prices of production, as more concrete regulators of market prices p_m.

With respect to the rate of profit, Marx's argument in the first two volumes of *Capital* is that all prices are proportional to their labour time. Under these circumstances, he shows that profits are created in the sphere of production. Subsequently, he introduces the exchange in terms of prices that deviate from their proportional relation to the labour time. The transformation problem in Marx is about an already complex form of value to another even more complex form of value. Consequently, we refer to two kinds of prices: direct prices and prices that contain equal profit rates, that is prices of production (p^*). The difference between the two types of prices lies in that in direct prices the surplus value (or profit) is proportional to the variable capital, whereas in prices of production the profit is proportional to the total invested capital. Thus, we have two different allocations of total surplus value. This investigation of the relation between the movement from one set of prices to another more concrete set of prices is called the transformation problem.

5.9.1 Marx's Solution

Following Marx and assuming with him a given and uniform real wage and length of working day, if exchange takes place in terms of values (direct prices), it will follow that the rates of surplus value are equal across sectors.[16] In such a case, however, the capital-intensive sectors would extract less surplus value than the labour-intensive ones, since they employ less labour time.[17] Such a distribution of surplus value lies outside the logic of capitalism, a system of production that is oriented towards the extraction of the maximum possible profit. Consequently, the owners of capital invest, and, therefore, increase the capital–labour ratio in the effort to increase their profits. If by investing, they would just reduce their profit, then there would be no motive whatsoever to invest. Therefore, what is consistent with the nature of capitalism is the equalisation of the rates of profit which means that surplus value is reallocated in different sectors of production in proportion to the invested capital. Industries with high organic (or value) composition of capital $(c/v)_j$

[16]Given v and l, we will have $(s/v)_i = (s/v)_j$ where i and j are two different industries.

[17]Consequently, there would not be any motive for investment in constant capital.

extract a higher proportion of surplus value from what they really produce, while the converse is true for the labour-intensive industries. In capitalism, it is the rate of profit that is tendentially equalised between sectors and not the rate of surplus value. Competition and the mobility of factors of production imply the inter-sectoral equalisation of the individual rates of profit to an average (r), which implies unequal rates of surplus value. Thus, we can write for the rate of profit of industry j:

$$r_j = \frac{s_j}{c_j + v_j} = \frac{s_j/v_j}{c_j/v_j + 1}$$

It is obvious that to have equalisation of the rates of profit between sectors with different value compositions of capital (c/v), it follows that there will be deviations of prices of production of commodities from their values, because the profit of each sector will differ from its surplus value.

For the sake of simplicity and following at the same time a whole theoretical tradition, we examine a simple economic system with three departments of production.[18] The department producing means of production (I), the department producing means of consumption for the working class (II) and the department producing luxury goods for capitalists (III). For the economy to be reproduced on the same scale, the fulfilment of the following conditions is required:

$$(1)\ c_1 + v_1 + s_1 = c_1 + c_2 + c_3$$
$$(2)\ c_2 + v_2 + s_2 = v_1 + v_2 + v_3$$
$$(3)\ c_3 + v_3 + s_3 = s_1 + s_2 + s_3$$

The left-hand side of the equations represents the supply of each department from the three departments, while the right-hand side refers to the demand of each department. We observe that the output of the first department c_1 is demanded from the first sector, while c_2 is the demand for means of production of the second department and c_3 is the demand for means of production of the sector of luxury goods. Considering the output of department II, we see that v_1, v_2 and v_3 represent the demand of the workers of each of the three departments for consumer goods, while the output of department III, that is, s_1, s_2 and s_3, represents the demand of capitalists of the three sectors for luxury goods, which do not constitute inputs for the other sectors. The three equations form the conditions of simple reproduction, a rather theoretical situation, since expanded reproduction is what characterises the evolution of capitalist economies.

Marx proposed the following transformation of values into prices of production. First of all, he estimated the rate of profit in terms of values according to the formula:

$$r = \sum s_i / \sum (c_i + v_i)$$

[18]By department we mean a collection of industries producing the same category of commodities.

The (value) rate of profit is used for the estimation of Marx's prices of production in the three sectors of a simple reproduction economy:

$$p_1 = c_1 + v_1 + r(c_1 + v_1) = (1 + r)(c_1 + v_1)$$

$$p_2 = c_2 + v_2 + r(c_2 + v_2) = (1 + r)(c_2 + v_2)$$

$$p_3 = c_3 + v_3 + r(c_3 + v_3) = (1 + r)(c_3 + v_3)$$

Marx argued in the third volume of *Capital* that prices of production can be determined from a system of equations such as that seen earlier and that though they are expected to be (in general) different from the respective values ($p_j \neq \lambda_j$), these differences can be determined on the basis of the organic composition of capital. For example, if the organic or value composition of capital of a department is greater than the average organic composition of the whole system, then it follows that the price of production of this department will be higher than its value and *vice versa*. In the special case, where the value composition of capital of a department is equal to the average, its price of production will be equal to its value.

This deviation of prices of production from their values led Adam Smith to abandon his labour theory of value (see Chap. 3), while David Ricardo argued that these deviations are relatively small and their direction predictable. For Ricardo, the quantity of labour that is used for the production of commodities continues to be the fundamental determining factor of prices (see Chap. 4). Marx argued that these deviations not only do not contradict the labour theory of value but also are expected and are fully compatible with it.

The transformation of direct prices to prices of production implies a redistribution of the value and the surplus value that have been produced between the different branches of production. From the moment, however, that we simply have a formal change in distribution, the total magnitudes that have been created in the sphere of production remain invariable. Thus, the following two so-called "invariance conditions" must hold:

- The sum of direct prices must be equal to the sum of prices of production.[19]
- The sum of surplus value (in terms of direct prices) will be equal to the sum of profits.[20]

These two "invariance conditions" have been imposed as a test of logical consistency of Marx's labour theory of value. The idea is that value and surplus value are created only in production and that their various forms that they take on in circulation (price or profit) cannot change their initial size.

[19]Formally, we have $\sum(c_i + v_i + s_i) = \sum(c_i + v_i) + r\sum(c_i + v_i)$, a condition that can be easily proved if we substitute the rate of profit by its equal.

[20]To prove this equality, we substitute the rate of profit in the relation $\sum s_i = r\sum(c_i + v_i)$.

5.9.2 The Critique of von Bortkiewicz

In 1907 Ladislaus von Bortkiewicz (1868–1931) pointed out that Marx, in the third volume of *Capital,* committed two mistakes:

(1) In Marx's transformation which can be written as $p_j = (1 + r)(c_j + v_j)$, what is really transformed is the value of output of good j, while the inputs that are required for its production, that is c_j and v_j as well as the profit rate continue to be expressed in terms of values. In other words, we cannot have two different price systems in the same economy and at the same time.
(2) In Marx's transformation the conditions of simple reproduction, which according to Bortkiewicz must hold, is a necessarily requirement for the consistency of Marx's scheme.

In his effort for constructive criticism, Bortkiewicz expressed the system of simple reproduction in terms of prices of production in the following way:

$$(1 + r)(c_1 p_1 + v_1 p_2) = (c_1 + c_2 + c_3)p_1$$

$$(1 + r)(c_2 p_1 + v_2 p_2) = (v_1 + v_2 + v_3)p_2$$

$$(1 + r)(c_3 p_1 + v_3 p_2) = (s_1 + s_2 + s_3)p_3$$

Hence, we have a system of equations that is being used for the solution of the transformation problem, where p_1, p_2 and p_3 are the prices of production that we want to estimate. We have a system of three equations and four unknowns, the three prices and the rate of profit. Bortkievicz suggested to take the price of the third sector as a numéraire and to solve the system for relative prices. Thus starting from values we arrive at production prices and the rate of profit expressed in terms of prices of production.

Bortkiewicz's solution is not without its problems since of the two invariance conditions only the second holds true.[21] The first invariance condition does not hold, since, in general, the sum of values will not be equal to the sum of prices of production except for the uniform organic composition of capital case.

An interesting consequence of Bortkiewicz's solution is that the rate of profit depends exclusively on the conditions of production of the first two sectors. Changes in the third department do not change the rate of profit of the economy. The proof of this proposition requires linear algebra. However, we may, intuitively examine its logic. If we suppose that for some reason the price of the first sector increases, then this will affect the prices of all the three sectors of the economy as well as the average rate of profit, since the output of the first sector is an input to all

[21]In fact, from the three equations of Bortkievicz we get: $(1 + r)(p_1 \sum c_i + p_2 \sum v_i) = p_1 \sum c_i + p_2 \sum v_i + p_3 \sum s_i$. By assuming that $p_3 = 1$ and after some manipulation, we get: $r(p_1 \sum c_i + p_2 \sum v_i) = \sum s_i$

the three departments. The same holds true for the second department. In the third department, however, any change in its price will confine its effects within the sphere of the third department only and will not affect the prices of the other departments and the rate of profit. The idea is that the output of the third sector is not an input to the other two sectors.

5.9.3 *Shaikh's Solution*

It seems that the transformation problem found a solution in Shaikh's (1978, 1984) iterative approach.[22] According to Shaikh, Marx's solution was just the first step to the full solution, which, however, he could not derive since the theorems of linear algebra necessary for this purpose were discovered many years later.[23] It is important to point out that Marx's transformation refers to the formal change of value from direct prices to prices of production. Consequently, the transformation problem is mainly an exercise in logic, since nothing essential changes. Marx starts from the analysis of values of commodities in terms of labour (which remains the same throughout the procedure, see the middle columns in Table 5.1), which he transforms then into direct prices by assuming that ½ labour hour equals 1€ or 1 labour hour equals 2€. These direct prices are in turn transformed into prices of production. Marx's transformation, however, is not complete in the sense that inputs are still expressed in terms of direct prices. Based on the usual numerical example of Bortkiewicz, we start with the labour values that remain invariable throughout the transformation, $c + v + s = \lambda$ and indicated in Table 5.1.

We formulate the direct prices $c + v + s = p^0$, which are shown in Table 5.1, part 1A. On the basis of direct prices, we estimate the value (in terms of direct prices) rate of profit:

$$r^0 = \sum s_i^0 / \sum (c_i^0 + v_i^0) = 400/1350 = 29.6$$

With the aid of r^0 we estimate the first step prices of production according to the formula:

$$\left(c_i^0 + v_i^0\right)\left(1 + r^0\right) = p_i^1$$

[22]Shaikh presented a conceptual and mathematical solution to this problem in his doctoral dissertation in 1973. Morishima (1973) and Okishio (1974) arrived at a similar mathematical solution; however, even in this case, the transformation problem is basically a conceptual and not just a computational problem as in the cases of Morishima and Okishio.

[23]We refer to the Perron–Frobenious theorems. For a summary presentation of these very important theorems, see Pasinetti (1977).

Table 5.1 Transformation of values to prices of production

	€c	€v	€c + €v	c	v	s	λ	€p^I	€π	r %	ψ %
1A	450	180	630	225	90	60	375	750	120	19.1	–
	200	240	440	100	120	80	300	600	160	36.4	–
	100	180	280	50	90	60	200	400	120	42.9	–
Total	750	600	1,350	375	300	200	875	1,750	400		
1B	450	180	630	225	90	60	375	817	187	29.6	1.09
	200	240	440	100	120	80	300	570	130	29.6	0.951
	100	180	280	50	90	60	200	363	83	29.6	0.907
Total	750	600	1,350	375	300	200	875	1,750	400		
2A	490	171	661	...	...	...	...	817	156	23.5	...
	218	228	446	...	...	...	...	570	124	27.9	...
	109	171	280	...	...	...	...	363	83	29.6	...
Total	817	570	1,387	...	...	...	...	1,750	363		
2B	490	171	661	...	...	...	...	834	173	26.2	1.02
	218	228	445	...	...	...	...	563	117	26.2	0.983
	109	171	280	...	...	...	...	353	73	26.2	0.973
Total	817	570	1,386	...	...	...	...	1,750	363		
...	...	...	...					...	...	...	...
	504	168	672	...	...	...	...	840	168	25	1
	224	224	448	...	...	...	...	560	112	25	1
	112	168	280	...	...	...	...	350	70	25	1
Total	840	560	1,400	...	...	...	...	1,750	350		

and present them in Table 5.1, part 1B. These are the prices of production that Marx calculates and they essentially constitute the first step of an iterative procedure that Marx started, but did not continue. There are references in *Capital* according to which it can be deduced that though Marx was aware of the semi-finished character of his procedure, he deemed however that it was not necessary to continue with a burdensome arithmetic to derive the full solution, which in any case would not be qualitatively different from that of the first step prices of production.[24] In fact, were we to continue Marx's procedure, we would also have to transform the inputs according to the new prices. For this purpose, we formulate multipliers, which are derived from the ratio of the prices of the current step to the prices of the previous, that is we get:

$$\psi_1 = p_1^1/p_1^0 = 1.09,\ \psi_2 = p_2^1/p_2^0 = 0.951,\ \psi_3 = p_3^1/p_3^0 = 0.907$$

We re-evaluate the inputs multiplying each and every one of them by the corresponding multiplier and we get:

[24]It is interesting to note that classical economists and especially Ricardo were aware of the existence of the feedback effects of an initial change. However, they thought that the first (direct) result is the most important one, while those that follow it do not change the qualitative character of the analysis. It is important to note that the intuition of classical economists is in general right, although there may be interesting exceptions as Sraffa (1960) has shown very elegantly.

$$\left(\psi_1 c_i^0 + \psi_2 v_i^0\right) = \left(c_i^1 + v_i^1\right)$$

We observe that the multiplier of the third department plays no role in the transformation procedure. Moreover, we notice that the rates of profit of each sector will differ, indicating the presence of disequilibrium. Equilibrium is restored by re-estimating the general rate of profit

$$r^1 = \left[\sum p_i^1 - \sum (c_i^1 + v_i^1)\right] / \sum (c_i^1 + v_i^1) = 26.2$$

A new set of (second step) production prices is then calculated, see Table 5.1, part 2B:

$$(c_i^1 + v_i^1)(1 + r^1) = p_i^2$$

Subsequently, we construct multipliers, as the ratios of the prices of production of the second step to the prices of production of the first step and we re-evaluate the inputs according to these new multipliers:

$$\psi_i = p_i^2 / p_i^1$$

The multipliers of the second step are much closer to one than in the first step and if we continue this procedure, we end up, once again, with profit rate differentials. We re-estimate the average rate of profit and then the prices of the third step and so forth. The procedure terminates, when the multipliers become equal to one. The properties of the numerical example, which happen to be also the properties characterising real economies, guarantee the existence of a single system of prices of production.

We observe that only the first invariance condition holds in each step but not the second. Shaikh (1984) argues that this inequality is expected and that it is interpreted on the basis of the difference between the circuit of capital and the circuit of revenue. In the circuit of capital all output is invested, while the circuit of revenue refers to the capitalist luxury consumption or unproductive consumption in general. Thus, in the circuit of capital, represented by the first two sectors, the deviations of prices of production from values (direct prices) remain within the circuit itself giving rise to profits or losses to each individual sector and ensuring that the invariance conditions hold true for the aggregate variables. When, however, the analysis is carried out with the third department which constitutes the circuit of revenue of capitalists, the deviations of values from prices of production and thus of profits from surplus value do not appear in the circuit of (productive) capital and, therefore, we cannot account for them. It is important to stress that the size of the difference is a function of the size of the luxury sector. The greater the size of this sector, the greater the difference between value magnitudes and price magnitudes. Clearly, the difference say between profits and surplus value is maximised in conditions of simple reproduction (as in the numerical example shown earlier)

and it is minimised as the economy grows (which means that not all surplus value is consumed by the capitalists, but a part of it is reinvested in the production process). In the case of maximum expanded reproduction, where all profits are invested productively, and, therefore, are not consumed by capitalists (there is no luxury goods sector), the economy's growth rate equals the profit rate.[25] Then, both invariance conditions hold.

The analysis has been too theoretical until now and the question that is posed is what happens in reality, that is, how much do direct prices differ from the prices of production and from market prices? In addition, how different are the value magnitudes of some fundamental variables (rate of profit, rate of surplus value, etc.) from the price of production and market price magnitudes. Today we know that, for a number of examined economies, direct prices, as well as Marx's prices of production, are extremely close to the fully transformed prices of production and, therefore, that the transformation problem has more of a theoretical, rather than practical, meaning. Table 5.2 summarises the differences between the various sets of prices in the US and Greek economies.[26]

The transformation problem continues to be part of the research of many economists and various new propositions and even solutions have been suggested adding new interesting elements and dimensions to this old problem. It seems, though, that such a discussion has reached a saturation point, since in our view, this is an area where the major issues have been settled in a satisfactory way both theoretically and empirically. A crucial issue in these discussions, which however has not been brought about so far, is that both direct prices and prices of production are (weighted) average magnitudes and that they are not necessarily the most immediate

Table 5.2 Measures of deviation and correlation for values and prices: Greece and USA

	Direct prices/ market prices (1)	Prices of production/ market prices (2)	Direct prices/ prices of production (3)	Marx's prices of production/prices of production (4)
	GR/USA	GR/USA	GR/USA	GR/USA
MAD (%)	23.1/12.2	14.3/13.6	18.7/16.9	9.3/5.5
MAWD (%)	21.6/12.5	15.4/14.6	18.1/17.4	7.6/6.7
NVD (%)	25.1/13.7	20.4/16.8	23.0/16.8	8.0/8.8
R^2	0.942/0.974	0.939/0.982	0.950/0.971	0.990/0.995

The source of data is Ochoa (1984, 1989) and Tsoulfidis and Maniatis (2002)

[25]The economy in other words is along Von Neumann's ray of economic growth (Morishima 1973).

[26]The data for the USA are the average of five benchmark years, while for the Greek economy are for a single (1970) year. Regarding the statistics of deviation used, that is, Mean Absolute Deviation (MAD), Mean Absolute Weighted Deviated (MAWD) and Norm Vector Distance (NVD), the smaller they are the smaller the deviation, while for the coefficient of determination R^2 the higher it is the smaller the deviation. The interested reader should consult a book in statistics for the above definitions.

regulators of market prices. This does not mean that the average magnitudes are not good approximations to market prices, but rather indicates that there are even more concrete centres of gravitation, which we need to estimate. For example, we already know from the classical economists that in agriculture and mining the regulating values and prices of production are those that are formed on the marginal lands and so one cannot just use the average direct price in agriculture or mining as the best centres of gravitation of market prices. The idea of marginal conditions is generalised in Marx by discussing, for example, the case of manufacturing, where the regulating direct price and price of production will be, in general, different from the average. In fact, the regulating conditions and the corresponding direct prices and prices of production are identified with those firms or capitals of an industry, where there is acceleration or deceleration of capital accumulation. This crucial aspect of Marx's work has received very little attention so far, precisely because the concept of regulating capital and the associated notions of regulating value and price of production are extremely difficult to operationalise. As a consequence, the empirical analysis up until now has been restricted exclusively to average direct prices and average prices of production, and although these are very good approximations to market prices they are not the most appropriate ones.

5.10 Marx on Competition

Marx intended to devote a whole book on the question of competition. In fact, volumes I and II of *Capital* are written on the assumption that there is no competition between capitals. Although Marx's plan was not fully accomplished, there are sporadic remarks on competition throughout his entire work. For example, there is a specific chapter on competition (Chap. 10) in volume III of *Capital*, while, in general, competition between capitals is assumed throughout *Capital* III, since there is a tendential equalisation of the profit rates to the economy's average. Marx's analysis of competition has many similarities with that of Smith and Ricardo. Nevertheless, there are also significant differences. The major difference is that competition in Marx is a derived concept and not the starting point of the analysis (Shaikh 1980b; Semmler 1984) which is the production for profit as an end in itself. Competition among capitals in Marx's perspective follows the analysis of the laws of capitalist accumulation. As the units of capital strive to expand their market share, increase production, and realise surplus value, they must take actions to confront the efforts of other units of capital engaged in similar efforts. Consequently, one cannot start with an analysis of competition before the analysis of the laws of accumulation. Marx is explicit about that when he notes:

> A scientific analysis of competition is not possible before we have a conception of the inner nature of capital (*Capital*, vol. I, p. 316).

He further notes that competition among capitals is the mechanism by which the laws of capitalist accumulation operate and become visible:

> Competition makes the immanent laws of capitalist production to be felt by each individual capitalist, as external coercive laws. It compels him to keep constantly extending his capital, in order to preserve it, but extend it he cannot, except by means of progressive accumulation. (*Capital*, vol. I, p. 592).

For Marx, competition is envisioned as a turbulent and inherently violent process that resembles, in many respects, actual "war" (Marx 1847). The market share of firms, for example, is like the territory of countries engaged in war, while technical change is like the arms race, since it is through technical change that firms can lower their cost and prices, attack their competitors and gain a larger share in the market for themselves (Shaikh 1980b).

In Marx's work, there is a clear distinction of competition between and within industries. For example, he notes:

> What competition, first in a single sphere, achieves is a single market-value and market price derived from the various individual values of commodities. And it is competition of capitals in different spheres, which first brings about the price of production equalising the rates of profit in the different spheres. The latter process requires a higher development of capitalist production than the previous one. (*Capital*, vol. III, p. 180)

In short, competition leads to the establishment of a common rate of profit with different equilibrium prices across industries, and a uniform price with differential profit rates between firms in the same industry. In what follows, we present the salient features of these two aspects of competition in Marx's work and their synthesis which is the concept of regulating capital.

5.10.1 Competition Between Industries

The first consequence of the analysis of competition between industries is the tendential equalisation of the inter-industry rates of profit. Firms in each industry are assumed to sell their commodities at market prices that tend to incorporate the economy's average rate of profit. The rationale for the formation of a general rate of profit between industries stems from the following sequence of events. Industries with a rate of profit above the general rate of profit attract most of the capital and thus they experience an acceleration of capital accumulation; by contrast, industries with a rate of profit below the general rate of profit experience a deceleration of accumulation. This process is a dynamic one in the sense that all industries grow diachronically, the difference being that industries with higher profit rates will grow faster than their demand so as to bring down their market price, thereby reducing their profit rate to the level of the general rate of profit. The converse will be true for industries with a lower than the average profit rate; capital accumulation grows at a rate lower than that of demand, giving rise to a higher market price elevating the rate of profit closer to the economy's general rate of profit. Marx puts it as follows:

> Competition levels the rates of profit of the different spheres of production into an average rate of profit through the continual transfer of capital from one sphere to another.

> The fluctuations of profits caused by the cycle of fat and lean years succeeding one another in any given industry within given periods must, however, receive due consideration [...] Experience shows, moreover, that if a branch of industry such as say, the cotton industry, yields unusually high profits at one period, it makes very little profit, or even suffers losses, at another, so that in a certain cycle of years the average profit is much the same as in other branches. And capital soon learns to take this experience into account. (Marx 1894, p. 208)

This kind of turbulent equalisation of profit rates implies that the dispersion of the rates of profit around the average takes place quite regularly and never comes down to zero, which is equivalent to saying that the two rates of profit eventually become equal to each other, but rather that at any moment in time, the two profit rates are unequal and, after a period of "fat and lean years" the differences on average tend to cancel each other out.[27]

The second consequence of competition between industries is that the equalisation of sectoral profit rates implies that the level of profit margins on sales (or on cost) is directly related to capital–output ratios. This result is derived in a straightforward manner from the definition of the profit rate. Thus, we have:

$$r = \frac{S}{K} = \frac{S/Q}{K/Q} = \frac{m}{K/Q} \text{ and } m = r\frac{K}{Q}$$

where r is the rate of profit, S is the total profits, K is the fixed capital stock, Q is the gross output or total sales, m is the profit margin on sales and K/Q is the capital–output ratio.

This formula shows the direct relationship between profit margin on sales and the capital–output ratios. If there is an equalisation of profit rates in the economy, the unequal capital–output ratios imply unequal profit margin on sales between industries. Formally, for industries i and j we have:

$$m_i = r\left(\frac{K}{Q}\right)_i \text{ and } m_j = r\left(\frac{K}{Q}\right)_j \text{ or } \frac{m_i}{m_j} = \frac{(K/Q)_i}{(K/Q)_j} \text{ and}$$

$$m_i \gtreqless m_j \text{ as } (K/Q)_i \gtreqless (K/Q)_j$$

Thus, a necessary consequence of the tendential equalisation of the profit rates is that the profit margins on sales tend to be proportional to the relative capital–output ratios.[28]

A third consequence of competition between industries is that for industries with a high capital–output ratio and thus high entry and exit costs, variations in demand

[27]For a formal presentation of the long-run equalisation of profit rates as a gravitational process, see Flaschel and Semmler (1990).

[28]For an empirical test of this as well as of other core propositions of alternative theories of competition with respect to the determinants of the profit margins on sales, see Ochoa and Glick (1992), Tsaliki and Tsoulfidis, (1998).

will be reflected more in variations in capacity utilisation and less in price variations through the acceleration or deceleration of capital accumulation. In other words, when demand changes, industries with a high capital–output ratio tend to absorb demand variations through the necessary adjustments in the rate of capacity utilisation and employment, and less through price changes. Thus, industries with higher capital stock or capital–output ratio will have smaller profit margin variations for any given percentage variation in sales. Formally, we have:

$$\left(\frac{dm}{m}\right)_i \gtreqless \left(\frac{dm}{m}\right)_j \text{ as } K_i \lesseqgtr K_j$$

Similarly, the percentage change of the rate of profit is expected to be smaller in industries with higher capital–output ratios. The idea is that if more of the variability in demand is absorbed in output than in price variations, it follows that the rate of profit will be less variable in high capital–output ratio industries than in the low ones.

The profit rate variations will be smaller in industries that display smaller profit margin variations, which, as we have already shown, are industries with the larger investment requirements and, by extension, with the higher capital–output ratios.

5.10.2 Competition Within Industries

We have shown that competition between industries equalises the rate of profit through unequal prices of production. Competition between firms within an industry is the second moment in Marx's (1894, pp. 138–139, 178–186, 197–198, 641–645) analysis of competition. Here, firms are viewed as large units of capital, which fight their competitive struggle over market share by undercutting price and eliminating competitors:

> The battle of competition is fought by the cheapening of commodities. The cheapness of commodities depends, ceteris paribus, on the productiveness of labour and this again on the scale of production. Therefore, the larger capitals beat the smaller. (Marx 1867, p. 626)

Although Marx was writing in the nineteenth-century, his analysis begins with large units of capital, which are already engaged in the "battle" of reducing production costs by increasing mechanisation. Innovations leading to techniques with lower cost make possible the reduction of the selling price, thereby increasing the market share of innovators. Imitators cannot follow immediately for they are stuck with their fixed capital, which must be kept in operation for a certain period of time for their owners to realise its value. Since they increase their capital per unit of output produced, innovators will temporarily reduce their profit rates. However, as they reduce the selling price of their commodity and expand their market share, their profit margin on sales increases and gradually their rate of profit becomes the highest in the industry. Eventually, all producers sell the same commodity for approximately the same price, that is, "the law of one price" prevails: "Competition

can only make the producers within the same sphere of production sell their commodities at the same price" (Marx 1894, p. 865). As in the case of the equalisation of profit rates, the equalisation of price within an industry only means that all firms in an industry sell at approximately the same price.

The tendential equalisation of price within industries implies differential profit rates between firms in the same industry. If all firms sell at the same price, it follows that firms with lower costs will tend to earn higher profit rates than those with higher costs. Thus, the tendential equalisation of profit rate across industries is consistent with a hierarchy of firm rates of profit within industries. This situation would continue to exist because some of the elements of production are not easily reproducible (location, climate, natural resources, etc.) and because of unequal firm innovation (Shaikh 1980b; Semmler 1984).

5.10.3 Regulating Capitals

The preceding discussion on competition between and within industries leads to the following seeming paradox: how is it possible for rates of profit to be equalised between industries while at the same time firms within industries earn different profit rates? Marx's answer to this paradox is that the equalisation of profit rates applies not to all capitals within an industry but only to specific capitals. This analysis, however, is not developed in any clear way in the chapters on competition (*Capital*, vol. III, Chaps. 8–10), where the exposition remains at a very high level of abstraction for it deals only with the average conditions of each industry. In Chap. 10 of the third volume of *Capital*, Marx develops a detailed analysis of competition within industries; however, he does not integrate this discussion with that of competition between industries. There are only sporadic remarks at the end of Chap. 10, which give some hints for a possible synthesis of these two moments of competition. The question before us now is "how do we determine the type of firm, or better the method of production whose rate of profit is tendentially equalised across industries?"

Marx's assumption in the aforementioned chapters on competition is that the average conditions of production are those at which expansion or contraction of accumulation takes place, and where therefore the rate of profit tends to be equalised across industries. Shaikh (1995, 2008) has argued that this assumption is only provisional. At a more concrete level of analysis, which occurs in the chapters on differential and absolute rent (*Capital*, vol. III, pp. 640–737), Marx argues that in agriculture the contraction or expansion of accumulation takes place on the worst land.[29] Since better lands are either not accessible – they are already cultivated and earn differential rent – or are not easily reproducible, it is the rate of profit of the worst land that tends to be equalised. Hence, the worst land forms the regulating condition of production, that is, the condition where the expansion or contraction of supply takes place.

[29]Agriculture is treated as a kind of industry with non-reproducible means of production.

With manufacturing, the regulating conditions of each sector are not always associated with the same type of conditions as in agriculture. However, regulating conditions in manufacturing are determined by exactly the same method, that is, by the type of capital in which expansion or contraction of accumulation takes place. Marx calls the method of production accessible to new capital "the regulating capital". This should not lead to the conclusion that all firms adopt this method of production immediately, since firms operate fixed capitals of different vintages and managers have different expectations. Consequently, firms do not easily switch from one method of production to another. However, new capitals are expected to enter into the method of production, which can be duplicated and in which, furthermore, the expected rate of profit is attractive enough. The production method which is targeted by the new entrants is usually the most recent in the industry and not the older or the most profitable. The older methods of production *ceteris paribus* will have a rate of profit lower than the average, whereas the most profitable methods of production may not be easily reproducible or their reproduction may be associated with a certain degree of risk, which new entrants may not wish to undertake.

However, over "a cycle of fat and lean years", that is a complete business cycle, there is a tendency for the rate of profit to equalise among regulating capitals between industries. Prices of production then are constructed around the rate of profit of regulating capitals, which gives rise to differential profitability within industries. The average profit rates of the regulating capitals across each industry are those that will be tendentially equalised.

The notion of regulating capital forms the synthesis of the two moments of competition between and within industries. The rate of profit earned on regulating capital is therefore the measure of new investment's return and determines the rhythm of accumulation in industries. If two regulating capitals have different rates of profit, the investment will flow differentially and will not just stop flowing in the industry with the lowest rate of profit because of uncertainty and differences in expectations. It is important to point out that the regulating conditions of production do not necessarily specify a single rate of profit but rather a narrow spectrum of rates of profit. This is true even in the case of a single regulating condition of production, because there are still differences in management, demand, etc. which may give rise to profit rate dispersions. Consequently, at any given moment in time, the rates of profit between regulating capitals across industries are not equal and only in the long run is there a tendential equalisation of the regulating rates of profit to an average (Christodoulopoulos 1996; Tsoulfidis and Tsaliki 2005; Shaikh 2008).

5.11 The Falling Tendency of the Rate of Profit

Marx's starting point for the analysis of the evolution of the profit rate is the nature of capitalist production, which is oriented towards the extraction of the maximum possible profit and the expansion of the production activity; in other words,

capitalists are dependent on profit. Capital in the effort to acquire the maximum possible profit must struggle in two fronts.[30] In the first front, capital fights against labour for the maintenance of wages to the lowest possible level and the increase in productivity. There is no doubt that although capital exhausts every possibility for the increase of the length of the working day and the intensification of work, it faces physical and legal limits that cannot be easily superseded. It is worth noting that in many cases capital does not seek to reach these limits, since this might be proved disadvantageous for profits. Taking this into account, it follows that the most effective way for capital to increase profits is the further subdivision and routinisation of the labour process. This on the one hand degrades labour, and, therefore, keeps wages down, while on the other hand, it makes the mechanisation of the labour activity and its replacement by fixed capital possible.

In the second front, capital fights against other capitals over the expansion of their market share by reducing the average cost and so the price of the product. This is possible once again through the introduction of fixed capital, which ensures higher mark-ups, and eventually a higher profit rate. The cost of the introduction of fixed capital is that the resulting rate of profit is lower than before; however, since the innovating firms expand their market share their profit rate will be higher than that of their competitors.

For the sake of simplicity, our analysis focuses on the fixed capital advanced and thus the definition of the rate of profit (r) is simplified to the formula: $r = s/C$, where s = total profits and C = total fixed capital. The process of mechanisation is reflected in the replacement of variable capital (v) by fixed capital and the rise in the ratio of C/v. The fixed capital advanced is the capital that embodies the new more effective techniques that increase the productivity of labour. In other words, the technical composition of capital, that is, the capital–labour ratio (C/l) has a tendency to increase, which is followed suit by the organic composition of capital (C/v).[31] The rationalisation and, at the same time, the capitalisation of the production process lead to the reduction of prices of commodities, because of the increase in the productivity of labour. This amounts to a fall in the rate of profit as this is shown in the following formula:

$$r = \frac{s}{C} = \frac{s/v}{C/v}$$

It is important to point out that the rise in the ratio C/v counteracts any increase in the rate of surplus-value (s/v) and necessarily leads to a falling tendency in the evolution of the profit rate. In fact, the formula of the rate of profit can be rewritten as:

[30]The presentation that follows draws on Shaikh (1978).

[31]For a comprehensive discussion of the various (technical, value and organic) compositions of capital, see Shaikh 1990.

$$r = \frac{s}{C} = \frac{s}{l}\frac{l}{C} = \left(\frac{s/v}{1+s/v}\right)\frac{l}{C}$$

where $l = s + v$, that is, the total labour time (l) is equal to the surplus (s) and necessary (v) labour time. It can be shown that regardless of the rate of increase in the rate of surplus value (s/v), the term in the parenthesis in this formula increases at a decreasing rate since its upper limit is 1. As a matter of fact, the term s/v reaches its maximum rate of increase, when $v \to 0$, and s/v tends to infinity. Consequently, we have:

$$\left(\frac{s/v}{1+s/v}\right) = \frac{\infty}{\infty}$$

What counts in a situation such as this is the rapidity with which the numerator and the denominator reach infinity. For this purpose we apply L'Hôpital's rule and we get:

$$\lim_{v\to\infty} \frac{(s/v)'}{(1+s/v)'} = 1$$

Consequently, the rate of profit necessarily falls since the term $(s/v)/(1 + s/v)$ has unity as its upper limit, while the very nature of capitalist production leads to a rising C/l, or what amounts to the same thing to a falling ratio of l/C, that is, to a fall in the maximum rate of profit (the rate of profit that we get for $v \to 0$).[32] The falling tendency of the maximum rate of profit implies that the general rate of profit (whose magnitude depends on the level of v) fluctuates within an interval with a falling upper limit. In short, the general rate of profit with the passage of time starts to display a falling tendency, for it is depressed from above by the falling maximum rate of profit.

5.12 General Rate of Profit and Economic Crisis

As we have already noted, the great economists before or after Marx (Smith, Ricardo, J.S. Mill, Jevons, Walras, J.B. Clark, Keynes, Schumpeter) accept, yet for different reasons, that the long run tendency of the rate of profit is a falling one and this fall eventually leads to the stationary economy. The rationale, however, which they have offered is far from convincing. For example Ricardo notes:

[32]From the formula of the rate of profit $r = s/C = (s/v)/(C/v)$, we get for $v \to \infty$, $(s/v)/(C/v) = \infty/\infty$. However, after the application of L'Hôpital's rule, we get: $(s/v)'/(C/v)' = s_{max}/C$, clearly $s_{max} = l$.

> The farmer and manufacturer can no more live without profit, than the labourer without wages. Their motive for accumulation will diminish with every diminution of profit, and will cease altogether when their profits are so low as not to afford them an adequate compensation for their trouble, and the risk which they must necessarily encounter in employing their capital productively. (Ricardo, *Works*, I, pp. 121–122)

This, however, is a trivial case, where the rate of profit is zero or near zero. In the usual case, the fall in the rate of profit in and of itself does not necessarily lead to a fall in the investment activity, since it is expected to be higher than the long-term rate of interest; and, therefore, the motive for investment continues to exist.

For Marx, however, a falling rate of profit (in the short run) is consistent with any stage of the economy. For example, he notes:

> A fall in the rate of profit and accelerated accumulation are different expressions of the same process only in so far as both reflect the development of productiveness. (*Capital*, vol. III, p. 241)

If we restrict the analysis to the downward stage of the economy, it is important to stress that the fall in the rate of profit in itself does not necessarily lead to the manifestation of economic crisis inasmuch as the profit rate, *ceteris paribus*, will exceed the long-term rate of interest, and, therefore, there does not appear to be any particular reason for the slowdown in investment activity. Marx, however, argued that the level of the profit rate in and of itself does not lead to stagnation. What is important is the cumulative long-run effect of the falling tendency of the rate of profit on investment and on the mass of net profits in real terms, which past a point generates the conditions for the manifestation of economic crisis. He pointed out:

> There would be absolute over–production of capital as soon as additional capital for purposes of overproduction = 0. The purpose of capitalist production, however, is self–expansion of capital, *i.e.*, appropriation of surplus–labour, production of surplus value, of profit. As soon as capital would, therefore, have grown in such a ratio to the labouring population that neither the absolute working–time supplied by the population, nor the relative surplus working–time, could be expanded any further (this last would not be feasible at any rate in the case when the demand for labour were so strong that there were a tendency for wages to rise); at a point therefore, when the increased capital produced just as much, or even less, surplus–value than it did before its increase, there would be absolute over–production of capital; *i.e.*, the increased capital $C + \Delta C$ would produce no more, or even less, profit than capital C before its expansion by ΔC. (*Capital*, vol. III, p. 251)

This causal relationship between the rate of profit and the mass of net profits and the manifestation of economic crisis is usually lost in the writings of many modern Marxist economists. For example, Foley (1986) notes:

> If the rate of profit were indeed falling consistently, why would the capitalist system not adapt to this fall through a gradual reduction in the rate of accumulation. Such a gradual reduction might not be welcome to capitalists, but it is not obvious that it must lead to the characteristic phenomena of capitalist crisis that we examined earlier. In other words, this explanation for capitalist crisis has to produce some systematic reason why a fall in the rate of profit leads at certain moments to sharp and discontinuous adjustments in economic activity. (Foley 1986, p. 153)

The work of Shaikh (1992), however, explicitly recognises the systematic relationship between the profit rate, the mass of profits and the manifestation of

crisis. In the formal model that he presents, capitalists possess a given propensity to save but the falling rate of profit leads to a stagnant mass of profits and to the crisis. This result can be cast in an alternative way. Specifically, starting with the usual formula of the rate of profit r, we get:

$$r = s/C \text{ or } s = rC$$

The total differential of this gives:

$$\Delta s = r\Delta C + C\Delta r$$

We divide through by ΔC and we get:

$$\frac{\Delta s}{\Delta C} = r + \frac{C}{\Delta C}\Delta r$$

Taking r as the common term, the equation can be rewritten as follows:

$$\frac{\Delta s}{\Delta C} = r\left(1 + \frac{\Delta r}{\Delta C}\frac{C}{r}\right)$$

The term $\Delta s/\Delta C$ indicates the way in which profits change with every change of capital stock or the change in profits for each unit of investment ($I = \Delta C$). It is obvious that the change in profits for each unit of investment is equal to zero only if the capital elasticity of the rate of profit is equal to -1, a condition that requires a falling rate of profit. This occurs at the point where total profits are maximised, that is at the point where the percentage change in capital stock ($\Delta C/C$) is equal to the percentage change in the rate of profit ($\Delta r/r$) in the opposite direction. As the economy reaches this point, the motivation for new investment fades away, because any profits from new investment are offset by the cost of that investment, thereby holding the mass of profits stagnant. The persistent lack of new investment and the rising unemployment rate form the two characteristic phenomena of depression. It is important to stress that this process is slow and takes place only in the long run. The fall in the rate of profit for a few years does not necessarily imply the slowdown in investment activity, and only if the rate of profit is falling for a protracted period of time will the mass of profits stagnate and the economy display the phenomena of crisis.

5.13 Summary and Conclusions

Karl Marx's economic analysis is mainly contained in three books: the three-volume *Capital*, the *Theories of Surplus Value*, also in three volumes, and the *Grundrisse*. In these works, Marx tries "to lay bare the law of motion of modern

society". Marx observes that capitalism is a historically specific system characterised by generalised commodity exchange and so, naturally, the starting point of his inquiry is the analysis of the commodity whose exchange value aspect (and not its use value aspect) is the dominant one. Exchange value is the external measure of the intrinsic property of commodities, which he calls *value*, defined as the total amount of socially necessary abstract labour time embodied in a commodity. In the process of exchange, one commodity, because of its possession of a set of useful properties, is chosen to function as the universal commodity against which all other commodities are compared and exchanged. The commodity that has historically performed the function of universal commodity more successfully than any other is gold. The ratio of the value of a commodity to the value of gold gives the direct price of the commodity. If the value of gold decreases (*e.g.*, because of discoveries of new gold mines and technological change), the general price level, other things being constant, increases, and *vice versa*. This rudimentary theory of money, which is derived from a straightforward generalisation of the theory of value, can be expanded to include both monetary systems that are convertible to gold and those that are not, and thus becomes relevant to new developments after 1971. At the same time, Marx's theory of value (together with the use of mathematical analysis and input–output data) has been shown to rather accurately predict market prices, and thus could become a viable alternative to neoclassical price theory.

The analysis of the universal commodity and money leads to an investigation of the capitalist process of production. This process is described by the circuit $M - C$ $(LP, MP) \ldots P \ldots C' - M'$, according to which capitalists invest an amount of money (M) in order to buy a set of commodities (C) consisting of commodity labour power (LP) – that is, the worker's capacity to work – and other means of production (MP), for the purpose of production (P) of a new set of commodities (C'), which when sold they expect to realise a sum of money greater than that of the initial investment, $M' > M$. This extra money is what really motivates the whole circuit of capitalist production as it is repeated on an expanded scale. The difference $M' - M > 0$, which Marx calls *surplus value*, stems from labour power, a special commodity characterised by its property of producing more value than the value of commodities that the worker buys with his money wage and consumes in order to reproduce his capacity to work. In contrast, the value of the means of production is either transferred to the final product all at once (as in the case of raw materials) or gradually through depreciation (as is the case with the plant and equipment) (*Capital,* vol. I). The distinction between *labour* and *labour power* is Marx's greatest discovery and contribution to political economy, because through this distinction the source of surplus value can be explained on the basis of equivalent exchanges. Marx argued that surplus value is created in the sphere of production by labour. The production sphere has primacy over the circulation sphere because the latter is supported by the surplus value produced in the former. Furthermore, the circulation sphere modifies and changes, within strictly specified limits, some of the results of the sphere of production. For example, surplus value in the sphere of circulation is redistributed to the various sectors of the economy in the form of profit according to its degree of capital intensity; however, the sum of the profits

cannot exceed the amount of surplus value produced. It is important to stress that there is no guarantee that the circuit of capital will necessarily be completed, as it can be interrupted at any stage by a number of unexpected factors. Thus, uncertainty and expectations are immanent in Marx's analysis of capitalism. Furthermore, the whole circuit begins and ends with money, a characteristic that allows the introduction of credit and also the hypothesis that savings may differ from investment, a difference that sets the stage for the development of an alternative to Keynes's theory of effective demand rooted in the process of capital accumulation.

According to Marx, the hallmark of the individual behaviour of capitalists is the pursuit of profit as a purpose in itself, which forces them into two kinds of competition: the first with workers in the labour markets over wages and conditions of work, and the second with other capitalists in the commodity markets over the expansion of market share at the expense of their competitors. Capitalists cope with these two types of competition through the introduction of more fixed capital. As a consequence, mechanisation of the labour process is used to raise the productivity of labour. The introduction of fixed capital both increases the scale of operation needed for minimum efficiency and reduces the unit cost of production. The latter implies that by reducing their prices innovating firms are able to expand their market share at the expense of less efficient firms. Thus, the process of capital accumulation leads to a small number of top firms controlling an increasing share of the total market. This is the reason why concentration of capital is the expected outcome implied by the nature of capital and by the operation of competition, which by no means diminishes over time. On the contrary, the very cause of mechanisation – the pursuit of profit – continues to exist even with fewer firms, as competition among them intensifies. Meanwhile, the ever increasing minimum-efficiency scale of operation requires higher investment that firms, especially the small ones, cannot undertake on their own, and thus there is pressure to merge, in order to avoid becoming the target of a hostile takeover. The resulting growth in the scale of production through the amalgamation of capitals is called *centralisation of capital* and is another aspect of the operation of competition (*Capital* vols. I, III). If there is a grand prediction that has been historically validated, it is Marx's law of increasing concentration and centralisation of capital.

Another grand prediction by Marx, which is also consistent with the available historical evidence, concerns the law of the falling rate of profit. This law is derived from the very purpose of capitalist production, which is the extraction of profits as an end in itself. As mentioned earlier, the realisation of this goal entails mechanisation of the production process through the introduction of fixed capital. On the one hand, this raises both the productivity of labour and profits for the firms that remain following concentration; on the other hand, however, the increase of fixed capital relative to labour leads to a falling profit rate. Marx noted that the fall in the rate of profit exerts a negative effect on the mass of real profits and, at the same time, a positive effect through the accumulation of capital. So long as the positive effect exceeds the negative, the mass of real profits expands at an increasing rate in a long wave-like pattern. Because new investment is a function of the rate of profit, it follows that a falling rate of profit at some point will necessarily slow down the

rate of growth of new investment, thereby slowing down the rate of increase in the mass of real profits. As this tendency continues, there will be a point at which the two (positive and negative) effects will cancel each other out and the change in the mass of profits will become zero. This means that the investment of the previous period will not contribute at all to an increase in profits and thus capitalists will have no interest in new investment. This is the point of "absolute over-accumulation of capital" that marks the onset of economic crisis. Its consequence is a slowdown in investment and rising unemployment. As more and more firms are led into bankruptcies and real wages fall, one can also observe the creation of new institutions, the emergence of new methods of management, and the diffusion of technological change. The combination of these processes results in a rising mass of profit (and a temporarily rising rate of profit) and sets the course for the reestablishment of the necessary conditions for another wave of expansion and contraction. Thus capitalism is both a growth- and crisis-prone system, as has also been documented in the literature on long economic cycles (see Shaikh 1992; Papageorgious and Tsoulfidis, 2006).

Marx's impact on economic thought has not received the recognition it deserves because of his view of the historical character of capitalist society and his vision of socialism. Thus, unfortunately, when orthodox economists discuss aspects of Marx's work, they generally do so to point out its alleged weaknesses rather than its strengths.

Questions for Thought and Discussion

1. Define the terms: abstract labour, private labour, social labour, socially necessary labour time (first and second sense), skilled and unskilled labour.
2. Discuss Marx's labour theory of value.
3. Compare Marx's law of value with Smith's invisible hand.
4. Define the terms: direct price, price of production, market price, general rate of profit.
5. Discuss the nature of the transformation problem and its significance for Marx's theory.
6. Discuss the details of the iterative procedure for the solution to the transformation problem.
7. Do the two invariance postulates hold in the iterative solution? Why or why not?
8. Discuss Marx's argument for the falling rate of profit? Is the fall in the rate of profit inevitable?
9. Why does a falling rate of profit in the long run lead to a crisis since the rate of profit will always be higher than the rate of interest?
10. What is meant by free competition in Marxian economics? How does this concept differ from the concept of perfect competition?
11. Discuss the "law of one price" and the "law of equal profitability" in Marx's analysis of competition.
12. Discuss some of the phenomena of competition and their explanation.

Notes for Further Reading

The most authoritative among Marx's biographies is perhaps that of Mehring (1936), but there are also other recommended ones such as McLellan (1973) and more recently, Wheen (2001) among others. McLellan (1976) has also written a biography of Engels. The notions of dialectics and historical materialism are not easily comprehensible especially for a student of economics. The best thorough introduction to these concepts, among the sources that we are aware of, is perhaps the first two chapters of Heilbroner's (1980) book. The reader is encouraged to look also at the relevant entries in the *New Palgrave Dictionary* and the *Dictionary of Marxist Thought*.

For the theory of value, the reader should look at Shaikh's articles (1977, 1982, 1984) and at some point must read thoroughly *Capital* I, (Chaps. 1–3). Clearly, this is the path that we followed in our presentation. The reader must be careful with other presentations. Rubin's (1928) theory, for example, is fine with the only difference that he finally claims that value is created in production and it is validated in exchange. This is a line of thought that many post Marxist authors have followed. However, the drawback of such an approach is that it disregards the two senses of the socially necessary labour time that we outlined in our text. According to the first sense the socially necessary labour time is estimated as the ratio of the labour time actually spent to the total labour time divided by the number of commodities and in socially necessary labour time in the second sense according to which the direct prices of commodities changes to account for the ebbs and flows in demand. For a critique of such approaches based on Rubin see Shaikh (1982). It is important to point out that the distinction between the two senses of socially necessary labour time allows of the empirical testing of the labour theory of value with data from input–output tables, while one could also use time series data and econometric analysis.

The transformation problem as we showed in the text is mainly a conceptual and not a computational problem (see Shaikh 1977, 1984). We described the solution to this problem through the iterative procedure, which really assumes that the level of output, the real wage and the technique of production remain the same throughout the transformation. The transformation problem can also be solved with the use of matrix algebra, while the two invariance postulates remain valid in the case of an economy growing at the maximum growth rate (Shaikh 1973, 1984; Morishima 1973; Morishima and Catephores 1979). Recent years have seen the development of other approaches that seek to settle the issue. The reader interested in these efforts can consult Duménil (1983), Foley (1986) and Moseley (2000), who are supporters of different variants of the so-called "new solution" to the transformation problem. These authors cast doubt to the assumption of a given real wage and they try to relax it. The result is to

(continued)

arrive at a new solution where the invariance postulates hold true only for the net products and not for the gross products. A variant of this approach is the Temporary Single System Interpretation (TSSI) of Kliman and McGlone (1988), Freeman and Carchedi (1996). In the TSSI even the technique of production is variable! It seems that this last variant of the transformation problem has become part of the literature because its proponents have made every possible effort in terms of attracting attention. Lately, this view has been supplemented by a few papers that essentially argue that the empirical verification of the labour theory of value is impossible because there is no way of knowing the actual market prices (Diaz and Osuna 2006). For theoretical critiques of the TSSI (see Veneziani 2004; Mohun 2004; Rieu 2007, 2009) and for an empirical critique (Tsoulfidis and Paitaridis 2009). In our view, the major issues of the transformation problem were settled a long time ago, and whether there is any problem worth pursuing is the question of the attainment of equilibrium dynamically. The iterative procedure is, after all, somehow static, since the question of the process of equalisation of the rate of profit (or of the growth rate) is taken as given and one wonders about the exact path of variables and processes through which prices gravitate towards equilibrium. In this respect, we suggest the articles by Flaschel and Semmler (1990) as well as by Duménil and Lévy (1993), which could become the starting point for further research.

The question of competition in Marx is of extreme importance and it seems that current issues of actual competition can be informed by Marx's theory. Competition in Marx is a dynamic process of rivalry between firms and has no relevance to perfect competition or some imperfection of that. The papers that we suggest include Clifton (1977), Eatwell (1982), Rosdolsky (1977), Shaikh (1980b) and Semmler (1984). In the recent decades, there have been many efforts to empirically verify Marx's notion of competition. Included among those efforts are Glick (1985), Ochoa and Glick (1992) Tsaliki and Tsoulfidis (1998). The concept of regulating capital which integrates competition between industries with competition within industries was developed for the first time in Shaikh (1995) and has been further explored in Botwinik (1994). Christodoulopoulos (1996) has used it to verify the inter-industry equalisation of profit rates in selected industries internationally, whereas Tsoulfidis and Tsaliki (2005) found gravitation of interindustry profit rates towards the average regulating rate of profit. Shaikh (2008) returned to the issue of regulating capital adding new theoretical insights and supplements them with fresh empirical evidence from US industries. It seems that marginal conditions are critical to the study of behaviour of economic units in both Marxian and neoclassical economics, as we see in the next chapters.

For the question of the law of the falling rate of profit (FROP), the reader is recommended to study mainly the chapters in *Capital III* (Chaps. 11–13) in

(continued)

connection with Shaikh (1978). Marx also discusses the law of the FROP in the Grundrisse (for a succinct summary see Rosdolsky 1977). The so-called Okishio (1961) theorem according to which rational capitalists would not introduce new techniques that resulted in a falling rate of profit has been repeated by Pajis (1981) among others but the answer to this, as argued in Shaikh (1980b), is that capitalists are forced to choose the more mechanised techniques that are associated with the lower rate of profit because this is the only way to survive in competition

The empirical research verifies the hypothesis of the FROP for all of the OECD countries. The interested reader can go over Amstrong *et al*. (1990) or Shaikh and Tonak's (1994) book for a summary of the research in many countries. The trouble with these studies (not with Shaikh and Tonak) is the identification of the cause of the fall in the rate of profit. As we know, Marx's argument was that both the rate of surplus value and the organic composition rise. In the long run, however, the organic composition of capital rises faster than the rate of surplus value resulting in a long-run falling tendency in the profit rate. However, in various studies, known as "profit squeeze", it is claimed that the "rate of surplus-value" is falling through time, while the "organic composition of capital" is approximately constant or rising slowly, leading to a falling rate of profit. The trouble with these studies, as Shaikh and Tonak (1984) argue, is that they take the orthodox measures of profit-to-wage ratio as being equivalent to the rate of surplus value and the ratio of capital to wages as if it were equivalent to the organic composition of capital. In estimating Marxian categories, the issue of productive and unproductive labour is of critical importance and one must account for it. Once this happens, we get Marx's predictions (see for example Shaikh and Tonak 1984, *inter alia*). Along these lines, one can test the incidence of under-consumption, through the adjustment of the rate of profit by capacity utilisation rate. It has been shown in the relevant empirical studies (*e.g.,* Shaikh 1992) that when the actual rate of profit is adjusted for capacity utilisation (that is on the assumption that demand and supply are equal to each other), its falling tendency is maintained (if not strengthened) and so one cannot attribute the economic crisis to the lack of sufficient demand.

We argued that the falling rate of profit leads inevitably to an economic crisis and we showed the precise mechanism. However, this is not the only explanation of economic crises. There is another quite popular explanation based on underconsumption that has been expounded by Sweezy (1942) and others. There are also dis-proportionality crisis theories, and the reader is recommended to consult Sweezy's (1942) book, and for a critique of all these theories, we recommend Shaikh's (1978b) article.

Appendix

A.1 The Mathematics of the Falling Rate of Profit

We argued that Marx's thesis for the falling rate of profit is that the organic composition of capital as a result of mechanisation of the production process, increases at a rate higher than that of the rate of surplus-value (s/v), and, therefore, leads to a long run falling tendency in the rate of profit. This result can be shown starting from the formula of the rate of profit which can be rewritten as:

$$r = \frac{s}{C} = \frac{s/l}{C/l}$$

where $l = s+v$, that is, the total labour time (l) is equal to the surplus (s) and necessary (v) labour time. The advantage of this formula is that it sets limits to the variation of the rate of profit. For example, we derive that regardless of the rate of increase in the rate of surplus value (s/v). The numerator of the above formula has as an upper limit the one, and the rate of profit for $v \rightarrow 0$ (*i.e.*, "workers leave on thin air") is equal to the l/C, that is, the maximum rate of profit (the rate of profit for $s = l$). The mechanisation process leads to a rising C/l ratio or what amounts to the same thing a falling maximum rate of profit. The latter implies that the general rate of profit (whose magnitude depends on the level of v) fluctuates with an interval with a falling upper limit. In short, the general rate of profit with the passage of time starts to display a falling tendency, for it is depressed from above from the falling maximum rate of profit. However, this in itself is not an adequate proof of the falling rate of profit, and one must show that the limit of the rate of profit is zero (Kurz 1998, p. 113, and 2010).

For the proof of this proposition, let $C' = C/l$, $s' = s/l$, $v' = v/l$ or $v' = 1–s'$ and the rate of profit can be rewritten as:

$$r = \frac{s'}{C'}$$

Assume now that C' increases at a rate equal to α, whereas the variable capital decreases at a constant rate equal to β. By using time subscripts, we can write for the evolution of each of these variables as follows:

$$C'_t = C'_0(1+\alpha)^t \text{ and } v'_t = v'_0(1-\beta)^t$$

and the evolution of s' is residually determined, that is, $s'_t = 1 - v'_0(1-\beta)^t$. The rate of profit therefore can be rewritten as follows:

$$r_t = \frac{1 - v'_0(1-\beta)^t}{C'_0(1+\alpha)^t} = \frac{(1-\beta)^{-t} - v'_0}{C'_0(1+\alpha)^t(1-\beta)^{-t}}$$

as t increases without bounds the numerator and the denominator of the rate of profit increase to infinity, so we end up with an indeterminate form, ∞/∞. Thus, we can write,

$$\lim_{t\to\infty} \frac{(1-\beta)^{-t} - v'_0}{C'_0(1+\alpha)^t(1-\beta)^{-t}} = \frac{\infty}{\infty}$$

We apply L'Hôpital's rule, which gives:

$$\lim_{t\to\infty} \frac{-(1-\beta)^{-t}\ln(1-\beta)}{C'_0(1+\alpha)^t\ln(1-\alpha)(1-\beta)^{-t} - C'_0(1+\alpha)^t\ln(1-\beta)^{-t}\ln(1+\beta)}$$

which simplifies to the following:

$$\lim_{t\to\infty} \frac{-\ln(1-\beta)}{C'_0(1+\alpha)^t\ln[(1+\alpha)/(1-\beta)]} = 0$$

A.2 The Incremental Rate of Profit and its Components

In the analysis of competition in Marx's capital, we faced the following paradox. On the one hand, the rates of profit were equalised across industries and on the other hand, there was a stratification of the rates of profit between firms within industries. How can these contradictory observations be reconciled? The idea is that the average rate of profit is the average of all firms that comprise the industry. And an industry consists of firms that use very advanced technology and excellent location and firms whose technology is old. Certainly, investment flows would be directed neither towards the old type of capitals because of low profits nor towards the very new, precisely because they have not been tested adequately, so there is too much risk involved. Besides, there are problems in investing in these kind of capitals simply because these capitals are not easily reproducible, for example, things such as patents, location near a source of raw materials and the like.

Classical economists were aware of these limitations in the flows of capital; perhaps the best example is the case in agriculture where the most productive pieces of land are already cultivated and they are not available to new entrants, so the new entrants enter not to the average quality of land since it is not available but rather to the worst type of land because only that is available. Classical economists therefore considered as the average rate of profit not simply the arithmetical average but rather the type of capital where expansion or contraction of accumulation takes place.

Turning to manufacturing, the regulating conditions of each sector are determined by exactly the same method; that is, by the type of capital where expansion or contraction of accumulation takes place. This concept is similar to what business people call the capital, which embodies "the best generally available method of

production", and is often called "the best-practise method of production". This should not lead to the conclusion that all firms adopt this method of production immediately, since firms operate fixed capitals of different vintages and managers have different expectations about the direction of demand and profitability. Consequently, firms do not easily switch from one method of production to another. However, new capitals are expected to enter into the method of production, which can be duplicated and, furthermore, the expected rate of profit is attractive enough. The production method which is targeted by the new entrants is usually the most recent in the industry and not the older or the most profitable. The older methods of production *ceteris paribus*, will have a rate of profit lower than the average, whereas the most profitable, methods of production may not be easily reproducible or their reproduction is associated with certain degree of risk, which new entrants may not wish to undertake. However, over "a cycle of fat and lean years", that is, a complete business cycle, there is a tendency for the rate of profit to equalise among regulating capitals between industries. The profit rates of the regulating capitals across each industry are those that will be tendentially equalised.

The rate of profit earned on regulating capital is, therefore, the measure of new investment's return and determines the rhythm of accumulation in industries. If two regulating capitals have different rates of profit, the investment will flow differentially and will not just stop flowing in the industry with the lowest rate of profit because of uncertainty and differences in expectations. It is important to point out that the regulating conditions of production do not necessarily specify a single rate of profit, but rather a narrow spectrum of rates of profit. This is true even in the case of a single regulating condition of production, because there are still differences in management, demand, *etc*., which may give rise to profit rate dispersions. Consequently, at any given moment in time, the rates of profit between regulating capitals across industries are not equal, and only in the long run, there is a tendential equalisation of the regulating rates of profit to an average. Anwar Shaikh (1995, 2008) argued that the rate of profit that tends to be equalised between industries is not necessarily the average rate of profit, but rather the rate of profit that corresponds to the regulating conditions of production within an industry; that is, the "incremental rate of return on capital" (henceforth *IROR*) and he approximated it by taking into account the following considerations: Investment flows are conditioned more by short-run rate of return such as the incremental rate of return than the rate of profit over the lifetime of investment. Hence, he expresses current profits (S_t) that accrue to a firm as the sum of profit from the most recent investment (ρI_{t-1}) and profits that accrue to the firm from all the previous investments (S^*), which is equivalent to saying "the current profits in the absence of new investment". Consequently, we write:

$$S_t = \rho I_{t-1} + S^*$$

If we subtract profits of the past period from both sides of the this equation, we get:

$$S_t - S_{t-1} = \rho I_{t-1} + (S^* - S_{t-1}) \text{ or } \Delta S_t = \rho I_{t-1} + (S^* - S_{t-1})$$

The term in parenthesis is expected to be very small in comparison with the term ρI_{t-1} and for practical purposes it can be ignored. The justification is the view that the shorter the evaluation horizon, the closer the current profit will be on carried-over vintages S^* to the last period's profit on the same capital goods (S_{t-1}). Moreover, since uncertainty and ignorance increase with the passage of time, it is reasonable to assume that the short-run (up to a year) is the relevant time horizon. After all, current profits are fraught with many ephemeral factors, and we know that abnormally high or low profits direct investment accordingly, which in turn gives rise to new uncertainty and thus profits or losses, and so forth. With these considerations in mind, it is reasonable to assume that expectations about future returns to investment are nearsighted; that is, expectations depend on the short-run rate of return. Consequently, the current rate of return on new investment will be

$$\rho_t = \Delta S_t / I_{t-1}$$

that is, the change in profits of each industry divided by the investment in the previous period. The above configuration provides a practical way to identify the *IROR* in the case that we do not have data on the best practise technique and the firm that utilises it over the years. Consequently, the motion of the *IROR* determines whether or not there is a tendential equalisation of profit rates for the regulating capitals.

Alternatively, we can derive the *IROR* from the simple definition of the rate of profit $r=S/K$ or $S=rK$, whose total differential gives: $\mathrm{d}S = r\mathrm{d}K + K\mathrm{d}r$. We divide by $\mathrm{d}K$ and we get

$$\frac{\mathrm{d}S}{\mathrm{d}K} = r + K\frac{\mathrm{d}r}{\mathrm{d}K} = r\left(1 + \frac{\mathrm{d}r}{\mathrm{d}K}\frac{K}{r}\right) = IROR^{33}$$

The term (drK/dKr) in the parenthesis is the elasticity of profit rate for which the following holds:

$$\text{if } \left(\frac{\mathrm{d}r}{\mathrm{d}K}\frac{K}{r}\right) \gtreqless 0 \text{ then } IROR \gtreqless r$$

It can be shown that the *IROR* is a variable that encapsulates the operation of a series of other variables such as the profit and wage shares, productivity of labour, capacity utilisation and capital–output ratio. In order to show the operation of all these variables, we start from the definition of total income (*Y)* as

$$Y = rK + wL$$

[33]The connection between the *IROR* and the term $\mathrm{d}S/\mathrm{d}K$ is based on the definition of the capital stock, which for discrete time can be written: $K_t = (1–\delta)K_{t-1} + I_{t-1}$, where δ is the depreciation rate. For the sake of simplicity, let $\delta = 0$ and so we get $\Delta K_t = I_{t-1}$. Thus, $\mathrm{d}S/\mathrm{d}K = \Delta S_t/I_{t-1} = IROR_t$

whose total differential gives:

$$\mathrm{d}Y = r\mathrm{d}K + K\mathrm{d}r + w\mathrm{d}L + dwL$$

We divide throughout by dK and get:

$$\frac{\mathrm{d}Y}{\mathrm{d}K} = r + K\frac{\mathrm{d}r}{\mathrm{d}K} + w\frac{\mathrm{d}L}{\mathrm{d}K} + \frac{\mathrm{d}w}{\mathrm{d}K}L = r\left(1 + \frac{\mathrm{d}r}{\mathrm{d}K}\frac{K}{r}\right) + w\left(\frac{\mathrm{d}L}{\mathrm{d}K} + \frac{\mathrm{d}w}{\mathrm{d}K}\frac{L}{w}\right)$$

$$= IROR + w\left(\frac{\mathrm{d}L}{\mathrm{d}K} + \frac{\mathrm{d}w}{\mathrm{d}K}\frac{L}{w}\right)$$

After some mathematical manipulation, we get:

$$IROR = \frac{\mathrm{d}Y}{\mathrm{d}K} - w\left(\frac{\mathrm{d}L}{\mathrm{d}K} + \frac{\mathrm{d}w}{\mathrm{d}K}\frac{L}{w}\right) = \frac{\mathrm{d}Y}{\mathrm{d}K}\left[1 - w\left(\frac{\mathrm{d}K}{\mathrm{d}Y}\frac{\mathrm{d}L}{\mathrm{d}K} + \frac{\mathrm{d}K}{\mathrm{d}Y}\frac{\mathrm{d}w}{\mathrm{d}K}\frac{L}{w}\right)\right]$$

or

$$IROR = \frac{\mathrm{d}Y}{\mathrm{d}K}\left[1 - w\left(\frac{\mathrm{d}L}{\mathrm{d}Y} + \frac{\mathrm{d}w}{\mathrm{d}Y}\frac{L}{w}\right)\right] = \frac{\mathrm{d}Y}{\mathrm{d}K}\frac{K}{Y}\left[1 - w\left(\frac{\mathrm{d}L}{\mathrm{d}Y} + \frac{\mathrm{d}w}{\mathrm{d}Y}\frac{L}{w}\right)\right]\frac{Y}{K}$$

or

$$IROR = \frac{S}{Y}\left[1 - w\frac{L}{Y}\left(\frac{\mathrm{d}L}{\mathrm{d}Y}\frac{Y}{L} + \frac{\mathrm{d}w}{\mathrm{d}Y}\frac{Y}{w}\right)\right]u\left(\frac{Y}{K}\right)^{*}$$

From the foregoing analysis, we observe that the *IROR* is directly related to the profit share (S/Y), the rate of capacity utilisation ($u = (Y/K)/(Y/K)^*$), the growth rate of productivity of labour $(dY/dL)/(Y/L)$ and to the normal capacity output–capital ratio $(Y/K)^*$. In addition, the *IROR* is inversely related to the share of labour income (wL/Y) and the elasticity of wage with respect to income (dwY/dYw).

Chapter 6
The Structure of Classical Theory

In reality, supply, and demand never coincide, or, if they do, it is by mere accident, hence scientifically = 0, and to be regarded as not having occurred. But political economy assumes that supply and demand coincide with one another. Why? It is so to enable the study of phenomena in their fundamental relations, in the form corresponding to their conception, that is, to study them independent of the appearances caused by the movement of supply and demand. The other reason is to find the actual tendencies of their movements and to some extent to record them.

(*Capital*, III, pp. 189–190)

Demand depends only on its supply.

(Ricardo, *Principles*, p. 365)

6.1 Introduction

Classical economics is a term invented by Marx[1] to characterise all economists beginning with William Petty (1623–1687) in England and Pierre Le Pesant de Boisguilbert (1646–1714) in France, and ending with Ricardo in England and Simonde de Sismondi (1773–1842) in France. According to Marx, the focus of classical economists was the determination of the surplus (value), defined as the difference between the value of total output produced and the value of (labour and non-labour) inputs used in production. The evaluation of inputs and outputs is in terms of prices, determined mainly by labour times. Furthermore, classical economists share the view that labour employed in production (in general) is responsible for the creation of surplus. Hence, some caution should be applied because Marx excluded, from his characterisation of classical economics, major economists such as Thomas Malthus (1706–1834) and John Stuart Mill (1806–1873), who not only questioned the validity of the labour theory of (exchange) value but also were eager

[1]In fact, Marx uses the term "classical political economy" (see Marx 1859, p. 52 and 1865, pp. 174–175).

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_6, © Springer-Verlag Berlin Heidelberg 2010

to replace it with the ephemeral forces of supply and demand. Marx has also used the term 'vulgar economists' to refer to those whose analysis was based on the surface phenomena of supply and demand.

The term classical economics has also been used by Keynes to include all economists that accept Say's law of markets (for details see below). In this sense, economists like Alfred Marshall and Arthur Cecil Pigou (1877–1959), although they lived during Keynes's time, are nevertheless characterised as classical economists (Keynes 1936, p. 3).[2] Of course, some of them who have been characterised as classical economists would oppose vehemently the idea that they accept Say's law, but Keynes anticipating this plea noted,

I doubt if many modern economists really accept Say's Law that supply creates its own demand. But they have not been aware that they were tacitly assuming it

> [. . .] The explanation is to be found, I suppose, in the tacit assumption they every individual spends the whole of his income either on consumption or the buying, directly, or indirectly newly produced capital goods. But, here again, whilst the older economists expressly believed this, I doubt if many contemporary economists really believe it. (Keynes 1937b, p. 223)

Keynes's characterization of classical economists of all the economists prior to his *General Theory* is not necessarily a manifestation of arrogance, but rather an effort to emphasise the fact that all the economists before him thought that savings determine investment and in doing so essentially accepted Say's law regardless that they even argued against it!

Today, most historians of economic thought would accept Marx's definition. However, there are important exceptions. For example, Eagly (1974) includes the Physiocrats and Marx in his definition of the classical school, and his characterisation stretches as far as to include the founders of the neoclassical school (see Chap. 7). According to Eagly, the unifying principle shared by all these economists is the so-called 'wage fund doctrine', that is, the capital advanced in order to pay for wages. The underlying idea in the wage fund doctrine is that the production process takes place prior to sales and so the capitalists must have set aside a certain amount of money for wage payments. It goes without saying that the wage fund was particularly important in the nineteenth century and its importance diminished with the passage of time, inasmuch as the credit system can generate enough funds to keep businesses going with their current expenditures before they sell their produced output.

From the above, it becomes clear that in order for the definition of classical economics to be meaningful it should not be based on the calendar but essentially on ideas. This is equivalent to saying that there must be a core of ideas that are shared by the economists who form that particular school of economic thought. For this purpose, we adopt (with some qualifications) the view expounded by Garegnani (1977 and 1984), Eatwell (1977 and 1983), Kurz and Salvadori (1995) *inter alia*, whereby in the classical school all economists who share a common set of data are included, or exogenously determined variables (to be explained below). In this

[2]The same economists today are called neoclassical economists.

sense the classical school would encompass economists such as Quesnay, Smith, Ricardo, Marx, Malthus, and J.S. Mill.

The classical school of economic thought is not necessarily an approach of the past but finds application also in the present, as this can be judged by the virtue of the fact that in the recent decades an increasing number of economists base their research on the same set of data (or independent variables) as the old classical economists did. In our view, this definition of the classical school is still operational and even modern economic problems can be addressed with it. Of course, we cannot say that all the classical economists shared exactly the same set of data with exactly the same philosophy, but it is certain that these data can be identified in Smith but more so in Ricardo and Marx. Furthermore, the classical analysis can be cast in terms of a linear model of production with the aid of which one can estimate equilibrium prices and address other questions and in so doing lay the foundation for an alternative and, at the same time, more promising microeconomic theory. In what follows, we discuss the 'long period method of analysis' and the data of the classical theory; Say's law and the determination of the level of output. The linear model of production and the determination of the relative prices of commodities follow and the chapter concludes with the famous Malthus–Ricardo debates on the possibility of the crises of overproduction.

6.2 The Long Period Method of Analysis

The major problem that the classical theory of value and distribution was confronted with was the relation between the creation of surplus and the functioning of the system of prices that allows the appearance of surplus in the forms of profit, rent, interest, *etc*., whereas the real wage (basket of goods that workers normally purchase) appears in the form of money wage. In the classical approach, the surplus is defined as the difference between the commodities produced and those that are required for the reproduction for society. In mathematical terms, the surplus is equal to the vector of gross output minus the vector of intermediate inputs and real wages (Eatwell 1981; Semmler 1984). The surplus (S) therefore is equal to the vector of gross output (X) after the subtraction of the vector of intermediate inputs (II) and the vector of real wages (W). In symbolic terms we have:

$$S = X - (II + W)$$

This difference is called surplus since this is a disposable quantity that can be spent any way that society wants to in order to reproduce itself. The manner in which surplus is allocated into production and consumption determines the scale of society's reproduction into stationary, expansionary, or contractionary. Since surplus is found in a natural form (*i.e*., as a vector of heterogeneous goods) we cannot refer to its specific forms (*e.g*., profit, rent, interest, taxes, *etc*.) unless we use a system of prices. The question that comes to the fore is how does the surplus

produced appear in the value form? The answer is connected to the estimation of the rate of profit as the ratio of surplus produced to the capital employed, which includes the means of production and wages. Formally,

$$r = \frac{S}{II + W}$$

Hence, we have the following problem: the variables in the definition of the rate of profit are cast in terms of the vectors of heterogeneous goods; as a result the elements of each vector cannot be added and also vectors cannot be divided. Consequently, there must be a way that the above heterogeneous quantities of goods can be homogenised, that is be converted from vectors to scalars; such a conversion can be done with the aid of a theory of value.

According to Sraffa (1951), the problem was set out in a straightforward way by Ricardo in his *Essay on the Influence of a Low price of Corn on the Profit of Stock* published in 1815 (See *Works* IV), where he struggled to offer a solution. He hypothesised an economy with only one sector, whose output is the same commodity as its input. A commodity such as corn can be used both as raw materials and, at the same time, as a means of consumption. Consequently, since we are essentially in a one-commodity world there are no prices and the rate of profit is estimated in purely physical terms. Furthermore, Ricardo argued that free competition ensures that the prices of other commodities must be determined in such a way as to give a rate of profit, which is equal to that in the corn sector of the economy. This is how Ricardo determined in principle the economy's general rate of profit and the equilibrium prices in the total economy.

The apparently restrictive assumption of a single commodity as being an input and output at the same time and that workers subsistence wage consists exclusively of this commodity (corn) made Ricardo to seek other more realistic solutions to the problem of evaluation of inputs and outputs. He therefore, turned to the labour theory of value (relative prices), in an effort to homogenise the vectors of goods produced and the inputs that go into their production. Thus, if all of the above variables are expressed in value terms, then it follows that the numerator in the formula of the rate of profit can be thought of as the sum of profits, while the denominator, in the same formula, can be thought of as the value of the means of production and of the means of reproduction of workers. Hence, it might be asked why not measure everything in terms of normal prices? The trouble with such a measurement is that in order to determine normal prices we need to know the general rate of profit and to know the general rate of profit we need to know normal prices, so we find ourselves going round in circles. This is the reason why classical economists in their theory of value kept the determination of natural prices separate from the rate of profit. This is especially true of Ricardo and Marx (see Chaps. 4 and 5). Hence, we will not proceed further with any specific theory of value, a subject that we have already pursued in detail in our discussion of Smith, Ricardo, and Marx. Our goal in this chapter is to clarify the structure of classical theory and more specifically the relation between the theory of value and the theory of total output.

The classical analysis assumes that free competition will tendentially equalise the inter-industry profit rates to the general one of an economy. The mechanism for this tendential equalisation is the acceleration or deceleration of capital accumulation. For example, if an industry makes a rate of profit above the economy's general one, the accumulation of capital in this industry accelerates and the expansion of its output reduces the price to the level that gives the general rate of profit. The converse is true for industries that make a lower rate of profit than the general one. In this case the deceleration of accumulation and the reduction of output raise the price of the product to a level where it incorporates the general rate of profit.[3] This position that the economy gravitates towards where prices, outputs, and rate of profit are at their normal levels is called long period equilibrium and the analysis of such positions is called the long period method. It is important to point out that the long period method does not necessarily refer to a long period of calendar time but rather to the analytical time, which is required until the economy attains its normal position. In this chapter, we focus on classical economics and its method of analysis, which is based on the following set of data:

1. The size and composition of output
2. The technique of production, or the fact that we know with precision the necessary inputs of raw materials and quantities of labour per unit of output
3. The real wage

Clearly, none of the old classical economists specified the core of his theory in any explicit way that resembles the above specification. The use of a core is only implicit in the writings of the classical economists and became explicit mainly with the help of the writings of Sraffa (1960), Garegnani (1979 and 1984), Eatwell (1977 and 1983) and Kurz and Salvadori (1995).[4] On the basis of the above data, we can estimate the set of relative prices that correspond to the maximal rate of profit or alternatively determine the set of relative prices that minimise the cost of production of the economy that produces a given level of output. It is important to point out that the above data refer to the general case where we have goods that are reproducible. In case that we have non-reproduced means of production, such as land for example, we need to know the various qualities of available land and the stocks of non-renewable resources, such as mineral products. In this case, we can estimate rent together with the rate of profit (Kurz and Salvadori 1995). In what follows, we examine the details of each datum of the classical theory of value and distribution.

[3]This adjustment mechanism, where prices react on outputs and outputs react on prices is called *cross dual* and its details have been discussed in Chap. 2, while its formal presentation can be found in Flaschel and Semmler (1989).

[4]Eatwell (1977) argued that in the determination of the rate of profit and the equilibrium of relative prices there is no need to assume any specific type of returns to production. That is to say, there is no need to assume that returns to production are constant in order to determine the rate of profit. The idea is that there is no interrelation between outputs and prices or between outputs and inputs, since output is taken as an independent variable.

6.2.1 Given Output

This is an assumption that is found in works by the old classical economists, for example when they examine cases where cost conditions change; *i.e.*, increase in the price of raw materials, taxes, and the like. In such a situation, classical economists argued that the owners of capital must increase their outlays on production in order to be able to produce the same amount of output. For example, Ricardo, in various parts of his work, uses what today is called comparative static exercises. In particular, he starts from an equilibrium situation, and then he assumes an exogenous change (*e.g.*, a new technique of production, increase in wages, *etc.*) and studies how the economy would gravitate towards another equilibrium (price) position. Such circumstances arise particularly in the chapters of taxation, where the introduction of taxes leads the economy from one set of equilibrium prices to another with the technology, the real wage, and the level of output remaining the same. So differences may arise because of changes in the profit and in the growth rate, but these require a much longer time to work themselves out. Marx also makes the same assumption; for example, in Chap. XI in *Capital* III, where he examines the effects of wage fluctuations on prices of production, or in *Theories* III. The same is true for the schemes of reproduction in *Capital* II, where Marx assumes that the social product is given in a similar way to that of Quesnay in the *Tableau Economique*.

The assumption of a given output does not imply that there cannot be a theory of output in the classical setting. It only means that output determination requires an altogether different theory from the theory of value. This is a characteristic feature of the classical approach, which is markedly different from the neoclassical. As we will discuss in Chap. 5, prices and outputs, in the neoclassical approach, are interrelated, and are therefore determined simultaneously. The separation of the theory of value from the theory of output should not be seen as a weakness of the classical approach. On the contrary, such a separation makes it necessary to supplement the classical theory of value and distribution with a theory of output. It has been argued that the Keynesian theory of output determination, which (as we will discuss in Chap. 8) is totally independent of the theory of value, could be integrated with the classical theory of value and distribution. Although this is a much-praised goal, the results up until now are far from satisfactory (Eatwell 1983; Flaschel and Semmler 1989; Kurz and Salvadori 1995).

6.2.2 Given Technique

Classical economists assumed that for the estimation of the rate of profit and the equilibrium of relative prices associated with it, the technique of production should be considered as given. Once again, this should not be interpreted to mean that classical economists ruled out technological change; on the contrary, technological

change holds the centre stage in their inquiry, as it can be seen by referring, for example, to Smith's discussion of the division of labour (see Chap. 2), Ricardo's famous chapter on Machinery (see Chap. 3) and Marx's discussion about mechanisation (see Chap. 4). More specifically, classical economists assumed that technological change takes place at a slow pace and fundamental changes appear only after the passage of long periods of time. Only after the passage of relatively long periods of time may one observe substantial changes in the input–output coefficients and the labour input-coefficients, which in turn may lead to empirically significant changes in prices and the distributive variables. Thus, the determination of the equilibrium of prices and the general rate of profit at different distant periods of time can be used as a way to measure the technological change that has taken place. Empirically, it has been found that the labour content of commodities decreases over time, which is another way of saying that productivity increases, and, therefore, technological progress has been achieved (Carter 1981). The latter can be measured with the aid of a system of prices that takes the technique in use in each particular period of time as a datum. Comparisons then of input–output coefficients as well as of labour and capital input coefficients become possible. Such comparisons in our view are much more meaningful than, for example, those that assume continuous substitutability of factors of production as is the case with the neoclassical approach of economics that we study in Chap. 7.

It is important to stress, once again, that the idea of a given technology that is necessary to estimate the equilibrium prices does not signify that there is no technological change but rather indicates that the changes in input–output coefficients are so slow over time that one needs the passage of a rather long period of time in order to notice important changes which will lead to an altogether different set of equilibrium prices. We know that small changes in the input–output structure of the economy do not affect the equilibrium prices in any empirically significant way.[5] In other words, the equilibrium prices are robust to changes in the input–output structure of the economy. Furthermore, we do know that a small change in the price of inputs does not induce the entrepreneurs to switch to a new technique of production. As Leontief (1986) argued, techniques of production are like 'cooking recipes' and as such the proportions between inputs must be right for successful results.

6.2.3 Given Real Wage

Classical economists thought that the real wage is determined prior to the determination of relative prices and the general rate of profit. Thus, for a given period of

[5]We would say that relative prices and the rate of profit are inelastic to the changes in the input–output or labour-input coefficients.

time it is reasonable to assume the real wage as a datum.[6] This view is found, for example, in the Physiocrats who claim that

> the level of wages, and consequently the enjoyments which the wage-earners can obtain for themselves, are fixed and reduced to a minimum by the extreme competition which exists between them. (In Meek 1963, p. 194)

A thesis like the above which is repeated in the writings of the classical economists led Marx to the following generalisation:

> the foundation of modern political economy, whose business is the analysis of capitalist production, is the conception of the value of labour power as something fixed, as a given magnitude – as indeed it is in practice in each particular case. (Marx, *Theories* I, p. 45)

Whereas in *Capital* III Marx writes about the independent determination of the real wage:

> Moreover, dragging competition into this problem does not help at all. Competition makes the market-prices of labour rise or fall. But suppose supply and demand of labour are balanced. How are wages then determined? By competition. But we have just assumed that competition ceases to act as a determinant, that its influence is cancelled due to equilibrium between its two mutually opposing forces. Indeed, it is precisely the natural price of wages that we wish to find, *i.e*., the price of labour that is not regulated by competition, but which, on the contrary, regulates the latter. (*Capital III*, p. 684)

Clearly, real wage in the analysis of Marx is taken as an independent variable. Once again, this assumption does not mean that the real wage does not change. On the contrary, classical economists discussed in detail the effects of a change in real wage on prices and rate of profit. However, they thought that the real wage (or the standard of living of workers) tends to change rather slowly over time, so it is realistic to assume it as a datum in the estimation of relative prices and general rate of profit. Alternatively, and also equivalently in mathematical terms, one could assume the other distributive variable, that is, the rate of profit as given and then estimate the relative prices and the real wage. This idea is attributed to Sraffa (1960); the trouble with the equilibrium rate of profit is that one does not know it. It could be argued that the rate of interest could be used as a proxy for the rate of profit, however the problem once again is to what extent can such an equilibrium interest rate be estimated.[7] Furthermore, Leontief (1986) opted for taking the money wage as given and estimated absolute prices with the aid of which he could form a price index to deflate, so to speak, the money wage and arrive at the real wage. Lately, Gehrke and Kurz (2006) and Kurz (2009) used the notion of

[6]The new Keynesian literature provides much more theoretical justification and empirical evidence in favour of the rigidity of real wages and also money wages. A rigidity which classical economists would not at all attribute to market imperfections.

[7]As we will see it is doubtful whether there is an equilibrium rate of interest in the sense of natural rate of interest. We know that Marx definitely ruled out such a case, whereas Keynes's position is ambivalent. It is important to note that this is still a hotly debated issue in macroeconomic analysis.

'proportional wage' which is equivalent to saying that the wage share or profit share may be taken as given in the estimation of equilibrium prices. This is certainly an interesting possibility that can find theoretical support in the writings of Ricardo (proportional wage) and Marx (given rate of surplus value) and is also consistent with the available empirical evidence. We know that the profit share is a mean-reverting variable that does not display large fluctuations and thus may constitute an interesting alternative that may supplement the assumption of a given real wage and contribute to the formation of a more general determination of equilibrium prices, when the output level and technique of production is given.

6.3 The Determination of the Level of Output, or Say's Law of Markets

The classical theory of output determination is known as Say's law of markets. Here, caution should be applied, since Say's law appears quite frequently in the macroeconomic textbooks and usually its presentation is narrowed to the catch-phrase 'supply creates its own demand'. Jean Baptiste Say (1767–1832) expressed an idea that was based on the principle that all individuals are both producers and consumers, since each producer intends to spend the surplus of his product on the purchase of other products. Thus, the production of each producer is essentially the demand for the products of other producers. In the case of two producers, the supply of the first producer is at the same time the demand for the products of the second producer and *vice versa*. As a result, apart from the equality situation, there is the possibility of excess supply of good x, or excess supply of good y but it is impossible to have excess supply of both goods. According to Say there is a normal price of exchange between the two goods. If, for example, good x is sold at a price below normal, this means that we are at disequilibrium which can be only temporary because the lower than the average rate of profit will discourage accumulation and the contraction of supply will eliminate the excess supply of x; thereby raising the market price towards its normal level. In this case, however, Say argues that the price of y should be above its normal, and, therefore, there will be a shortage of y whereby the acceleration of accumulation will bring down the market price to the normal price; thereby eliminating all excess profits and restoring equilibrium. In short, in Say's law, the supply of goods is, at the same time, demand for other goods and so the possibility of a general glut of commodities is ruled out.

Ricardo adopted Say's idea and tried to generalise it. According to Ricardo the production of goods represents either demand from the producer or the producer's demand for the commodities of other producers. For example, Ricardo notes:

> M. Say has, however, most satisfactorily shown that there is no amount of capital, which may not be employed in a country, because demand is only limited by production. No man produces, but with a view to consume or sell, and he never sells but with an intention to purchase some other commodity, which may be immediately useful to him, or which may

> contribute to future production. By producing, then, he necessarily becomes either the consumer of his own goods, or the purchaser, and consumer of the goods of some other person. (Ricardo, *Works,* I, p. 290)

The question that comes to the fore is what happens if part of the product is saved. In this case, Ricardo argued that the very purpose of the production of commodities is their investment. In Ricardo's analysis the act of savings is at the same time an act of investment (Garegnani 1978–1979). Moreover, in the analysis of Ricardo and of the other classical economists there is no procedure for the equalisation of savings and investment because for them, simply, savings were automatically invested. This view is completely different from that of the neoclassical economists, where the equality of savings and investment is brought about through variations in the rate of interest. A role for the rate of interest, which is quite different from that in classical economics, where the rate of interest allows the owners of finance capital to appropriate a portion of the total surplus produced. In Keynes, the equality of saving and investment comes about through variations in income, and the causality runs from investment to savings and not the other way around; an idea the details of which are analysed in Chap. 9.

6.4 The Linear Model of Production

We can cast this set of exogenously given variables in a modern context by expressing them in terms of an input–output table. Thus, in a particular period of time, we assume the level of output of each particular industry together with the technological coefficients and the vector of labour input coefficients (*i.e.*, the technique of production) as being given. If we combine the technique of production with the real wage, that is, the quantities of goods that are consumed by a typical worker, we can uniquely determine the general rate of profit and the vector of positive relative prices associated with it. More specifically, we have:

$$p = pA + pba_o + rp(A + ba_o)$$

The notation is as follows:
p = a row vector ($1 \times n$) of relative equilibrium prices
ς = a matrix ($n \times n$) of technological coefficients
b = a column vector ($n \times 1$) of goods that workers consume (their real wage)
α_o = a row vector ($1 \times n$) of labour coefficients
r = the rate of profit, a scalar.

The right-hand side of the above equation shows the outlays of the owners of capital on the production of a unit of output plus profits. Thus, the price of a unit of output is equal to the unit cost of the product, which includes the outlays on raw materials (pA) and wages (pba_o), together with the normal rate of profit times

the amount of money that has been invested $rp(A + ba_o)$.[8] The above equation solves for

$$p(1/r) = p(A + ba_o)(I - A - ba_o)^{-1}$$

The last expression is an eigen equation. We can use the familiar Perron-Frobenious theorem of non-negative square matrices according to which the maximal eigen value of the system corresponds to the economy's average rate of profit associated with a set of positive relative prices (Pasinetti 1977, mathematical appendix).

The important element in the preceding analysis is that in the determination of relative prices, we assumed the level of output, the real wage and the technology as given. This by no means implies that these data are invariable over time or they do not interact with each other. For example, changes in distribution lead to changes in prices, which in turn can affect the structure of demand and the final output. The central point, however, is that relative prices are determined without accounting for these interrelations. The study of technological change and of produced output is possible at a later stage provided that we have determined the equilibrium prices of the system. For the determination of prices in the classical system it is not required to know the manner in which the equilibrium quantities were determined. This means that in the classical analysis the theory of value is totally independent of the determination of output; the two theories are disconnected (Flaschel and Semmler 1990). In the neoclassical theory, as we will discuss in Chap. 7, prices and quantities are both unknown variables and are determined at the same time.

6.5 The Malthus–Ricardo Controversy on Say's Law

Malthus's view on the deficiency of total demand, as a result of the general glut of commodities, is inspired by his population theory. As the increase in population is accompanied by a fall in wages, the increase of productive capacity of the economy is accompanied by a reduction in demand. Malthus, however, continued to accept Ricardo's – and to a certain extent the classical economists' – view that savings can be automatically regarded as investment. As a result, the increase in savings (or the fall in consumption) leads automatically to an increase in the productive capacity of the economy, while the current total demand falls short of what is required to absorb the additional amount of total output produced. The value of a commodity can be reduced to the sum of wages and profits.[9] If all profits are saved, which is equivalent

[8]For the sake of simplicity we assume a circulating capital model with wages paid in advance. Of course, a more complicated model could be used and the interested reader can consult the relevant literature.

[9]Rent is price determined rather than the other way around.

to saying that profits are automatically invested, although workers are assumed to spend all their wage, it is not enough to purchase all the output produced; consequently, Malthus, was led to the conclusion that the economy is driven to a stage where the excess supply of labour coexists with the excess supply of capital, a situation which forces a simultaneous fall in the wage and the rate of profit. Malthus characterised this situation as that of a general glut of commodities.

Ricardo disagreed with such an argument that led to the absurd conclusion of the simultaneous fall of wages and the rate of profit. These two variables in Ricardo's analysis compete with each other; consequently, the fall in the rate of profit is necessarily accompanied by a rise in wages and *vice versa*. In the correspondence between the two friends (*Works* IV) it seems that Ricardo's thesis prevailed because of the purity of his labour theory of value in combination with Malthus's weakness to clearly explain his point of view. The only thing that Ricardo accepted from Malthus's argument was the idea that it is possible to have a partial excess supply of goods, which cannot, however, be generalised to the economy as a whole. Keynes a century later in formulating his own theory of effective demand came across this debate and sided with Malthus, whom he considered as the precursor of his theory of 'effective demand'. According to Keynes,

> the complete domination of Ricardo's [approach] for a period of a hundred years has been a disaster for the progress of economics. (Keynes 1956, p. 33)

and a few pages below he continues.

> if only Malthus, instead of Ricardo had been the parent stem from which nineteenth-century economics had proceeded, what a much wiser and richer place the world would be today. (Keynes 1956, p. 36)

There is no doubt that if Malthus had a theory of effective demand, the dominance of Ricardo in their debate was a setback for the development of an economic theory. A closer examination of the correspondence between the two friends reveals that Malthus accepted the automatic equality between saving and investment (Corry 1959; Garegnani 1983), which is another way to say that neither Ricardo nor Malthus had a theory for the determination of output. For both of them savings precede investment and that their equality does not necessitate variations in income as in Keynes or the interest rate as in neoclassical economics. For both Malthus and Ricardo savings are identified with investment.

Malthus, however, developed an interesting argument, which was based on the assumptions, that the owners of capital save all their profits and that workers consume all their wages (Rubin 1929, p. 298). If all profits are invested, then there are additional goods for which there is no corresponding demand, which of course is a necessary condition for investment to be profitable. The additional demand, Malthus argued, cannot come from workers whose wages consist only of a portion of the total value produced. If we extend this situation into the future then it follows that as capitalists invest all their profits the problem of the lack of sufficient demand can only get worse. Contemplating possible policy solutions to this problem (see Bleaney 1976, pp. 54–55) might include:

1. A more equitable redistribution of land[10]
2. Expansion of domestic and international trade
3. The maintenance of a class of unproductive consumers

From the list of available policies the first obviously is not only hard to apply but also is not in Malthus's philosophy of inalienated and sanctioned property rights. Furthermore, such a policy does not solve the problem of finding all the extra demand needed to absorb the excess output produced endogenously (according to Malthus) by the system. The second policy proposal certainly leads to an expansion of demand but such an expansion through trade creation is only a local and certainly an insufficient solution to the problem of excess supply. The third policy proposal is much more promising and naturally, one would turn to the unproductive nature of government and its expenditures which would certainly expand demand, while supply remains the same or even reduces. However given Malthus's philosophy against big government, the sought *deus ex machina* was found in the unproductive expenditures of landlords. Malthus argued that since landlords do not produce, it follows that their rent incomes when spent generate extra demand without affecting the supply. Malthus, in other words, argued that the role of landlords not only is not passive as the other classical economists thought, but rather active and socially useful. In fact, Malthus argued that their unproductive expenditure on the maintenance of a class of non-productive workers (servants, *etc*.) is what sustains the level of demand at high levels and aligns it to equality with the economy's productive capacity.

According to Malthus the landlords in spending their incomes to hire servants whose wages become the extra demand required to absorb the excess output produced. Consequently, landlords constitute a social class which performs a socially useful function, by consuming, and, therefore, demanding goods on the one hand, and by not increasing the amount of output produced on the other. Malthus, did not question whether the rent of landlords is sufficient to buy the extra goods produced. The major problem with Malthus's analysis is his notion of investment. For Malthus investment means more employment, which is accompanied by some increase in the raw materials invested. Malthus does not account for the fact that the production of investment goods require the operation of a separate sector in the economy that produces this type of goods. As a result, when profits of the consumer goods sector are invested, it follows that the demand for investment goods increases. Employment in the investment goods sector increases and the wages of the workers become the additional demand needed to absorb the additional amounts of consumer goods. Should Malthus have made the distinction between branches of social reproduction, then his basic question, that is, the source of the demand for the additional consumer goods, could be answered. However, the distinction in sectors of production of consumer goods and investment goods was

[10]The idea is that as more people become owners of land overall demand for consumer goods is expected to increase.

not made in any clear way by Malthus. By contrast, the Physiocrats and Marx distinguished between sectors of production in the *Tableau Economique* and in the schemes of reproduction, respectively; consequently, they developed the necessary analytical tools to address questions about the source of additional demand. Thus, in his critique of Say's Law, Malthus could not really show any shortcomings, since he essentially accepted that saving automatically become equal to investment and on the basis of this equality it is impossible to prove that there is a problem of lack of demand for the total output produced.

Marx also criticised Say's law of markets in various parts of his work. His main critique, however, is in the *Theories* II where he points out some weak aspects in Ricardo's views. For example, Marx argued that the purpose of production is not to satisfy the needs of the consumers, (that is the use-value aspect of the commodity), but rather its exchange value.

> It must never be forgotten that in capitalist production what matters is not the immediate use-value, but the exchange value and, in particular, the expansion of surplus value. (Marx 1969, pp. 493–495)

In this sense, the circuit of commodity is $C–M–C'$, where the producer sells his commodities for money ($C–M$) and with the money that he gets purchases other commodities ($M–C'$). This circuit can break down for a variety of reasons; for example, some producers hoard their money and so the crisis of overproduction from one sector may spread to the entire economy. It is important to note, that this analysis refers only to the possibility of crisis, which is endemic in the nature of the capitalist system. However, this view is too simple and Marx continued his analysis with a more advanced possibility theory of crisis based on the circuit of industrial capital $M–C\ldots P\ldots C'–M'$, where M indicates that the production process starts with money advanced to purchase C, that is means of production and labour power, P stands for production and the dots indicate the time needed for the inputs to be transformed to new products C', which are of higher value than those invested and this is manifested in M', which is higher than the initial M, by the amount of surplus (value). Hence, Marx shows that the production process starts and ends with money and this implies that savings and investment may not necessarily be equal to each other. In spite of the fact that Marx's analysis is rich in insights, he continued to accept the notion that savings determine investment, which is equivalent to saying that Marx did not have a theory for the determination of the output level. More specifically, in the circuit of (industrial) capital, each capitalist expects that the output produced will not only be sold but also sold at a price that gives at least the general rate of profit. This goal can be achieved for all producers, if aggregate demand is at a particular level. If not a series of phenomena may appear such as overproduction, crises or a fall in the general rate of profit. Whatever happens this is an area of the actuality theory of crisis, which deals with the actual behaviour of the economy. The question at hand, however, is whether or not aggregate demand, a monetary variable, is at a sufficient level to allow for the normal operation of the economy. The answer to the question presupposes a theory for the determination of aggregate demand.

6.6 Summary and Conclusions

In the classical approach, there is no theory for the determination of output which is taken as a datum together with the technology and the real wage. And on the basis of these data or independent variables classical economists could determine the equilibrium (long period) prices, without reference to supply and demand schedules as in the case of neoclassical economics. In this approach there is also an assumption of the normal position of the economy, which implies full utilisation of productive capacity. There is no reason whatsoever for the economy to underutilise (in the long run) its productive capacity, since production is, at the same time, demand.

It is important to point out that in the classical analysis and in Say's law there is no assumption about the full utilisation of the labour force. The unemployment of the labour force coexists with the normal utilisation of the economy's productive capacity.[11] The existence of unemployment appears in the clearest way in Ricardo (1951, Chap. 1) or in Marx in his notion of the industrial reserve army (*Capital* I, Chap. 25). According to classical economists the reduction in wage reduces unemployment through capital accumulation and the variation in population. This process is characteristically different from the neoclassical theory of the demand for labour, which is based on factor substitution. More specifically, classical economists assumed that in the agricultural sector of the economy there was expected to be an excess supply of labour, which would serve as a regulator of the socially determined subsistence wage. In short, labour is not something that is produced by businesses; consequently there is no interest whatsoever on the part of business to fully utilise the available labour!

Say's law does not refer to price changes as a requirement for the equalisation between supply and demand. This does not mean that there cannot be any excess supply in some sectors of the economy but that this will not spread to all industries. Excess supply in some industries is counterbalanced by a shortage of supply in some other industries and so forth. Say's law was criticised by Marx who argued about the possibility of a general crisis of overproduction. Nevertheless, Marx continued to accept that savings are responsible for investment. Only Keynes discredited Say's law by reversing the causality between investment and savings.

Questions for Discussion and Thought

1. Discuss the different meanings of classical economics for Marx and Keynes. How do Sraffian economists use the term classical economics?
2. What are the data of the classical theory and how are they used in the determination of the natural prices of commodities?
3. Do the data of the classical theory change? Explain.
4. What is Say's law of markets?

[11]This conclusion is in stark contrast with the presentation that is usually found in macroeconomic textbooks, where it is claimed that classical economists assumed full employment of labour!

5. Discuss Malthus's theory of general gluts. Was Malthus right? Explain.
6. What did Keynes think about this debate? Was Keynes right?
7. What did classical economists assume with regard to the degree of capacity utilisation and unemployment of labour? Why?

Notes for Further Reading

The core of the classical theory was identified for the first time in Sraffa's work in the 1960s. Garegnani, in his dissertation (written under Sraffa's supervision) discussed the object of analysis of the classical economists and their method of inquiry. A paper of his, published in 1976 and reproduced in Eatwell and Milgate (1983, Chap. 7), explained the classical approach in detail. Eatwell (1977) discussed the core of the classical theory and he made a lot of interesting points about the assumption of returns to scale in Sraffa. A more detailed presentation is found in his book with Milgate (1983) see especially introduction and Chap. 3. The same book includes two important articles by Garegnani in Chaps. 2 and 4 that refer to a host of issues. Among them are included the theory of effective demand and the debates between Ricardo and Malthus on the possibility of general gluts. It is worth pointing out that included in the classical approach are the contributions of John Von Neumann (1945) and of economists of the famous Kiel School (Ladislaus von Bortkievicz, Georg von Charasoff, Wasilly Leontief, and Adolph Lowe) in Germany mainly during the inter-war period (for details, see Kurz and Salvadori 1995).

The classical approach is not without its critiques. Among them the names of Paul Samuelson (1989), Samuel Hollander (1999), and Mark Blaug (1999) figure prominently. Blaug (1999) disagrees with the so-called Sraffians in their definition of classical economics. They do not do what he calls a *historical reconstruction* (*i.e.*, 'to recover the ideas of past thinkers in terms that they, and their contemporaries, would have recognised as a more or less faithful description of what they had set out to do'.) but rather a *rational reconstruction* (*i.e.*, 'the tendency to view history as a relentless march of progress from past errors to present truths'.). Kurz and Salvadori (2001) have argued that there is also the textual evidence in favour of the classical approach. Thus, the discussion of the core, *etc*. can be supported on both logical and textual grounds.

The debates between Malthus and Ricardo have been discussed extensively by Keynes (1936, Chap. 23). Here, one should be particularly careful because while the wordings of Malthus and Keynes may look alike they nevertheless have entirely different meanings. By way of an example, Malthus's concept of effectual demand has no connection whatsoever to Keynes's notion of effective demand. For Malthus, effectual demand is '[t]he quantity of a commodity wanted by those who are willing and able to pay the costs of its production' (Malthus 1827, p. 111, cited in Dome 1994, p. 23). For Malthus, the cost of

(continued)

production of a good is identified with '[t]he advance of the quantity of accumulated and immediate labour necessary to their production with such a percentage upon the whole of the advances for the time they have been employed as is equivalent to ordinary profits' (Malthus 1827, p. 111, cited in Dome 1994, p. 23). If the usual rate of profit is viewed as the rate at which capitalists are willing to maintain their scale of production constant, then the notion of effectual demand in Malthus can be taken as that which equates the rate of profit of an industry with the natural rate of profit. As a result, the effectual demand in Malthus differs essentially from that of Keynes, because the effectual demand in Keynes has no relationship whatsoever with the rate of profit but with the quantity of goods and services that is in fact purchased. As for the notion of effectual demand in Smith its difference from that in Malthus is in that, for Smith, the natural price is equal to the natural incomes of all factors of production including rent of land, while in Malthus as in Ricardo the price of agricultural products is determined on the marginal land where there is no rent. It is true that Malthus and Keynes assign special role to the demand; however, Keynes rejected Say's law by no means of output determination.

Corry (1959), however, has shown that Malthus had accepted Say's law, and, therefore, he cannot be regarded as a precursor of Keynes, while Paglin (1973) claims exactly the opposite. The points that have been raised by Garegnani (1983), in particular, the idea that both Malthus and Ricardo share the view that savings are automatically invested gives an end to the debate. However, the under-consumption argument that is found in Malthus deserves some attention. In fact, his problem about the generalised glut of commodities and the lack of adequate consumption demand is resolved, according to Bleaney (1983), once we bring into the analysis the sector producing investment goods and the demand emanating from the workers employed in this sector. The lack of adequate demand (that comes from a social class regardless if they are workers, capitalists, or landlords) is a constant feature of all under-consumptionist theories, since they tend to view the purpose of production as consumption, which of course is not true.

In this book, we do not refer to Malthus's so-called population law, which is shared by Ricardo and other old classical economists but certainly not by Marx, because it does not feature in any direct way in our presentation. The books by Foley (2001) and Dome (1996) give a critical account of Malthus's population theory.

The question of the degree of capacity utilisation and its relationship to full employment does not appear explicitly in the writings of the classical economists. However, there is no doubt that in their analysis, classics assumed that the economy utilises fully (normally) its resources – including labour. (Garegnani 1983 and Winston 1979). However, the normal utilisation does not imply the full (100%) employment of labour (Tsoulfidis 2008b). The notion of unemployment is present in the writings of classical economists. In Marx, for

(continued)

example, we have the industrial reserve army of labour and its various expressions, while in Ricardo we have a detailed discussion of the conditions that lead to the creation of technological unemployment.

A final point relates to the object of analysis of classical economists, which is the determination of equilibrium prices as centres of gravity of market prices. It has been argued that in this determination there is no need to refer at all to labour values because they differ from normal prices (Steedman 1977). One argument that is repeatedly stated is that despite the fact that labour values differ from equilibrium prices, they do not do so in any empirically significant way. The research so far has shown that both normal prices and labour values are surprisingly close to market prices (Ochoa 1989; Shaikh 1994; Tsoulfidis and Maniatis 2002; Tsoulfidis 2008a, *inter alia*). For a rather different view see Steedman and Tomkins (2000). The works of Mariolis and Tsoulfidis 2007 and 2009 deals with the issue of income distribution and its effects on relative prices.

Appendix A

A.1. The Input–Output Analysis

The input–output analysis is a technique, which is used in the study of relations between industries. The major characteristic of input–output tables is that the total input of each industry must be equal to its total output. Quesnay's *Tableau Economique* and also Marx's schemes of reproduction form the basis of modern input–output tables.

The columns of an input–output table describe the inputs of each industry to itself and to the others. The column sum of an input–output table gives the total cost of production. Costs include the rewards of primary inputs also; that is, wages, profits, depreciation, taxes, *etc*., in short, the value added. The rows of an input–output table refer to the sales of an industry to itself and to the other industries (see also Chap. 1). A portion of the total output produced is absorbed by the final demand; that is, consumption, investment, government expenditures, and exports.[12]

[12]The treatment of imports is a complicated matter and the interested reader should consult the more specialised literature.

Table A.1 Input–outputs

Outputs / Inputs		Intermediate	Demand		Final demand	Total output
	Industry 1	Industry 2	...	Industry n		
Industry 1	x_{11}	x_{21}	...	x_{1n}	y_1	x_1
Industry 2	x_{21}	x_{22}	...	x_{2n}	y_2	x_2
...	...	...	...	...	...	...
Industry n	x_{1n}	x_{2n}	...	x_{nn}	y_n	x_n
Value added	v_1	v_2	...	v_n		
Total	x_1	x_2	...	x_n		

The classical assumption of a given technology implies that if for some reason the output of industry j increases, x_j, it follows that the inputs x_{ij} of the industry j must increase proportionally. The technological coefficients are determined by the following relationship $a_{ij} = x_{ij}/x_{j,}$ where a_{ij}s are the technical coefficient or input– output coefficient. Once computed, input–output coefficients are treated as constant (fixed). Moreover, the column sum of technological coefficients for a viable economy must less than one, which is equivalent to saying that the value of output produced must not exceed the value of inputs that were used in its production.

The input–output coefficients can be converted from a descriptive invention to a useful analytical tool with the aid of linear algebra. In the interest of brevity we restrict the analysis to two industries. Thus we have:

$$x_1 = a_{11}x_1 + a_{12}x_2 + y_1$$

$$x_2 = a_{21}x_1 + a_{22}x_2 + y_2$$

where y_1 and y_2 are the final demands (consumption, investment, (net) exports, and government expenditures) of industries 1 and 2, respectively. Hence, we have a system of equations which can be written in terms of matrices as follows:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

and in compact form $x = Ax + y$ which solves for the output vector

$$x = (I - A)^{-1}y$$

the matrix $(I - A)^{-1}$ is known as the Leontief inverse whose columns show the input requirements, both direct, and indirect, on all other producers, generated by one unit of output.

A.1.1 Price Determination

In what follows we introduce prices p_1 and p_2 which correspond to the outputs of industries 1 and 2, respectively. The total revenues from the sales of each sector will be:

Revenues of industry 1: p_1x_1
Revenues of industry 2: p_2x_2

The total cost of each industry is:

Cost of industry 1: $(p_1a_{11} + p_2a_{21})x_1$
Cost of industry 2: $(p_1a_{12} + p_2a_{22})x_2$

We further assume a uniform profit rate r and we get:

$$p_1 = p_1a_{11} + p_2a_{21} + r(p_1a_{11} + p_2a_{21}) = (1 + r)(p_1a_{11} + p_2a_{21})$$

$$p_2 = p_1a_{12} + p_2a_{22} + r(p_1a_{12} + p_2a_{22}) = (1 + r)(p_1a_{12} + p_2a_{22})$$

or in matrix form:

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = (1 + r)\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$

The problem now is to find the rate of profit and prices which are consistent with the givens of this economy. Clearly, the profit rate will correspond to the maximum eigen value of our hypothetical economy while equilibrium prices will correspond to the associated eigenvector. If we therefore write the above system in compact form, we will have:

$$p'(1/1 + r) = A'p' \text{ or } p\gamma = pA,$$

where the eigenvalue $1/1 + r$ and where the eigenvector p corresponds to the vector of positive relative prices.

A.1.2 A Numerical Example

For the better understanding of the preceding analysis let us take a realistic input–output table in order to present a series of questions that we referred to above. The data are presented below:

Table A.2 Aggregated input–output table

Outputs Inputs	Agriculture	Industry	Services	Final demand	Total output
Agriculture	8,969	32,919	484	28,810	71,189
Industry	8,180	84,296	16,661	135,380	244,519
Services	3,034	18,286	9,350	84,204	114,875
Wages	40,190	33,947	28,858		
Indirect taxes	137	31,567	8,864		
Other value added	10,679	43,504	50,658		
Total output	71,189	244,519	114,875		

The input–output coefficients of the above input–output table are estimated if we divide each of the inputs of each sector by the total output produced. Thus we have:

$$A = \begin{bmatrix} 8969/71189 & 32919/244519 & 489/114875 \\ 8180/71189 & 84296/244519 & 16661/114875 \\ 3034/71189 & 18286/244519 & 9350/114875 \end{bmatrix}$$
$$= \begin{bmatrix} 0.125 & 0.134 & 0.004 \\ 0.114 & 0.344 & 0.145 \\ 0.042 & 0.074 & 0.081 \end{bmatrix}$$

We observe that the sum of no column or row of matrix ς exceeds one and as a result the economy produces surplus and thus it is capable of reproduction (for details see Passineti 1977).

A.1.3 The Marxian Theory of Value and Direct Prices

If we symbolise the row vector of values of produced commodities by λ, the row vector of direct labour coefficients by a_o the indirect labour, that is, the labour contained in the inputs which are used in the current production of commodity j by $\lambda_j a_{ij}$ and the value of depreciation, that is the wear and tear of the fixed capital invested in every production period, by $\lambda_j d_{ij}$, then we will have:

$$[\lambda_1, \lambda_2, ..., \lambda_n] = [a_{o1}, a_{o2}, ..., a_{on}] + [\lambda_1, \lambda_2, ..., \lambda_n] \begin{bmatrix} a_{11}+d_{11} & a_{12}+d_{12} & ... & a_{1n}+d_{1n} \\ a_{21}+d_{21} & a_{22}+d_{22} & ... & a_{2n}+d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}+d_{n1} & a_{n2}+d_{n2} & \cdots & a_{nn}+d_{nn} \end{bmatrix}$$

or in compact form

$$\lambda = \alpha_o + \lambda\ (A + D)$$

where D is the matrix of depreciation coefficients. The above equation solves for:

$$\lambda = a_o[\mathrm{I} - (A + D)]^{-1}$$

Using the above input–output numerical example together with the corresponding matrix of depreciation coefficients and the vector of the labour input coefficients, that is,

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0.012 & 0.019 & 0.029 \\ 0.006 & 0.004 & 0.002 \end{bmatrix} \text{ and } a_o = [\,0.592 \quad 0.198 \quad 0.422\,]$$

and after the appropriate substitutions we get:

$$\lambda = [0.784\ 0.547\ 0.567]$$

The row vector λ gives the quantity of homogenised labour contained (directly and indirectly) in the output of each sector. The notion of value in Marx is, as we know, monetary. Thus we have to transform the above quantities of direct and indirect labour time to direct prices, which we symbolise by d_j. For the estimation as well as for the comparison of direct prices with the market prices, we consider that the market price of each unit of output of a sector is equal to 1 monetary unit. We stipulate the following normalisation condition $d\lambda x = \lambda ex$ and so the vector of labour values is transformed to direct prices as follows:

$$d = \lambda \frac{e \cdot x}{\lambda \cdot x} = [0.784\ 0.547\ 0.567] \cdot 1.68$$

so we get:

$$d = [0.846\ 1.025\ 1.064]$$

We observe that the vector of direct prices is extremely close to the vector of market prices, which we symbolise by e and which is the row unit vector. This can be identified by using one of the usual measures of deviations:[13]

- The mean absolute deviation is 0.147 or 14.7%.
- The mean absolute weighted by the output deviation is 0.098 or 9.8%.

A.1.4 Prices of Production

In Marxian analysis prices of production are defined as the prices that incorporate the economy's general profit rate. For their estimation, we need the real wage of the workers, that is, the basket of goods that the workers spend their wage money on, *i.e.*,

$$w = pb$$

where w is the wage money and b is the $(nx1)$ vector of real wage goods. In addition, we take into account the following matrices that correspond to the above

[13]For the definitions see footnote 37 in Chap. 5.

numerical example:

$$<t> = \begin{bmatrix} 0.0019 & 0 & 0 \\ 0 & 0.1291 & 0 \\ 0 & 0 & 0.0771 \end{bmatrix}, \; K = \begin{bmatrix} 0 & 0 & 0 \\ 1.136 & 0.480 & 1.8918 \\ 0.0272 & 0.0410 & 0.08083 \end{bmatrix} \text{ and}$$

$$b = \begin{bmatrix} 0.0046 \\ 0.0124 \\ 0.0118 \end{bmatrix}$$

where $<t>$ is the diagonal matrix of indirect tax coefficients, K is the matrix of fixed capital stock coefficients and b is the vector of wage goods. Thus prices of production are defined as

$$p = pba_o + pA + pD + p<t> + rpK$$

where ba_o is a new matrix that represents the quantity of commodity i which is required for the consumption of workers in order to produce commodity j. The above relation after some manipulation gives the following eigenequation:

$$p(1/r) = pK[I - (ba + A + D + <t>)]^{-1}$$

According to Perron–Frobenious theorem only the maximal eigenvalue is associated with a unique positive left-hand side eigenvector which gives the vector of relative prices. In terms of our numerical example we get the rate of profit $r = 0.136$ and the normalised vector of prices of production which will be:

$$p = [1.054 \; 0.971 \; 1.027]$$

Once again we invoke the statistics of deviation:

- The mean absolute deviation is 0.036 or 3.6%.
- The mean absolute weighted deviation is 0.032 or 3.2%.

Chapter 7
The Structure of the Neoclassical Theory

In fact, the whole world may be looked upon as a vast general market made up of diverse special markets where social wealth is bought and sold. Our task then is to discover the laws to which these purchases and sales tend to conform automatically. To this end, we shall suppose that the market is perfectly competitive, just as in pure mechanics we suppose to start with, that machines are perfectly frictionless.

(Walras, *Elements of Pure Economics*, 1874, p. 84).

Repeated reflection and inquiry has led me to the somewhat novel opinion, that value depends entirely upon utility. Prevailing opinions make labour rather than utility the origin of value; and there are those who distinctly assert that labour is the cause of value.

(W.S. Jevons, *The Theory of Political Economy*, 1871, p. 1).

The fact is, that labour once spent has no influence on the future value of any article: it is gone and lost forever. In commerce, bygones are forever bygones.

(W.S. Jevons, *The Theory of Political Economy*, p. 164).

7.1 Introduction

In the structure of the classical theory, we know that utility and subjective evaluations of the usefulness of commodities play no role in the determination of relative equilibrium prices and the economy's general rate of profit. Hence, we need to distinguish between the use-value of classical economists and the utility of neoclassical economists. Use-value, for classical economists, meant the property of a good to satisfy human needs (real or imaginary), whereas utility – a concept that was developed gradually in the mid to the end of the nineteenth century – means the satisfaction that an individual derives from the consumption of a good or the use of a service.[1] Since utility refers to the consumer and the intensity of satisfaction that he or she derives from a good or a service, it follows that utility has an apparently

[1]The idea of utility was based on the theories of Jeremy Bentham (1748–1832), who defined utility as the ability to reduce pain or increase pleasure.

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_7,

subjective character. By contrast, use-value refers to the properties of a commodity to satisfy social needs; it has an altogether different character.

In the next section, we discuss the emergence of the neoclassical theory and its associated marginal revolution in economics. Subsequently, we deal with the development of the structure of the neoclassical theory restricting ourselves to its absolutely essential building blocks. We continue with the determination of prices and outputs in the pure exchange model and, within this model, we introduce a simple version of Walras law. The model for production follows but it includes only non-produced means of production (*i.e.*, labour and land). The generalisation of production with produced means of production (*i.e.*, capital) presents insoluble problems, which we deal with in the next chapter. Finally, we summarise and make some critical remarks about the nature of the neoclassical theory.

7.2 The Silent Marginal Revolution

In our discussion of the structure of the classical theory, we found that the core of this theory is based on a set of data, which can be objectively measured and on the basis of these data we can determine the normal prices of the economy. One wonders whether there was something wrong with this theory and if not then what led to its replacement by the neoclassical theory based on an apparently subjective concept such as utility. We know that this is an open question and there are no definitive answers. Usually, historians of economic thought single out one reason and they find evidence against it. In our view, the replacement of the theory of value based on the labour time by the neoclassical was not immediate and took many decades until the classical ideas were set aside. In their first intellectual struggle over the theory of value, neoclassical economists had already established themselves in the 1870s and switched the research agenda towards their theory of value based on utility and gradually incorporating around it issues of public finance and international trade.

In order to deal with this issue, we need to bear in mind the intellectual atmosphere that was developed in the middle of the nineteenth century. We know that during this time period there was a rising concern with the labour theory of value; this was particularly true in the UK during the post-Ricardian period, when the Ricardian labour theory of value found strong support among a number of socialists and, in general pro-labour, economists (*e.g.* Robert Owen, William Thompson, Thomas Hodgskin, *inter alia*), who furthermore extended the labour theory of value to its logical, and what is worse (in the sense of attracting discontent from the establishment) to some normative, and, therefore, socially dangerous conclusions. More specifically, these economists argued that since labour creates the value of commodities, it follows that the exchange of commodities should be proportional to labour times. Hence, economists of the post-Ricardian period criticised Ricardo (and the other classical economists) for accepting the capitalist *status quo*, in the sense that commodities in capitalism do not precisely exchange

according to their labour times, and they claimed that since the value of commodities is created by labour alone, it follows that the capitalist's profit and landlord's rent are direct deductions from this value, which naturally belongs to workers. In short, there are no moral justifications for the profits and rents, and, therefore, the total sum of values of commodities in a society is created by labour and should belong to workers. This rather normative variant of the labour theory of value gave rise to an anti-capitalist movement. A problem of this sort that was developed within the framework of the labour theory of value (or relative prices) together with the inability of the proponents of such a theory to provide satisfactory answers to certain thorny questions (*i.e.*, the role of demand, the quantity of capital employed, *etc.*) led to the 'disintegration of the Ricardian School'. Gradually, many economists abandoned the idea of natural price being determined by permanent forces and instead the gist of their analysis became the idea that the prices of commodities are determined by the ephemeral forces of supply and demand, that is, by competition.[2]

Some historians of economic thought characterise this succession as a silent non-revolutionary process (Blaug 1983; Hollander 1985). No matter how long it took for this process to fully unfold, its very purpose (stated or not) was to set aside the more realistic classical approach because of its disturbing political implications, especially those emanating from the labour theory of value (price) and its association with Marxism and socialism. The idea that the value of commodities is determined by their labour content was too challenging for a system that underwent a structural transformation. Industrial capitalists, up until the middle of the nineteenth century, were directly involved in the production process in their incessant pursuit of expanding profits as a purpose in itself. The labour theory of value contributed to the understanding of the source of profit as well as the source of incomes for the merchant and the landlord classes. As a result, the labour theory of value was the product of, and at the same time contributed to, the intellectual atmosphere for about a century. However, the growth of corporation and the subsequent concentration and centralization of capital that took place during the depression of 1873–1896 changed the structure of the economy as well as the role of the capitalist. The capitalist's direct involvement in the production process and other related activities was limited and the management of the newly created large-scale enterprises was transferred to a small group of owners or professionals. As a consequence, the capitalist class was transformed, to a great extent, into a mere recipient of profit incomes by virtue of property rights in a way similar to that of landlords. Naturally, under these new conditions, it became clear that profit income could not find justification in a labour content explanation of equilibrium prices, other than some form of exploitation of labour. This was already explicitly stated by

[2]Among the economists who supported such a view were Malthus and Nassau Senior (1760–1864). Marx called the followers of this view "vulgar economists", since for them competition or, what amounts to the same thing, the forces of demand and supply determine the equilibrium prices of commodities. In particular, a shortage (surplus) of a good, *ceteris paribus*, leads to high (low) price.

John S. Mill (1848) in his *Principles* – a text that continued to be popular until the turn of the nineteenth century – where he explains that the 'cause of profits is that labour produces more than is required for its support' and concludes that 'profits arise [...] from the productive power of labour' (Mill 1848, pp. 416–417).[3] Such views were regarded as socially dangerous for the *status quo* and their dissemination should not be allowed due to their social implications. John B. Clark (1847–1938), for example, reflecting the sentiment of his time illustrates very vividly the socially dangerous consequences that the labour theory of value may exert and notes:

> The indictment that hangs over society is that of 'exploiting labour'. 'Workmen' it is said, 'are regularly robbed of what they produce. This is done within the forms of law, and by the natural working of competition.' If this charge were proved, every right-minded man should become a socialist; and his zeal in transforming the industrial system would then measure and express his sense of justice. If we are to test the charge, however, we must enter the realm of production. (Clark 1908, p. 4)

This is not to say that the first neoclassical economists were insensitive to social problems and that they did not propose solutions for such problems. For example, J.B. Clark favoured minimum wage legislation, in those cases where the real wage lied below the marginal product of labour (Prasch and Sheth 1999); Walras was in favour of the nationalisation of land and advocated that the rent which would be collected and could be used to replace taxation (see Niehans 1990), Knüt Wicksell (1851–1926) was a radical who proposed a fairly revolutionary programme of income redistribution from the rich to the poor, an idea that was reasoned out from the strict application of the principle of diminishing marginal utility of income (see the next section). Furthermore, the first neoclassical economists argued for government intervention in case of externalities and in the USA favoured antitrust legislation. It is important to stress, however, that the first neoclassical economists were always under the spell of the classical economists. For example, the anti-trust legislation was not based on application of the efficiency criteria of the neoclassical theory but rather on wealth-transfer concerns which were not different than those of the classical economists (Hunt 2000, 248–250). The same is true with regard to public finance issues, Schumpeter for example points out,

> Smith's book on public finance [...] was to become the basis of all the nineteenth-century treatises on the subject until, mainly in Germany, the 'social' viewpoint – taxation as an instrument of reform – asserted itself'. (Schumpeter 1954, p. 186)

The first neoclassical economists, *i.e.*, the triad Jevons, Menger and Walras initially and subsequently Marshall, J.B. Clark and Böhm-Bawerk, contributed to the creation of a new intellectual atmosphere in which the classical system was found to be unsatisfactory and its replacement by a theory that would legitimise property and emphasise the merits of an exchange economy became imperative although not necessarily urgent. It is important to point out that these ideas were

[3]Smith had also expressed a similar view (see Chap. 3).

developed in the 'Victorian Era', which was a period of steady economic growth and so the demands for a realistic economic theory from policy makers were much more elastic. Whereas, in the period of depression of 1873–1896 both classical and neoclassical theories were in agreement with respect to no government intervention.

The task of the gradual replacement of classical theory by the neoclassical one was accomplished by the architect of the neoclassical economics Alfred Marshall, who was very conscious of the status of the economic discipline of his time and the requirements for its future direction. He realised that more than a century of dominance of classical theory could not just be overthrown in a short period of time and that for the construction of a new theoretical perspective one needs to plan and above all to compromise with the hitherto dominant theory.[4] Some of the corrective compromises that he proposed included the following:

1. The labour theory of value should be reduced to a cost of production theory, with Ricardo being credited as a forerunner of this 'cost of production' concept, his only weakness resting in that he was not fully attentive to an analysis of the demand side of the market.
2. The distinction between productive and unproductive labour should be abandoned at some future and more appropriate time (Marshall 1890, p. 54)
3. The notion of competition as a dynamic process of rivalry between firms should give way to the idea of perfect competition and
4. The classical economists' notion of economies of scale which are the result of competition and division of labour that evolve over historical time must be replaced by the static economies of scale, where time is conceptual (Marshall 1890, Chaps. 9–13).

The lack of realism in this analysis was compensated for by transforming the neoclassical approach into the image of the natural sciences and especially physics. The extensive use of mathematics and also of neutral language that one finds in the writings of the major representatives of this approach served to underscore this purpose (Mirowski 1984).

Hollander's (1973, 1979, 1985) view is much more extreme as he contends that there was no break between neoclassical and classical economics. In Hollander's view, Smith, Ricardo (even Marx), Walras and Marshall are all general equilibrium theorists. There are two strands in this general equilibrium theory, the first starting from Walras and Marshall to Samuelson, Arrow and Debreu today, in which the priority is over price determination and that all relevant economic variables are determined simultaneously, and a second strand that starts with Smith, Ricardo, J.S. Mill and continuous with Marx, Sraffa and includes even Keynes. In this strand income distribution takes priority over pricing, where economic variables are determined sequentially starting from a predetermined real wage. As for Blaug,

[4]The motto of his (1890) book "natura non facit saltum" is quite revealing of Marshall's philosophy and approach to economics.

he rather holds an agnostic position by not finding any of the suggested reasons for the 'marginal revolution' persuasive enough (see also Dome 1992, pp. 86–89).

As a consequence, with the exception of the general dissatisfaction of the classical theory of value and distribution, we certainly cannot pinpoint any specific single reason that gave rise to the marginal revolution; at the same time, we cannot say that some of the above mentioned reasons should be completely ruled out. We can certainly say that the dominance of the neoclassical theory was not easy and that from the very early stage it faced difficulties and problems of internal consistency that accompany it to the present time.

7.3 Salient Features of the Neoclassical Theory

In case we accept that equilibrium prices are determined through the forces of demand and supply, that is, through competition, then a series of vexed questions and problems arise. The major problem is that as supply and demand are determined by different factors; it follows, they are not homogeneous, and, therefore, not comparable to each other. This heterogeneity of supply and demand was pointed out by J.S. Mill (1806–1873), who in his critique of Nassau Senior (1790–1864) argued that a theory of price determination through demand and supply must express both demand and supply in terms of homogeneous units and since it is only then that the two variables are rendered comparable. With respect to demand, the idea was that it depends on utility and, specifically, on what in modern terminology would be called 'marginal utility'; that is, the utility derived from the consumption of an additional unit of the good in question. It was claimed that as the consumption of a good increases, the satisfaction that a consumer derives from the consumption of successive units of the good in question progressively diminishes. This is known as the 'law of diminishing marginal utility'.[5] Consequently, the consumer would be willing to pay a lower price for higher quantities of the same good. Thus, we may construct a typical demand curve, that is, a schedule between prices and quantities, with a negative slope precisely because it reflects the law of 'diminishing marginal utility'. The next step is to aggregate these individual demand curves to arrive at the market demand curve.

While the derivation of the demand curve was relatively easy we cannot say the same thing about the supply curve. Economists had accepted the idea that supply is determined by cost. Specifically, wages are determined by the cost of reproduction of workers, rents are determined by the bargaining power of the landlords vs. the capitalists (farmers), whereas profits are determined by the scarcity or the sacrifice of the entrepreneur. Senior, in particular, argued that the sacrifice of the

[5]Marginal utility is present in the writings of the three pioneers of the neoclassical theory. Jevons called the marginal utility as "final degree of utility", Walras (1874, p. 145) called it "rareté" (*i.e.*, "the intensity of the last want satisfied is precisely the same thing as scarcity" and Menger characterised it 'the satisfaction of least importance' (see Dome 1992, p. 77).

entrepreneur, for example to set up a factory (instead of consuming his savings unproductively) and to operate it is compensated by profit. In this sense, the heirs of the entrepreneur receive rents, since they did not abstain from anything and according to Senior abstainism is the source of profits (Anikin 1975, p. 271). Consequently, with regard to profits the situation was unclear until the emergence of the neoclassical school and the contributions of the three pioneer economists, that is, Stanley Jevons (1835–1882) in England, Carl Menger (1890–1921) in Austria and Léon Walras (1834–1910) in Switzerland.[6] These three economists are considered as the pioneers of a revolution which in economics has come to be known as the 'marginal revolution'. Research, however, has shown that no such revolution, in the sense of a break with the past, has actually taken place (*e.g.*, Blaug 1983). The basic ideas of the above three economists had already been exposed by Augustin Cournot (1801–1877), Johann Heinrich von Thünen (1780–1850), Jules Dupuit (1804–1866) and Herman Gossen (1810–1858). It is important to point out that these economists did not develop a fully elaborated theory of equilibrium price determination, but they only had rich insights which could be used for the development of such a theory. In fact, their insights were limited to the marginal utilities and the construction of demand curves and also their optimization analysis according to which total utility is maximised when the marginal gain from the consumption of a good is equal to the marginal sacrifice. The so-called marginal revolution (of Jevons, Menger and Walras) was really a long-lasting process rather than a short period of time so as to call it 'revolution' proper. In fact, it took many decades for the marginal ideas to become part of the established economic theory. In other words, the emergence of the ideas of the 'marginal economists' constitutes an example of a non-revolutionary change in the history of science.

The ideas of the 'marginal economists' gradually formed an integrated theory, which has been called neoclassical economics. The term has been invented by Thorstein Veblen (1857–1929) in his effort to distinguish the ideas of Alfred Marshall (1842–1924), who argued that the ideas of the 'marginal economists' are merely the evolution of the classical economists and in this sense they are neoclassical economists. By contrast, the ideas of Jevons, who argued that the 'marginal economists' had developed an altogether different theorization and therefore their approach signifies a *break* from the approach of the classical economists. Aspromourgos (1991) argues that the term neoclassical economics is perhaps not the most appropriate, since it does not do justice to the originality of the ideas of the marginal economists. There is no doubt that there is continuation in the development of economic ideas as economists (although not (Aspromourgos 1991)) deal with the same object of analysis; however, one should not emphasise the similarity, but the difference, since by definition anything new is neoclassical.

[6]It is interesting to note that despite the fact that these economists worked independently of each other they nevertheless reached similar conclusions. Jevons published his work in 1871, the same year that Menger published his own, while Walras published his work in two sequels: the first in 1874 and the second in 1877.

The major contribution of the great neoclassical economists, that is Jevons, Menger and Walras, was that they managed to express the cost of production of commodities in terms of negative utility or disutility and in so doing to express cost in terms of a common unit of measurement. More specifically, wages are no longer theorised as the cost of reproduction of workers – clearly, a classical idea – but rather as the disutility that a worker suffers by offering his labour services. Profits are viewed as the disutility of the entrepreneur, who by abstaining from consumption manages to save resources in order to invest them and profits are the compensation for this sacrifice. Rent, on the other hand, was viewed as the compensation for the disutility that the landlord suffers by giving his land to the entrepreneur (farmer) for a certain period of time. If the cost is expressed as disutility, then it can be balanced by the utility of demand. As a result, for the first time, an adequate interpretation of the equilibrium price through the forces of demand and supply was given, since both of these forces could be evaluated through utility (positive or negative). It is important to stress that the notion of utility is intrinsically connected to individuals, since the individuals are those that acquire utility or suffer disutility.

Since the individual becomes the centre of analysis, the question that comes to the fore is how does the individual balance utility with disutility? The answer of the neoclassical economists is that the individual seeks to maximise his utility and in this pursuit he must decide on the quantities of goods that he is going to consume. These decisions depend on the initial endowment (*i.e.*, the stock of goods) that every individual possesses, and the part of the endowment he is going to offer in exchange for other goods. Clearly, the individual suffers disutility when he offers his endowment and enjoys utility in the consumption of goods that he gets through the offer of his endowment. The allocation of endowment is an optimisation problem which may be solved through calculus.

Such an exercise is within a theory of price and quantity determination. In the classical theory as we know these two variables (prices and quantities) are determined separately, while in the neoclassical theory they are determined simultaneously through the forces of demand and supply.[7] In what follows, we start with the model of pure exchange (of two goods) and generalise to the economy with production using the factors of production (labour, land and capital). The generalisation from two goods to any number of goods is conceptually straightforward.

7.4 The Model of Pure Exchange Economy

We start from the pure exchange economy, where the individuals seek to maximise the utility that they acquire from the consumption of different goods. The usual real life example is that of 'prisoners of war camps' where the individuals receive

[7]In the classical analysis output is taken as a datum for the determination of equilibrium prices. This separation makes possible the development of an altogether different theory of output.

parcels of goods through the Red Cross, and because of the differences in preferences, individuals exchange the goods that they do not want so much with others who want it more. As individuals increase their consumption, their total utility increases, but successive units of one good give less satisfaction to the individual ('the law of diminishing marginal utility' holds). As the marginal utility diminishes with the consumption of more units of a good a point is reached that marginal utility becomes zero, at this point total utility is at maximum. If the individual consumes more units marginal utility becomes negative (disutility) and total utility starts to fall.

The first neoclassical economists thought that the utility that an individual acquires is measurable and for this purpose, they invented an imaginary unit of measurement which they called 'util'. Since the utility of each and every individual can be measured, then the first neoclassical economists concluded that society's total utility is a measurable magnitude. This idea, at first sight, was totally innocuous and furthermore it was resolving the problem of what is behind the demand curve. On further consideration, however, the idea of measurable or cardinal utility in combination with the law of diminishing marginal utility gave rise to the radical idea of a totally equitable distribution of income, as a means to achieving society's maximisation of welfare. The rationale is as follows: assuming that individuals have (approximately) the same utility function and if the law of diminishing marginal utility applies to individuals, then the maximisation of society's utility is achieved in an absolutely egalitarian society. This conclusion is derived by strict application of neoclassical principles. For instance, the Swedish economist Wicksell argued that if we take income from the rich and we give it to the poor, then the utils lost by the rich people will be much less than the gains in utils of the poor people, and the redistribution of income will stop at the point where the marginal utilities of all people are equal to each other, that is, at the point where the society's total utility will be maximised. Clearly, such an idea whose logical conclusions and social extensions were competing in radicalism with the ideas of utopian socialists could not last for long. Francis Ysidro Edgeworth (1845–1926) argued that the idea of equitable distribution of income was based on the false assumption that all individuals have the same utility function and he argued that individuals have completely different utility functions. The upshot of his argument was that it is not meaningful, and, therefore, not permissible to compare utilities between people.[8] Although there is a support for the social *status quo* in Edgeworth's argument, nevertheless it was obvious that there was a problem with the so called 'cardinal utility' where the subjectivity of individual comparisons was open to criticisms. Naturally, such an approach could not last for long and the next generation of neoclassical economists made an effort to replace it by a more consistent (with the neoclassical theory) notion of utility.

[8]Jevons (1871, p. 15) argued against interpersonal comparisons of utilities, he nevertheless compared utilities between individuals and so did Menger and Walras along with their followers. Pareto argued that although interpersonal comparisons of total utility are possible, however, one cannot compare marginal utilities (see also Dome 1992, p. 83).

Indifference curves were invented by Edgeworth and refined by Irving Fisher (1867–1947) and Vilfredo Pareto (1848–1923) initially; John Hicks (1904–1989) together with Roy George Allen (1906–1983) later argued that what we find with the utility function $u = u(x_i)$ is essentially not by how much the total utility changes from the consumption of various quantities of goods x_i, $i = 1, 2, \ldots, n$, but by a combination of goods that are preferred over others.[9] These economists argued that the absolute level of utility is without meaning, and that the ranking of various combinations is what counts in economic theory. Thus, the scale may change; for example, through a monotonic transformation, however, the ranking of combinations of goods remains invariable.

The essential problem in the neoclassical theory is the full utilisation of available means of production, which amounts to the idea that the market is in equilibrium. In the pure exchange economy the individuals are assumed that have made their planned exchanges and that the stock of all available resources are fully utilised, while there are no free goods. In the case of exchange with production, the question that is posed is whether the demand for the services of the factors of production is sufficient to fully employ all of them. Consequently, when we refer to demand for the product, whose production requires the services of the factors of production, the question that comes to the fore is whether the demand is adequate enough to employ all the factors of production that are disposed of in the economy.

From a macroeconomic perspective neoclassical economics refers to the determination of demand and not of supply as is commonly thought. The problem is to find the extent to which the total demand is adequate to fully utilise the initially given stock of goods. As a consequence, our analysis in what follows will be on the demand side starting from the pure exchange economy, where the behaviour of individuals (as consumers) becomes the focus of analysis. In the pure exchange economy[10] the data of the neoclassical theory are:

1. Consumers' preferences as they are depicted in the utility functions which are characterised by a special form.
2. The initially given endowments (stock) of goods for which we know their absolute size and the way through which they are allocated to individuals. If for some reason the allocation of the initial endowments changes, say because of a change in preferences, then the prices of commodities change.

In such an economy, it is obvious that we cannot start having the income of individuals given, because income is determined through prices. The only 'income' we can hypothesise is that which comes from the initial endowment of the individuals. On the basis of a given set of prices, the individual's income (m) is the product of his endowment of goods (q_i) times their respective prices (p_i), thus we

[9] In the theory of cardinal utility, the relevant function is written as follows: $u = u(x_i) = u_1(x_1) + u_2(x_2) + \cdots + u_n(x_n)$. Where the total utility, u, is the sum of utilities derived from each good x_i, $i = 1, 2, \ldots, n$.

[10] The realistic case of the economies with production are extensions and further elaboration of the model of the purely exchange economy and their analysis is postponed until the next section.

have, $m = \Sigma p_i q_i$. In other words, we suppose that the individual sells his endowment in the market to get income and then goes again in the market and buys goods that he needs, which might include goods that he had before. Furthermore, we assume zero cost of transactions.

We could continue on the footsteps of the pioneer neoclassical economists based on cardinal (instead ordinal) utility and derive the demand curve. Walras, for instance, hypothesised not only that utility is cardinal but moreover that the utility derived from one good is independent from the utilities derived from the consumption of other goods, and that the individual's demand for a commodity is a function of the prices of all commodities. Furthermore, total utility is maximised when the marginal utility of income spent on each and every good is equal. The agents of the economy hold only this combination of goods that maximises their utility (Walras [1874] 1954, p. 284; Walker 2007, p. 280).[11] Walras, for instance notes:

> Given two commodities in a market, each holder attains maximum satisfaction of wants, or maximum effective utility, when the ratio of the intensities of the last wants satisfied, or the ratio of their raretés, is equal to the price. (Walras [1874] 1954, p.125)

In order to be in line with the recent literature, let us suppose an ordinal utility function of two goods in the following general form $u = u(x, y)$, with two goods x and y whose consumption gives a certain level of utility. The movement along the same indifference curve entails substitution of one good for another keeping the total utility constant. The notion of substitutability in consumption (and, as we will examine in the next sections, in production) is absolutely necessary in the neoclassical theory for the derivation of the so-called well-behaved demand and supply curves. Furthermore, the law of diminishing marginal utility must hold. The indifference curves which fulfil certain properties (see Silberberg 1978; Layard and Walters 1981, *inter alia*) cover the whole space between the axes x and y and as we move northeast we attain higher indifference curves, and, therefore, preferred combinations of goods x and y. As we mentioned the consumer is assumed to sell his initial endowment of goods in the market and to spend all the income (m) he receives in order to buy goods (Fig. 7.1a). The consumer's budget will be, $m = p_x x + p_y y$. Where x and y are the quantities of the two goods, whereas p_x and p_y are their respective prices. The individual consumer uses his resources to obtain some other combination of goods x and y, which depends on the prices of goods x and y. Consequently, we end up with the combinations of goods x and y, which represent a straight line (in each straight line, only one combination is relevant, the one tangential to the indifference curve), whose slope is equal to the relative prices of goods x and y and its exact location depends on the level of income.

For example, when the price for good y is p_2 individuals want the combination of goods (x_2, y_1), when the price of good y falls to p_1, then consumers would move to a higher indifference curve and attain the combination (x_1, y_2), and so forth. If we plot these different combinations of optimal prices on the axes of Fig. 7.1b we generate a

[11] Walras is the economist who revealed better than any of his contemporary economists the underlying utility calculus that remained latent in the analysis of price determination.

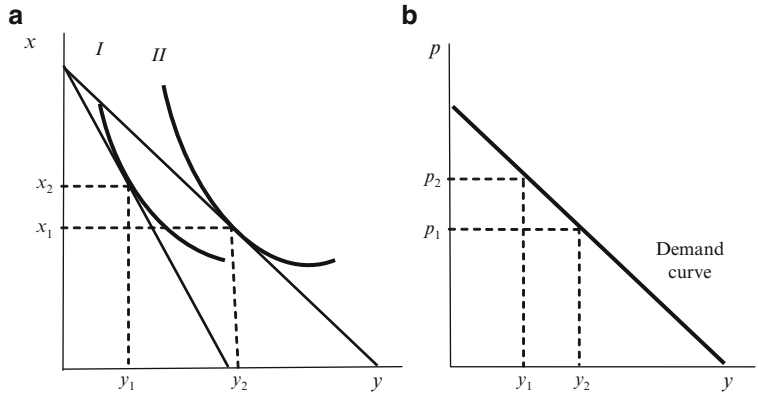

Fig. 7.1 The derivation of the demand curve

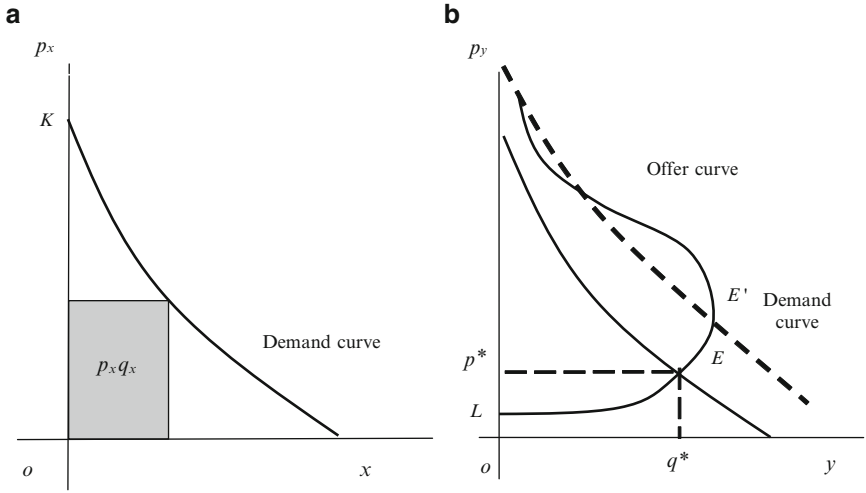

Fig. 7.2 Demand and offer curves

demand curve with a negative slope indicating the inverse relationship between price and quantity demanded of good *y*.

The demand curve, which we present in Fig. 7.2a, at the same time, implicitly includes an offer curve. The idea is that, in an economy of two goods *x* and *y* the price of good *x* is the reciprocal of the price of good *y*, and *vice versa*, that is $p_x = 1/p_y$ and $p_y = 1/p_x$. The demand for the good *x* is a function of the price of good *y*. If the price of good *x* in terms of *y* is high, for example, at a price such as at point *K* of Fig. 7.2a, then none would be willing to buy (that is to offer quantities of good *y*). Consequently, at point *K*, the demand for *x* is zero and consequently the supply of *y* will be zero, point *L* in Fig. 7.2b. As the price of *x* falls, individuals buy

more units of good x, which means that they would be willing and able to offer the required units of good y. When the price of x tends to zero, then the offer of good y tends to zero, since good x can be obtained at no cost. The offer curve of good y becomes asymptotic to the axis of prices. As a consequence, when we know the demand for one good, we essentially know the offer curve of the other good.[12] The demand for x entails the offered quantity of good y. The individual out of his total endowment of good y offers a portion of it, in order to obtain a portion of good x. This offered quantity of good y is estimated as the product of the price of x times the quantity demanded ($p_x q_x$), in terms of Fig. 7.2a, it will be equal to the shaded rectangle. Consequently, the offer of good y is equal to the area found under the demand curve for each particular price of good x.

Let us now suppose an individual possessing an initially given quantity of the two goods and wants to change their composition. The total available quantity of y is q_y from which he retains a portion z_y for his own use (individual demand) and supplies the rest to exchange it against good x. Thus, we have, $o_y = p_x q_x$. The offer curve of one good is, at the same time, the demand for the other good. In fact, the offer of good y, at the same time, constitutes the demand for good x.[13] Thus, from the demand curve, we can derive the offer curve of the other good. As Wicksteed (1844–1927) argued, the supply curve is simply the other side of the demand curve.

> But what about the 'supply curve' that usually figures as a determinant of price, co-ordinate with the demand curve? I say it boldly and baldly: there is no such a thing. When we are speaking of a marketable commodity, what is usually called the supply curve is in reality the demand curve of those who possess the commodity. The so-called supply curve, therefore, is simply a part of the total demand curve [. . .] The separating out of this portion of the demand curve and reversing it in the diagram is a process which has its meaning and legitimate function [. . .] but it is wholly irrelevant to the determination of the price [. . .] It is the single combined curve [total demand] that tells us what the equilibrium will be. The customary representation of cross curves confounds the process by which the price is discovered with the ultimate facts that determine it. (Wicksteed 1914, p. 13)

In Fig. 7.2b, we bring together the offer of good y along with its demand, we observe that a price higher or lower than p^* initiates changes that restore equilibrium, for instance, at a price higher than p^* the quantity offered is greater than that demanded and the price of good y falls, the converse is true at a price lower than p^*. The shape of the offer curve may give rise to no equilibrium at all or multiple equilibrium points. In fact in Fig. 7.2b we draw a second demand curve which intersects the offer curve at three points. From these three potential equilibrium

[12] J.S. Mill is the first who expressed the idea that the demand curve for one good is essentially the supply of another. The graphs that are presented here are in some deviation from those published in the appendix of Walras's ([1874] 1988, p. 109–112 and Appendix 1), where the price is on the horizontal axis and the quantity on the vertical axis. Creedy (1999, p. 193) claims that Walras's graphs were unclear, inasmuch as their axes were not labelled (in the French text) and when the book was translated into English the axes were misleadingly labelled.

[13] As Walras stated with reference to commodities A and B, "to say that a quantity Da of (A) is demanded at the price of Pa is, *ipso facto*, the same thing as saying that a quantity Db of (B), equal to DaPa, is being offered" (Walras 1954 [1984], p. 88).

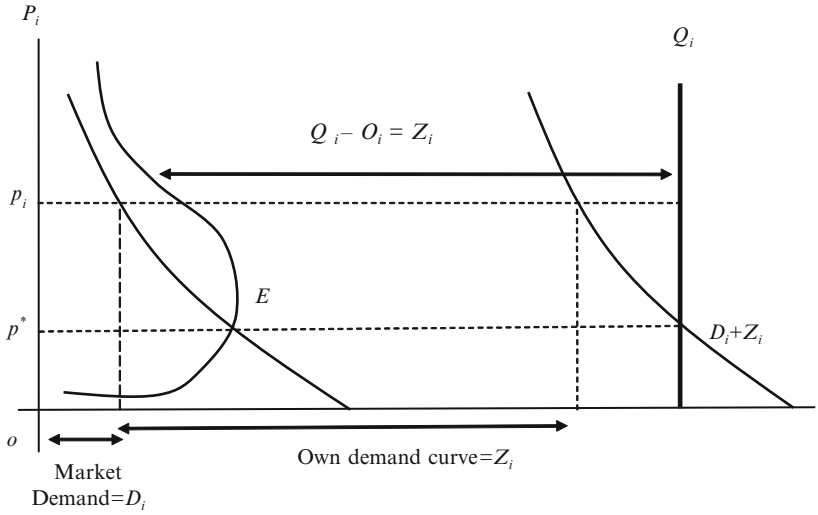

Fig. 7.3 Equilibrium in the market for good *i*

points, the middle one is certainly unstable and having to choose between the other two, we may conclude that the point which is closer to the previous equilibrium point is more likely to be the (stable) new equilibrium point, we also cannot rule out the cases of tangencies. These possibilities are not explored in any detail by Walras – neither in his book, nor in his 1893 article that has been appended to his book (Walras 1954, lessons 7 and 8, appendix 1) presumably because these cases are considered extreme and thus with no economic interest.[14]

Having derived the demand and offer curves for one individual we can derive the market demand and offer curves by horizontal summation. Clearly, the total offer cannot exceed the initially given endowment of the good in question. The demand and offer curves are put on the same graph (see Fig. 7.3), and we determine the equilibrium point E, where p^* and x^* are the equilibrium price and quantity, respectively.

The intersection of the offer curve of good i (where $i = x, y$) with the demand of good i determines the equilibrium price and quantity combination in the market for good i. It is interesting to note that the equilibrium in one market implies the equilibrium in the other market. If, for some reason, the actual price is above the equilibrium price, then supply exceeds demand and the price falls until it becomes equal to the equilibrium price. The converse is true in case where the actual price is below the equilibrium price. Consequently, for each individual the initially given quantity of good i is equal to the sum of his own demand (private demand) and the quantity that he has available for exchange (offer). In other words, the initial stocks

[14]For relevant discussions see Miller (1957), Daal and Jolink (1993, Chap. 4) and Creedy (1991 and 1999).

of good i are either offered for exchange or they are retained by the individual for his own use. The private demand and supply are functions of price. If we add the equations of all the individuals, we will have:

$$Q_i = Z_i + O_i$$

Where, Q_i is the total endowments of good i, Z_i the own demand which is a function of price and O_i the total offer which is also a function of price (Fig. 7.3). Suppose now that the price is p_{i}, the own demand is Z_i and the market demand is D_{i}. Thus, we could draw an own-demand function, that is, the difference between the offer function and the fixed endowment (see Fig. 7.3). In this analysis what is important is the market demand, that is, the total demand for the commodity of all the individuals in the economy, regardless of whether they offer the good i for exchange or retain a portion of it for their own use. In other words, the total demand is the horizontal summation of all market demand curves (D_i) as well as of all own demands (Z_i). The total demand curve for good i intersects the total fixed endowment at the same price p^* which is determined by the intersection of the individual demand and the individual offer curves. The reason is that only at p^* is the total demand of all individuals in the market, that is $D_i + Z_i$, equal to the total endowment of good (Q_i).

According to this approach we start with the behaviour of individuals and investigate the way in which each individual reacts to prices. We continue by forming the demand and offer curves for each particular individual. If we add up horizontally the demand and offer curves of all individuals, we determine the equilibrium prices and quantities through the equality of offer and demand curves. When we add up these offer curves for the economy as a whole, we arrive at the equilibrium prices.

It is interesting to note that in neoclassical economics relative prices reflect relative scarcities; this essentially stems from the idea of demand for commodities and the given endowments of commodities. Thus the notion of scarcity should not be taken literally but rather in the sense that prices are determined as a relation between demand and fixed endowments.

The analysis until now is based on two types of data: The preferences of consumers and the initially given endowments of goods. There are, however, some other factors which must be taken into account. The first concerns substitutability between goods as we approach the equilibrium point. Let us suppose that there is no substitution, and that the goods are demanded in constant proportions that are independent of prices, which means that the indifference curves are of 'L' shape. This would mean that the supply and demand curves would be inelastic. Since with no substitution between goods there is zero elasticity of demand, the demand curve would be vertical. This would imply that the total demand is less than or equal to the available endowments. In the first case, the equilibrium price would be equal to zero; in the second case, the price of the good would be indeterminate. Consequently, the curvature of the demand curve is due to the hypothesis of substitutability between goods for securing a given level of utility.

In the preceding analysis each individual is considered a price taker. In other words, we hypothesised a perfect competition environment, an assumption that characterises neoclassical economics, in contrast with classical economics, where competition refers to the flows of capital and firms compete with each other by reducing their cost and prices in their effort to expand their market share. In neoclassical economics, competition is viewed as a state in which individuals behave as price takers in their efforts to form demand and supply curves, and from there onwards the determination of the equilibrium price and quantity ensues. Consequently, perfect competition is a condition *sine qua non* for the operation of the pure exchange economy.

7.4.1 A Formal Presentation

Formally, the discussion can be set in the following way (see Creedy 1991, 1999). Let us consider the two individuals A and B, who possess goods x and y, respectively. Let us further suppose that the two demand curves are linear. Starting with individual B his demand for good x, x_d is given by:

$$x_d = a - b(p_x/p_y)$$

Turning to individual A endowed with good x, his demand for good y, y_d is given by:

$$y_d = \alpha - \beta(p_x/p_y)^{-1}$$

Clearly, the demand of A for good y must be associated with a reciprocal supply of good x by A. If, therefore, A demands y_d of y at a relative price of p_y/p_x the corresponding supply of good x will simply be:

$$x_s = y_d(p_y/p_x) = \left[\alpha - \beta(p_x/p_y)^{-1}\right](p_x/p_y)^{-1}$$

Setting $p = p_x/p_y$ we may write

$$x_s = \alpha p^{-1} - p^{-2}$$

Let s_x be the endowment of good x held by A. Then it follows that the demand for x by A is given by $s_x - x_s$. The total demand curve is given by:

$$x_D = x_d + (s_x - x_s) = s_x + a - b(p_x/p_y) - \alpha p^{-1} - \beta p^{-2}$$

Setting $x_D = s_x$, we can solve for the equilibrium price:

$$a - bp - \alpha p^{-1} - \beta p^{-2} = 0 \text{ or } bp^3 - ap^2 + \alpha p + \beta = 0$$

Clearly, students of economics are accustomed to solutions of such equations as the above for the determination of equilibrium pairs of prices and quantities, which may not be unique but multiple. In fact, in our simple model of linear demand equations, we may have three possible solutions from which only one is economically meaningful. From the above discussion two corollaries follow:

1. A change in the stock of good x held by person A, s_x, other things held constant, would have no effect on the equilibrium price. While this is mathematically true it is certain though that a change in s_x will imply a change in all the parameters of the model.
2. Equilibrium in the market for good x implies equilibrium in the market for good y.

7.4.2 Walras Law

The above analysis leads us to the generalisation of analysis from two goods to n goods of the economy. Consequently, if we define the excess demand for good i, E_i as the difference between total demand for good i (*i.e.*, the sum of market and own demand, $T_i = D_i + Z_i$) and the total endowment of good i, then we will have:

$$E_i = T_i - Q_i$$

we define the excess demand for good i of a single individual as follows:

$$e_i = t_i - q_i$$

We know that the value of all endowments, that is, the total income will be equal to:

$$\sum_{i=1}^{n} p_i q_i = m = \text{income}$$

and the value of expenditures for goods will be:

$$\sum_{i=1}^{n} p_i t_i = m = \text{expenditures}$$

We suppose that the individuals spent all their income. This is known as the non-satiation hypothesis. Consequently, we write:

$$\sum_{i=1}^{n} p_i t_i = \sum_{i=1}^{n} p_i q_i \quad \text{and} \sum_{i=1}^{n} p_i (t_i - q_i) = 0$$

Consequently,

$$\sum_{i=1}^{n} p_i e_i = 0$$

Since, p_i's are positive, it follows that some excess demands may be positive or negative to balance each other out at equilibrium, where the amount that an individual spends is equal to the amount that he has sold. Summing over all individuals we arrive at the excess demand in the society as a whole

$$\sum_{i=1}^{n} p_i E_i = 0.$$

This has come to be known as Walras's law (Walras 1874, p. 170). In other words, the value of excess demands is equal to zero, a result which stems from the fact that individuals spend all their income and do not satisfy their demands. Hence we refer to the sum of excess demands and not to each individual excess demand. In fact, each individual excess demand can be different from zero, but when we aggregate the values of all excess demands (positive and negative) in the economy the net result is zero. In short, Walras's law can be satisfied even when there is no equilibrium even in a single market; all that is required is for the sum of excess demands to be equal to zero. This is a fundamental proposition of the neoclassical theory and without it there cannot be a proof of the existence of equilibrium in the economy. It is important to clarify the following points:

1. Walras's law should not be confused with Say's law, since it does not claim that all the markets are in equilibrium, but only that the sum of the value of excess demands of all the markets will be equal to zero. That is, in some markets the demand will be greater than the supply and in some other markets the demand will fall short of supply, but the sum of the value of excess demands (positive and negative) will be equal to zero.
2. Walras's law does not imply that saving is equal to investment, since there is no proposition in Walras's law to claim that in any single market there will always be equilibrium. In contrast, Say's law claims that savings and investment by definition will be equal to each other.

A corollary of Walras's law is that if we have n markets and if we suppose that $n-1$ markets are in equilibrium, that is the sum of excess demands in these $n-1$ markets is equal to zero:

$$\sum_{i=1}^{n-1} p_i E_i = 0,$$

then it follows that the n^{th} market, $p_n E_n = 0$, will be in equilibrium. In other words, we will have a system of linear equations and the only thing that we need to show is that $p_i \neq 0$, that is, we have a sum of positive constants, in our case prices, which show linear dependence. Finally, the presentation of Walras's law above is general enough so as to include the case of production by including the relevant markets for factors of production. The case of an economy with production becomes the focus of our analysis in the next section.

7.5 From Pure Exchange to Production

We refer to the structure of neoclassical theory in detail because such an important issue is not discussed with the required attention in books of microeconomics where the reader is introduced directly to the graphics and the optimization techniques. On the other hand, in the texts about the history of economic thought, the emphasis is placed on the contributions of individuals without their synthesis to a single approach. It is important to stress that microeconomic theory is an area that is not subjected to the rapid changes that one observes in macroeconomics. The innovations in microeconomics are really more about techniques of presentation and much less about substantial issues. In short, unlike macroeconomics in microeconomics there has been achieved a consensus among economists. At the same time, we avoid the detailed presentation of either the theory of demand or the theory of the firm, since these issues are treated in detail in the microeconomic textbooks. Let us suppose an economy in which the individuals express their preferences for the various goods through their utility functions. Moreover, the individuals dispose of the quantities of their available resources which consist of quantities of labour and land measured in hours and acres, respectively. The allocation of endowments is known to the individuals from the beginning. The technology is one of constant returns to scale and it is known to the individuals. Consequently, the data of the neoclassical theory are:

1. Utility functions (or preferences) of individuals
2. Initial endowment of factors of production (land and labour)
3. Technology

With these data we seek to analyse the production and exchange of goods as an extension and further elaboration of the neoclassical theory of pure exchange

economy. The introduction of production in the neoclassical theory indicates a kind of disconnection between the demand (for goods) and the initially given quantities of factors of production. Consequently, instead of the existence of a direct relation between utility and disutility of goods that are consumed there is a disconnection, which means that the notion of production must be formed in such a way as to be capable of being subsumed itself under the model of generalised exchange. As Cassell (1866–1945) remarked:

> The demand for commodities is indirectly a demand for factors of production. (Cassel 1918, p. 90)

For example, the demand for good x constitutes essentially supply of labour, which in turn constitutes indirect supply of good y. Similarly, the demand for good y is converted to demand for the services of factors of production (land and labour), which are used in the production of such a good, whereas the factors of production are being offered from those that demand good x.[15] Thus, between the consumption of goods and the services of the factors of production that are in the disposal of the individuals mediates production, which is called 'black box', precisely because its content is not known.[16]

Nevertheless the basic relation between the maximising of utility subject to the resource endowment continues to be true. The only novelty introduced moving from the pure model of consumption to the model of production is the mediation of the black box, where the available resources (factors of production) are transformed to consumer goods. In reality, production is actually an indirect exchange of the initially given endowment of resources. Alternatively, factors are demanded because of the commodities they produce. In other words, factors are demanded not for their intrinsic worthiness to the other consumers, but rather because they can be converted to consumable goods via production, and it is these utility-yielding goods which are desired by consumers. According to Walras the analysis of production is essentially an analysis of indirect exchange of the services of the factors of production. Consequently, in neoclassical analysis if production is an extension and further elaboration of the model of pure exchange, then the analysis of production must be cast in terms of utility and disutility and in the way in which the decisions are taken by the rationally behaving agents. As in the case of pure exchange, the decisions are taken to the point where the marginal benefit from renting out a factor of production is equal to the marginal sacrifice for parting with the factor of production.

The difference with the analysis of pure exchange economy is in that the endowments of resources include the productive services of the (non-produced)

[15]Factor services (of land and labour) are not demanded by other consumers in any direct way, simply because factor services in and of themselves offer no utility reward and thus are not directly "demanded" by consumers.

[16]This is in sharp contrast to Smith and Marx, who describe production as a process that involves time and its discussion is central to their theories. See for example Smith's chapter on the division of labour and Marx's *Capital,* I whose subtitle is "The Process of Capitalist Production".

means of production, that is, the services of land and labour. The analysis according to Walras is similar to that of the pure exchange economy with the difference that the goods that individuals demand via their preferences are not available and the goods that individuals offer in terms of services of the factors of production are not those that are demanded in terms of commodities. Hence, these asymmetries need to be corrected by aligning the demand for final goods with the supply of services of the factors of production. This is possible by introducing, in the data of the neoclassical model, the technology which describes the way in which the demand for factors of production is used in the production of goods and services.

In what follows, we suppose an economy where we have an initially given quantity of land and labour. The agents of the economy express their preferences through their utility functions, which convey the fact that the services of labour as well as of land give negative utility. We must connect the supply of factors of production and the production of goods that are demanded and the derived demand for the factors of production. This connexion is possible through the mediation of technology. Moreover, we need to hypothesise that competition forces the entrepreneurs to choose from the set of possible techniques only those that minimise the cost of production, which amounts to the maximisation of profits. In other words, firms decide upon two interconnected questions:

- The choice of technique that minimises the cost of production of goods x and y.
- The choice of goods x and y as well as the quantities that will be produced.

This procedure secures that in a capitalist economy the price of each good will be equal to its cost of production, which is defined by the total sum of wages and rents that must be paid for the services of labour and land. Consequently, the price of good x, $p_{x,}$ will be equal to:

$$p_x = \rho l_x + w n_x$$

Where l_x is the quantity of land which is used in the production of a unit of good x; ρ is the rent per unit of land measured say in acres; w the wage per unit of labour measured say in hours and n_x the quantity of labour used for the production of a unit of good x.

If the cost of production is different from the price of the product, then counteracting forces are developed that restore the equality. For example, if the price of good x is higher than its cost of production, then the excess profits attract firms in the production of good x, with the result that its supply increases and its price falls to the point where it equals to the cost of production. If, on the other hand, the price of good x is lower than its cost of production then it follows that losses are being made in the production of this good. The subsequent exit of producers reduces the supply of good x and so its price increases to the point that it becomes equal to the cost of production.

These quantities of the factors of production could have been chosen as the cheapest from a variety of available techniques. In what follows, as in the analysis of the pure exchange model, the agents of the economy reveal their choices for each

vector of prices $[w, \rho, p_x, p_y]$ presented to them. For reasons of clarity of presentation and economy in space our focus is on the behaviour of a single individual, who is endowed only with the services of labour and who expresses, for each price vector, his intentions with respect to the quantity of labour which he is willing to supply. Consequently, at this point, we have the disposition of the services of land and the quantities of goods x and y, which he is willing and able to purchase. At the same time, however, the supply of the services of land and the quantities of goods that are demanded must be co-ordinated in such a way so that the total supply is equal to the total demand.

The procedure through which the equality of total demand and supply is achieved is as follows. The total demand for the services of land is determined by the demand for the good x times the quantity of land used in the production of good x. For simplicity's sake, let us suppose that 1/5 of an acre is being used for a unit of product x. If we suppose the production of 20 units of product x, then this implies that the demand for land will be equal to $(1/5)(20) = 4$ acres of land. In a similar fashion, we reason out the required quantities of the services of land in order to produce good y. We add up these two demands for land and we get the total demand,

$$D_l = l_x D_x + l_y D_y$$

where l_x and l_y are the quantities of land per unit of product x and y, while D_x and D_y the quantities demanded of goods x and y respectively. Similarly, for the demand for the services of labour we have:

$$D_n = n_x D_x + n_y D_y$$

Hence, we arrive at what Marshall (1890, pp. 317–318) calls *derived demand* for labour, which stem from the demand for goods x and y. As in the analysis of indifference curves with two goods (x and y), now the factors of production (land and labour) are used in different combinations for the production of a given quantity of output. The curves that are formed by these different combinations of land and labour are called isoquants and they are similar to the indifference (or isoutility) curves. However, there is an essential difference: in the isoquants we have measures expressed in real quantities that are produced from different combinations of factors of production, whereas, in the analysis of consumption the indifference curves because of their subjective character are measured only relatively and not cardinally.

The role of substitution so crucial in the analysis of the pure exchange model is also extremely important in the analysis of production. In fact, substitutability is the property that gives rise to the specific shape of the isoquant curves. If there were no substitution, then the isoquants would be L-shaped. The choice of technique in the production is expressed in terms of cost minimisation, where the cost is equal to the value of services of the factors of production that are used in the production process (Fig. 7.4).

Together with the isoquants, we have the isocost curves which are given by the price equation $p_x = \rho l_x + w n_x$. The discussion is similar to that of budget line and

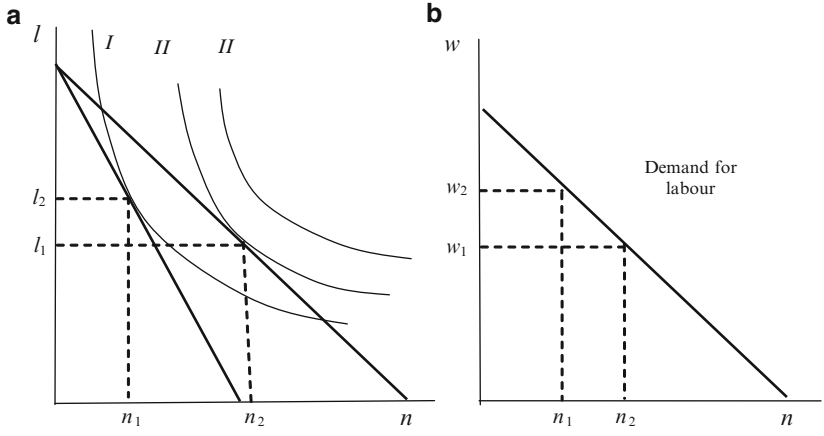

Fig. 7.4 Isoquants and non-produced means of production

indifference curves, that is, the further out to the right we extend, the higher the level of output that can be produced by the combination of the two factors of production. If, for instance, we want to produce a given level of output, then we must reach the lowest isoquant curve which is just tangent to the isocost line. Consequently, the slope of the isocost curve, which is equal to the relative price of land and labour, leads to the appropriate (cost minimising) combination of labour and land. If the relative prices of land and labour changes then we would arrive at a different combination of the two factors of production. The property of substitutability between the factors leads to the formation of a demand function for the services of factors of production with some elasticity.

The idea of substitutability between the services of factors of production is the result of a given quantity of factors of production; consequently, the increase of the one implies the decrease of the other. The problem can be expressed as follows: Let us suppose that we have a given quantity of land and employ more and more labour. In the beginning, the produced output increases linearly, since a small number of workers uses only a portion of the available land. Consequently, the optimal land-labour ratio, l/n, is maintained. As the quantity of employed labour increases with the given quantity of land, a point is reached where all land is being used and if the number of workers continues to increase, then the output per worker (productivity) falls. The cause of this fall must be found in that we are forced to use a land-labour ratio which is no longer optimal. Consequently, the marginal product of labour begins to diminish.

The substitution procedure is intrinsically connected to the construction of demand functions from the services of factors of production and the substitution of the factors of production presupposes that all the factors of production are fully utilised (employed). If there were no full utilisation, then simply we could not choose the optimal ratio and maintain it. When we drew the demand curve, we supposed the substitution between labour and land and that one factor of production

(land) was fully utilised. After all, if land were not fully utilised there would be no need for substitution.

Let us now suppose that the demand for labour D_n falls short of the supply of labour ($D_n < O_n$). We further suppose a fall in the wage in order to bring equilibrium in the labour market. This fall in the wage leads to a fall in the prices of the two goods. If we suppose the general case, where goods x and y are produced in different proportions of land and labour it follows that the reduction of prices of goods x and y will not be proportional. We suppose that the production of good x is land-intensive while the production of good y is labour-intensive. Consequently, the fall in wage will affect the production of good y more than the production of good x. Consequently, the relative price p_x/p_y between the two goods will rise since the price of y will fall by more than the price of x. This change will lead to an increase in demand for good y, because it is relatively cheaper and in a reduction of the demand for good x. Consequently, the demand for labour in the production of labour-intensive good y increases and the demand for labour for the production of good x falls. The net result will be an overall increase in the demand for labour.

Until now we assumed an economy of fixed input–output coefficients, an assumption that was made by the first neoclassical economists especially by Walras.[17] However, in neoclassical theory we know that substitutability is continuous and if the relative factor prices change it follows that firms would move to another technique. Thus in our case, since the wage falls it follows that a new labour-using and land-saving technique will be adopted. Consequently, the increase in the demand for labour will be reinforced since the new technique uses more labour in the production of both goods. The above lead to a demand curve for labour, which is derived from the characteristics of the maximisation of utility from the good x via the technology. Thus we can construct a demand curve for labour which is derived from the characteristics of maximisation of consumer's utility as well as of the technology. The technology transforms the demands for goods to *derived demands* for the services of factors of production. Consequently, we can define an equilibrium point in the labour market, since we suppose the equality between price and cost of production. This implies that the quantity supplied will be equal to the quantity demanded. The analysis until now has shown that the model of the economy with non-produced means of production is essentially an extension and further elaboration of the model of the pure exchange economy.

Many neoclassical economists argue that the supply curves are technological phenomena. In reality, however, this is not so, since the offer curves constitute a relation between price and quantity. The price is a measure of disutility of the disposition of the factors of production that are going to be employed in the production of the good, in the sense that the supply curves are produced by technology and that the technology is supposed to have a special form. The offer

[17]Walras assumed fixed coefficients of production ("coefficents de fabrication") as a first approximation and in the later editions of his *Elements* allowed for the possibility of variable coefficients (Jaffé 1983, Chap. 11; Kurz and Salvadori, 1995, p. 432.). Marshall (1890) also assumed fixed input coefficients in Note XXI appendix.

curve constitutes a schedule of subjective cost. Marshall (1890, p. 282) called it the real cost and he meant the subjective cost, in the sense that he referred to the disutility (sacrifice) of the disposition of factors of production.

We observe that in the neoclassical analysis prices and quantities are intrinsically connected, whereas in the classical model, the output is a datum in the analysis of the determination of prices. In the neoclassical model prices and quantities are functionally determined and cannot be separated. The structure of the neoclassical theory gives the impression that economics refers to scarcity, but essentially this is not so, since neoclassical economics refers to the optimising behaviour of the agents. Consequently, we have a relation between price and quantity. The price is determined by the forces of demand and supply and by the supposition that there is full utilisation of resources in the economy. This does not necessarily mean that all the available services of labour are fully employed but rather that only the disposable services of labour are employed. There are services of labour, which could have become available for employment, but they are not, their owners (workers) withhold them because at the going wages they would rather increase their leisure instead of their working time.

The same argument applies with respect to the services of land (some of it may be held say for gardening). Again there is an offer curve for the services of land, which is derived in a way similar to that of labour. If land is not fully employed this is because the owners of land think that this is the best they can do, since an increase of supply will reduce the price of land. It is true that in questions of this sort it is hard to come up with good answers. The analysis could be expanded to encompass capital goods, the characteristic of which is that they are produced means of production, that is, they possess characteristics of consumer goods and so their price is determined by the cost of production and also the characteristics of factors of production and their prices are determined by their marginal productivity. The question is, to what extent are these two characteristics of capital goods compatible with each other? The neoclassical view argues that there is no problem with the dual role of capital goods and that the analysis of land can be straightforwardly generalised. This point of view is shared by economists at Cambridge Massachusetts and the exact opposite view is held by economists at Cambridge England, who initiated the famous debate in the decade of 1960s. The details of the arguments on both sides are discussed in the next chapter.

7.6 Summary and Conclusions

Our discussion of the neoclassical theory began with the historical circumstances that gave rise to this school of economic thought. We argued that the neoclassical theory has the determination of natural or normal prices as its object of analysis. We introduced the pure exchange economy, which is entirely hypothetical, in as much as no economy can survive without production. In this pure exchange model, we had as givens the preferences or utility functions of individuals as well as their

initial endowment of goods. We assumed a simple version of pure exchange economy of two goods and that individuals maximise their utilities subject to the constraint of their endowment and in doing so we could derive the individual demands from which we could correspond the individual offers and putting them together we could determine the equilibrium pairs of prices and quantities in each market. On the basis of this analysis we derived Walras law and explained its differences from Say's law.

The next step was to leave the hypothetical world of pure exchange and move to the more realistic but still hypothetical model of non capitalist production. We say non-capitalist production because the issues associated with the introduction of produced means of production, that is capital, are postponed to the next chapter. So in this chapter we discussed production abstracting from capitalism by restricting ourselves to non-produced means of production, that is, labour and land. In this model of production we argued that the fundamental relation of the pure exchange model continue to hold true. The difference is that in this model there is the mediation of the 'black box' of production, where individuals offer the services of the factors of production and they demand goods. The first neoclassical economists argued that the demand for goods decide about the offer of the services of the factors of production. In other words individuals do not demand directly the services of labour or land but through their demand for goods. And the services of the factors of production are transformed to goods that are being demanded. We observe, that production is the *indirect exchange* of the initially given resources. Walras observed that the analysis of production is simply the analysis of indirect exchange of factors of production, and he further suggested that one could even

> abstract from entrepreneurs and simply consider the productive services as being, in a certain sense, exchanged directly for one another instead of being exchanged first against products and then against production services. (Walras 1874, p. 225)

Moreover, Walras showed that the analysis of the exchange of two goods in a competitive economy, can be extended to include the whole theory of production and distribution of social wealth. Consequently, if production in the neoclassical approach is viewed as continuation and extension of the model of pure exchange, then it follows that it must be conducted in terms of utility and disutility as well as their optimization subject to constraints.

With the exception of Marshall all the first neoclassical economists developed first the notion of pure exchange and then the notions of non-capitalist production and exchange. Wicksell (1934) noted that the analysis of pure exchange and non-capitalist production is unrealistic. The idea is that these models help to the development of the notions and methods of analysis of a capitalist society. In other words, they constitute the prelude to our object of analysis, that is, the determination of equilibrium, relative prices and the rate of profit associated with these. The usual, mostly mistaken, critiques of the neoclassical theory include:

1. In the neoclassical theory everything depends on something else and because of the interdependence we cannot determine the normal prices of goods and services.

In reality, however, in neoclassical theory there is a strictly determined causal relation that is based on the maximisation of utility of individuals subject to the constraint of endowment and it determines the relative prices in a perfectly competitive economy. There is no doubt that despite the complexity there is a fully determined causal relation. For example, Walras notes:

> [. . .] the utility curves and the quantities possessed constitute the necessary and sufficient data for the establishment of current and equilibrium prices. From these data we proceed, first of all, to the mathematical derivation of individual and aggregate demand curves in view of the fact that each party to an exchange seeks the greatest possible satisfaction of his wants. And then, from the individual or aggregate demand curves, we derive mathematically the current equilibrium prices since there can be only one price in the market, namely the price at which total effective demand equals to total effective offer. (Walras 1874, p. 143)

On further consideration, we should bear in mind that the logical consistency of the neoclassical model is the characteristic that made possible its wide acceptance and even its dominance over the classical model.

2. The neoclassical theory refers essentially to exchange and not to production

This is not true, since in neoclassical theory deals with production although such an analysis is translated to indirect exchange. For example, the problems of the labour process can be defined in terms of maximising utility and in terms of the choices of businesses and workers (see for example Steedman 1981). The emphasis on exchange does not mean that neoclassical economists undermined production, it only means that even production is viewed as a process of indirect exchange, where consumers demand the services of the factors of production not directly but only through their demand for goods. What is essential about neoclassical theory is that production is subjugated to exchange, not that neoclassical theory does deal with production.

3. The neoclassical theory is static

The neoclassical theory often gives the impression that time is spirited away. The truth though is that time often is left out of analysis for reasons of simplicity and not that time is not important or that is not accounted for in the neoclassical theory. In the next chapter where we deal with the model of production with capital, we will see how time is incorporated in the neoclassical analysis even from the writings of the first neoclassical economists. One could criticise the neoclassical theory for the way time is being treated but not that the neoclassical analysis is necessarily timeless.

4. The neoclassical theory is a subjective theory

This is true since utility in neoclassical theory is subjectively determined by the consumer, and not by the intrinsic properties of the good, in the sense that scarcity is something, which is subjectively conceived by the individual and not necessarily an objective fact.

In our presentation of the structure of the neoclassical theory we referred exclusively to the non-produced means of production (labour and land) and we considered that the produced means of production (capital) have a similar theorization. In reality, however, the embodiment of capital in the neoclassical theory creates a number of problems that led to the famous capital controversies of the 1960s about which we discuss in next chapter.

Questions for Discussion and Thought

1. Discuss the difference between use-value and utility.
2. Derive the demand curve from ordinary utility analysis.
3. Derive the demand curves for factors of production.
4. Does Marx's characterisation of 'vulgar economics' refer to neoclassical economics? Explain.
5. Discuss the data of the neoclassical pure exchange model. Is this model realistic?
6. How does one determine equilibrium prices in the pure exchange model?
7. In what sense, if any, does the neoclassical model of production differ from the pure exchange model?
8. Present the basic ideas of the model with production with non-produced means of production.
9. Discus Walras's law. In what ways does Walras law differ from Say's law?
10. The neoclassical theory of value claims that the prices of goods reflect relative scarcities. In this sense the neoclassical theory of value should not differ from the labour theory of value since scarcity requires more (labour) effort which entails a higher price. Discuss.

Notes for Further Reading

In discussing the structure of the neoclassical theory of value and distribution we avoided the usual presentations, where one finds a list of names where each and everyone had contributed something towards a theory and yet this theory is not explicitly stated. We also refer to the major neoclassical economists and give a flavour of their main contributions; however, our attention is on the possible generalisations that constitute the neoclassical theory. In this effort, we wanted to avoid the usual microeconomics presentation, where the reader is wrestling with the mathematics and at the end misses the very essence of neoclassical theory. In this approach the articles by Eatwell 1977; Garegnani 1983 and also the books by Gramm and Walsh 1981 as well as Kurz and Salvadori 1995 are recommended. The website on the History of economic thought of Fonseca Ussher (2009) is extremely helpful for the current as well as for the other chapters of the book. We also strongly recommend Creedy's (1991, 1993) elegant mathematical presentation as well as discussion of the neoclassical pure exchange model which were used in our formal treatment of this model. The book by Ferguson and Gould (1971) and

(continued)

the much more advanced microeconomic text by Silberberg (1978) are good standard presentations of the neoclassical theory of value. The same is true of the book by Krepps (1990), which however is extremely mathematized.

The theory of utility (cardinal or ordinal) constitutes the heart of the neoclassical approach. An excellent description of this idea is found in Schumpeter (1953, pp. 1053–1066). For the origins of neoclassical theory we recommend Aspromourgos (1991). As for the constituent elements of the neoclassical theory we follow Walras's (1874) approach and for a summary presentation, Jaffé (1954). The approaches of the other neoclassical economists of the marginal revolution (Menger and Jevons) inspite of their importance for the total approach, nevertheless did not prevail and are simply mentioned in the texts of the history of economic thought. In a comparison between Marshall and Walras, we observe that although Marshal accepted the same set of data as Walras, the Marshallian approach is more of the partial equilibrium type as opposed to the Walrasian one which is a general equilibrium approach. Marshall, more than any other neoclassical economist, popularised and promulgated the neoclassical ideas. It is interesting to note that the notion of utility is not used by Marshall as one would expect. In fact, Marshall used this word only once in his *Principles* to express his disagreement with Smith. In fact, demand in Marshall depends on a host of variables other than utility and it does not seem that he was so much interested in the question of what exactly was behind the demand curves. In this chapter, we tried a simple presentation of Walras law, restricting ourselves to a pure exchange economy assuming of course that the inclusion of production only adds complexity to the analysis without necessary improving our understanding of the operation of the law.

At this point it is worth referring to the controversy between Mirowski (1984 and 1989) and Hollander (1985, 1989). Mirowski (1984) argues that Jevons and Walras, with the subsequent aid of Edgeworth, Pareto and others, broke with the traditional classical analysis. It is important to mention that Mirowski (1984) would rather exclude Menger from the group of neoclassical economists and classify him along with the Austrian economists. Mirowski, in the same article, further argues that the break with the past was not provoked by the theory of utility, but by something that economists usually do not pay attention to and that is the successful penetration of mathematics in economic theory. Unlike the classical economists the early neoclassical economists are oriented more to the formal presentation of their results and less with the economic theory *per se*. For a counterargument see Hollander (1985, 1989).

It is interesting to note Schumpeter's (1954, p. 242) admiration for Walras, whom he considers as the greatest economic theorist, and his system of equations as 'the Magna Carta of economic theory'. In similar fashion, Morishima in his *Economics of Walras* praises Walras as one of the greatest

(continued)

economists (Morishima admires him more than Ricardo, Marx and Keynes) who was interested in questions of income distribution between social classes and in this sense his analysis bears similarities with Marx's and on some policy issues with Keynes's. To our view, these Walras's parallels are exaggerated; true, Walras expressed radical ideas (see *e.g.*, Niehans 1990, p. 209) favouring abolition of taxes and state revenues collected through the nationalisation of land and also supported state intervention, but these suggestions were not the result of his economic analysis.

Chapter 8
Theory of Capital and Cambridge Controversies

[T]he production function has been a powerful instrument of miseducation. The student of economic theory is taught to write Q = f(L, C) where L is a quantity of labour, C a quantity of capital and O (Q) a rate of output of commodities. He is instructed to assume all workers alike, and to measure L in man-hours of labour; he is told something about the index number problem involved in choosing a unit of output; and then he is hurried on to the next question, in the hope that he will forget to ask in what units C is measured. Before ever he does ask, he has become a professor, and so sloppy habits of thought are handed on from one generation to the next.

(Joan Robinson, 1953)

[. . .] One could measure capital in pounds or dollars and introduce this into a production function. The definition in this case must be absolutely water-tight, for with a given quantity of capital one had a certain rate of interest [. . .]. The work of J. B. Clark, Böhm-Bawerk and others was intended to produce pure definitions of capital, as required by their theories, not as a guide to actual measurement. If we found contradictions, then these pointed to defects in the theory, and an inability to define measures of capital accurately. It was on this –the chief failing of capital theory– that we should concentrate rather than on problems of measurement.

(Piero Sraffa, Interventions in the debate at the Corfu Conference on the 'Theory of Capital', 1958)

8.1 Introduction

In our analysis of the structure of the neoclassical theory, we stated that the theory is, usually, advanced in three stages: In the first stage, the discussion is limited to pure exchange, where the individuals (or households) are endowed with various commodities and their differences in preferences induce them to exchange these goods in their effort to maximise their utility. Walras's contribution was that he managed better than any of his contemporaries to incorporate the (new) utility theory into an explicit model of a pure exchange economy. In such a model, given the preferences of individuals and the initial endowment of goods, we form the demand of each and every individual and then, by aggregating the demand curves of all individuals, we get the total social demand. The model of pure exchange

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_8,

economy is used only for instructive purposes and is restricted to showing the attainment of general equilibrium; a more realistic analysis, besides exchange, should include production.

The next step therefore was to generalise the pure exchange model to one with production. The transition was not simple and straightforward and along the way there were asymmetries and obstacles that had to be overcome. We started this generalisation by assuming that individuals, besides the goods that they possess, are also endowed with factors of production. The analysis was restricted to non-capitalist production, that is, a model with non-produced means of production (*i.e.* land and labour) as a transitory step towards a fully fledged capitalist production. We postponed the analysis of the capitalist production for this chapter because of its difficulties and the controversy that surrounded it. This problem with the analysis of the produced means of production was pointed out from the first neoclassical economists (Walras, Böhm Bawerk, Marshall and Wicksell, *inter alia*) who tried to offer plausible solutions to this conundrum, which we briefly examine below. Our attention, however, will be on the famous Cambridge Capital Controversies that involved economists such as Piero Sraffa, Joan Robinson, Piero Garegnani and Luigi Pasinetti at the University of Cambridge in England and Paul Samuelson and Robert Solow at the Massachusetts Institute of Technology in Cambridge, Massachusetts.

8.2 Production with Produced Means of Production

In the non-capitalist production, we assumed that individuals had given their preferences and also their endowments, a portion of which they offer in exchange for goods or other endowments. For example, individuals do not offer all of their potential labour services but retain some portion to use for leisure activities, while some of their land services may also be kept, say, for gardening purposes. The difference with the analysis of pure exchange economy is in that the endowment of resources includes the productive services of the (non-produced) means of production, that is, the services of land and labour. We found that the analysis of an economy with non-produced means of production is an extension of the analysis of pure exchange. There are some asymmetries, which, however, are easily resolved. For example, the goods that individuals demand are not exactly comparable with the services of the factors of production offered by the individuals. Individuals do not consume the services of the factors of production in any direct way. Thus, the services of the factors of production that the individuals are endowed with must be transformed to an offer of goods that will match the demand for goods. Hence, there is the need of connecting the demand for final goods and the supply of services of the factors of production. This is possible if we add, in the data of the neoclassical model, the technology which describes the way in which the demand for factors of production is used in the production of goods and services. The analysis of production with non-produced means of production did not really present unresolved problems.

Naturally, one would expect that the analysis of production with non-produced means of production could be straightforwardly generalised and include produced means of production, that is, capital goods. This is not the case, however, as we know from the first neoclassical economists, and the generalisation is not without its problems. The reason is that the measurement of the means of production must fulfil two requirements:

1. Their measurement units must be suitable for cost minimization
2. Their measurement must be independent of prices

Clearly, in the case of non-producible means of production, measuring arable land in terms of acres of uniform fertility and labour in terms of undifferentiated hours of work posits no (significant) problems regarding the aggregation of these two factors of production and the determination of their equilibrium prices, since the quantities of both factors of production fulfil the above two requirements.[1] Turning to capital, though, we realise that the two requirements are hard to fulfil because capital goods are heterogeneous, and to aggregate them, one needs a common unit of measurement, which cannot be (equilibrium) prices because this is precisely what we need to estimate in the first place.

It is important to stress at this point that a popular misconception has been created out of this discussion, according to which capital cannot be measured at all. This is not true though, since capital can be measured in market prices and there are various ways to deal with the problems of its evaluation. The real issue in the measurement of capital is that it cannot be evaluated in equilibrium prices, which are consistent with the neoclassical theory of value and distribution. This is a problem specific to the neoclassical theory of value. We know that such a problem does not arise in the classical theory of value because capital can be measured in terms of labour values and also in terms of prices of production. The idea is that prices of production and the rate of profit are determined given the size and composition of output, the level of real wages and the state of technology.

According to the neoclassical approach, both the profit rate and the normal prices are determined by the forces of supply and demand. The crucial issue in neoclassical analysis is that in the models of production and exchange employing only non-reproducible means of production (land and labour), goods are divided into two categories. The first includes the non-reproducible means of production whose price is determined by the forces of supply and demand. The second includes consumer goods whose equilibrium price is determined by their cost of production. If in the neoclassical model we hypothesise produced means of production (capital), then their price is determined in two ways. First, as with consumer goods the price of capital goods is determined by their cost of production. Second, the price of capital is the capitalised income generated during its useful life and so is determined by the

[1]Of course, there are other problems with the aggregation of land or labour but these are manageable and one can form an acceptable index of the quantity of arable land and an index of the quantity of labour (see Kurz and Salvadori 1995, Chaps. 10 and 11, *inter alia*).

forces of supply and demand. In other words, the price of capital is over-determined, we have more equations than unknowns and so there is no unique solution.

Piero Sraffa (1951) pointed out that there was an inherent measurement problem in applying the neoclassical theory of value and income distribution, because the estimation of the rate of profit requires the prior measurement of capital. The problem is that capital – unlike labour or land, which can be reduced to homogenous units stated in their own terms (for example, hours of the same skill and intensity or land of the same fertility) – is an ensemble of heterogeneously produced goods, which must be added in such a way as to enable a cost-minimising choice of techniques. From the various alternatives, neoclassical theory chooses to measure capital goods in value terms; that is, the product of physical units (buildings, machines, *etc*.) times their respective (equilibrium) prices. Joan Robinson (1953), inspired by Sraffa's teaching and early writings, and later Sraffa himself (1960), argued that the measurement of capital requires the prior knowledge of equilibrium prices, which in turn requires an equilibrium rate of profit that cannot be obtained unless we have estimated the value of capital (Sraffa 1960, p. 38).

8.3 Production with Capital and the First Neoclassical Economists

The measurement of capital and its inclusion in the neoclassical theory of value was not easy and the first neoclassical economists made valiant efforts for its appropriate theorization. The problem in this theorization is that the determination of the rate of profit (interest) and the associated equilibrium prices require the specification of capital in the economy as a quantity of value; of course, this is contradictory, because the value of capital cannot be estimated without knowing the rate of profit and the associated with it equilibrium prices. This is the fundamental contradiction in the very core of the neoclassical theory of value that the first neoclassical economists had noticed and tried to resolve (Kurz 2008). Walras, for instance, pointed this out and was conscious of the difficulties of finding a consistent solution; Wicksell also sensed this problem and gave an intellectual struggle for a solution and although he went a step further than his precursors, nevertheless his advancements were far from providing a fully satisfactory solution, as he himself admitted.

Starting from Stanley W. Jevons, we find that in his incomplete studies of capital (*e.g. The Theory of Political Economy* [1871], 1911), he essentially accepted the Ricardian notion of capital as a stock of goods that can be consumed either directly (like food) or indirectly (like machines).[2] In Jevons's conceptualization, capital is

[2]"The views which I shall endeavour to establish on this subject are in fundamental agreement with those adopted by Ricardo; but I shall try to put the Theory of Capital in a more simple and consistent manner than has been the case with some later economists" (Jevons 1871).

intrinsically connected to the passage of time between the input of labour and the final act of consumption.

> Capital, as I regard it, consists merely in the aggregate of those commodities which are required for sustaining labourers of any kind or class engaged in work. A stock of food is the main element of capital; but supplies of clothes, furniture, and all the other articles in common daily use are also necessary parts of capital. The current means of sustenance constitute capital in its free or uninvested form. The single and all-important function of capital is to enable the labourer to await the result of any long-lasting work, – to put an interval between the beginning and the end of an enterprise. (Jevons 1871, p. 214)

To measure this length of time, Jevons introduced the concept of an 'average interval of investment', which anticipated, in many respects, the Austrian notion of 'average period of production' that we deal with below.

Formally, let $Q = F(t)$ be the output of a process lasting t years, having a positive first derivative $F'(t) > 0$, which is another way of saying that 'roundaboutness' makes inputs more productive. The interest rate is equal to the growth rate of output resulting from waiting (Jevons 1871, p. 267),[3] which is equivalent to writing that $r = F'(t)/F(t)$ or in Jevons's words

> the interest on capital is [...] the rate of increase of the produce divided by the whole produce (Jevons 1871, p. 267).

Of course, $F(t)$ as a function of time can be optimised and its maximum is attained if $F'(t) = 0$ and $F''(t) < 0$, the negative second derivative indicating the law of diminishing marginal productivity and the falling rate of profit associated with it. In fact, the notion of long run falling rate of profit was, as we saw, a standard topic of the classical economists and continued, with the same status, in the writings of the major neoclassical economists.

> It is one of the favourite doctrines of economists since the time of Adam Smith, that as society progresses and capital accumulates, the rate of profit, or more strictly speaking, the rate of interest, tends to fall. The rate will always ultimately sink so low, they think, that the inducements to further accumulation will cease. This doctrine is in striking agreement with the result of the somewhat abstract analytical investigation given above. Our formula for the rate of interest shows that unless there be constant progress in the arts, the rate must tend to sink towards zero, supposing accumulation of capital to go on. There are sufficient statistical facts, too, to confirm this conclusion historically. The only question that can arise is as to the actual cause of this tendency. (Jevons 1871, p. 94)

Jevons came too close to the notion of marginal productivity of capital, however, he could not provide with a complete demonstration of the relationship between the rate of interest and the production period from the viewpoint of a producer's maximising behaviour. This point can be shown by invoking from financial mathematics (or Stigler 1941) the usual formula of the present value (V) of a stream of income in continuous time,

[3] Wicksell calls 'natural', the rate of interest which is equal to the growth rate of output.

$$V(t,r) = F(t)(1+r)^{-t}$$

With the rate of interest given for maximisation, we set $V(t)\prime = 0$ and so

$$F(t)'(1+r)^{-t} - F(t)(1+r)^{-t}\ln(1+r) = 0$$

and

$$\ln(1+r) = \frac{F'(t)}{F(t)}$$

The left hand side is the instant interest rate and the right hand side is the natural interest rate. Thus through the above equation, we can determine the t that maximises the present value of output, when the interest rate is given. Of course, we could maximise the above function for the rate of interest, r, assuming t as given.

Böhm-Bawerk (1889) a student of Menger made the study of the interest rate and capital the area of his specialisation. He began by adopting his teacher's ranking of goods into consumer goods and capital goods. The capital goods are used as means of production of consumer goods and so are called goods of 'higher order' that take time to mature (Kurz and Salvadori 1995, p. 177), whereas the consumer goods are immediately available for consumption and are called 'goods of the first order'. Clearly, the time dimension pervades both types of goods. Having made time the distinguishing characteristic of goods, Böhm-Bawerk introduced the notion of the 'average period of production' in order to give quantitative worth to capital independently of prices. The average period of production (APP) is the weighted average of all intervals of time during which the quantity of labour is spent to obtain a certain amount of output, where the weights are given by the respective quantities of labour. By way of an example, in order to produce a certain capital good, we applied 150 labour units (L), 5 years (Y) ago and 50 labour units 1 year ago, the total amount of past labour is 200 labour units and the average period of production (τ) will be:

$$\tau = \frac{150L \times 5Y + 50L \times 1Y}{(150+50)L} = \frac{800LY}{200L} = 4 \text{ years}$$

This estimation of the APP raises a number of interesting questions:,

- Why is land excluded from the estimation?
- Why is the specific average selected and not others?
- Why is the interest rate not used?

As for the first question the answer is simple and has to do with the different physical units of measurement of land and labour; as a consequence, the two factors of production must be expressed in homogenous units of measurement (same valuation) and this is certainly difficult practically, but manageable in principle.

As for the choice of the average this is not a really a difficult problem and all one has to do is stick to one average (weighted, geometric, *etc.*) all these averages will give approximately the same answer and ranking. Finally, the inclusion of the rate of interest is no problem, so long as we assume a simple interest rate, the idea being that the interest rate will be included in both numerator and denominator and its effects will be cancelled out. In terms of the above numerical example and neglecting symbols for labour units and years, for the case of the simple interest rate we can write:

$$\frac{150(1+5i)+50(1+i)}{(150+50)}=\frac{(150+50)(1+\tau i)}{200}\ \text{and}\ \tau=\frac{4i}{i}=4\ \text{years}$$

Although the simple interest rate does not affect the initial findings of the APP, the same is not true with the compound annual interest rate.[4] The latter not only complicates the estimations but moreover presupposes the measurement of capital which in its turn requires the knowledge of the rate of interest, and so we enter into a vicious circle from which there is no way out.[5]

Let us now experiment with our simple, but fair, numerical example and the compound annually interest rate. The new APP will be:

$$\frac{150(1+i)^5+50(1+i)}{200}=\frac{200(1+i)^\tau}{200}\ \text{and}\ \tau=\frac{\ln\left[150(1+i)^5+50(1+i)\right]-\ln(200)}{\ln(1+i)}$$

$$\text{and for } i=10\%, \tau\approx 4.13$$

In this particular example the difference in the two APP is not too large (about 3.25%). Thus the APP depends also on the rate of interest, and the relationship is no longer linear and so we cannot tell *a priori* whether the APP of one good is higher than another (other things equal) unless we know the rate of interest. It is fair to say, though, that apart from theoretical reasons (the vicious circle problem), for all practical purposes one does not expect the compound interest rate to give altogether different results in terms of rankings of the APPs, unless the APPs are too close to each other in the first place and so their ranking may not remain the same.

Let us now express the productivity of labour as a function of the average period of production, $q=f(\tau)$, the usual assumptions hold, that is the function is, twice differentiable with $f'(\tau)>0$ and $f''(\tau)<0$. The wage rate, w, is determined in the

[4]We may recall from the third chapter that Ricardo in explaining the deviations of relative prices from relative labour times expended in the production of commodities requiring different production periods for their completion includes compound interest rate explicitly.

[5]Niehans (1990) argues that the negligence of the interest rate is a "remarkable feature" in Böhm-Bawerk's conceptualisation and makes some hints that in Vienna surprisingly enough economists did not know much about the compound interest rate. The truth, however is that the inclusion of compound interest rates creates insoluble theoretical problems of measurement of capital in physical units, something that Niehans (1990, p. 229) also acknowledges.

labour market and profits (Π) are residually determined by the difference between productivity and wage rate, that is, $\Pi = q\text{-}w$. Furthermore, in order for the entrepreneur to complete the production period, τ, his capital outlays will be equal to the product of the APP times the wage rate, τw. This is equivalent to saying that capital in Böhm-Bawerk is conceived as a wage fund. It follows therefore that the rate of return on capital will be,

$$r = f(\tau)(w\tau)^{-1} - w$$

Thus, the problem for the entrepreneur is to find τ, given w, such that r is maximised. The first order condition of the above function is:

$$f'(\tau)\tau = f(\tau) - w$$

and so the maximum (optimum) interest rate will be

$$r = \frac{f'(\tau)}{w}$$

In this expression Böhm-Bawerk, unlike Jevons, managed to express the interest rate as the marginal product of capital relative to the wage rate. Furthermore, turning to the economy as a whole, we see that while the wage rate is a variable the labour force is considered to be fixed. Thus if by $\kappa = K/L$ we represent the capital–labour ratio of the economy, then it follows that the equilibrium wage will be the one at which all labour is being employed during the period of production by the entrepreneurs. Then $w = \kappa/\tau$ and by substituting in the first order condition we get:

$$f'(\tau)\tau^2 = f(\tau)\tau - \kappa$$

Where the left-hand side of the equation stands for total profits, that is, the product of unit profits $f'(\tau)\tau$ times τ. Total output is $f(\tau)\tau$ and total wages are equal to $\kappa = w\tau$. The above equation essentially states that equilibrium is attained at the point where, given the total employment in the economy, the wage is at a level where labour is fully employed.

Wicksell's contribution to the theory of capital is partly based upon Walras's theory to the extent that capital is viewed as a heterogeneous set of goods and partly upon Böhm-Bawerk's idea about the APP. Wicksell analysed how the structure of capital goods depends not only on the number of units of investment, but also upon the length of time during which the inputs are invested. He asserted that

> the importance of the time-element in production was never properly appreciated by Walras and his school. The idea of a *period* of production or of capital-investment does not [...] exist in the Walras-Pareto theory; in it capital and interest rank equally with land and rent;

> in other words, it remains a theory of production under essentially non-capitalistic conditions, even though the existence of durable, but apparently indestructible instruments, is taken into account (Wicksell 1901/1934, I, p. 171).

Wicksell argued that capital should be treated in a way similar to that of labour and land which is equivalent to saying that capital should be transformed to a homogeneous, and, therefore, measurable factor of production. We know that despite difficulties labour and land can be measured in homogeneous physical units and that their marginal productivity determines the relative rewards of these factors of production. The trouble with capital is that it cannot be expressed in common units of measurement which are amenable to optimisation. At the same time, Wicksell was critical to the Walrasian 'solution' to the problem at hand, which was to express capital as a bundle of heterogeneous goods, because in this case we derive the rate of profit for each type of capital good and not the general (uniform) rate of profit.

Wicksell (1934) argued that the homogenisation of capital can be achieved once we express it in terms of the original and indestructible factors of production of labour and land together with the time dimension. Capital therefore is viewed as a derived concept resulting from the accumulated labour and land,

> a single coherent mass of saved-up labour and saved-up land. (Wicksell [1901] 1934, p. 150)

Assuming, for a moment, that land is a free good, we are left with accumulated (dead) labour versus current (living) labour with the former being more productive than the latter. In general, the productivity of accumulated labour or, what amounts to the same thing, the productivity of capital depends on the stages of production (roundabout production), while the time element of production is responsible, other things being equal, for the increase in efficiency, which is a condition *sine qua non* for the rate of profit (interest). According to Wicksell

> Capital is saved-up labour and saved-up land. Interest is the difference between the marginal productivity of saved-up labour and land and of current labour and land. (Wicksell, [1901] 1934, p.154).

Having conceived the physical quantity of capital as a sum of dated quantities of labour and land, Wicksell could proceed with the estimation of the value of the marginal product of each of the two factors of production which when added up would give the total value of the capital stock. In other words, the value of the capital stock is the amount invested in labour and land compounded at the interest rate over the average investment period (Kurz and Salvadori 1995, p. 442).[6] Wicksell was sceptical about his definition of capital in value terms as he found that the marginal productivity of capital was not always equal to the rate of interest (Kurz 1990, p. 84). We deal with some of the details of such discrepancies in the next sections.

[6]For a detailed formal presentation see Kurz and Salvadori (1995, p. 442).

8.4 Samuelson's Surrogate Production Function

We have already known from Robinson's (1953) article that the use of an aggregate measure of capital in a production function in an effort to deal with macroeconomic and income distribution problems is highly problematic. The defence of the neoclassical theory of value and distribution was undertaken by Paul Samuelson (1915–2009), who argued that it is possible to develop capital theory *without* conceiving capital as an aggregate or homogeneous good (as for instance Jevons, Böhm-Bawerk and Wicksell do) by making use of mathematical techniques. Samuelson in his 'Parable and Realism' stirred a fierce debate, the essence of which revolved around the fundamental premises of the theories of value, distribution, and growth, each of which depends upon an aggregate production function, where the inputs or factors of production for capital and labour are aggregated in some fashion prior to the determination of the rate of profit (interest) and the wage rate.

According to neoclassical theory, the price of each factor of production is determined by its marginal contribution to production; furthermore, there exists substitutability between the factors of production that give rise to diminishing returns. As a consequence, the rate of profit (or interest) is the price of capital and as such reflects capital's relative scarcity. More specifically, a relative abundance of capital, in combination with the law of diminishing returns of a factor of production (whereby the greater use of an input will imply a lower marginal product, other things being equal) will give rise to a low rate of profit (interest). The opposite would be true in the case of a relative scarcity of capital. Capital income would amount to the product of the rate of profit times the amount of capital employed.

Crucial to the understanding of the substance of the debate is the issue of the measurement of the quantity of factors of production and especially of capital. This measurement must be consistent with the requirements of the theory and, at the same time, must be measured in units which are economically meaningful. Thus, the measurement of factors of production should be in terms of units which are independent of prices, since prices are the unknowns that we want to estimate. On the other hand, the quantity of a factor of production must be expressed in such a dimension, so as to be suitable for a cost minimization choice of technique.

These conditions are reflected in the shape of the isocost and isoquant curves. Figure 8.1 presents such curves, where on the vertical axis we set the amount of land (non-reproducible means of production); while on the horizontal we measure the amount of labour. Thus in the case of labour and land, their quantity can be measured independently of prices; for instance, acres of (same quality) land; while the quantity of labour is being measured by the amount of (homogenised) working hours. We observe that both measurements are independent of their price. When we talk of a quantity of land measured in acres and of labour measured in hours, we mean that both variables can be expressed in such a way as to form appropriate indexes of their quantity which becomes amenable to cost minimization for the production of a given quantity of output. In other words, the units of measurement of the two variables can

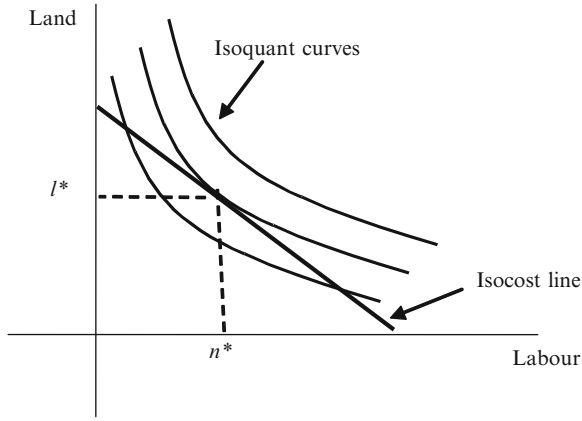

Fig. 8.1 The optimal combination of land-labour

be set in the economically meaningful dimension.[7] Having made the appropriate choice of the units of measurement of the two non-produced factors of production, we can then choose the cost minimising technique, that is, the point of tangency of the isocost line with the highest attainable isoquant curve. The tangency point determines the optimal combination of the two factors of production (l^*, n^*) that produce a given level of output.

While the analysis of non-produced means of production does not present any insurmountable problems, when we come to produced means of production and especially capital, we encounter the following characteristics:

1. Capital is reproduced and thus differs from both land and labour
2. Capital depreciates
3. Capital's price is equal to its production cost

The first two features of capital are self-evident; the third one deserves further examination. First of all the price of the capital good is determined by the relation

$$P_K = \frac{X_K}{r}$$

where P_K = price of capital good, X_K = annual net (of depreciation) income of the capital good and r = rate of profit. In this case, capital should be viewed as a perpetual bond (consol) that gives rise to an annual net income forever. For the sake of simplicity, we assume that this annual income is uniform.

[7]If, for instance, instead of labour time we used the height or weight of workers, we would manage to homogenise the "quantity" of labour, however, the result of such an optimization (of height or weight) would be devoid of economic meaning.

In equilibrium, the demand price of capital goods should be equal to their cost of production. Since this condition should apply to all capital goods, it follows that the profit on the cost of production of each particular capital good should be equated with the average rate of profit of the economy,

$$r = \frac{X_K}{C_K}$$

where C_K is the cost of production of capital goods. Therefore, the condition for the existence of a uniform rate of profit is identical to the condition that the price of capital goods should be equal to the cost of their production.

The capital goods have a second characteristic, belonging to the initially given endowment of goods and factors of production possessed by each individual, therefore the price or remuneration for their services of capital goods X_K must be determined in the capital market. Therefore, the ratio of rentals of capital goods to their production cost must be equal to the general rate of profit. This is a unique characteristic of capital goods. For example, in the case of labour and land, the value of fixed rates in their respective markets is determined by the forces of demand and supply. The prices of consumer goods are determined by the cost of their production, while the prices of capital goods must be determined (Eatwell *et al*. 1990, p. 12) by supply and demand in the capital markets and their production costs in the commodity markets. In other words, we have two conditions that define the same set of prices. In mathematical terms, we say, that the system is over-determined and as such there is no unique solution.

In order to show the problem with the evaluation of capital, let us suppose an economy with only two goods, a consumer and a capital good. There is a single technique for the production of capital goods, which is also used in the production of the consumer good. In equilibrium we know that the price of consumer goods must be equal to the cost of their production,

$$P_C = wL_C + rK_C P_K$$

where P_C = price of consumer goods; w = the nominal wage; L_C = labour employed in the production of consumer goods; K_c = the capital employed in the production of consumer goods; P_K = price of capital. While for the price of capital goods we have:[8]

$$P_K = wL_K + rK_K P_K$$

[8]A more precise way to state the above is to write $P_i = wL_i + K_i P_K + rK_i P_K$, where $i = C, K$. For reasons of simplicity, we omit other elements of production costs such as circulating costs and so we are left with the wage costs (wL_i) and profits ($rK_i P_K$,). The results of our analysis are not qualitatively affected by this simplification.

The notation is similar to that of consumer goods, with the difference that the index K refers to capital goods. If we take the price of consumer goods as a common denominator (or numeraire), then we form the following system of equations:

$$1 = wL_c + rK_cP_\kappa$$

$$P_K = wL_K + rK_KP_K$$

We solve for the price of capital goods from the second equation of the system and thus we have,

$$P_K = \frac{wL_K}{1 - rK_K}$$

By replacing in the first equation of the above system of equations, we get:

$$1 = wL_C + rK_C\left(\frac{wL_K}{1 - rK_K}\right)$$

Solving for the wage rate

$$w = \frac{1 - rK_K}{L_C(1 - rK_K) + rK_CwL_K}$$

We arrive at the wage–rate of profit frontier

$$w = \frac{1 - rK_K}{L_C + r(K_C/L_C - K_K/L_K)L_CL_K}$$

which is a quadratic equation. If we put the restriction that the capital–labour ratio is the same in both sectors, $K_C/L_C - K_K/L_K = 0$ then the w–r relation simplifies to the following linear equation:

$$w = \frac{1 - rK_K}{L_C}$$

which is shown in Fig. 8.2a, where the maximum wage, that is, the wage for $r = 0$ is equal to $1/Lc = w_{\max}$. While for $w = 0$, we get the maximum rate of profit which is $1/K_K = r_{\max}$. The above relation is generalised to a variety of techniques, where a technique is defined as a different capital–labour ratio. Theoretically, we can have any number of techniques ($w = f(r)$) and these are illustrated in Fig. 8.2b

If we differentiate the relationship $w = (1 - rK_K)/L_K$ with respect to r and by omitting the indices we can write:

$$\frac{dw}{dr} = -\frac{K}{L} = \text{constant}$$

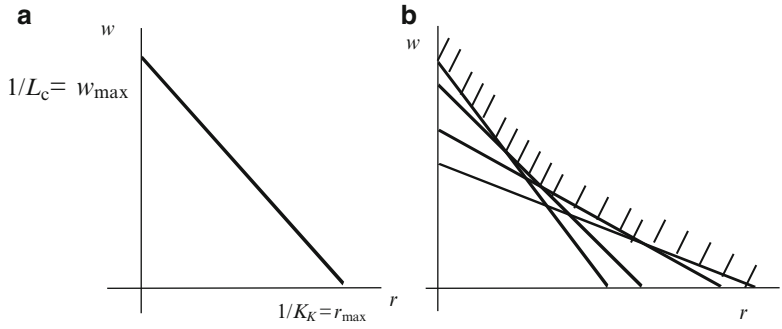

Fig. 8.2 The wage–profit rate line and *w*–*r* frontier

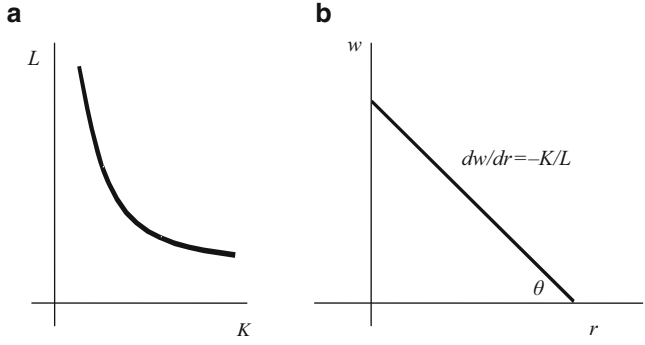

Fig. 8.3 The isoquant and the wage-rate of profit curves

This means that if the wage–rate of profit frontier is linear then the slope of the line is equal to the capital–labour ratio. If we assume a uniform capital–labour ratio for all sectors, it is apparent that the relationship between rate of profit (interest rate) and wage is linear. This relationship is the dual representation of the already known (Chap. 6) diagram of the isoquant curve.

On the left-hand side graph of Fig. 8.3, we have capital and labour and the isoquant curve represents the locus of points of different combinations of capital and labour to produce a given quantity of output. The tangents to this curve represent the price ratios of the two factors of production, which in this case are the wage and the profit rate. The tangent of the angle θ of the right hand side graph of Fig. 8.3 represents the capital–labour ratio. It therefore follows that the two graphs are equivalent representations of a choice of techniques. We now assume more than one technique of production which we depict on the upper part of Fig. 8.4a, while in the bottom part of the graph we depict the value of capital.

Producers, owing to competition, choose the cost minimising techniques or profit maximising techniques, which are also the same. Therefore, for a given wage, producers choose the technique with the highest rate of profit or for a given rate of

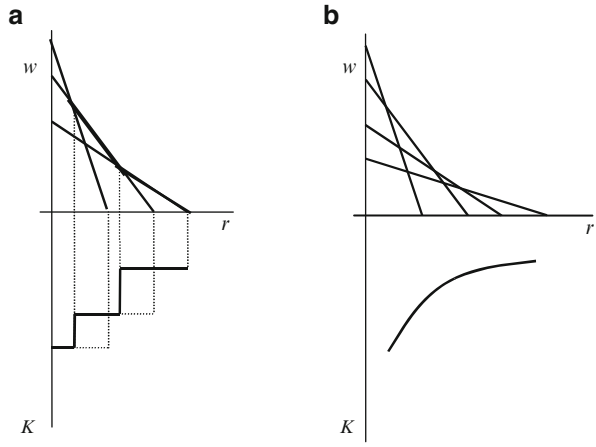

Fig. 8.4 The demand for capital curve

profit producers choose the technique with the maximum wage. As a consequence, entrepreneurs choose points on the outer envelope formed by the outer segments of each line from the three techniques. In short, entrepreneurs choose points on the factor price frontier. Since we have straight lines and hence only one point of intersection per pair of lines, it follows that the relationship between the value of capital and the rate of interest (profit rate) is an inverse one. This relationship is depicted in a stepped form curve in the bottom of Fig. 8.4a. If, however, the number of techniques increases indefinitely as in Fig. 8.4b, then the relationship between the value of capital and the rate of profit becomes a continuous function.

These were the relationships that led Paul Samuelson to the conclusion that if we have a uniform capital–labour ratio across sectors, or if we have a one-commodity world, which amounts to the same thing, the profit rate is determined by the relationship between cost of production and demand for this single capital good. Moreover, the rate of profit and the value of capital are inversely related, a result which is fully consistent with the neoclassical theory, where the payments for the services of the factors of production reflect their relative scarcity.

8.5 From the One-Commodity World to the Real Economy

Samuelson argued that the conclusions arising from the analysis of the one economy world are generalised to economies that produce n number of commodities.[9] One wonders, what was the reason that led economists to adopt this non-realistic

[9]Samuelson criticised the Marxian model for lack of logical consistency, because labour values and prices of production are equal to each other only if one makes the unrealistic assumption of

assumption? The problem lies in determining the relationship between price-cost of production, profit rate, and the price determined in the market. If we have only one good then we can write:

$$P = wL + rPK$$

if we set $w = 1$, then for any $P > 0$ there will be a profit rate such that $r > 0$. As a consequence, in the case of a one commodity we can determine its price, as is determined by the market forces and so the rate of profit is determined indirectly.

If we have two goods, their relative prices must correspond to a uniform rate of profit.[10] If we add more goods, then there are more relative prices, which should correspond to a uniform rate of profit. The question is whether and to what extent we can generalise the findings of a one commodity economy to the actual economies. If, for instance, we have two goods, then the following relationship holds:

$$w = \frac{1 - rK_K}{L_C + r(K_C L_K - K_K L_C)}$$

which is not linear but convex or concave. The curve is concave, if the second derivative of this relationship is positive and is convex if the second derivative is negative. We set $(K_C L_K - K_K L_C) = D$ and then differentiate the above relation with respect to the rate of profit (interest) and get:

$$\frac{dw}{dr} = \frac{(1 - rK_K)'(L_C + rD) - (L_C + rD)\prime(1 - rK_K)}{(L_C + rD)^2} = -\frac{-K_K L_K}{(L_C + rD)^2}$$

While the second derivative gives:

$$\frac{d^2 w}{dr^2} = \frac{\left[(L_C + rD)^2\right]' K_K L_K}{(L_C + rD)^4} == \frac{2DK_K L_K}{(L_C + rD)^3}$$

We observe that the shape of the wage–profit rate curve depends on the sign of the determinant

$$D = K_C L_K - K_K L_C = \left(\frac{K_C}{L_C} - \frac{K_K}{L_K}\right) L_K L_C$$

uniform capital intensity. The irony is that the Marxian theory of value (as we showed) does not need to make such an assumption; nevertheless, Samuelson adopted this very assumption of a uniform capital labour ratio to save the logical consistency of the neoclassical theory.

[10]If the two goods are produced by the same capital–labour ratio, it follows that we essentially have a one commodity economy.

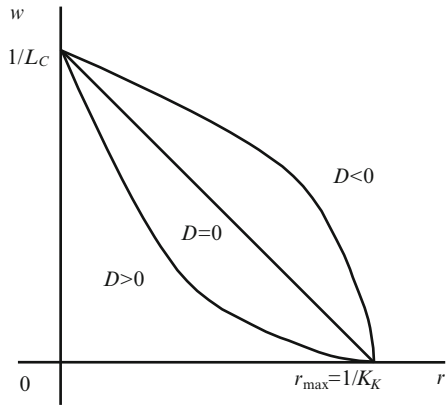

Fig. 8.5 Wage–profit rate curve and the *K*/*L* ratio

We distinguish the following cases:

- If $D > 0$, that is, $\frac{K_C}{L_C} > \frac{K_K}{L_K}$ with $\frac{d^2w}{dr^2} > 0$ and *w*–*r* curve is concave.
- If $D < 0$, that is, $\frac{K_C}{L_C} < \frac{K_K}{L_K}$ with $\frac{d^2w}{dr^2} < 0$ the *w*–*r* curve is convex.
- If $D = 0$, that is, $\frac{K_C}{L_C} = \frac{K_K}{L_K}$ with $\frac{d^2w}{dr^2} = 0$ the *w*–*r* is a straight line.[11]

The three cases are illustrated in Fig. 8.5

If we have three goods, we will have two curvatures and an inflexion point, with four goods will have three curvatures and two inflexion points, and so forth. The maximum number of curvatures depends on the number of sectors. Through the factor price frontier we managed to show the technology and all possible payments of factors of production without having to refer to a production function. The question is whether we end up with the same results.

We now focus our attention to the case of two sectors and compare the results obtained in the case of a sector, whose technique is portrayed in Fig. 8.6.

The area under the curve is the per capita physical output. If $r = 0$, then all output goes to labour, and so $w = w_{\max}$. Assuming that $w = 0$, then all output goes to capital, and so $r = r_{\max}$. The profit rate is estimated from the relation:

$$r = \frac{\pi}{K/L}$$

Therefore, the capital–labour ratio is written:

$\frac{K}{L} = \frac{\pi}{r} = \tan\vartheta$, and, therefore, $\frac{dw}{dr} \neq -\frac{K}{L} = \text{constant}$

Thus, $\tan\theta$ is not constant at any point on the curve, but changes with every change in the distribution of income.

[11] This is the case of a one good economy or an economy, where the capital–labour ratio is identical across sectors.

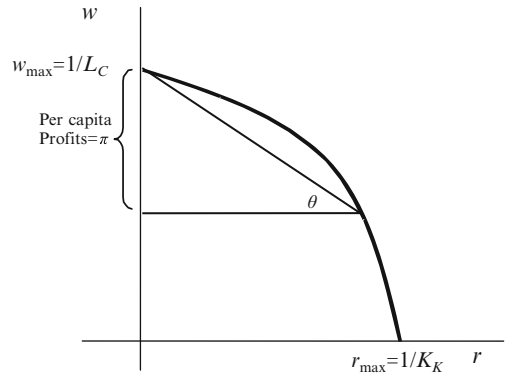

Fig. 8.6 The *w*–*r* curve with two goods

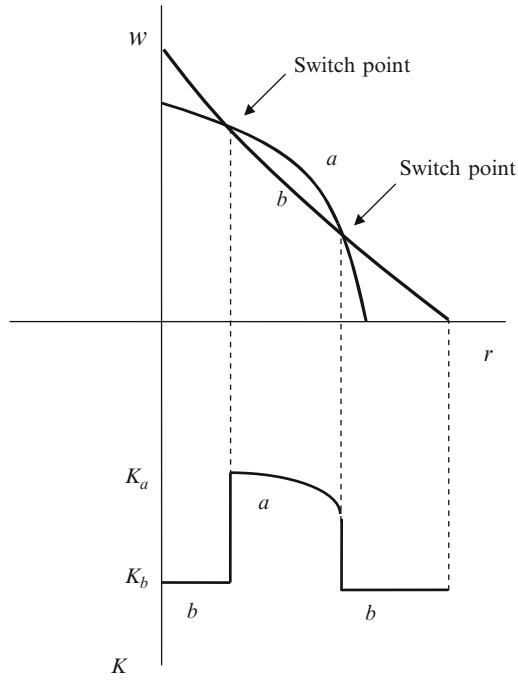

Fig. 8.7 Switch points and the demand for capital

We now assume a two goods and two techniques economy. We already have shown that if we had only one good and two or more techniques these would be shown as straight lines. In the case of two goods, the wage–rate of profit curve would display curvature (see Fig. 8.7). We observe that as we move from one technique to another, the value of capital and the profit rate can be related in a number of ways and not necessarily in the one required by the neoclassical theory; that is, the negative and monotonic relationship between quantity of capital and

the rate of profit. If we have more techniques and more goods, the indeterminacy of the relationship between the quantity of capital and the rate of profit is simply reinforced.

On the top graph of Fig. 8.7 we observe two switch (intersection) points which would be impossible to have in case of a single commodity. However, the choice of technique, as in the case of production of one good, makes the selection of techniques along the price factor frontier. Thus, for a given profit rate the technique with the highest wage is chosen or for a given wage the technique with the highest rate of profit is chosen.

Let us assume that the profit rate is very low. In such a case, technique *b* is chosen, that is the technique with a capital–labour ratio higher than that of technique *a*. When we reach the switch point, the two techniques are equally profitable and thus we are indifferent as to which technique is chosen. As the profit rate increases, *a* becomes the most profitable technique, and, therefore, is preferred to technique *b*. In other words, as the profit rate increases instead of choosing the technique with the lowest capital–labour ratio (as required by the neoclassical theory) we choose the technique with the highest capital–labour ratio! Until we reach the next switch point, where again we are indifferent as to which technique we use (*a* or *b*). For an even higher rate of profit, we have a reswitching of techniques to *a*. That is, we choose the technique with the highest capital–labour ratio and leave the technique with the lowest capital–labour ratio. Theoretically, we can have more than one switch point as shown in Fig. 8.8.

It goes without saying that if we produce more than one good the linear wage–profit rate relations no longer holds, while, at the same time, the probability of occurrence of switching points increases, raising questions on the consistency of neoclassical theory.

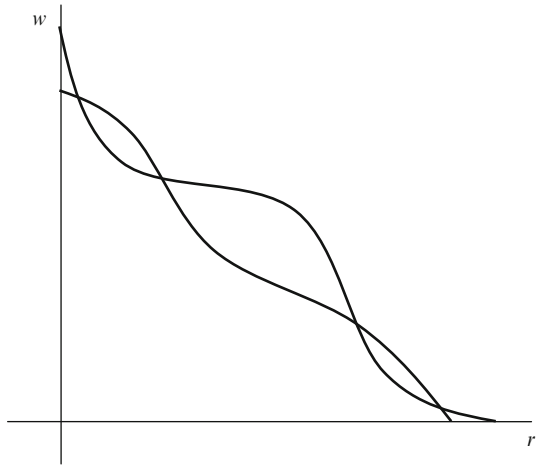

Fig. 8.8 Wage–profit rate curves with many goods

8.6 Wicksell Effects

Wicksell effects refer to changes in the value of capital, as a result of changes in the distributive variables (*i.e.*, w and r). They are called *price Wicksell effects*, when the technology is given and the changes in the distributive variables affect the value of capital. The *real Wicksell effects* refer to the changes in the value of capital as an effect of income distribution and the resulting changes in technology, that is, the real Wicksell effects refer to the switch points along the factor price frontier. The two effects reflect the influence of w and r over time on inputs, but the real Wicksell effects reflect additional changes in the methods of production, that is, the changes affecting the productive capacity and not only the value of production. These results can be traced through the following relationship:

$$y = w + rk$$

where $y =$ per capita production of consumer goods and $k =$ capital–labour ratio. The total differential of the above relations gives:

$$dy = dw + rdk + kdr$$

dividing through by dk, we get an expression for the marginal product of capital:

$$\frac{dy}{dk} = \frac{dw}{dk} + r + k\frac{dr}{dk}$$

Joan Robinson viewed the above result as crucial for the neoclassical theory of distribution, because the marginal product of capital is not equal to the rate of profit, unless we assume that:

$$\frac{dw}{dk} = -k\frac{dr}{dk} \quad \text{or} \quad dw = -kdr \quad \text{and} \quad \frac{dw}{dr} = -k = \frac{K}{L}$$

In other words, the derivative of the wage with respect to the profit rate is equal to the negative capital–labour ratio, which corresponds to the case of a one good economy or an economy, where the capital–labour ratio is uniform across sectors. In this case, assuming the output produced as fixed and changing the technique, we generate a number of lines forming the price factor frontier, that is, the optimal technique in a perfectly competitive market, given the wage and the profit rates.

On the basis of the above analysis the demand for capital schedule can be constructed. It is important to note that we assume perfect substitution between production factors. If we now focus on the above differential and solve for the change in the capital–labour ratio, we will have:

$$dk = \frac{1}{r}dy - \frac{k}{r}dr - \frac{1}{r}dw$$

where, $r^{-1}dy$ is the capitalised value of physical productivity corresponding to the real Wicksell effect, which is also called *Ricardo effect*, and $kr^{-1}dr - r^{-1}dw$ is the price effect. Depending on the sign of dk/dw the Wicksell effect will be positive or negative. Thus, if $dk/dw > 0$, then we say that the results are 'normal', while if $dk/dw < 0$, then the results are 'abnormal', for the neoclassical theory. Obviously, the terminology is biased in the sense that the neoclassical theory is considered as *a priori* consistent and so if the results are in agreement with this theory, then this is good for the results, which in this case are called 'normal'. If, however, the results do not agree with the theory, the blame is placed not on the theory but on the results which, this time, are called 'abnormal'.

The capital–labour ratio $k = (y-w)/r$ is the tangent of the angle θ (see Fig. 8.6). If now the tangent θ increases every time that r increases, which means that the factor price frontier is convex (or what amounts to the same thing, $D < 0$) the Wicksell effect is negative or 'abnormal', precisely because the value of capital increases with each increase in the rate of profit.

If now $D > 0$ this implies a concave factor price frontier, which is equivalent to saying that the capital–labour ratio falls as the rate of profit rises, the value of capital falls as the rate of profit rises and the price Wicksell effect is positive or 'normal' and reinforces the real Wicksell effect. We observe that even though we cannot rule out the re-switching of techniques, where the factor price frontier is concave the results that we derive are consistent with the neoclassical theory.

If finally the w–r line is linear, *i.e.*, $D = 0$, then the price effect is zero and the profit rate equals the marginal productivity of capital.

The importance of this effect (re-switching of techniques) is that the conclusions, which are valid in the case of a one-commodity economy, are not generalised. Practically this means that we cannot, for instance, construct on the basis of information from the national income accounts the capital–labour ratios and then estimate the demand function for capital for the entire economy. And if the demand for one factor of production cannot be formed then because of interrelationships the demand functions for the other factors cannot be constructed and the whole theory is in question.

8.7 Summary-Conclusions

Piero Sraffa argued that it is not in general possible 'to find in the 'period of production' an independent measure of the quantity of capital which could be used, without arguing in a circle, for the determination of prices and of the shares in distribution' (Sraffa 1960, p. 38). The subsequent capital controversies have shown, at least in theory, that the overall production functions (used extensively in neoclassical theory) are a myth, while the theory of marginal productivity theory of income distribution, according to which (under conditions of perfect competition) prices reflect the relative scarcity of goods and factors of production and each production factor is paid according to its marginal contribution to total production, is theoretically ill-founded.

It is important to point out that the capital theory critique does not affect the classical theory of value and income distribution, because the classical theory does not claim that relative prices of factors of production reflect relative scarcities; additionally, this theory assumes one of the distributive variables, usually the real wage, as a datum which in combination with the given technology and output level determines the relative equilibrium prices together with the equilibrium rate of profit. Furthermore, the evaluation of heterogeneous capital goods can be achieved in terms of labour values; hence there might be a problem of consistency because variables estimated in terms of labour values will differ from those estimated in terms of equilibrium prices (production prices). This, however, is mainly an empirical question and research hitherto has shown that the two types of prices are surprisingly close to each other and that variables such as the rate of profit, capital output ratio and the like, when estimated in labour values and then equilibrium prices, are approximately equal to each other (Shaikh 1984, 1997; Ochoa 1989; Cockshott and Cottrell 2000; Tsoulfidis and Maniatis 2002; Tsoulfidis and Rieu 2006, *inter alia*).

Therefore, in the classical approach to the theory of value and distribution the determination of the real wage (or the rate of profit) is logically prior to the determination of equilibrium price. By contrast, in the neoclassical approach, the wage and profit rates should be determined in labour and capital markets, respectively. Finally, the analysis shows, at least mathematically, that there is no monotonic relationship between the profit rate and the capital–labour ratio. These results raise questions about the idea that the yield of capital is inversely proportional to its scarcity, which is one of the key propositions of the neoclassical theory, that is, the prices of factors of production reflect their relative scarcities.

The capital controversy had an initial effect on neoclassical economics, but soon it was forgotten to the point that the new generation of neoclassical economists either dismisses it or simply does not know about it. As a result, both theoretical and empirical neoclassical research makes use of aggregate production functions, where capital is still used along with labour in the determination of output and the marginal products of these inputs are estimated on the assumption of substitutability between factors of production, as if the capital critique was *much ado about nothing*. The critique of Sraffian economists, however, continued and a number of publications began to surface, which may revive theoretical questions that puzzled, as Leontief (1986, p. 410) once remarked, 'the best Cambridge economists in England and the United States that pitted for many years the sharpest minds of Cambridge, Massachusetts against the brightest theoretical lights of Cambridge, England'.

The reaction of neoclassical economists in the critique of their theory was to develop models of general equilibrium, where prices of goods are determined by the forces of demand and supply (Arrow and Hahn 1971). According to this model the decisive factor is the scale of production and there is no separation between the theories of price and income distribution. The prices of the services of labour and capital are determined as the prices of all other goods. Since the distributive variables (wage and profit rate) are determined endogenously it follows that their

relationship is not necessarily competitive, and this is so because both variables change toward the same direction with every change of the exogenously specified parameters. In this general equilibrium framework, the determination of equilibrium prices requires equilibrium in all future markets. In this sense there is no systematic relationship between the rate of profit and capital–labour ratio. As a consequence, the general equilibrium models do not have the disadvantages of the neo-classical theory proper, where the increase (decrease) in capital–labour ratio is reflected monotonically on the profit rate, which is essentially an indicator of the relative scarcity of capital.

Questions for Thought and Discussion

1. What problems does the theorization of capital create to the Neoclassical theory of value and distribution?
2. How did Jevons try to resolve the problem of the measurement of capital? Was his solution satisfactory?
3. Present and critically evaluate Böhm-Bawerk's treatment of the problem of the measurement of capital.
4. Discuss the advantages of Böhm-Bawerk's treatment of the measurement of capital over Jevons's treatment.
5. How did Wicksell deal with the problem of the measurement of capital? Was his method of measurement satisfactory?
6. Is the measurement of the quantity of capital a problem for the classical theory of value and distribution? Explain.
7. Discuss Samuelson's solution to the problem of measurement of capital.
8. What is meant by real and price Wicksell effects?
9. What are the consequences of the reswitching of techniques in the usual macroeconomic analysis?
10. How does the theory of capital affect the Marxian labour theory of value?

Notes for Further Study

A summary history of the concept of capital in neoclassical economics would be a volume in itself. For this reason we discussed the essential ideas of the first neoclassical economists, mainly Jevons, Böhm-Bawerk and Wicksell. The inventiveness of these economists to define capital using the labour time as the physical units of measurement is quite impressive. It seems that the first neoclassical economists, while they discarded the theory of value based on labour time for reasons that we explained in Chap. 7, nevertheless, had no problem in reintroducing the labour time for the measurement of the quantity of capital. We showed that the new theory of value could not corroborate with the old doctrine and the problems associated with the measurement of capital consistent with the neoclassical theory of value and distribution persist until today.

(continued)

The debate on capital theory started with the article of Robinson's (1953) article which was inspired by Sraffa's (1951) introduction to *Ricardo's Works*. The controversy, however, essentially began after the publication of Sraffa's (1960) book. Samuelson (1962) assumed the defence of the main proposals of the neoclassical theory in the context of a one good economy and then claimed that his conclusions apply to economies with many goods. It is interesting to note that the assumption of a one good economy is equivalent to the assumption of uniform capital–labour ratio across industries. Parenthetically, Samuelson attacked the Marxian value theory for its alleged inability to explain relative prices. However, if one applies Samuelson's heroic assumption of equal capital intensity across all industries to Marx's labour theory of value, then all of Samuelson's criticisms of Marx become irrelevant. This irony was not unnoticed by the British participants in the capital debates. Samuelson's assumption was attacked for lack of realism by Garegnani (1970, 1976, 1990) and Pasinetti (1966) among others, who showed that once we hypothesise different capital intensities across industries, the neoclassical results do not hold.

The idea is that as relative prices change as an effect of changes in income distribution the subsequent revaluation of capital can go either way, and it is possible for an industry that is capital-intensive in one income distribution to become labour-intensive in another (see Pasinetti 1977 and for a formal presentation as well as with empirical findings see Shaikh 1998; Mariolis and Tsoulfidis 2009). As a consequence, we no longer derive Samuelson's straight-line wage–profit rate frontiers, which are consistent with the cost-minimising choice of technique and which give rise to well-behaved demand-for-capital schedules. In the presence of many capital goods and various capital intensities across industries it follows that the wage–profit rate frontiers are nonlinear and may cross over each other more than once, which means that for a low rate of profit one may choose a capital-intensive technique. As the rate of profit increases, the technique with lower capital intensity may be chosen, and for a higher rate of profit the initial technique of higher capital intensity may be chosen again. We observe that a capital-intensive technique may be chosen for both low and high rates of profit, a result that runs contrary to the neoclassical theory of value and income distribution. Under these circumstances we cannot determine a well-behaved demand for capital schedule and so the whole neoclassical theoretical construction is under question. Samuelson (1966) despite the intellectual struggle that he gave admitted, at the end, the shortcomings of neoclassical theory. For more details about the capital theory critique, we recommend the entries of New Palgrave volume on the *Theory of Capital* (1990) edited by Eatwell *et al*. (1990), where one can find articles of those engaged in this long debate starting with the first article by Garegnani and continuing with a host of other authors contributors.

(continued)

To the reader who is interested in a critique of capital theory from a different perspective we highly recommend the article by Shaikh in the same volume. An early version of Shaikh's (1990) article appeared in the *Review of Economics and Statistics* (1974). In this article (in the form of note) Shaikh argued that technical progress as measured by the usual production function is the result of an econometric specification of an identity. This critique was levelled against the Cobb-Douglas production function and Solow's (1957) famous article. Solow's (1974) lukewarm reply was in the same journal. Shaikh grappled again with the issue of the production function and the measurement of technical progress in 2003. This time however his critique was not restricted to the Cobb-Douglas function but was generalised to all production functions and furthermore he developed a measure of technical progress along the classical model. The articles by Felipe and Fisher (2003), Felipe and McCombie (2001, 2005, 2006) are along the same lines and shed new light to the issue at hand. This line of critique was not pursued by the economists of the classical tradition to the extent that one would expect, given that it directly addresses a very popular and, at the same time, sensitive issue of the neoclassical analysis, *i.e.*, the production functions.

The critique of capital theory had some implicit hints against the coherence of Marxian theory, which soon became explicit and were expressed by Steedman (1977). The idea is that if it is true that a change in income distribution results in changes in the capital intensities of industries then Marx's analysis of transfers of value from labour intensive sectors to capital intensive sectors as well as his explanation of the relation between labour values and prices of production become questionable. This criticism would be valid in all respects if it was not that the assumptions underlying the re-re-switching are unlikely to appear in real economies. Empirical research so far has shown that the form of the wage–profit rate curves is almost (quasi-) linear, something which has been shown for the economies of West Germany (Krelle 1979), USA (Ochoa 1984; Leontief 1986; Shaikh 1998), Brazil (Da Silva 1990), Greece (Tsoulfidis and Maniatis 2002), South Korea (Tsoulfidis and Rieu 2006). An objection that could be raised to these experiments is that one is interested not just for the selected technique(s) but rather all the available techniques at a specific point in time (Kurz and Salvadori 1995, p. 450). Such a grand scale experiment was carried out by Han and Schefold (2006) using data of four benchmark input output tables for each of nine OECD countries. A total of 36 input-output tables was tested and re-switching was found only in one case and only 3.65% of the tested cases displayed non-monotonic (perverse) behaviour. Thus, re-switching of techniques is mathematically possible; it nevertheless does not arise so frequently empirically. In fact, there is some mathematical evidence that the properties of the actual input-output data preclude the possibility of re-switching as a generalised phenomenon (see also Schefold 2009).

(continued)

This does not mean however that the neoclassical theory is free from internal inconsistencies. To the contrary, these are still nested in its very core. In the meantime, however, the criticism of Cambridge will continue to preoccupy historians of economic thought (not all), this does not mean that modern economists take into account the capital theory critique seriously. Thus, in the recent years the production functions continue to be used and in the new growth models production functions include besides the physical capital, the human capital which tends to displace labour out of the picture. In the recent years, we observed one case where the issue of capital critique resurfaced in a mainstream journal (see Cohen and Harcourt 2003).

Chapter 9
Between Competition and Monopoly

I believe that like the rest of us you have had your faith in supply curves shaken by Piero. But what he attacks are just the one-by-one supply curves that you regard as legitimate. His objections do not apply to the supply curve of output – but Heaven help us when he starts thinking out objections that do apply to it!
(Brief note that Joan Robinson wrote to Keynes in 1932, Keynes, CW, vol. XIII, p. 378).

9.1 Introduction

In this chapter, we discuss the basic elements of the neoclassical theory of the firm and competition. We begin with the evolution of the notion of competition as a dynamic process of rivalry of firms in their struggle for dominance and continue with the neoclassical notion of competition as an "end state" and we discuss the different types of returns to scale. Sraffa demonstrated that neither the increasing returns to scale nor the decreasing returns to scale are consistent with the assumption of perfect competition in the determination of the supply curve in the industry. The only assumption that is consistent with perfect competition is the case of constant returns to scale, which, however, leads to implausible results. Pierro Sraffa, in his articles (1925, 1926 and 1930), concluded that the way out of this conundrum is to side step perfect competition and adopt in its place the notion of monopolistic or imperfect competition. His suggestion was pursued by economists in Cambridge England (mainly J. Robinson and Richard Kahn) during the 1930s. In the same time period, in Cambridge–Massachusetts, monopolistic competition revolution (mainly E. Chamberlin, J. Bain) was in its making. These developments in both Cambridges faced criticism from the economists of Chicago University. Thus, during the 1930s, there has been a revolution in microeconomic analysis known as "imperfect competition", which was taking place, at the same time, with the macroeconomic revolution of Keynesian economics.

In this microeconomic revolution, economists were divided into two camps. The first comprised the proponents of monopolistic competition, who were arguing that

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_9, © Springer-Verlag Berlin Heidelberg 2010

the actual economy was characterised by monopolistic elements that give rise to distortions and who tried to theorise these elements and also correct them by proposing specific anti-trust and regulation policies. We shall call these economists "imperfectionists". In the second camp, there were economists mainly from the Chicago University, who claimed on both methodological and empirical grounds that there is no such a thing as "monopolistic" or "oligopoloistic" competition and that the actual economic life is not in an empirically significant deviation from the ideal model of perfect competition. Naturally, this camp of economists may be called "perfectionists".[1] In the ensuing debates, the "perfectionists" view dominated the "imperfectionists". Fierce as it may have been, from the debate between the economists in the two camps we recognise that, at the end, they both understood the importance of perfect competition. The imperfectionists used the perfect competition concept as a yardstick to gauge the extent to which real economic life differs from the perfectly competitive state, while the perfectionists argued that there are no significant differences between the actual and the perfectly competitive economy.

It is ironic that this process of return to perfect competition began initially as an attempt to escape from perfect competition through the introduction of realistic elements in the economic analysis of the firm. These efforts led to the development of industrial organisation, as an entirely new field of economic research, and to regulation policies that regarded the various market forms as deviations from an ideal model of the perfectly competitive economy, which should be the prototype of actual economic life.

9.2 Neoclassical Theory and Perfect Competition

The analysis of competition in the neoclassical theory is contained in the model of perfect competition, which describes the ideal conditions that must hold in the market so as to ensure the existence of perfectly competitive behaviour from the typical firm and, by extension, the characterisation of the market or industry as competitive or not. The model of perfect competition describes a market form that consists of a large number of small – relative to the size of the market – buyers and small producers selling a homogeneous commodity. Both buyers and sellers have perfect information on the prices and the costs of each good. Moreover, there is a perfect mobility of the factors of production. The result of the above conditions is that the producers and consumers – because of their large number and small size – are incapable of influencing the price of the product. As a consequence, the price of the product becomes a datum and the behaviour of the firms becomes completely

[1]Nowdays the perfectionists in the USA are also called purists, known as "freshwater" economists because of the lakeside universities where they happened to congregate, while the impefectionists that come from coastal universities are also known as "saltwater" pragmatists.

passive, that is, firms display a price taking behaviour considering only the optimal quantity that they will produce. The criterion is the maximisation of profit, which is achieved, when the selling price of the good is equal to its marginal cost of production.

The intensity of competition is directly proportional to the number of producers and, in general, the structure of an industry. In this "quantitative notion of competition", the firm is conceived as the legal entity that hires the services of the factors of production and combines them to supply goods in the market. It is important to note that the firm does not own any factors of production; it merely hires the services of the factors of production that are offered by their owners, that is, the individuals. The larger the number of firms that operate in an industry the more vigorous is their competitive behaviour and, by extension, we have the establishment of a uniform rate of profit across industries. By contrast, the smaller is the number of firms, the more oligopolistic and monopolistic is the behaviour of the firm in the market and the higher are the inter-industry profit rate differentials.[2] In this non-competitive state of equilibrium, some prices are above the marginal cost, so society as a whole suffers losses from the underproduction and the underutilisation of disposable productive resources. In the neoclassical microeconomic theory, if the firm or the industry displays profits above the normal, for a fairly long period of time, they are attributed to imperfections in the operation of the market and thus in the existence of some degree of monopoly.

We say that firms in perfect competition are price takers, but at the level of general equilibrium, we want to determine the prices that change as a result of the action of some firms. The question, however, is if each and every firm is a price taker, then how do prices change? The usual answer is that prices change exogenously; for example, consumers' preferences change, which lead to the increase (or decrease) in demand. In other words, if there is a deficit (or surplus) of the output produced, it is equivalent to saying that all firms face a negatively sloped demand curve meaning that firms in and of themselves cannot increase their price without reducing their market share. In other words, firms in this case operate as if they were in conditions of monopolistic competition. As a consequence, perfect competition exists only in the conditions of equilibrium. It is important to stress that perfect competition is a mathematical assumption imposed by neoclassical economics to determine equilibrium and not a market form that arises from historical observation of the way in which firms are organised and compete with each other.

Similar conclusions are drawn from Walras's conception of attainment of equilibrium through the mediation of the auctioneer. We know that the participants in this model act independently of each other and simply react to the prices announced by the auctioneer, who is supposed to know all the facts. Clearly, if the participants in the Walrasian model act differently, then the attainment of equilibrium is problematic. As a consequence, perfect competition is a *sine qua*

[2]In the limit there is only one seller or buyer, a fact that signifies the total absence of competition.

non assumption in both Marshallian and Walrasian models of equilibrium. One corollary of the above is that some theories of competition, which were developed in the past, were eventually rejected not for their lack of realism, but precisely because they were out of the analytical framework of neoclassical economics, which is oriented towards equilibrium.

In neoclassical economics, competition is defined by the way in which technology is being used. More specifically, competition secures that the agents of production (that is firms) will tend to choose the lowest unit cost and price to maximise their profits and reduce the market share of their competitors. Thus, competition will combine technology with the behaviour of the firms in the market. Unlike classical, neoclassical economists view production not as a process but rather as a result derived from a functional relationship between inputs and outputs. The production functions are assumed to be continuous and differentiable up to the desired degree. The techniques that are used in production are usually assumed as continuous; nevertheless, the neoclassical analysis is not affected, if we have fixed input–output coefficients and L-shaped isoquant curves. Thus, the production functions in neoclassical analysis may take on various forms, such as fixed proportions or the direct opposite of it which is that of perfect substitutability between factors. The assumption of substitutability between inputs is represented with the aid of a concave production function. The proportions between inputs are convex for every single combination of inputs. Hence, we have the isoquant curves, discussed in the previous chapter, according to which a given level of output can be produced by a variety of input combinations. The curves that we derive are convex to the origin as shown in Fig. 9.1. The negative slope of the isoquant curves represents the diminishing marginal rate of substitution of one factor of production from the other. The isoquants cover the positive quantrant, exactly as in the case of indifference curves, with the difference that the isoquants are measurable, that is, they are amenable of absolute, not only relative, measurement.

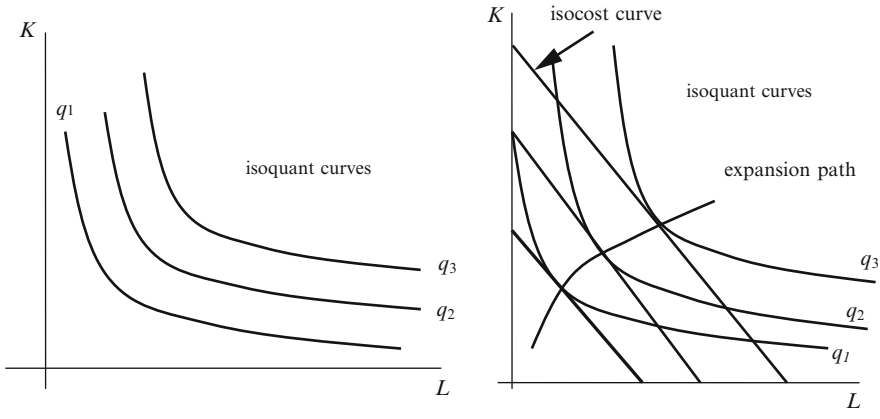

Fig. 9.1 Isoquant curves

As in the case of consumer behaviour, where choices are made at the point of tangency of the highest attainable indifference curve to the income constraint, so in the case of production, the producer chooses the combinations of capital and labour to the point where the isoquant is tangent to the isocost curve, that is the curve $C = rK + wL$, where r and w are the rewards of the services of capital (K) and labour (L) respectively, and C is the total cost of production. By using the different isocost curves, we can form the expansion path that connects all the points of tangency of isoquants and isocost curves and represents, therefore, the optimal technique in use, that is, the technique with the minimal cost of production in the case of the different proportions of inputs.

From the above it becomes clear that the givens of the neoclassical theory, that is, the preferences of individuals, the endowments as well as the technology, when combined, impose a type of competition which cannot be different from perfect competition. Firms, that is, the carriers of choice of technique maximise their profits at the point where the value of the marginal product of each and every factor of production is equal to its price. The issue that we will deal with is the level and the composition of output of a firm as well as the method of production. The analysis of the firm bears many similarities with that of the consumer. For example, the isocost curves correspond to the income constraint and the isoquants to the indifference curves.

There are two major differences, between the pure exchange model and that of production. The first is that individuals and not firms own the available resources (endowment). Firms simply hire the services of the factors of production owned by the individuals and through the production process transform them into commodities. The second difference is that the isoquants, unlike isoutility (indifference) curves, are objective, that is, isoquants depend on the level of technology; technology is not about a free choice (as in the case of individuals) but rather is imposed upon the firms through competition.

9.3 Economies of Scale

The role of the firm in the neoclassical theory of production is that of the organisation of production process through the hiring of the services of the means of production (which are owned by individuals) and transform them into goods and services and subsequently sell them in the market. In other words, firms organise a process according to which the demands of individuals for goods and services are transformed to respective supplies of goods and services. Firms are viewed as price takers and do not know *a priori* the price at which they are going to sell their products. The size of the firm is directly proportional to its market share, and therefore, returns to scale are particularly important in determining the level of production of a firm.

It is worth mentioning that the concept of economies of scale as it develops within the neoclassical theory and especially in Marshall (1890, Chaps. 9–13) is

static, that is, it does not arise over time, but rather at a particular moment in time. More specifically, one estimates the level of output in each increase in inputs and according to the answer; the economies of scale are distinguished to the following three categories:

- Increasing returns to scale arise, when inputs are doubled and output increases by more than double.
- Decreasing returns to scale arise, when inputs are doubled and output increases by less than double.
- Constant returns to scale arise when inputs are doubled and output doubles as well.

It is important to stress that the returns to scale imply a change in inputs and a subsequent change in output. In this sense, in the neoclassical analysis the returns to scale are derived from a unified analysis of cost. This is quite a different derivation of the returns to scale of the classical economists, whose analysis is dynamic, and therefore the variables involved are dated and evolve during time. Thus, the case of increasing returns to scale is described in Smith's famous exemplar of a pin factory. The difference from the Marshallian and, by extension, neoclassical analysis is found in which Smith's economies of scale have a dynamic dimension resulting from the division of labour, which in turn depends on the growth process of the total economy and not on the individual initiatives that are assumed at the level of production units or even at the level of industry. In other words, for the classical economists, economies of scale can only be dynamic and particularly in Smith economies of scale in industry are only increasing.

Decreasing returns to scale in the classical analysis are associated with the theory of rent. For example, Ricardo refers to the law of diminishing productivity of land, a law which is the result of the rising population and the subsequent rising demand for food that forces the cultivation of less productive parcels of land, leading to a rising average cost of production. Diminishing returns to scale according to Ricardo are counteracted in part by the technological progress; nevertheless, in the long run the rise in population offsets the technological progress with the net result of the diminishing returns on land. If, however, one does not account for the technological progress and accounts only for the increase in population then we end up with diminishing returns in production, but this result is in deviation to Ricardo's dynamic analysis. Furthermore, within the static analysis the assumption of diminishing returns to scale is questionable for it presupposes that one of the factors of production is fixed. In fact, when we double the inputs, it is always possible to repeat the production process with the optimal use of resources without reducing the output produced. Consequently, when we refer to diminishing returns to scale, we essentially presuppose that one of the factors of production remains fixed, and, therefore, as the other factors increase the proportions of inputs that are used differ from the optimal. The question that comes to the fore is; why should firms produce at a range of output associated with diminishing returns when they can produce at the optimal level of output associated with constant returns to scale. In other words, there is no motive what so ever for a firm to move away from the minimum cost of

production associated with constant returns to scale and produce at a range of output associated with a higher cost of production and decreasing returns to scale.

Sraffa (1925) pointed out that increasing or decreasing returns to scale in the classical analysis are derived from quite different economic phenomena. Increasing returns, for example, are derived from the process of accumulation and technological change, associated with the division of labour and the extension of the market. Decreasing returns were derived from the limited availability of land, and were an important component of the theory of income distribution, being the foundation of the theory of rent.

The case of constant returns to scale is quite reasonable and is found quite frequently in economic analysis; for example, it is adopted by classical economists and Marx. Marshall on the other hand while he accepts whenever there is pressure on the raw materials that are being used in industry there is a tendency for rising prices, nevertheless he observes that because the cost of raw materials is only a small fraction of total cost it then follows that they cannot in and of themselves affect the scale of production. Walras in the first edition of his book (1874) also assumed fixed input coefficients and constant returns to scale. In the second edition of his book (1877) he allowed for more substitutability between inputs. Finally, the empirical research has shown that at least in manufacturing the average cost curves have a wide range of output associated with constant returns to scale.

Clearly, Marshall was worried about the case of increasing returns to scale as an assumption that does not fit to the neoclassical static paradigm and this is the main reason that he distinguishes between the economies of scale that are internal to the firm and to those internal to the industry and external to the firm.

9.4 Cost Curves

We know from introductory microeconomics that the cost curves of a firm are derived from the production function and the expansion curve (Fig. 9.1b). In the beginning the firm is producing at the falling cost part of the usual U-shaped average cost curve. The shape of these cost curves has to do with the average fixed cost which is supposed to follow a rectangular hyperbola shape which when added to the average variable cost gives rise to the typical U-shaped average cost curves. If we furthermore suppose perfect competition, the profit maximising firm for the particularly given price selects the output at the point where $P = MC$ and in the long run at the point where $P = d = AR = MR = MC = \min AC$ (see Fig. 9.2), where d is the demand curve faced by the firm and the other notation is usual.

In the short run we may have $P > P^*$, which means that firms in the industry make excess profits. The result is that firms from other industries are attracted and as the number of firms increases the supply increases and the price of the product falls. If, on the other hand, $P < P^*$, the firms realise losses and so we expect an exit of firms from the industry, a reduction in supply and an increase in price. Finally, we have the case where $P = P^*$, which gives the equilibrium, given that the firms

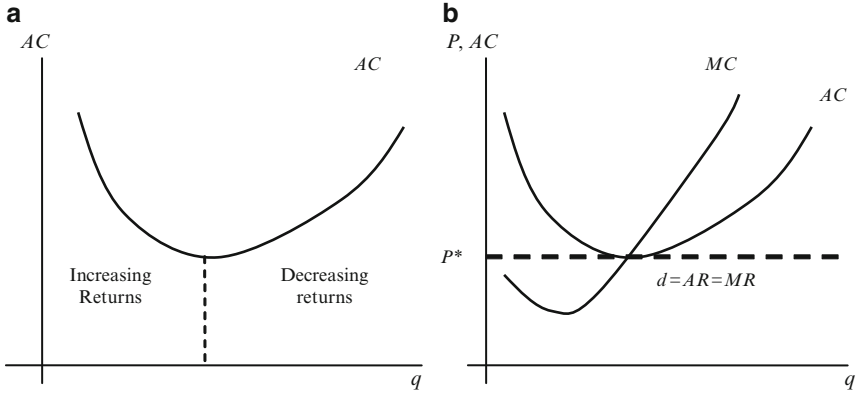

Fig. 9.2 Returns to scale and equilibrium

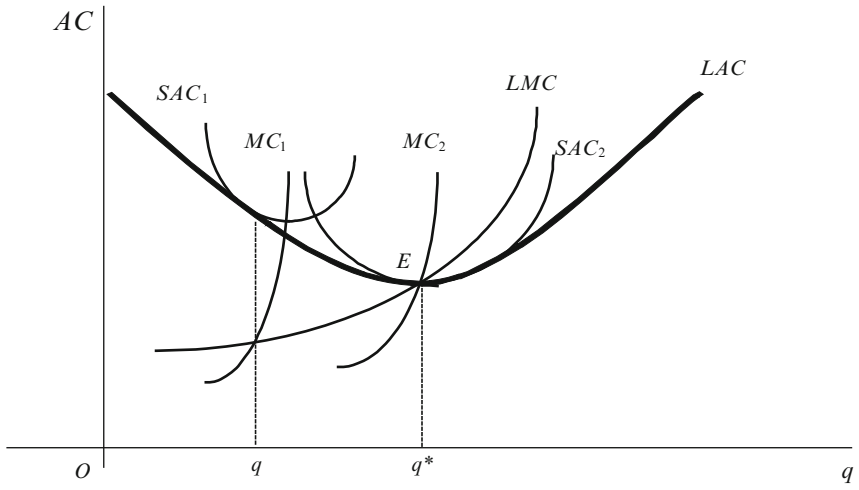

Fig. 9.3 Short run and long run cost curves

that operate simply make normal profits and there are no motives neither for entry of firms from other industries nor for exit of firms that already operate in the industry.

It is important to note that the AC curve has the same shape in both the short run and the long run (Fig. 9.3).[3] In the short run, the average cost curve of the firm is

[3]This shape of the AC curve is attributed to Viner (1931), who in the beginning envisaged the envelope curve as being tangent to the minimum point of each short run AC curve. He then asked his research assistant Voughn to draw the so described graph. Voughn realised that when the curve is going through the minimum cost of each and every average cost curve it cannot be tangent to them, whereas the envelope curve which is tangent to each short run average cost curve is not

drawn under the assumption of a fixed production capacity. In the long run the firm has the capacity to change the initial proportions between the factors of production in an effort to achieve their optimal combination. We define the long run average cost of a firm from the points of equilibrium achieved by the firm for different levels of output. We realise that the points of tangency are not the minimum points of the short run average cost (*SAC*) curves and this can be contemplated theoretically by recalling that the *SAC* are constructed under the assumption of no optimum use of the available inputs at each output level. In the long run, however, this optimal combination is achieved for the given output. Point *E* is the minimum cost, which, nevertheless, is the highest point from this which is achieved in the long run if all the productive factors are used optimally. Hence, we have the well known envelope curve which is attributed to Viner (1931), that is, the long run average cost curve (LAC) is a frontier or an envelope for the short run cost curves. The *LAC* curve owes its shape to the succession of increasing returns to scale, to the point of constant returns to scale (corresponding to the optimal firm size) and past this point, to diminishing returns to scale. The plausible question is why this optimal size is not reproduced as the scale of production increases, given that in the long run there is no fixed cost to prevent this from happening. The usual answer is that there are diminishing returns to the entrepreneurship, each firm is run by a president and as the size of the firm increases it becomes more and more difficult for the same person to run effectively the firm.

Let us refer to the long run position of the economy where point *E* indicates the optimal combination of all inputs. The size of the firm is determined from the minimum point of the average cost curve which is associated with a given level of production. We claim that the supply curve of the industry is the sum of the supply curves of the firms that form the industry. In other words, the supply curve of the industry is equal to the sum of the marginal cost curves of the firms for levels of output past the minimum point of the average cost curve. A precondition of the above is that we know the exact position of equilibrium of each and every firm, which is characterised as a relation between increasing and decreasing returns to scale.

John Clapham, an economic historian at Cambridge, found the discussion on economies of scale less than satisfactory for he thought there is distance between the theoretical discussion and the economic reality. His article of "empty economic boxes" impressed the economists of the time because in that he pointed out the distance that separates Marshall's theoretical discussion on the economies of scale and the well-known shape of the average cost curve and the difficulties of economists in using these ideas in empirical research. More specifically, he argued that we cannot know what percentage of the performance of a firm is attributed to the economies of scale and what percentage to innovations (Clapham 1922, p. 129). Simply put, Clapham essentially claimed that economists could not ascertain the

tangential to the point that represents the minimum (short run) cost of production (Chamberlin 1933, p. 235).

type of economies to scale. For this reason he characterised the economic theories that could not be demonstrated empirically as "empty economic boxes". Since we cannot discern the type of economies of scale and thus their characterisation is an extremely difficult or even an impossible task, then, following this theoretical deficiency, some plausible questions follow; as for example, what kind of measures should governments follow in designing their policies with respect to taxation or the provision of subsidies and incentives in general as components of an economic policy.

In the ensuing debates, it was argued that the incongruence between Marshall's theory of variable returns to scale and empirical observation is solely attributable to the undeveloped nature of statistical analysis and not to any weakness of the theory. We could say that this is the usual response that one gets by applying an empirical critique, which in and of itself could not overturn or create a significant theory. Empirical critique, as it repeatedly has been pointed out, can, at best, ascertain correlations between the variables and not verify causal relations, that is, it cannot derive theoretical relationships between the variables at hand. This does not mean that the empirical critique is redundant. On the contrary, the empirical critique may enhance our understanding of the underlying relationships between the variables and to reveal relationships hitherto unknown.

9.5 Sraffa's Critique of the Marshallian Theory of the Firm

Sraffa's criticism focused on Marshall's hypothesis of returns to scale in production and the assumptions of the competitive firm. The assumption of increasing returns to scale for a large range of output implies that the average cost curve of the firm displays negative slope over a large part of its range and that the marginal cost curve lies always beneath it. Two are the reasons for the decreasing average cost; the first is related to the average fixed cost of the firm which, naturally, as the output expands decreases asymptotically, and thereby, since average fixed cost is a part of average total cost, the total average cost curve tend towards a negative shape. The second reason has to do with the more efficient use of the resources. Between the two reasons only the second is associated with a diminishing marginal cost, whereas the first reason leaves the marginal cost unaffected. With this description of the cost structure, if we assume the case of increasing returns to scale, which are internal for the perfectly competitive firm, then there will be a continuous pressure on the (perfectly) competitive firm to expand its size until its absolute dominance in the market.[4]

[4]Marshall's view is that the life of the firm is intrinsically connected to the active life of the entrepreneur which is not long enough to allow the firm to dominate absolutely in the market! (Marshall 1890, p. 317).

In particular, Sraffa argued that in the case of increasing returns to scale, which are internal to the firm, there would be a continuous motive by the firm to expand its production until it can supply the whole market. Clearly, such a hypothesis of returns to scale *prima facie* contradicts the notion of perfect competition for it leads to monopoly. Marshall had also noticed this inconsistency, for example, the case of increasing returns internal to the firm that lead to monopoly was detailed in (1920, p. 666, n. 3). He credited this idea to Cournot and, as an act of intellectual honesty, characterised the increasing returns case as "Cournot's dilemma" (Marshall 1890, p. 380, n. 1). This is the reason why Sraffa pointed out that the case of increasing returns to scale "was entirely abandoned, as it was seen to be incompatible with competitive conditions" (Sraffa 1926, pp. 537–538).[5] The only case of increasing economies of scale which is consistent with the requirements of perfect competition is when these economies of scale are external to the firm and internal to the industry, a case, however, which is rarely met in real economies (Sraffa 1926, p. 540). Furthermore, this type of returns to scale cannot be limited to a single industry, and sooner or later its effects are diffused throughout the economy. The problem in this case is that the Marshallian partial equilibrium framework is inadequate to deal with the complexities emanating from the subsequent development of strong interactions between industries (Sraffa 1926, pp. 538–539).

The same is true *a fortiori* with the economies of scale which are external to the firm and to the industry, since the interactions across industries are expected to be much stronger, and, therefore, reinforcing the case for abandoning the analysis of partial equilibrium. Turning to the diminishing returns to scale and perfect competition, it follows that since firms buy their inputs in competitive markets they face no restrictions whatsoever in the quantities that they buy, and, therefore, there is no reason for the increasing part of the usual U-shaped average cost curves. Hence, the structure of the theory of perfect competition does not allow for the case of increasing cost, as the scale of production increases, simply because there is no mechanism to force firms to abandon the minimum cost of production and move to higher cost of production.

Consequently, the only assumption that remains is that of constant returns to scale, which give rise to the constant part of the average cost curves (Sraffa 1926, p. 540). Thus, Sraffa through a critique of the Marshallian theory of the firm was led to a description of the average cost (graphically presented as a line parallel to the horizontal axis) similar to that of the classical economists. This is the reason why he notes:

> In normal cases the cost of production of commodities produced competitively [. . .] must be regarded as constant in respect of small variations in the quantity produced. And so, as a simple way of approaching the problem of competitive value, the old and now obsolete theory which makes it dependent on the cost of production alone appears to hold its ground as the best available (Sraffa 1926, pp. 540–541).

[5]It is interesting to note that the discussion of the economies of scale and the perfectly competitive firm was totally dismissed by Stigler (1937, p. 708) as "too vague to be meaningful at present".

Hence, Sraffa endorses the theory of value of classical economists, where the price is determined by the cost of production, and not by the intersection of demand and supply curves. More specifically, in the case of perfect competition since the average and marginal cost curves will be identical to each other and since, in equilibrium, the given price (the demand curve) will coincide with the marginal cost (or supply) curve, it follows that equilibrium is not determined uniquely and so the size of the firm is indeterminate.

There are two alternatives out of this conundrum; first, abandon partial equilibrium analysis and adopt the general equilibrium; second, abandon the perfect competition model and adopt monopolistic competition. The first alternative is the best but it is extremely difficult to pursue in any satisfactory way

> [T]he conditions of simultaneous equilibrium in numerous industries: a well-known conception, whose complexity, however, prevents it from bearing fruit, at least in the present state of our knowledge, which does not permit of even much simpler schemata being applied to the study of real conditions. (Sraffa 1926, p. 542)

Sraffa concluded that the second alternative, that is, the imperfect (or monopolistic) competition model, might offer a simple and, at the same time, viable solution. In this second alternative, while it is possible to maintain the partial equilibrium framework together with the large number of participants whose products are differentiated, at least, in the eyes of consumers (Sraffa 1926, p. 542).

Consumers' preferences do not easily change because they are determined by factors, such as the marketing of the product, the personal acquaintance and the loyalty of customers to a specific firm that lasts for long. Thus, he proposed the replacement of the assumption of perfect competition by that of monopoly:

> It is necessary, therefore, to abandon the path of free competition and turn in the opposite direction, namely, towards monopoly (Sraffa 1926, p. 542).

In short, the theory of firm cannot be built on the assumption of perfect competition, because in actual competition firms cannot sell any quantity they produce at a given price. The production is not limited by cost, but rather by demand.

The initial reaction of neoclassical economists was to assume certain fixed characteristics in the operation of the firm that give rise to diminishing returns to scale. Thus, they argued that entrepreneurship is a characteristic which does not increase with the size of the firm and so there will be diminishing returns to this factor of production.[6] The logical consequence of this argument according to Kaldor is that we are led to the idea that the optimal size of the firm is determined by the working time of the entrepreneur, in other words we have one entrepreneur firms. Another way to address Sraffa's critique was to assume general equilibrium where entrepreneurial talents not only are unequally distributed in the economy but moreover there is a fixed supply of them which is equivalent to saying that there are diminishing returns to this factor of production. For this case Kaldor's (1935)

[6]According to Kaldor the same firm cannot be ran by two entrepreneurs Kaldor (I, 1935, p. 41).

counterargument was that the entrepreneurial abilities are required only in the initial stage of productive activity of the firm. Once general equilibrium is achieved then there is no longer need for the entrepreneurial abilities because simply the optimal production process is repeated from the less talented businessmen. Consequently, the entrepreneur with special talents is needed only in the case where the firm is out of equilibrium. From the moment that equilibrium is achieved then there is no role for the entrepreneur because past a point his abilities are transmitted to the lower echelon of the firm. Clearly, these efforts on the part of the neoclassical economists to save the Marshallian theory of the firm were not convincing.

Another effort to rescue the neoclassical theory of the firm was undertaken six decades later by Samuelson (1990). His argument was based on the idea that once we assume general equilibrium and perfect competition some resources are fixed and so the increase in production of a good may imply the decrease of production of the other good and so we are led to diminishing returns (Samuelson 1990, p. 269). The trouble with this view, however, is that Sraffa's analysis is focused on the level of industry and criticises the method with which one may construct the supply curve of each industry assuming perfect competition (Eatwell 1990, p. 281). Thus, general equilibrium is out of the scope of Sraffa's analysis.

9.6 Model Differentiation: Robinson Vs. Chamberlin[7]

Up until now we showed that Sraffa's critique was about the various types of returns to scale and the assumptions of the perfectly competitive firm. Sraffa's contribution was not so much about the increasing returns to scale, which are internal to the firm, but rather about the strongest cases of diminishing and constant returns to scale. As for the diminishing returns to scale, he argued that they were only possible if the firm drifted further away from the optimal combination of resources and there was no particular reason in a perfectly competitive environment for firms to abandon such an optimal position, *i.e.*, to move away from the minimum cost to a higher cost of production, unless we assume a fixed factor of production, an assumption which is inconsistent with the notion of perfect competition and also partial equilibrium analysis.[8] Consequently, only the case of constant returns to scale was found to be "consistent" with the requirements of perfect competition and partial equilibrium analysis. In this case, however, the marginal cost curve would coincide with the average cost curve and so for a given price, or what amounts to the same thing, a horizontal demand curve, it is impossible to determine the precise size of the firm and its supply decisions. Furthermore, consistent, results may be

[7]This section and the next are based on Tsoulfidis (2009).

[8]In fact, we know from the works of Penrose (1959) and Chandler (1977) that even for management, which is usually invoked as the fixed factor, there is not any particular reason that it cannot efficiently adjust to a larger scale of output.

plausible in the case of increasing returns to scale internal (or external) to the industry and external (external) to the firm, two cases which are rarely met in reality. In such unlikely situations, however, Sraffa argued that the partial equilibrium framework is inadequate to capture the possible complexities that are being developed and the general (not the partial) equilibrium analysis becomes appropriate to deal with the strong interactions that are expected to be developed between industries.

Sraffa concluded that a simple and, at the same time, viable solution to the logical inconsistencies of the perfectly competitive model in the case of increasing returns to scale might be the development of the imperfect (or monopolistic) competition model. The idea is that in this model one maintains the hypothesis of the large number of firms together with the partial equilibrium analysis and the difference from perfect competition is that the product is differentiated, at least, in the eyes of consumers. In short, the theory of the firm cannot be built on the assumption of perfect competition, because in actual competition firms cannot sell any quantity they produce at a given price. In the real world, production is not limited by cost, but rather by the downward-sloping demand curve.

Sraffa's suggestion to abandon perfect competition inspired the development of imperfect competition in Cambridge UK by Joan Robinson, who misses no opportunity to admit her intellectual debt to Sraffa's contribution. Chamberlin, on the other hand, claims that he was actually the first that formulated the analysis of monopolistic competition in his doctoral thesis that he defended in 1927 and published 6 years later (Chamberlin [1933], 1962, p. 5, n. 4).[9] Consequently, Chamberlin contends that his analysis not only was independent from Robinson's but moreover had no connection whatsoever to Sraffa's 1925 and 1926 articles and the pertinent literature. Furthermore, he claimed that his conceptualisation of monopolistic competition and the associated idea of product differentiation stems from Frank William Taussig (1859–1940) in his debate with Arthur Cecil Pigou (1877–1959) over railway rates differentiation.[10] As a result, he argues that the sources of his inspiration are quite different from those of Robinson and so people mistakenly identify his "monopolistic competition" with Robinson's "imperfect competition". He concedes though that this identification has been so much established in the literature that perhaps it is futile to make any effort to change it (Chamberlin 1982, p. vii). The truth however is that Sraffa had published the essential points of his 1926 article a year earlier in Italian, and that his ideas had been around for some years. Sraffa's publication in 1926 became possible after the advice and encouragement of Edgeworth, who had read the 1925 article in Italian

[9]Schumpeter (1954, pp. 1150–1151) mentions that in his private communication, Chamberlin claimed that the subject of his dissertation was already in his mind in the year 1921, while a student at the University of Michigan.

[10]Reinwald (1977, p. 530), however, notes that Taussig's product differentiation was the result of joint production which was not associated with Chamberlin's differentiation of the same commodity and that the debate took place almost 20 years earlier to have any effect in the late 1920s or 1930s.

and certainly was in close contact with Allyn Young (1876–1929), the supervisor of Chamberlin's thesis at Harvard. Young, on his part, was well informed about the ideas circulating on both sides of the Atlantic and besides Edgeworth he was also in contact with many of the other renowned economists of the London School of Economics, where he taught during the short period 1927–1928. Kaldor mentions that Young succeeded Cannan in the London School of Economics and taught there until his sudden death in March 1929. Thus, Chamberlin may be right, when he claims that his contribution was independent; however, we are allowed to speculate that his ideas were not independent at all of the intellectual milieu on both sides of the Atlantic, although he was not fully aware of the details of these developments in economic theory (Kaldor 1980, p. xii).[11] It is interesting to note that Chamberlin acknowledges his intellectual debt to Young as "he encouraged with a lively interest in the project as it developed".[12] Furthermore, it has been repeatedly ascertained that great discoveries in the history of sciences may occur at approximately the same time. After all, scientific research is not carried out by a single researcher in total isolation, but many people in various places in the world may grapple with the same questions, and, therefore, it may come as no surprise that ideas disseminate among researchers; notwithstanding, that they themselves may not know the ramifications and exact routes of these ideas. In this sense, we fully share Samuelson's ([1967], 1986) view, which deserves to be quoted in full:

> Although we have abundant evidence, after 1933 as well as before, that Edward Chamberlin was a lone-wolf scholar with infinite capacity for formulating and pushing a problem to solution in his own way, still, no man is an island unto himself. If A has any sort of communication with B who has any communication with C, [. . .], there is no way to rule out mutual interaction between A and Z even if they have never met or had any direct contact. (Samuelson, *Collected Papers III*, [1967], 1986, p. 19)

Robinson's version of imperfect competition unquestionably was created as a solution to the conundrum propounded by Sraffa. In fact, we know that Richard Kahn in his dissertation in 1932 had already developed some ideas on monopolistic competition that Sraffa had sketched out in his 1926 article. In the same time period, Robinson managed to integrate some of Kahn's arguments with regard to the demand side of the market with the cost analysis of the time to a single theory of imperfect competition.[13] More specifically, in this analysis rising costs were excluded by the formal conditions of perfectly competitive firms and given the partial equilibrium setting, the only viable and immediate solution was a downward-sloping demand curve for the industry and the firms within the industry. Robinson's analysis of the imperfectly competitive firm was carried out on strict

[11]The source of our information about the intellectual atmosphere in those years is Nicholas Kaldor, Young's student in the period 1927–1928 (Kaldor 1935, pp. viii and xii).

[12]It is clear from the introduction of Chamberlin ([1933], 1962, iv), that this quotation does not look like a usual generous acknowledgment of supervisee to his supervisor, but is rather a laconic but at the same time, revealing description of Young's involvement in the dissertation (see also Blitch 1983).

[13]Robinson recognised her intellectual debt to Kahn (see Marcuzzo 2001).

neoclassical principles, inasmuch as she used the exact same tools of the perfectly competitive firm. Consequently, her approach was an extension and further elaboration of Marshall's *Principles* and the neoclassical tradition in general. Robinson advanced her analysis to new areas of inquiry and to new issues such as the price discriminating monopoly that constitutes, even today, a standard topic in the economics of industrial organisation and the subsequent antitrust legislation.[14] She also arrived at radical conclusions regarding the presence of excess profits and capacity, and she developed the notion of labour exploitation based on application of the principles of marginal analysis. Her blunt marginal approach and the clarity with which she presented her views, soon had established her book as the basic text of microeconomic analysis for many decades not only in England but also in the USA. There is no doubt that Robinson has a theoretical starting point absolutely loyal to the Marshallian tradition and that her conclusions follow directly from a strict application of marginal analysis. More specifically, Robinson makes a clear distinction between industry and firm; thereby, couching her analysis in a partial equilibrium framework. Furthermore, she brings to the fore the industry demand curve and the associated with it marginal revenue curve. In fact, Robinson resurrected the marginal revenue and the marginal cost concepts that were laid dormant since the time of the French engineers (mainly Antoine-Augustine Cournot and Jules Dupuit). We know that Marshall used the total revenue and cost curves and his analysis was often vague and pedagogically difficult to follow. All these changed with Robinson's contributions that explicated the exact relationship between the average and marginal magnitudes and defined the point of optimisation by the intersection of the MR and MC curves. Her models became part of the standard microeconomic apparatus and are reproduced in modern microeconomic textbooks. In what follows (Fig. 9.4, below) we present, for comparison purposes, her model of imperfect competition:

On the left hand side graph of Fig. 9.4, in the short run, the downward-sloping demand curve and the U-shaped average cost curve are put together along with their respective marginal curves that determine the monopolistic equilibrium output (Q_m) and, through the demand curve, the respective equilibrium price (P_m).[15] In this case, we have excess profits equal to the shaded rectangular area shown on the left hand side of Fig. 9.4. In the long-run, the inflow of firms attracted by excess profits reduces the demand curve for each individual firm to the point that it becomes tangent to the AC curve and at the same time the new MR intersects the MC curve determining the long-run equilibrium pair of quantity (Q^*) and subsequently the equilibrium price (P^*). In this long-run equilibrium, we have $P^* = \text{AC} > \text{MC}$ and output produced (Q^*) falls short of the minimum AC output (Q_c) and so

[14]In fact, the Robinson-Patman Act of 1936 which was against price discrimination is based, largely, on Joan Robinson's theoretical developments of the issue (*e.g.*, Dilorenzo and High 1988).

[15]Marshall had discussed monopoly purely in terms of total sales and cost curves yielding the unique point of profit maximisation.

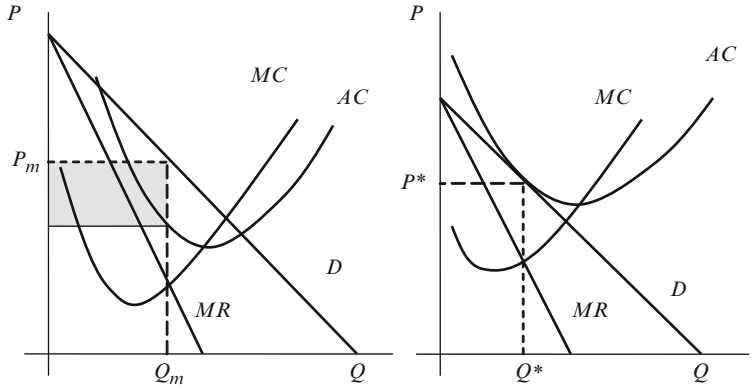

Fig. 9.4 Robinson's imperfect competition

there is excess capacity; moreover, since $P^* >$ MC, there is loss in consumer welfare.

Turning now to Chamberlin's notion of "monopolistic competition", we know that he did not take issue, at least explicitly, with Marshall's laws of return, but his theoretical model grew out from the effort to infuse realism to the established theory of perfect competition. During the period that the theory of monopolistic competition was emerging there was a tremendous development in the analysis of the firm and a variety of other characterisations of its behaviour, which can be best, called "pragmatic". In other words, in the interwar period there was an enormous upswing in pragmatism. Economists tried to incorporate into the analytical framework of the firm more practical ideas, which seemed to have been derived from the empirical characteristics of the nature of the firm, as they could be seen operating in the market. Chamberlin, in his persistent effort to inject realism into his model, did not strictly use marginal analysis (*e.g.*, [1933], 1962, pp. 191–193). In fact, Chamberlin objected to the determination of price via the equation of marginal revenue ("a joke", as he characterised it) and marginal cost because in this way one determines first the equilibrium quantity and then the price, something that for him was unrealistic, as he characteristically notes:

> A major deficiency in the marginal revenue technique is that it does not by itself reveal the price. This means that the discussion of equilibrium takes place primarily in terms of output; the category so neatly determined by the intersection of the two marginal curves, instead of in terms of price, the category with reference to which business decisions are most usually taken. (Chamberlin [1952], 1982, p. 275)

Figuratively speaking his typical analysis, where firms do not take into account the behaviour of competitors is depicted, for the sake of simplicity, in a set of two graphs displayed in Fig. 9.5, where the left hand side graph represents the short run case, where the downward-sloping demand curve (D) is put together with the average cost curve (AC) and the monopolistically competitive firm charges a

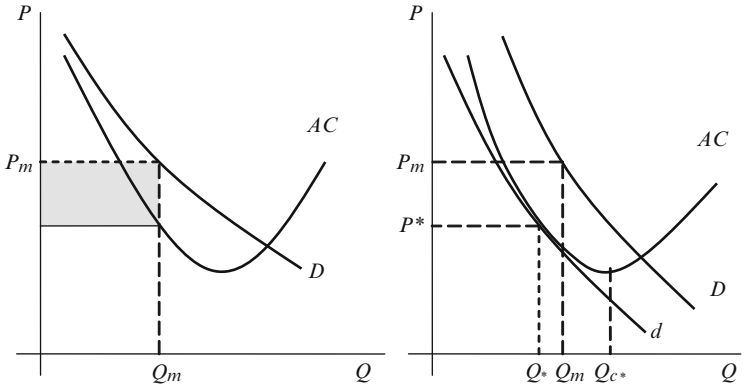

Fig. 9.5 Chamberlin's monopolistic competition I

price through a trial and error procedure (Chamberlin [1933], 1962, pp. 83–84) so as to secure a desired (maximum) amount of profits which is measured by the shaded rectangular area.[16] The excess profits, however, attract other similar firms and "since the total purchases must now be distributed among a larger number of sellers" (Chamberlin [1933], 1962, p. 84) the demand curve of each individual firm shifts inwards. The process continues until the demand curve becomes tangent to the average cost curve.

Chamberlin ([1933], 1962, p. 84) is reluctant to admit the similarity of his configuration to that of Robinson's. It is abundantly clear though that once we have the average curves the marginal ones are implicit and their equality determines the points of optimum decisions. These optimum points may differ somewhat in the short-run and perhaps Chamberlin is right to be a bit cautious about the maximisation of profits because the short-runs are fraught with uncertainties and by definition signify disequilibria situations; however, the long-run positions are identical in both Chamberlin and Robinson. This has also been pointed out by Shackle's attentive observation:

> Equilibrium of the firm is represented in Ms Robinson's language by the output at which the marginal cost curve cuts the marginal revenue curve from below: in professor's Chamberlin language, by the output at which the average cost curve has the same slope as the average revenue curve and does not lie above it [. . .]. Equilibrium of the group (the "industry") is represented in both languages by the tendency, for every firm; the average revenue and average cost curves [. . .] the equality of two functions of output and also equality of the first derivatives. (Shackle 1967, p. 63)

So the difference perhaps lies in the short-run, but on closer examination one discovers that Chamberlin's "trial and error" process realistic as it might be, if

[16]This pricing procedure has an affinity to the well established business practise known as full cost pricing, the salient feature of which is realism. In fact, full cost pricing was derived on the basis of a questionnaire addressed to the managers of manufacturing companies (Hall and Hitch 1939).

profits are to be maximised then the respective pair of price and output must be exactly the same to that determined by the intersection of MR and MC curves. The lack of these two curves in Chamberlin's analysis is what made his book less appealing to economists and the presence of these curves in Robinson's book made her ideas of optimisation much more accessible as teaching material. In fact, not only in economics, but also in other sciences optimality is obtained once the system is set in its marginal conditions. Nevertheless, Chamberlin ([1952], 1982) time again insists on the minor importance of the marginal revenue in the determination of prices:

> [. . .] my own book arose, not out of the marginal revenue curve, but out of the attempt to combine the two theories of monopoly and of competition into a single one which would come closer to explaining the real world, where, it seemed, the two forces were mingled in various ways and degrees. This idea does not appear in Mrs. Robinson's Imperfect Competition. [. . .] In my own attempt to blend monopoly and competition, the marginal revenue curve was discovered at an early stage and seen for what it is – a piece of pure technique unrelated to the central problem. (Chamberlin [1952], 1982, p. 274)

Hence, Chamberlin essentially makes an effort to get too much credit for the work and accumulation of knowledge about monopolistic competition in the decade of 1920s, if not a century earlier. Furthermore, by downplaying Robinson's "imperfect competition" he was essentially downplaying the importance of economists at Cambridge UK and their contributions to the microeconomic revolution. The truth is that Robinson with the term "imperfect competition" did not merely want to fill some gaps in the "intermediate zone" between pure monopoly and perfect competition, but rather she sought to underscore that the neoclassical theory of competition leads inescapably to conclusions completely opposite to those that it would like to derive. In other words, imperfect competition equilibrium is associated with excess capacity and also loss in consumers' welfare, since the equilibrium price exceeds the marginal cost. Furthermore, the models of "price discrimination" and "exploitation of labour" arising when the marginal (revenue) product of labour exceeds the marginal resource cost were the logical consequences of the neoclassical conceptualisation of competition and marginal productivity theory of income distribution, respectively.

The next step for Chamberlin in his quest for pragmatism and also differentiation from Robinson was his idea of two demand curves. Let us suppose that all firms in the group as well as those that may enter in the group have the same cost functions. The assumption that Chamberlin makes is that the individual demand curve is much more elastic that the demand curve of all the firms that comprise the group. Hence, the individual demand curve 'd' conveys the idea that a firm assumes that the other firms do not match its price reductions. By contrast, the demand curve D represents the share of the market curve which is drawn for the individual firm assuming that all other firms of the group match the price changes. The elasticity and exact location of the demand curve for the group depends, *ceteris paribus*, on the number of firms that comprise the group. In terms of Fig. 9.6, let us start with the left hand side graph, where the number of firms is supposed to be fixed, that is, there is no entry or exit of firms.

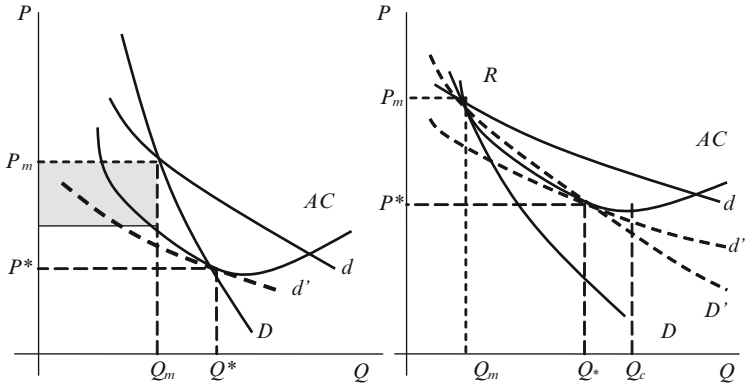

Fig. 9.6 Monopolistic competition II

Let us further suppose that the price set is P_m and each firm makes excess profits equal to the shaded rectangular area. This price, however, holds only in short-run equilibrium and each firm will have an incentive to lower its price by moving along the d curve and assuming that the other firms do not follow suit. But each firm has exactly the same incentives, which is equivalent to saying that all firms in the group cut their prices. As a result, the 'd' curve will be sliding down along the *D* curve to the point that the 'd' curve becomes tangent to the *AC* curve and also intersects with the *D* curve, and, therefore, any incentive to lower prices is eliminated.

On the right hand side graph, we allow for the number of firms to vary and, starting with a position of excess profits, as in the previous case, it follows that there is entry of firms to reap these excess profits and, as the number of firms in the group increases, it follows that the *D* curve becomes steeper to the point *R* of its tangency with the *AC* curve. Point *R*, however, is unstable, because there will always be an incentive for each individual firm to lower its prices assuming that the others do not follow suit and, once again, the d curve will be sliding down along the *D* curve until it becomes tangent to the *AC* curve. However, this tangency point is not yet an equilibrium proper, because at the price *P** the majority of firms makes losses and so they start leaving the group and in doing so the D' curve rotates to the right and in the limit it passes through the tangency point of the 'd′ and average cost curves. This is a stable equilibrium point attained in a more complex way than before (Chamberlin [1933], 1962, p. 93).

In evaluating these models one wonders how is it possible for a firm to assume consistently that its competitors will not react to a probable price change. The idea to incorporate the reaction of other firms into the analysis is a step forward in the microeconomic analysis, but to assume that firms follow a strategy that is falsified consistently is perhaps worse than assuming the independence in actions of the participating agents. And in this sense, Chamberlin did not really advance the analysis much beyond the well-known models of Cournot, Bertrand and Edgeworth. However, Chamberlin's idea of the two demand curves created an entire literature about "discontinuity" in the marginal revenue curve that leads to price

rigidities in the oligopolistic markets, whereby prices are determined by demand and supply (average cost). This approach made a lot of sense in the 1930s because it was explaining price rigidities that called forth for government intervention and also labour unions could demand higher wages without causing inflation. The idea was that the discontinuity in the MR curve allowed even substantial increases in the MC curve without affecting prices in any significant way.[17]

Apart from these differences, that is, the inclusion of strategy in the behaviour of individual firms and the two demand curves, the two economists (regardless, of what Chamberlin claims) use more or less the tools of marginal analysis, Robinson more explicitly than Chamberlin. The marginal analysis is what made Robinson's book widely accessible and established it as a textbook in microeconomics, whereby the lack of explicit marginal analysis is what made Chamberlin's book confusing and difficult although his insights about the demand side of the market were richer than those of Robinson. The MR is a concept that far from being a "joke" was also essential in Chamberlin's analysis.

9.7 The Rise and Fall of a Revolution

One of the surprising results of the analysis of monopolistic competition lies in the strengthening and also wider acceptance of perfect competition. We know that the idea of perfect competition appears in Cournot (1838), whose analysis was based on the maximising behaviour of the participating firms at the point of equality of marginal revenue and marginal cost. These concepts were also present in the writings of the other French engineers of the early nineteenth century. The often-cited didactic example of the inconsistencies that arise in the application of marginal principles has been advanced by Dupuit ([1844], 1969) and is related to the imposition of the correct price of crossing the bridge. We know that the MC of crossing the bridge is zero and so must be the optimal price (toll) of crossing the bridge. But for a price equal to zero, there is no private incentive to build bridges and a positive price (toll) on the other hand leads to resource misallocation and net welfare loss.[18] Cournot's and the French engineers' ideas, however, could not attract attention in the early nineteenth century because of the absolute dominance of classical economics and their view of competition as a process of rivalry and not as a static situation. The depression of 1873–1896 created the necessary conditions for the appearance of new ideas, and as it has been observed in dismal situations such as those of depressions, people, often, tend to distant themselves from the harsh reality and are ready to accept idealised situations. Clearly, such situations are

[17]Hence, we have the famous "kinked-demand" in oligopolistic markets developed by Paul Sweezy (1939) but also by Hall and Hitch the same year (see Appendix A9).

[18]The rule $P = MC$ was already well known but, at the same time, it was shown that it cannot be applied in cases of decreasing cost industries which necessitated government intervention.

those that are described in perfect competition and so Edgeworth (1881) found a fertile ground to promote the notion of perfect competition by developing its formal requirements.

Once again, this analysis could not gain broad acceptance not only because of its unrealistic assumptions but also because of the dominance of the ideas of classical economists. Marshall sought to circumvent the problem by assimilating the classical tradition with neoclassical economics. The classical dynamical process of competition gradually was to be translated into static terms, that is, the number of producers and the type of product may characterise the form of competition. However, even in Marshall's time, perfect competition was not fully formulated into an operational model and this job was accomplished, to a great extent, in Knight's (1921) book, which was essentially his dissertation written under Young's diligent supervision. Knight in his book described in a comprehensive and meticulous way the requirements of perfect competition that could be used in the real economy and in so doing he managed to operationalise and to popularise the concept. Nevertheless, Stigler (1957) argued that this detailed description of the requirements of perfect competition was responsible for the initial appeal of monopolistic competition in the 1930s and delayed the explicit incorporation of perfect competition in neoclassical economics. Meanwhile, the books of Chamberlin and Robinson sparked a renewed interest in the static analysis of market forms: key words such as monopoly, oligopoly, rigidity of prices, price discrimination, exploitation of labour, excess capacity and the like excited and activated the interest of economists and policy makers to eliminate these undesired features of markets. The depression of the 1930s, however, changed, once again the perception of the majority of economists about the role of these mega-corporations and there was a widespread belief that government intervention was necessary for the limitation of market power of big businesses that were also responsible, at least in part, for the depression. In fact, the usual argument (*e.g.*, Berle and Means 1932) was that prices in the US economy became increasingly stickier in the consumer goods industries due to the concentrated, and, therefore, monopolistic structure. These "sticky prices" undermined the already constrained purchasing power of consumers. The same phenomenon was also observed in the capital goods sector and so producers were less willing to invest in new plants and equipments. Price stickiness thus inhibited the recovery of both final product demand and investment demand; thereby, precipitating the depression. Naturally, such views offered the necessary economic rationale for government intervention in the markets. In fact, governments became increasingly more interested in correcting the operation of markets in the effort to bring them closer to the hypothetical perfectly competitive markets (Bishop 1964; Dilorenzo and High 1988). This is equivalent to saying that the actual markets were characterised by some degree of imperfection in their operation, and hence they were found in divergence from an ideal operation, which was identified with the notion of perfect competition.

In the interwar period, as a result of Sraffa's critique, the attention was directed to the development of the theory of monopolistic (or imperfect) competition and the suppression of perfect competition. More specifically, the theorisation of competition

in its imperfect form during the 1930s led to the development of the field of industrial organisation, which on the one hand encompassed the new theoretical refinements of the theory of the firm and forms of competition and on the other hand made an effort to give quantitative content in these forms. Meanwhile, data on prices, costs, output and concentration ratios started to be collected on an industry basis. It is important to point out that the systematic collection of such data begins at approximately the same time with the collection of national income and product accounts data compiled for the aggregate economy and macroeconomic purposes, lending further support to the idea of two parallel revolutions in economic theory that took place at approximately the same time period. We know that the Keynesian revolution continued successfully and after WWII until the late 1960s, however, we cannot say the same about the monopolistic competition revolution which did not last for long, with economists gradually rediscovering the notion of perfect competition. As Stigler notes:

> The theory of imperfect competition has raised questions which it cannot answer satisfactorily until the theory of perfect competition has been much more fully developed. [. . .] the chief work of economic theorists should for the present still be in the theory of perfect competition. (Stigler 1937, p. 707)

Meanwhile, many neoclassical economists in the USA (this time in Chicago) perceived the monopolistic competition revolution as a departure from the strict scientific analysis that economics ought to follow and, worse of all, as a critique of the actual market system which in turn necessitated government intervention in the economy. Stigler was very specific about the implications of monopolistic competition to the neoclassical theory of the firm. He described monopolistic competition and its implications in the following terms:

> The new theory, in other words, has become something of a destructing fad. It seems often to be an escape from the very hard thinking necessary to secure a satisfactory and useful theory of perfect competition. Sound theories of price and production are indispensable to the solution of even the simplest practical problems. Yet the majority of the writers on imperfect competition seem not to realize that almost all the important concepts they have taken from perfect competition are suspect. (Stigler 1937, p. 708)

Furthermore, Stigler (1937) claimed that the "newer literature of imperfect competition" is so complex that is incomprehensible for the legislator and the layman and so it is extremely difficult to find useful applications. In fact, both Stigler and Friedman systematically and forcefully opposed all efforts for further elaboration and possible improvement of the theory of monopolistic competition. An example of how much Stigler objected to monopolistic competition is his textbook in microeconomics ([1942], 1966), where there is no reference at all to Chamberlin and the notion of monopolistic competition, while Robinson is only mentioned *en passant* in the discussion of price discrimination. Stigler's opposition was based on the idea that such a direction of research in monopolistic competition would render economic analysis more case oriented, and, therefore, the lack of generalisations would make economic theory less scientific.

Friedman (1953) on the other hand, argued against monopolistic competition mainly on methodological grounds, *i.e.*, a model is judged according to its predictive content and not the realism of its assumptions.[19] In this context, he used the example of the price effects of an indirect tax imposed on cigarettes which could be predicted with sufficient accuracy using partial equilibrium analysis and perfect competition although the cigarettes industry possessed the characteristics of monopolistic or oligopolistic competition.

A characteristically different effort was that of Triffin (1941) who sought to reorient the theory of monopolistic competition away from partial equilibrium towards general equilibrium analysis. Friedman's response was immediate and he lamented that such a reorientation of the research essentially dispenses with the concept of industry and the firms that operate within it and places instead to the fore of the analysis the single firm in combination with the "whole economy collectivity". Friedman (1941) strongly opposed the general equilibrium approach; his argument was that most of the practical problems on which economists want to apply their theories are at the level of industries, not at the level of firms or of the economy as a whole. As a consequence, he argued that since industries are so important and are not accounted for in the analysis of monopolistic competition, it follows that monopolistic competition must go. Stigler ([1949], 1983) is also dismissive of Triffin's version of monopolistic competition characterising it as "ad hoc empiricism".

The empirical findings in the industrial organisation studies showed that when cross section data are used the results with respect to the profit rates of industries could go either above or below the economy's average; however, whenever economists use time series data spanning a time period long enough the results show no evidence of excess profitability across industries. These findings are consistent with the idea of the classical economists' tendential equalisation of interindustry profit rates to a general one and indicate that essentially neoclassical economists in these empirical analyses abandon the static framework of analysis and in effect use the classical disequilibrium dynamical process.[20]

Meanwhile, at the macroeconomic level, the welfare implications of monopoly for the economy as a whole were estimated to exert a negligible effect that did not exceed 1% (approximately, one-seventh of one per cent) of the GDP in the US economy (Harberger 1954; Schwartzman 1960). These results were in favour of the perfectionists, at the University of Chicago, who claimed that the actual economies do not differ in any empirically significant way from the ideal of perfect competition and thus there is no need for government intervention. The counterargument was advanced by Leibenstein (1966) who criticised the studies that sought to measure the welfare loss of monopoly on the grounds that monopoly permits

[19]On further examination, we find that the methodological rejection of imperfect competition was, in fact, first launched by Stigler (1949); however, Friedman (1953) popularised this methodological principle so much that he managed to associate it with his name.

[20]For a survey of the evidence, see Scherer and Ross (1990).

managers and workers to function at levels of productivity much lower than those in perfect competition. As a result, in monopoly we have the development of x-inefficiencies, which were not measured for example in Harberger's (1954) study. Should the x-inefficiencies be measured then the welfare loss of monopoly would be much larger than previously estimated. The rebuttal of Chicagoans was that if indeed there were x-inefficiencies due to monopoly, then profitability would decline and monopolies would not be able to attract new investment and would rather direct their own investment to other more profitable lines of business. From the above it becomes evident that the monopolistic competition revolution did not last for long and its initial outbreak and brief ascent only had as an unintentional effect the restoration of perfect competition and its placement from the fringes of economic analysis to the very core of microeconomic model-building.

9.8 Summary and Conclusions

In this chapter we examined the concept of competition of neoclassical economics and the way in which it dominated over the more realistic theory of competition of the classical economists. We argued that the hypothesis of perfect competition is a requirement of the neoclassical approach, which is oriented towards the attainment of equilibrium. In other words, perfect competition is a neoclassical theory generated concept rather than a theory generating concept; furthermore, the classical theory of competition was set aside not for its lack of realism but rather for its incompatibility with the static notion of equilibrium. We also discussed the short run and long run cost curves and the returns to scale of production as well as Clapham's empirical critique.

A different kind of critique was advanced by Sraffa who directed his efforts on the logical foundations of the neoclassical theory of the firm. The basic conclusion drawn from this critique is that if we cannot determine the precise size of the firm and its exact supply decisions, the logical foundations of the neoclassical theory of the firm are in question. Sraffa's critique of the theory of the perfectly competitive firm encouraged the research toward the development of a theory of monopolistic competition. In this theory, firms are assumed to possess some power over the market forces, and this gives rise to phenomena such as price discrimination, price rigidity, exploitation of labour, excess capacity, *etc*. The detailed analysis of these phenomena led to the development of industrial organisation as a branch of economic theory that deals with the structure of markets and the rationale of government intervention. Hence, once the market is identified and some monopolistic elements are found, the next step is to assess the quantitative significance of these elements and propose corrective economic policies. In these efforts, it became particularly important to define a precise benchmark that would inform and, at the same time, direct government corrective policies. This ideal model naturally was found in perfect competition, whose status not only was restored, but it was also elevated to a benchmark for evaluating market outcomes and to inform economists

and policy makers for the rationale and also the limits of government intervention in the markets. According to Blaug (1978, p. 417) Robinson and Chamberlin are those that initiated and propagated this revolution which preceded the Keynesian revolution. Thus, in the 1930s economic theory was split into micro and macroeconomics and the two disciplines were developed parallel to each other. The revolution in microeconomics, unlike that of macroeconomics, faced strong resistance from the beginning. More specifically, the "perfectionists", that is, the economists from Chicago University argued that the results of the operation of the actual economic system are not that far from those expected from the ideal system identified with pure or perfect competition. On the other hand, the "imperfectionists", that is, the economists from Cambridge (USA and UK) argued that the actual economic system was suffering from monopolistic characteristics that ought to be corrected by government intervention. To the extent that monopolistic competition was not a self-contained model, but rather many models of particular cases, it was relatively easy for economists of the calibre of Stigler and Friedman to prevail in these debates. From the protagonists of this revolution in monopolistic (or imperfect) competition, some (Sraffa, Robinson, Kahn, *inter alia*) lost faith in both the cause and the results of this revolution and shifted their interests to other areas of economic analysis and some others (Chamberlin, Triffin, *inter alia*), as is usual with revolutions, became part of the establishment and strove to struggle in preserving the paradigm that gave them prestige, regardless of the progress achieved.

In conclusion, the monopolistic competition revolution did not really last for long and had lost almost all of its vitality by the late 1930s or early 1940s. As a result, it was supplanted by perfect competition, for the next three decades. However, by the mid or late 1970s, a series of important events that shook the world economy – such as the two oil crises, the creation of cartels of the size and importance of OPEC, the rising inflation and the simultaneous slowdown in the level of economic activity – revealed the limitations of the paradigm of perfect competition. Economists, therefore, became more receptive to ideas such as returns to scale, uncertainty, strategic behaviour and the like that are in stark contrast to the perfectly competitive model and in the recent decades we are witnessing a second monopolistic revolution. The difference is that this time the monopolistic competition revolution does not develop as a series of unrelated models, but rather as a unifying concept that seeks to provide microfoundations to macroeconomics and in so doing to integrate micro and macroeconomics to a single economic theory. To what extent this second revolution in monopolistic competition will be successful is something that time will tell; however, the lacuna that this revolution seeks to fill is real and so the demand for theoretical advancement to a realistic direction is already established. This in turn posits a difficult dilemma to the Sraffian critique of perfect competition: will it adopt this type of monopolistically competitive behaviour, which this time is cast in a canonical model and in a general equilibrium setting, or will it criticise it in the effort to reveal its neoclassical essence and its possible internal contradictions? In the latter case, the return to classical economics and the

modelling of the classical dynamic conception of competition seems to be the preferred alley.

Questions for Thought and Discussion

1. Discuss Marshall's view of returns to scale. How does this view compare with that of classical economists?
2. Discuss Clapham's notion "of empty economic boxes"?
3. Explain the usual U-shaped average cost curves. Is the U-shaped average cost curve realistic?
4. Present Sraffa's critique of the Marshallian theory of the firm.
5. Discuss Chamberlin's claim that he is the originator of the concept of monopolistic competition and his analysis is independent of similar developments at Cambridge England.
6. What are some of the differences between Robinson's and Chamberlin's analyses of imperfect competition? Are these differences essential?
7. Discuss the evolution of the idea of perfect competition.

Notes for Further Reading

The discussion on returns to scale and the cost curves can be found in any introductory microeconomics book (*e.g.*, Ferguson and Gould 1975; Silberberg 1978). The more recent texts are characterised more for their mathematical treatment of these theories, nevertheless they do not add much in terms of essence. Sraffa's article, although it was published in the well-known and widely read *Economic Journal* (1926) and also the same article was republished in a textbook titled *Readings in Microconomic Theories* (1952), Sraffa's critique to the neoclassical theory of the firm passed unnoticed as this can be judged by the lack of rejoinders. On the other hand the discussions that took place in the pages of the *Economic Journal* in the 1930s did not really address the issues raised by Sraffa's critique. From the various presentations of Sraffa's argument, we distinguish the articles by Sylos-Labini (1990), Pasinetti (2003) and Eatwell (1989, 2008). At this point, it is important to mention the work of Allyn Young, who somehow stands behind all the important theoretical developments that took place during this period of time. Although Young did not write many articles, he nevertheless was renowned as a supervisor of some important doctoral theses; the witnesses that we have lend support to the view that his contribution to these theses were essential and not just typical. Among the dissertations that he supervised are included Frank Knight's who defended it in 1919 and published it (after modifications) as a book in 1921. In this book, we find for the first time a straightforward presentation of the requirements of perfect competition and also the beginnings of a critique that bears some similarities to that of Sraffa, which to our view deserves more attention than it has received so far. Clearly, Knight did not pursue this aspect of his work; after all he could not do it anyway given

(continued)

that he was teaching in Chicago University. Knight became known for his theories about risk and uncertainty, whereas the other aspects of his work have passed to a great extent unnoticed. Young a few years later supervised Chamberlin's dissertation in the mid 1920s which was published in 1933. Chamberlin's book is considered essential for the "microeconomic revolution" of the 1930s. For the work and life of Young we recommend the articles by Blitch (1983) and Machionatti (2003).

As we know the notion of monopolistic competition became the target of critique from a number of important economists across the Atlantic. Hicks (1939, pp. 83–85) discarded monopolistic competition because it was not cast in terms of an operational model. Ten years later George Stigler renounced the theory of monopolistic competition on methodological grounds. According to Stigler (1949) the predictions of the monopolistic competition model are approximately similar to those of perfect competition. Under these circumstances, the Occam's Razor principle dictates that one should choose the most parsimonious model and this is perfect competition. A few years later Milton Friedman (1953) based also on methodological grounds arrived at similar conclusions. Clearly, the monopolistic competition revolution of the 1930s was not as successful as the Keynesian revolution that took place about the same time. From the very beginning, monopolistic competition became the subject of criticism and if it was not for Chamberlin and the economists (Bain, Trifin) at Harvard University then the developments in the early 1930s would, perhaps, have been forgotten. Nowadays, we are witnessing a revival of Keynesian economics (see Chap. 15) that has led to a new theorisation of macroeconomics within which "microfoundations" is the key word. As a result, the monopolistic competition theory is widely accepted and this is because it is cast in terms of operational and more complete models that capture the realistic aspects of actual competition.

The History of Economic Thought books are usually very brief when it comes to the revolution in microeconomics. Schumpeter characterised Chamberlin's book "Monopolistic Competition" as "one of the most successful books of theoretical economics" (Schumpeter 1954, p. 151). Such a statement should be interpreted more like a collegiate gesture from the part of Schumpeter, who was teaching at Harvard at that time and less like an endorsement of the theory of Monopolistic Competition. We know that Schumpeter essentially developed a dynamic theory of competition inspired by the classical economists and Marx. Blaug (1984), on the other hand, accepts the notion of microeconomic revolution nevertheless he does not mention Sraffa and his contribution to this revolution. We cannot say the same thing for the book by Screpanti and Zamagni (2002) who are very detailed on this issue of Sraffa's critique and the subsequent monopolistic competition revolution.

Appendix A

Full Cost Pricing of Hall and Hitch

An altogether different direction is the research followed by Hall and Hitch (1939), who focused on two questions:

- If, and to what extent entrepreneurs in industrial firms follow the pricing rules by equating their marginal revenue to their marginal cost
- If, and to what extent the shape of the cost curves of the industrial firms looks like the usual U-shaped textbook cost curves

In a questionnaire drawn up on a large scale survey of industrial firms in England, Hall and Hitch found that the pricing rules of orthodox microeconomics was neither followed nor did the managers of the industrial enterprises know about the marginal revenue-cost pricing principle. One could legitimately argue that it does not matter what business people think that they do or they do not do but to whether or not they follow rules of behaviour anticipated by microeconomic theory. In a similar fashion, it is indifferent for the functioning of the economy, whether individuals know anything about Smith's "invisible hand". The real trouble, however, for the standard microeconomic theory is the second finding which is absolutely consistent with the first, that is, the shape of average cost curves in industrial firms is of increasing returns up to a point and then becomes constant as output increases. This shape of average cost curves is attributed to the average fixed cost, whose shape is one of the rectangular hyperbola and to a great extent determines the shape of the total average cost.

If by Q^* we represent the output that corresponds to the full utilisation of productive capacity of the firm, Hall and Hitch argued that when the demand for the product falls (or rises) then it is not the price that falls (or rises), as one would expect from the standard microeconomic theory, but rather the profit margin (see Fig. 9.7). The idea is that the firm faces many short-run changes in demand and is not necessary to respond through price changes but rather through changes in the utilisation of capacity. Hall and Hitch have concluded that businesses operate on a cost plus pricing rule according to which the price of their product is equated to the average cost of production at full utilisation of capacity plus a markup. The markup is of a relatively stable size and that all discussion of demand curves is meaningless in the business world.[21]

As for the supply curves they exist (albeit with problems, as we know from the Sraffa's critique) only for the perfectly competitive firms. For the other market forms; that is, monopoly, oligopoly, monopolistic competition and the like there are

[21]It is important to stress that this discussion mainly refers to industrial firms whose changes in capacity utilisation and stability of markups are related to their capital intensity (see also Chaps. 5 and 9).

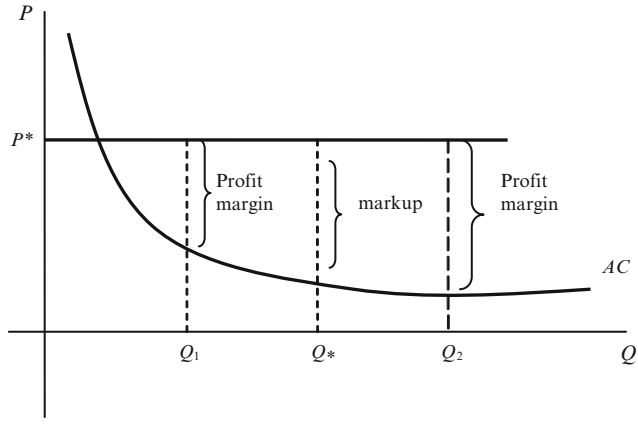

Fig. 9.7 Full cost pricing

no supply curves. For example, in the case of monopoly for a given price the monopolist does not know how much to supply, unless he knows his demand curve. In other words, the monopolist does not have a supply curve proper, and the same is true in the other market forms.

The crucial issue in Hall and Hitch is the determinants of markups, and to this question there is no easy answer because the markup does not depend on the behaviour of competitors, but rather affects the cost of production of the firm. One possible interpretation of the size of the markup refers to the Kaleckian (1969) idea of the degree of monopoly. The more monopolistic is the position of a firm the higher the markup and the price of the product. In our view, the full-cost pricing is not a theory proper but a description of business practises that seem to disregard the standard supply and demand analysis. This is the reason why that such a pricing policy is perceived by the neoclassical economics as evidence of monopolistic practises and power over the market forces. In classical analysis, however (as shown in Chaps. 5 and 6) the markup is explained endogenously (see also Semmler 1984 and Lavoie 2006, pp. 40–53).

Chapter 10
Keynes's General Theory

The outline of our theory can be expressed as follows. When employment increases aggregate real income is increased. The psychology of the community is such that when aggregate real income is increased aggregate consumption is increased, but not by so much as income. Hence employers would make a loss if the whole of the increased employment were to be devoted to satisfying the increased demand for immediate consumption. Thus, to justify any given amount of employment there must be an amount of current investment sufficient to absorb the excess of total output over what the community chooses to consume when employment is at the given level. For unless there is this amount of investment, the receipts of the entrepreneurs will be less than is required to induce them to offer the given amount of employment. It follows, therefore, that, given what we shall call the community's propensity to consume, the equilibrium level of employment, i.e., the level at which there is no inducement to employers as a whole either to expand or to contract employment, will depend on the amount of current investment. The amount of current investment will depend, in turn, on what we shall call the inducement to invest; and the inducement to invest will be found to depend on the relation between the schedule of the marginal efficiency of capital and the complex of rates of interest on loans of various maturities and risks.

(Keynes, *General Theory*, p. 25)

Many people are trying to solve the problem of unemployment with a theory which is based on the assumption that there is no unemployment.

(Keynes, *Means to Prosperity*, p. 305, CW, IX)

I sympathise, therefore, with the pre-classical doctrine that everything is produced by labour, aided by what used to be called art and is now called technique, by natural resources which are free or cost a rent according to their scarcity or abundance, and by the results of past labour, embodied in assets, which also command a price according to their scarcity or abundance. It is preferable to regard labour, including, of course, the personal services of the entrepreneur and his assistants, as the sole factor of production, operating in a given environment of technique, natural resources, capital equipment and effective demand. This partly explains why we have been able to take the unit of labour as the sole physical unit which we require in our economic system, apart from units of money and of time.

(Keynes, *General Theory*, p. 214)

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_10, © Springer-Verlag Berlin Heidelberg 2010

10.1 Introduction

John Maynard Keynes (1883–1946) studied at Eton and King's College in Cambridge, England, where he showed his talent not only in mathematics but also in classical and philosophical studies. Encouraged by Alfred Marshall (a family friend), he studied economics and after the completion of his studies and 2 years of service to the government of England, he was offered a teaching post at Cambridge. Keynes had written many books that established him as an economist with international reputation. However, he is known for his book published in 1936, titled, *The General Theory of Employment, Interest and Money* (henceforth *General Theory*). The book was written during the great depression and was about the great depression. Keynes calls his theory general, because it deals

> with the behaviour of the economic system as a whole, – with aggregate incomes, aggregate profits, aggregate output, aggregate employment, aggregate investment, aggregate saving rather than with the incomes, profits, output, employment, investment and saving of particular industries, firms or individuals. And I argue that important mistakes have been made through extending to the system as a whole conclusions which have been correctly arrived at in respect of a part of it taken in isolation. (Keynes, Preface to the French edition of the *General Theory*)

We observe that Keynes defines with precision a new area of economics, which now is being called macroeconomics and not only he distinguishes it from micro-economics but also calls the attention of the reader to the fallacy of composition, that is, what is true for the parts is not necessarily true for the whole. There is no doubt that the *General Theory* is recognised as an epoch-making book for it created a completely new view of the way in which the economy works and also provided the rationale for the active role of government.

Politics is another area where Keynes plays a prominent role. It was usual for him to assume government positions that required solution to difficult and urgent economic problems. In 1919, he was a member of the British delegation in Versailles for the signing of the peace treaty. He disagreed with the terms of the treaty and soon after he resigned from the delegation. When he returned to England, he wrote the book *The Economic Consequences of the Peace* (1919), in which he explained his disagreements. He argued that the terms of the treaty were extremely harsh for Germany to fulfil. As a result, he anticipated that the German economy would be led to a serious economic crisis which would likely develop into a world war. The tragic justification of his prognosis did not take a long time.

In the decade of the 1920s, he stayed in Cambridge where he was mainly involved in academic activities. He assumed the positions of the general secretary of the *Royal Economic Society* and the managing editor of the *Economic Journal*, one of the most prestigious journals in economics until today. Moreover, Keynes was one of the founding members and President of the *Econometric Society*, although he was not involved in econometrics, at least as we know it today; nevertheless, Keynes was already known for his knowledge of mathematics and especially statistics from his book: *A Treatise on Probability* (1921). Keynes also

had the time to engage in speculative activities in the foreign exchange and stock markets and he managed to create wealth for himself and the King's College, where he served as the president. In the mid-1920s, he criticised Winston Churchill for his decision to restore the gold standard and the pre-war parity of pound to gold. The results of this decision once again justified Keynes's critique, since England, in 1931, went off the gold standard. During World War II, Keynes served in the ministry of finance, where he played a key role in the shaping of economic policy of England. By the end of the war, he was leading the British delegation in the famous Bretton Woods agreements in 1944, which shaped the international monetary system. Keynes expressed his views about the global character of economic problems and proposed radical solutions such as world government and world money (with the currency unit called 'the bancor'). One of his recommendations was that countries with a trade surplus ought to invest in the deficit countries in the effort to promote stability and welfare in the world economy. However, his propositions were only partially adopted not because they were not sound but because the economic and political power had already shifted from the UK to the USA. Keynes is recognised as one of the most important economists and certainly the most influential of the twentieth century.

In what follows, we summarise the core ideas of the *General Theory* and, in this sense, we identify three of the major analytical components of his book. The first is the principle of effective demand, hence the 'psychological law' of marginal propensity to consume of less than one. The second is the marginal efficiency of capital and the third is this liquidity preference theory (Keynes, XIV, pp. 84–85). Keynes (1937) repeats the above three essential components of his theory in his summary of the *General Theory*. There is evidence, according to Patinkin (1990), that Keynes clung to this view until the end. The truth, however, is that Keynes did not give equal weight to all these three analytical components of his theory, but rather he ranked them according to their order of importance. This is the reason why the *General Theory* begins with his principle of effective demand, which is really Keynes's true theoretical innovation. Then he continued with the other two analytical components, which were not so original and really added to complete Keynes's principle of effective demand. In what follows, we preserve Keynes's order of presentation, starting with the theory of effective demand.

10.2 The Principle of Effective Demand

Keynes developed his principle of effective demand after a rather long time of gestation and controversies on central economic questions. One such question was subjected to a heated debate in the inter-war period and it was on the economic consequences of public works. In England, for example, there were two major views on the effects of public works on the overall level of employment. The first one was identified with the Treasury department, which is responsible for the

conduct of public policy, and attributed the causes of unemployment to the imperfections of the market system. In these imperfections were included government regulation, welfare institutions and, above all, strong unions that did not allow the market forces to operate properly, and rising unemployment was one of the dire consequences. If government were to intervene in the economy, this ought to be directed to the lifting of all inhibitions that prevent the efficient operation of market forces. In this view, there was no room for the government to assume any active role investing in public works. The idea was that public works entailed investment spending which would be financed through private savings. In other words, government, through taxation, would divert the given amount of total savings from the private to the public sector to finance the additional employment. According to the Treasury view these efforts would be counterbalanced by a reduction in the private savings and investment, which is another way to say that what would be gained in terms of public employment, would be lost in terms of private employment. In modern parlance private investment would be 'crowded out' by public investment, and the net effect in the level of employment would be at best nil.

The Liberal party that was led by Lloyd George claimed that government intervention to undermine union power and social institutions in general might create much more serious problems than those that it intends to solve. The liberal party proposed to expand the demand for labour through public works. Keynes approved of these policies; in his pamphlet (co-authored with Henderson) 'Can Lloyd George do it? An Examination of the Liberal Pledge' (1929) argued that the reduction in unemployment through public works has further employment-generating effects on the rest of the economy. The reason is that the newly employed workers spend the income that they earn on consumer goods, where employment should also increase and so forth. Keynes and Henderson (1929), however, did not clarify the exact mechanism that leads to the expansion of employment and also did not give any precise mathematical formula to explaining these employment-generating effects. Keynes's (undergraduate) student Richard Kahn (1931) in his article 'The relation of home investment to unemployment' not only developed the theoretical argument favouring public works, but also gave its mathematical formulation through the concept of employment multiplier. Kahn's idea was that in a depressionary situation the level of savings falls short of what is required for the full employment of labour. Consequently, as the level of activity increases through public investment, output increases and results in an increase in savings. Meanwhile, in the investment goods sector the additional income is being spent on consumer goods and so output there increases as well. So part of that output is also saved. We observe that savings are not fixed – as was assumed in the Treasury view – but rather a variable quantity, which could be affected by government intervention. Once this employment-generating process sets in it works itself out until the economy reaches the full employment level of output.

In Kahn's analysis the shortcoming in the Treasury's view was that in a recessionary situation the level of savings falls short of what is required for the full employment of labour. Thus, for Kahn as the level of activity increases through public investment, output increases and with that increases the amount of savings.

The limit to this virtuous cycle-like process, whereby public investment expands employment and output, which in turn increases savings and so forth, is until the economy reaches the full employment level of output. It is interesting to note that Kahn was trapped to the old mode of thinking according to which savings determine investment and so his employment multiplier was cast in purely neoclassical terms and as such it was not entirely new. The innovative element that Kahn contributed to economic theory was the mathematical formulation of the multiplier process as a sum of infinite terms of a declining geometric progression. The following year Warming (1932) criticised constructively Kahn's (1931) article by introducing an explicit consumption function in the economic literature of the multiplier (Skidelsky 1992, p. 451).

Keynes adopted Kahn's employment multiplier and he transformed it from a sheer description of changes that come from investment to a basic component of the theory of income determination and employment. More specifically, he adjusted Kahn's multiplier process to a relation between savings and investment. More specifically, we know that in classical economists[1] the central idea is that savings and investment are two aspects of the same act, for classics 'the money that is saved is automatically invested'. By contrast, in the neoclassical theory the equality of savings and investment is established through variations in the rate of interest. Keynes rejected the classical economists' view as unrealistic for the modern monetary economies, while he disagreed with the neoclassical view, where the interest rate plays the equilibrating role between savings and investment. For Keynes the equality of these two variables is brought about by variations in income (output):

> The novelty of my treatment of saving and investment consists, not in my maintaining their necessary aggregate equality, but in the proposition that it is, not the rate of interest, but the level of incomes which (in conjunction with certain other factors) ensures this equality. (Keynes 1937b, p. 250)

Furthermore, Keynes by borrowing Kahn's formulation of the concept of multiplier managed to develop an entirely new idea of the relationship between savings and investment, where the two variables are equalised through variations in the level of output produced.

10.3 The Income Determination Model

In terms of simple national accounts the above can be specified in the following way. The level of national income is determined by consumption and investment expenditures:

$$Y = C + I$$

[1]Keynes calls all the economists before him classical; however, he mainly means Alfred Marshall, Cecil Pigou, and Knut Wicksell (see Chap. 5).

Similarly, national income that is earned by households is either consumed or saved:

$$Y = C + S$$

Consequently in purely accounting terms it will always be true that:

$$S = I.$$

This is an accounting identity that holds independent of income.

With respect to the equality between savings and investment via the multiplier, let us suppose that investment increases by 100 monetary units and the marginal propensity to consume is 80% and remains constant, hence we have Keynes's 'fundamental psychological law' according to which

> men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income. (*General Theory*, p. 96)

Empirically, we know that in time series data the consumption expenditures and disposable income are directly related. The same stable relationship, however, does not seem to hold in cross section data, where the high income groups tend to consume a lower proportion of their income, the converse is true for the low income groups. As a consequence, there has been a voluminous literature on the characteristics of the propensity to consume in time series as well as cross section data. However, these studies did not change any essential aspects of Keynes's proposition, according to which income is the dominant determinant of consumption expenditures, whereas the interest rate and other possible variables only play a secondary role.

Hence, we observe that the decisions to save (*i.e.*, not to consume) are not related in any direct way to the decisions to invest. From introductory economics we know that the increase in investment by 100 monetary units leads to an equivalent increase in income Y in the investment goods sector, from this income 80 will be consumed and 20 will be saved. In the next period, these 80 monetary units become income for the producers of consumer goods, who in turn consume 64 and save 16. We observe that in the third period total savings are 36 and are smaller than investment, which are 100 monetary units.

The purpose of the above example is to lead us from an identity ($S \equiv I$) to a behavioural relation of saving and investment. In other words, the idea is that the public adjusts the decisions for expenditures with regard to the planned (normal) expenditures. The multiplier does not work instantaneously, so as the normal savings and income do not adjust automatically to investment. We observe that in every particular moment there is equality between the actual investment and savings. An equality which is *ex post* and comes about from the unwanted savings. Table 10.1 shows the difference of the *ex post* from the *ex ante* equality.

We observe that in the first period consumers save 100% of their income, whereas their normal behaviour indicates that only 20% of their income must be

Table 10.1 The process of the multiplier

Time	I	Y	C	Unplanned (*ex post*) S^u	Planned (*ex ante*) S^d
1	100	100		100	0
2		80	80	80	20
3		64	64	64	36
Total	100	244	144		

saved, since the marginal propensity to save (s) is residually determined. Consequently, normal behaviour requires that the public adjusts its expenditures with respect to the desired savings, which increase in each time period so as at some point total savings would be equal to total investment, that is $I = S^u + S^d = 100$. With the passage of time income will tend to 500, whereas consumption and desired savings to 400 and 100, respectively. Thus, we have:

$$Y = \frac{1}{s} I = kI = 5 \times 100 = 500$$

where $k = 1/s$ is the well-known formula for the investment multiplier, which according to Keynes:

> (It) tells us that, when there is an increment of aggregate investment, income will increase by an amount which is k times the increment of investment. (*General Theory*, p. 115)

The central idea of the *General Theory* is that the level of production is adjusted such that the normal (planned) savings is equal to the level of investment. Hence, we have an entirely new theoretical conception for the way in which savings are related to investment. We know that in neoclassical economics, savings are equalised to investment through variations in the rate of interest. In the *General Theory*, however, the equality between savings and investment comes about through variations in the level of income. This new conception is called principle of effective demand. More specifically, according to the principle of effective demand the level of output in the economy is determined by the level of monetary expenditures. The corollary of this principle is that savings and investment are equalised through variations in the level of output. Hence, we have two questions: the first relates to the role of prices (relative or absolute) and the analysis shows that prices play no role what so ever in the theory of effective demand. The second question relates to the causal relation between savings and investment. For Keynes, the independent variable in his analysis is investment expenditures which determine savings and not the other way around.[2]

[2]Hence we are assuming a super simple income determination model with no government and foreign sector. Of course, the sum of monetary expenditures (which in the general case include investment expenditures, government expenditures and exports) determine the amount of savings (savings, taxes and imports).

This is an entirely new idea on the basis of which Keynes constructed an altogether new theory.

> Saving at the prior date cannot be greater than the investment at that date. Increased investment will always be accompanied by increased saving, but it can never be preceded by it. Dishoarding and credit expansion provides not an *alternative* to increased saving, but a necessary preparation for it. It is the parent, not the twin, of increased saving. (Keynes 1939, p. 572)

Keynes's idea is counterintuitive, since common sense would indicate that in order to invest one must secure first the necessary savings. Keynes argued that this conceptualisation no longer holds in modern economies, where there is a well-developed and highly sophisticated financial system. Businesses first make their investment plans and then they are looking for the required financing. In other words, it is not necessary for businesses before they invest to have accumulated the corresponding amount of savings, which they will have to reduce latter on in order to finance their investment. Saving, in Keynes's theory of effective demand is not a stock but rather a flow variable. Saving is output that has been produced and has not been consumed yet. In this sense, savings although a real magnitude, and, therefore, constituent component of total output produced can take on monetary expression.

Saving is generated through the demand from the part of the public. More specifically, the public orders an amount of output through its expenditures. A part of the output produced is absorbed through consumer expenditures, which are equal to the marginal propensity to consume times the income. With regard to savings, that is the goods that are produced but are not consumed yet, the question is whether or not they will be absorbed from normal investment. Consequently, total expenditures (or injections) are those that determine the volume of output produced. Demand, overall, determines the general activity in the total economy. This appears also to be true in neoclassical economics. In Chap. 6, we showed that relative prices change in a way so as to adjust the demand to the level of endowment. The level of demand determines the supply and that demand is a function of relative prices. The neoclassical analysis is based on Say's law; by contrast, in Keynes there is the idea that demand determines output and, at the same time, the overthrown of Say's law as this is shown in the reversal of the causal relationship between savings and investment. In other words, only in Keynes we find the idea of investment as the independent variable of the system. The independence of investment from savings is due, according to Keynes, to the character of modern economies characterised by a well-developed credit system that makes possible the separation of finance from saving. The notion of finance refers to money, whereas savings refer to output that has been produced and has not been consumed yet. Investment does not require the reduction in the accumulated saving it only requires additional finance. There are no additional savings; there is only additional financing. Entrepreneurs with their investment expenditures may increase their income and their savings until the point that they equate savings to investment. So long as expenditures increase savings become profits and the entrepreneurs pay back the amount of money that they borrowed and so there is no increase in debt.

10.4 The Marginal Efficiency of Capital

In the Keynesian analysis in its simple version, that is, of a closed economy without government and foreign sector we know that in equilibrium savings are equal to investment and that variations in investment lead to variations in savings. But what determines investment? Keynes argues that investment, the major component of his theory of effective demand, depends on the marginal efficiency of capital (henceforth, MEC) in relation to the interest rate. Specifically, Keynes (1936) argues that when an entrepreneur buys investment goods in reality, he buys the right to a series of future incomes that expects to earn (during the useful lifetime of the capital good) selling the product after the subtraction of current expenditures. More specifically, Keynes notes:

> I define the marginal efficiency of capital as being equal to that rate of discount which would make the present value of the series of annuities given by the returns expected from the capital asset during its life just equal to its supply price. (*General Theory*, p. 135)

Let us suppose an entrepreneur who purchases an additional unit of a capital good (*e.g.*, a machine) and let us further suppose that the expected returns symbolised by Q_i, where $i = 1, 2, ..., n$ are the years. By P_s we symbolise the supply price of the capital good, which should not be confused with the current price of the capital good, but rather as Keynes notes:

> The price which would just induce a manufacturer newly to produce an additional unit of such assets, i, what is sometimes called its *replacement cost*. (*General Theory*, p. 135)

The MEC in Keynes's definition is equal to the discount rate r that satisfies the following condition:

$$P_s = \frac{Q_1}{1+r} + \frac{Q_2}{(1+r)^2} + \cdots + \frac{Q_n}{(1+r)^n} = \sum_{i=1}^{n} \frac{Q_i}{(1+r)^i}$$

where P_s and Q_i are given, as a result we have an equation of the n-th degree which solves for r.[3] If we suppose that $Q_1 = Q_2 = \cdots = Q_n = Q$, then the above sum forms a geometric series with $Q/(1 + r)$ as the first term and ratio the term $1/(1 + r) < 1$. Subsequently, we substitute in the well-known formula of the sum of n terms of geometric progression, which gives:

$$P_s = \frac{Q}{1+r}\left[\frac{1-(1+r)^{-n}}{1-(1+r)^{-1}}\right] = \frac{Q}{r}\left[1 - \frac{1}{(1+r)^n}\right]$$

[3]It is interesting to note that in the above relation there are so many r's as the roots of the equation that is being formed. From these roots, however, one and only one corresponds to the MEC and this might be an interesting exercise for the mathematically oriented reader.

If we now suppose that n $\rightarrow \infty$, we get:

$$P_s = \frac{Q}{r} \quad \text{and } r_s = \frac{Q}{P}$$

The above relation indicates that the MEC can be viewed as the annual yield that is expected to accrue to the entrepreneur from the additional investment of a unit of capital goods. Obviously, there is an inverse relation between the MEC and P_s.

10.4.1 The Falling MEC

Clearly, the definition of the MEC depends on expected and not on current or past profits and also these expected profits of a project are not evaluated against a stock of capital but rather against the flow of capital, that is, the increment of the existing capital stock, in particular the price of new equipment investment.[4] Thus Keynes argues that the MEC

> depends on the rate of return expected to be obtainable on money if it were invested in a newly produced asset; not on the historical result of what an investment has yielded on its original cost if we look back on its record after its life is over. (*General Theory*, p. 135)

It is interesting to note that the assumption of expected returns is absolutely necessary to Keynes in order to be consistent with his overall theory of effective demand, according to which the decisions to invest determine saving. If Keynes had assumed current or past profits instead of expected ones in his definition of the MEC, then he would have essentially accepted the idea that saving determines investment. Although the MEC depends on expected and not realised profits, which of course are fraught with uncertainty, Keynes was, nevertheless, absolutely certain about the falling MEC schedule, that he did not feel that there is a need for any detailed analysis. The gist of his argument on the falling MEC is contained in just a single paragraph:

> If there is an increased investment in any given type of capital during any period of time, the marginal efficiency of that type of capital will diminish as the investment in it is increased, partly because the prospective yield will fall as the supply of that type of capital is increased and partly because, as a rule, pressure on the facilities for producing that type of capital will cause its supply price to increase; the second of these factors being usually the more important in producing equilibrium in the short run, but the longer the period in view the more does the first factor take its place. Thus for each type of capital we can built up a schedule, showing by how much investment in it will have to increase within the period, in

[4]This is the reason why Pasinetti (1997, p. 207) approves Abba Lerner's use of the term marginal efficiency of investment instead of capital. Eisner (1997, p. 196) although in agreement with Lerner, nevertheless prefers to maintain both terms. In this paper we opted for the term MEC although we know that Keynes refers to the flow of investment and not the stock of capital (see also Chick, 1983, Chap. 6; LeRoy, 1983; Asimakopoulos, 1991, Chap. 4).

> order that its marginal efficiency should fall to any figure. We can then aggregate these schedules for the different types of capital [...]. We shall call this the investment demand–schedule of the marginal efficiency of capital. (*General Theory*, p. 136)

In the above succinctly written paragraph there are two intertwined arguments concerning the falling profitability. The first refers to short-run and the supply side of the market, where the investment expenditures of a firm imply that competition with other firms over resources gets more intense. However, the supply of resources is given in the short run; as a consequence, their price increases and profits decrease for each of the competing firms. Hence, Keynes assumes inverted L-shape unit cost curves, which imply that as competition gets more intense firms are bound to operate at the increasing part of their unit cost curves. For example, he notes

> [...] in the short period supply price usually increases with increasing output, on account either of the physical fact of diminishing returns or the tendency of the cost-unit to rise in terms of money when output increases. (*General Theory*, p. 328)

This argument, as Keynes notes, works more effectively in the short-run and weakens with the passage of time inasmuch as investment expands the capacity to produce.

The long-run argument refers to the demand side of the economy. Hence, Keynes's idea is that as a firm increases its investment and expands its output, it would become extremely difficult to keep its sales growing at the going price. Its sales can grow *pari passu* with its productive capacity only if the firm reduces its selling price. Consequently, expected profits fall and so does the MEC. It is important to stress, once again, that the supply and demand arguments in Keynes are not mutually exclusive; on the contrary, they may complement each other thereby reinforcing his overall argument for a falling MEC (Eatwell 1989). For the total economy, we simply add the behaviour of individual firms. Since for each particular firm there is an inverse relationship between the MEC and investment it follows that this is true for the economy as a whole. It is important to point out that for Keynes the fall of the MEC, in and of itself, does not automatically imply a reduction in investment expenditures. Everything depends on whether or not the rate of interest on loans is lower than the MEC. If for some reason the rate of interest is kept below the MEC, then there always exists an investment motive despite the falling MEC. This is the reason why Keynes, in Chap. 24 of the *General Theory*, argues for the 'euthanasia of rentiers', which can be achieved as the rate of interest approximates zero.

Keynes's analysis of falling profitability is too brief and certainly does not contain the subtleties that one finds, for example, in the classical economists.[5] This, however, by no means implies that there are no important insights and innovations. In fact, Keynes in Chap. 11 of the *General Theory* has some original contributions such as that the MEC is based on expected profits from current

[5]For a succinctly written summary of the views of the major economists on the falling rate of profit see Eltis (1989).

investment and the notion of uncertainty, a view which is consistent with the idea that the arrow of causality is running from investment to saving. The importance of these points, however, has passed unnoticed even by Keynes's major commentators (*e.g.*, Dillard 1948, Chap. 7; Hansen 1953, Chap. 5; Asimakopoulos 1991, Chap. 4). Keynes must also be blamed for that since he underestimates, in at least two instances, his own contributions by crediting the definition of the MEC to Irving Fisher. The first is in the *General Theory* (p. 140) and the second in 1937, in Fisher's *Festschrift* (*Collected Writings*, xiv, 101). The similarity, however, is only superficial and reminiscent of Keynes's style to find precursors of his views. For example, Keynes (*General Theory*, Chap. 23) refers to Malthus as the precursor of the theory of effective demand, and to Fisher as the precursor of the MEC. We know that neither Malthus nor Fisher share Keynes's view of investment determining saving and that the equality of saving and investment comes about through variations in output. Keynes's view is characteristically different to Fisher's and the neoclassical economists' who posited that the equality of full employment saving and investment is brought about by variations in the rate of interest. Furthermore, an identification of Keynes's theory of the MEC with that of Fisher's, as Garegnani (1978–1979) has pointed out, leads to two inconsistencies: first Fisher's expected profits are determined by marginal productivities of capital and labour; and second Fisher's 'MEC' presupposes full employment of both capital and labour. An argument that *prima facie* contradicts the quintessence of the *General Theory* according to which the cause of unemployment is the lack of adequate effective demand and that the price system left to its own devices cannot generate enough output to fully employ labour.

As for the marginal productivity theory of value and distribution, Keynes ruled out such a theory from his overall perspective of the way in which the actual capitalist economy works. For example, in the 1933 draft of several chapters of the *General Theory* Keynes (*Collected Writings*, xiii) introduces the distinction between a *real exchange economy* and a *monetary economy*. In the latter the presence of fiat money changes the law of production in a characteristically different way with respect to the former; that is, the real exchange or barter economy of classical and neoclassical economics. More specifically, Keynes resorts to the distinction, that was initially introduced by Marx (Chap. 6), between the simple commodity production (Keynes's *real exchange economy*) in which products are exchanged for the sake of consumption and a capitalist (Keynes's *monetary*) economy, where production of commodities is for the sake of profit in monetary terms.

This transition to the monetary economy involves the presence of fiat money which changes, in a fundamental way, the laws of production of the classical theory:

> The classical theory supposes that the readiness of the entrepreneur to start up a productive process depends on the amount of value in terms of product which he expects to fall to his share; *i.e.*, that only an expectation of more *product* for himself will induce him to offer more employment. But in an entrepreneur economy this is a wrong analysis of the nature of business calculation. An entrepreneur is interested, not in the amount of product, but in the amount of *money* which will fall to his share. He will increase his output if by so doing he

> expects to increase his money profit, even though this profit represents a smaller quantity of product than before. The explanation of this is evident. The employment of factors of production to increase output involves the entrepreneur in the disbursement, not of product, but of money. (*Collected Writings*, xxix, p. 82)

Keynes, a few years later, in the *General Theory*, continues to assume a monetary economy and explicitly rules out the marginal productivity theory of income distribution as this can be judged from the following:

> If capital becomes less scarce, the excess yield will diminish, without its having become less productive – at least in the physical sense [. . .] the only reason why an asset offers a prospect of yielding during its life services having an aggregate value greater than its initial price is because it is scarce [. . .]. (*General Theory*, p. 213)[6]

It has been argued (Dimand 1995) that Keynes perhaps was not aware of all the details of Fisher's analysis and that maybe he just did not find it appropriate to explain their conceptual differences in a book honouring Fisher's contributions. We know that Keynes disregarded Fisher's notion of the 'MEC' in his lectures, at a time as early as 1934 (Dimand 1995, 257) and that he admitted, in his correspondence with Harrod (August, 27 and 30, 1936), that his definition of the MEC is quite different from the works of classical economists and that it was 'vital for his analysis' a concept that he devised 'last of all, after an immense lot of muddling and many drafts' (*Collected Writings*, xiv, 85). Thus, although Keynes did not really present an analytically coherent argument, his desire for pragmatism led him to the conclusion that the MEC schedule was much lower in the 1930s than in the nineteenth century. There is no doubt that Keynes thought of the falling MEC as an already accomplished fact:

> Today and presumably for the future the schedule of the marginal efficiency of capital is, for a variety of reasons, much lower than it was in the nineteenth century. (*General Theory*, p. 308)

Hence, Keynes essentially adopts Smith's idea that the rate of interest, as a rule of thumb, can give us an approximate idea of both the level of the rate of profit and the direction of its long-term movement. Since in Keynes's time there were no national income accounts and certainly no time series data on profits and investment,[7] it seems that he was led to this conclusion by observing the evolution of the

[6]Garegnani (1977–1978) argued that the MEC is the 'Trojan Horse' of the price of capital goods through which the marginal productivity theory of distribution will undermine the statement: 'capital is getting now its marginal productivity (in some sense or other), is only valid in a stationary state. The aggregate current return to capital has no direct relationship to its marginal efficiency; whilst its current return at the margin of production (*i.e.* the return to capital which enters into the supply price of output) is its marginal user cost, which also has no close connection with its marginal efficiency' (*General Theory*, 139). For a related view see Minsky (1975, p. 96), while Pasinetti (1997, p. 218) argues that Garegnani's critique of the MEC is misplaced.

[7]The national income and product accounts data for the years up until the first decades of the twentieth century were created mostly retrospectively and after the publication of the *General Theory*, which essentially created both the need for such data as well as the conceptual framework for the estimation of variables such as income, investment, consumption, saving, *etc*.

rate of interest, as Smith did in his own time.[8] For example, in the *General Theory* (Chap. 16, p. 219) Keynes presents estimates of the long run average interest rate in the range of 2–2.5%, which is in fact equal to our estimates of the average interest rate on consols for the period 1900–1936, whereas for the entire nineteenth century the average interest rate on consols was around 4%.[9] There is no doubt that Keynes was aware of both the limitations of his theoretical analysis and the need to be backed up by empirical evidence. For example he notes: 'To develop the thesis (on the falling MEC) would occupy a book rather than a chapter, and would require a close examination of facts' (*General Theory*, p. 313).

It is interesting to note that the *MEC* refers to relations between prices and quantities. Hence, we have a demand function for investment that relates investment and the MEC. Keynes notes:

> The rate of investment will be pushed to the point on the investment demand-schedule where the marginal efficiency of capital in general is equal to the market rate of interest. (*General Theory*, pp. 136–137)

Such a relation does not appear in the theory of effective demand where prices do not play any role. The MEC schedule, however, refers to prices since it associates the level of investment expenditures to the internal rate of return and interest rate. If there is such a relationship, then a series of questions arise that challenge Keynes's position of chronic unemployment equilibrium. For example, what level of investment leads the economy to full employment? Why the economy does not succeed in achieving this level of investment? What are the factors that determine the level of investment? From the point that Keynes accepts the notion of the MEC, then there is the problem of determination of the rate of interest. Keynes's theory of unemployment equilibrium must also explain the reason why the rate of interest does not fall so as to lead investment to the level that is necessary for the full employment of labour. This is an issue that we deal with in the next section.

10.5 The Liquidity Preference Theory

Keynes's third major component of his *General Theory* is the determination of the rate of interest in order to explain the volume of investment spending. We know that, according to Keynes, investment depends on the MEC in relation to the rate of interest. The latter according to Keynes depends on the relation between the stock of monetary assets and their demand. The monetary assets are issued by the

[8]Clearly, Keynes regarded the rate of interest and the rate of profit (or the MEC) as distinct and strictly separate economic categories. In fact, Keynes criticised those economists (like Mises and Hayek) of "confusing the marginal efficiency of capital with the rate of interest" (*General Theory*, pp. 191–193 for a related analysis see also pp. 173–174).

[9]Data on the real interest rate on consols come from Global Financial Data (www.globalfinancialdata.com).

government according to its borrowing needs and the regulations of the money markets. There is no doubt that the monetary assets exceed the annual flow of money by far. Turning now to the liquidity preference or demand for money Keynes distinguished three motives:

- The demand for money for transaction purposes
- The demand for money for precautionary purposes
- The demand for money for speculative purposes

From the above three motives only the demand for money for speculation depends on the rate of interest, whereas the other types of demand for money depend mainly on income.

The crucial element in the analysis of the interest rate determination is the relationship between preference of the public for cash and for other less liquid assets. Clearly, the public considers that the advantage of holding cash is the flexibility in transactions, whereas its disadvantage is that no interest is earned. The converse is true for less liquid assets such as bonds, which they offer interest but are not liquid enough. If we suppose an individual who could possess either cash or bonds, his decision would be based on the current interest rate in comparison to that expected in the long run. If the current interest rate is above than the expected one in the long run, naturally the public anticipates falling interest rates and the price of bonds, *ceteris paribus*, increases. The public therefore prefers to buy bonds now in the hope of benefiting from their anticipated appreciation. If all the individuals in the markets are rational with perfect information the adjustment would be instantaneous and the market under investigation would be too volatile. Keynes, however, assumed that the market consists of a large number of heterogeneous individuals that each has all the money either in cash or in bonds and the exact proportion depends on expectations on the rate of interest. The demand for money is formed from all the individuals that each and every one of them chooses one of the two extreme behaviours.

In Fig. 10.1a, where the current interest rate, i, together with the expected in the long run, i^*, are on the vertical axis, whereas the quantity of money is on the horizontal axis. The demand for money is inversely related of the rate of interest, as this is depicted in Fig. 10.1a. If the current interest rate exceeds the long run interest rate ($i_2 > i^*$), the public expects that the rate of interest will fall and so, *ceteris paribus*, reduces its cash holdings and increase the amount of bonds in the hope of making capital gains from their appreciation. If the current interest rate falls short of the long run ($i_1 < i^*$), it follows that the current interest rate is expected to rise and thus the price of bonds to fall. Consequently, the public prefers the sale of bonds now since their price is expected to fall and so the demand for money will be increasing. It is worth stressing that the demand for money in Keynes is a function of two interest rates the current and the future expected one. So there is no single demand for money but rather many which depend on the expected interest rate and volatility in the current interest rate derives from volatility in the expected interest rate. In Fig. 10.1b, where we have depicted the long run interest rate and the current interest rate which is the result of the intersection of the supply of money and the

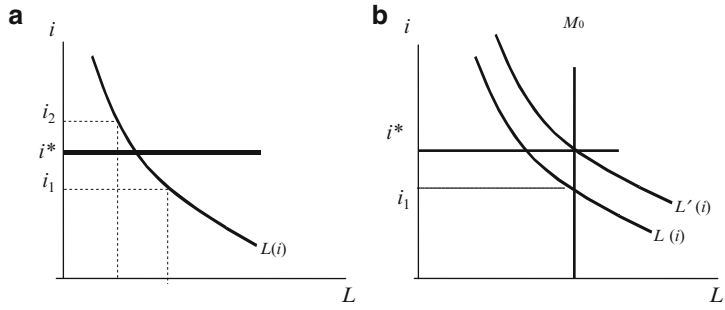

Fig. 10.1 Interest rate determination

demand for money. Suppose that the $i < i^*$ so the current interest rate is expected to increase a result that brings a lot of volatility in the money market and shifts the whole demand for money to the right.

In spite of the fact that expected interest rate, which is so crucial in the analysis of Keynes and yet one does not get definitive answers reading the *General Theory* or other Keynes's writings in the main, however, he argues that the long run interest rate depends on what the banking community considers accepted or natural long-run interest rate. In this sense, the determination of the long run rate of interest has many similarities, for instance, with the determination of the real wage in classical economics. It is important to point out the existence of a natural interest rate is a hotly debated issue even today. As we will see central banks decide about the supply of money on the basis of a natural interest rate whose existence is in question. We do know though that Marx and Sraffa, for instance, argued against the existence of a natural interest rate. Keynes's position on this issue is ambivalent.[10]

The demand for money displayed in Fig. 10.1b is compared with the supply of money (M_0), a variable which is within the control of the central bank and thus the equilibrium rate of interest is determined. Hence, the rate of interest is purely a monetary variable and operates as a centre of gravity for a whole spectrum of interest rates that are around it. The rate of interest then is compared to the MEC in the effort to determine the volume of investment (see Fig. 10.2, below). More specifically, entrepreneurs rank all their possible investment projects starting from the most profitable to the less profitable and then they decide how much to invest on

[10]In neoclassical economics the natural rate of interest would be associated with full employment. In Keynes, underemployment may coexist with macroeconomic equilibrium. Thus, Keynes in the *General Theory* (pp. 202–203, 242–243) rejected the idea of a unique natural interest rate. It might be remarked in passing that Keynes in the *Treatise of Money* (1930) had adopted Wicksell (1898) notion of a single natural interest rate identified with the marginal productivity of capital (see also Dillard 1948, pp. 197–198).

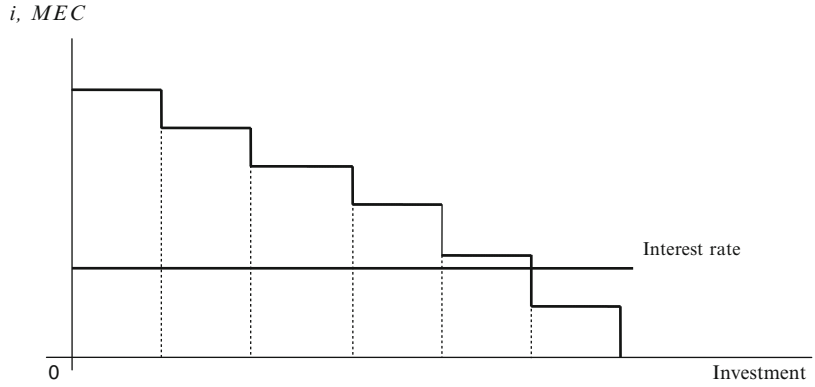

Fig. 10.2 Investment decisions

the point where their MEC from the most recent investment project is at least equal to the rate of interest.

Consequently, we see that there is a relation between the rate of interest, investment and savings however the arrow of causality runs in the opposite direction from that in the neoclassical analysis. More specifically, economists before Keynes argued that the equality between savings and investment is due to variations in the rate of interest, as we show in the next equation.

$$S(i,\ \overline{Y}) = I(i,\ \overline{Y})$$

In this sense, neoclassical economists could support this relation on the assumption of full employment output $\Upsilon = \overline{Y}$. Keynes argued that investment is a function of the interest rate, whereas saving is a function of both the interest rate and income. This description might give the impression that nothing is really new in Keynes. The new element that Keynes contributed to the analysis is the arrow of the causal relationship. In Keynes the rate of interest with given MEC determines the level of investment, which in turn determines the level of savings and subsequently via the multiplier the level of income. In Keynes we have the following chain of events:

$$(i, MEK) \rightarrow I \rightarrow S \rightarrow \frac{1}{s(i)} \rightarrow Y$$

The issue here is why the automatic forces in the economy do not set in motion the rate of interest so as to move towards the level which is consistent with the full employment of labour? Clearly, this possibility exists in the *General Theory* and Keynes had to show that this is only a theoretical possibility, because the rate of interest cannot fall to the required level, so as to stimulate investment spending to such an extent, so as to bring forth the full employment of labour.

Keynes argued that in reality the fall in the rate of interest may require substantial changes in the supply or in the demand for money and for Keynes only moderate changes in some variables may have an effect to the desired direction, which nevertheless are not strong enough to bring about full employment. If, by contrast, the changes in some variables are substantial, then the normal relations among the variables are no longer valid and so one cannot predict the final outcome. In order to illustrate Keynes's thesis let us suppose that because of the widespread unemployment the price level falls, and, therefore, the real supply of money increases and the transactions demand for money falls and so the interest rate falls stimulating investment spending and increasing employment. This is true, Keynes argued, to the extent that the deflation process is slow; if, by contrast, the deflation process is substantial, then the public feels insecure about the value of real assets and turns instead to the possession of money. Liquidity increases and the ensuing chaos would break down the above described favourable for the employment chain reaction of variables that restores equilibrium in the economy with full employment. Furthermore, Keynes considered the case of an active monetary policy, where the supply of money increases through the purchase of securities and interest rates fall, investment increases and unemployment decreases, and so forth. This effect can work so long as the changes in the money supply are moderate. If, however, the supply of money increases by much, Keynes argued that the liquidity preference function becomes particularly unstable[11] and so the direction of variation in the interest rates is uncertain (*General Theory*, p. 205).

10.5.1 Money Rate of Interest and Returns on Assets

Thus, Keynes ruled out the case of price variations to affect the rate of interest to a direction that would bring full employment. Keynes focused on the liquidity preference and the associated with it money interest rate. In this section, we examine the relation between the money rate of interest and the rate of return on other assets (*General Theory*, Chap. 17).[12] Our focus on the money rate of interest is justified from the view that in the final analysis it determines the level of the other interest rates. In conditions of equilibrium, all the interest rates tend to be equal to each other, a result that follows from the foregoing discussion on the MEC and the idea that the volume of investment is determined at the point where the MEC is equal to the rate of interest.

Keynes in the analysis of the rate of interest notes that this is based on an idea that he borrowed from Sraffa in his debate with Hayek in 1932 (*General Theory*,

[11]The volatility in the liquidity preference function depends upon degrees of expectations and it cannot be predicted to move in any particular direction.

[12]Hansen contends "that not too much would have been lost had it (Chap. 17) not been written" (1953, p. 159) and Keynes himself admitted "the obscurity of this chapter" (CW, XIV, p. 519).

p. 223). According to this idea the returns of every asset i (denoted by ρ_i) is equal to the sum of the three elements:

(1) The interest that is paid for this asset, r_i.
(2) The rate of change of the value of the asset, a_i.
(3) The expenditures for the possession of liquidity premium, l_i.

If we examine the various assets we observe that all of them display more or less the above characteristics. For example, let us suppose an asset such as bonds which display high liquidity, since there is a market for them and so they can be easily converted into cash.[13] Moreover, bonds pay interest (explicitly through a coupon or implicitly by their appreciation in the maturity date) and their price is subject to changes. For example, the same bond has different prices in different time periods. Other assets such as houses are characterised by high illiquidity since there is no market in which they could be sold relatively easily and with no loss of value. Consequently, the owners of such assets must look for higher returns in order to compensate themselves for their low liquidity. Moreover, assets of this kind are subject to capital gains or losses. The return of money (ρ_m) equals to its liquidity only, that is, $\rho_m = l_i$. The value of money remains constant $r_i = a_i = 0$, since money is the unit of account to evaluate all the other assets. The return on money (*i.e.*, the interest rate) can rise at a very high level but it cannot fall past a certain possible minimum, while the returns on other assets vary much easier in any direction and become even zero. In conditions of equilibrium the returns of all assets will be equal to each other, that is, $\rho_i = \rho_m$. This equality comes from the operation of competition and explains the reason why the rate of interest on money determines the interest rate on all other assets, the money rate of interest, as Keynes says, 'rules the roost', that is, the money rate of interest determines the level of the other interest rates for its two unique properties (*General Theory*, pp. 230–231):

- Zero or negligible elasticity of production (*General Theory*, p. 230).
- Zero or negligible elasticity of substitution (*General Theory*, p. 231).

Let us suppose an asset whose returns exceeds the money rate of interest, that is, $\rho_i > \rho_m$. Under these circumstances one would expect the expansion of production of good i, whose increased supply will lower its returns to the point that $\rho_i = \rho_m$. If we now suppose the converse that is $\rho_i < \rho_m$, then it does not follow that there will be expansion of 'production' of money so as ρ_m to fall. This is because of the first property of money, which is the zero (or negligible) elasticity of production. In other words, money is not a privately produced good, and, therefore, its stock in the economy is fixed.[14] Consequently, the rate of interest or the returns on money constitutes the lower limit of money for the returns of other assets.

[13]Of course there are restrictions such as that markets do not operate round the clock, there are holidays, *etc*.

[14]If money could be produced privately then as Dillard (1948, p. 201) claims the courses of depressions would be different, as indeed they are to some extent in gold-producing countries.

It is true that there is no asset that can replace money; this is the meaning of the zero elasticity of substitution of money. If, for example, there is a shortage of money, then it follows (in a monetary economy) that everyone will seek to find substitutes for money and there is no asset that approximates, in any satisfactory way, the properties of money. In such a situation, since monetary economies cannot function without money it follows that the rate of return on money will tend to increase without limits. This means that there is no asset that has lower returns from money, that is $\rho_i < \rho_m$ since this is a disequilibrium situation. The returns on money ρ_m form the maximum returns (ceiling) for the returns of the other assets which must approach the level of ρ_m.

Consequently, the returns of money because of the zero elasticity of substitution operates for the other assets as a ceiling and a zero elasticity of production operates as a floor, which means that all other interest rates vary within these two bounds determined by the returns on money. This analysis is in Chap. 17 of the *General Theory* and it is a long run analysis, after all a uniform interest rate (return) by definition requires a long run analysis.

Keynes from the moment that he brought in his analysis of the determination of the volume of investment through the MEC schedule and the rate of interest had to create some kind of inflexibility in his theoretical system in order to maintain his main proposition of unemployment equilibrium. In other words, Keynes had to fix one crucial price so as the variability of the other prices would not restore full employment. The fixed price he chose was the money rate of interest. The latter derives its characteristics firstly from the speculative demand for money, which is a short-run argument and secondly from the peculiar qualities of money which clearly is a long run argument. The short run and the long run arguments seem that they cannot be reconciled to a consistent theory for the determination of the rate of interest.

The problem is in Keynes's analysis of the determination of the equilibrium level of output which is accompanied by the unemployment of labour. Equilibrium implies that all the necessary adjustments have already taken place and there are no centripetal forces. If, however, the economy is in long-run equilibrium, then there is a problem in the foundations of the theory since the long-run interest rate remains unexplained. Keynes further argues that if there were not an asset such as money with the dual character then the interest rate on commodities would vary until the restoration of full employment. In other words, the presence of money (with its two attributes) prevents the rates of returns from falling and encouraging investment, and then Keynes argues that the price mechanism would drive the economy back to full employment. This argument is inconsistent with Keynes's overall analysis because he claims that the interest rate on money obstructs the rates of return of other assets to fall and to increase investment to the full employment level of output. So Keynes creates inflexibility in his analysis. He uses a fixed price to stop variable prices pushing the economy towards full employment. This fixed price is the money rate of interest. And the money rate of interest derives its characteristics from the speculative demand for money, which is a short run argument and the peculiar qualities of money which is a long run argument.

10.6 The Effects of Wage Reduction

A critique was launched directed at Keynes's argument of the given money wage; it was argued that why in a case of a prolonged recession a fall in the money wage would not lead the economy to a level of output associated with higher employment if not full employment. Keynes's answer was based on the observation that the money wage is such a fundamental variable that it cannot be subject to substantial reductions.

Keynes assumed the money wage rigidity without the development of a satisfactory theory of price level determination. In Keynes by definition we cannot have a quantity theory of money because such a theory presupposes as given the level of output that corresponds to full employment, while in Keynes the question is the determination of this level of output which corresponds to the full employment of labour.[15] As a consequence, the price level should be determined in a way different than the quantity theory of money. In this case, Keynes introduced the nominal wage, the level of which determines the price level. This hypothesis is characterised by realism, as this can be shown from the available time series data. In fact, it is a stylised fact that money wages and price level move together. The idea is that the money wage is such a central variable that cannot change in any substantial way as this may be the case with many commodities, some of which may change dramatically even in a single day. The level of money wage in the *General Theory* is used for the determination of the price level. The hypothesis of wage rigidity, on the one hand, helps Keynes to determine the general price level and on the other hand, however, this rigidity may be responsible for the level of output which falls short of that required for the full employment of labour. In other words, many economists understood Keynes's thesis as the presence of unemployment equilibrium.

Keynes (1936, Chap. 19) confronted this criticism by arguing that if we assume flexible money wages then the economy is going to be affected by the MEC, the investment multiplier and the liquidity preference. Thus, if money wages affect the level of output this may happen through three processes:

(1) The fall in the money wage increases profits and the marginal propensity to save and so the price of the investment multiplier falls. At the same time, the real income falls and so do the marginal propensity to consume and the price of the investment multiplier increases. The final result depends on the relative strength of each of the above two effects.
(2) The fall in the money wage leads to an increase in employment and thus in an increase of the MEC. In addition, the fall in the money wage leads to a fall of the general price level with a simultaneous fall in the MEC. Consequently, as in the above case the final outcome with respect to the MEC is ambiguous.

[15]It is interesting to note that in the *Treatise of Money* (1930) Keynes was in favour of the quantity theory of money but not in the *General Theory*.

(3) The fall in money wages reduces the transaction demand for money and increases the speculative demand for money; thus, the rate of interest falls and stimulates investment spending and so the level of output and employment increase.[16]

As a result, of the third case one may conjecture that a sufficiently large reduction in money wages may lead to the full employment of labour. Keynes, however, posited that this is only a theoretical result and it works in the desired direction only for moderate reductions in money wage which in turn elicit moderate changes in output and employment. A substantial fall in money wages, Keynes argued, might lead to quite opposite results and this because of the chaos that will be created in the economy and the resulting uncertainty which would disrupt the systematic relationships among variables:

> The chief result of this policy would be to cause a great instability of prices, so violent perhaps as to make business calculations futile in an economic society functioning after a manner of that in which we live. (*General Theory*, p. 269)

Thus a substantial fall in money wage will be conceived as a serious economic crisis. Consequently, the public prefers cash; the interest rate increases and so does the difference between actual output produced and full employment output. It is important to point out that Keynes does not completely rule out the effectiveness of wage cuts to establish full employment in case of authoritative governments of his time, such as those of Germany, Italy and the Soviet Union (*General Theory*, p. 269).

10.7 Keynes on Economic Policy

In the last two chapters of the *General Theory* Keynes describes his economic philosophy and also he comments on government's role in the economy. There is no doubt that Keynes believes that automatic market mechanisms in and of themselves are inadequate to solve the economic problems of a prolonged recession. According to Keynes, the government had to intervene all of the time to ensure that aggregate expenditures are in the right level, since depressions are caused by decreased aggregate expenditures, which is reflected in decreased investment. Keynes notes,

> The State will have to exercise a guiding influence on the propensity to consume partly through its scheme of taxation, partly by fixing the rate of interest, and partly, perhaps, in other ways. Furthermore, it seems unlikely that the influence of banking policy on the rate of interest will be sufficient by itself to determine an optimum rate of investment. I conceive, therefore, that a somewhat comprehensive socialisation of investment will prove the only means of securing an approximation to full employment. (*General Theory*, p. 378)

[16]This process has been called 'Keynes's effect'.

Hence, clearly Keynes suggests government investment as the primary means of increasing employment, after all this was a proposition that Keynes (1929) advanced quite early in his pamphlet with Henderson, with the difference that in the *General Theory*, Keynes put the policy recommendation on solid theoretical foundations. After the *General Theory*, economists could no longer argue that the economy would function best if left to its own devices.[17]

It is true that Keynes's policy prescription about the role of government was not revolutionary, because it was already being put into action by the US government before the *General Theory* was written. For this reason, many viewed the *General Theory* not so much as presenting a radical new solution, but as justifying actions already shown to be necessary. On further consideration, we find that even though the New Deal, for instance, was politically necessary, nevertheless economists still believed that such an intervention was not theoretically justified. Keynes's argument of an endemic market failure provided the theoretical justification for government intervention.

Another area of government intervention was related to rentiers, that is to say, all those that live on incomes from assets. Keynes argued against this type of incomes and he thought that gradually it must be eliminated in the effort to enhance profitability. Keynes's idea of the 'euthanasia of the rentier' is precisely a way to slow down the falling tendency in profitability and to encourage investment spending. It seems that this idea was developed much earlier than the time that the *General Theory* was written.

Another reason why the *General Theory* is revolutionary is that it emphasises problems and ideas that had been little considered before. For example, Keynes focused on aggregates such as aggregate employment, output and prices for the economy as a whole instead of employment, output and prices of individual industries. The *General Theory* relates to employment and output in the entire economic system, instead of individual businesses. The aggregate supply and demand functions are ideas that had not been given much attention in the past. Instead economists focused on the effects of supply and demand on individual goods or resources. This use of aggregates constitutes a shift in emphasis to macroeconomic questions. This is not to imply that economists before Keynes did not study aggregate variables and their interrelations. We know that classical economists and Marx cast their analysis in terms of macroeconomic aggregates, the difference; however is that they did not put these macroeconomic aggregates in terms of a theory to explain the level of output as Keynes did. Consequently, it would not be an exaggeration if we claimed that Keynes was the founder of what today is called macroeconomics. This claim is strengthened by the fact that Keynes focuses on macroeconomic variables and the control of these variables through government intervention gave rise to the creation of the system of national

[17]It is important to point out that Swedish economists (Myrdal, Wicksell, *inter alia*) had reached similar conclusion much earlier than Keynes, however these remained known in Sweden.

accounts, which was really developed after the publication of the *General Theory* and in particular after the World War II by Simon Kuznets (1971) and Richard Stone (1984).[18] In fact, only a few countries were collecting systematically data before World War II and the main reason was the prevailing philosophy at that time of no government intervention in the economy. Thus we can say that the *General Theory* was responsible for the system of national accounts and that the development of such a system further promoted the ideas of the *General Theory* and macroeconomics. Thus, even though Keynes and his economics have lost a lot of reputation after the 1970s, nevertheless it is certain that macroeconomics and the system of national accounts were brought about by Keynes.

We know that economic analysis had been traditionally concerned with the study of the long-run problems. In Keynes's early analysis, *Tract on Monetary Reform* (1923) and *Treatise on Money* (1930), one finds statements of the sort that 'we are always in the short-run' or 'in the long-run we are all dead', which often give the impression that Keynes's analysis is short-run. This, however, is not entirely true in the *General Theory* or in the latter writings, where the analysis is clearly couched on a long run setting and the conclusions that are reached call for government intervention.

10.8 Summary and Conclusions

Keynes's major contribution to economic theory is his 'principle of effective demand' which is a completely new view of the way in which the market economy operates. The 'principle of effective demand' is a set of propositions:

- The level of monetary expenditures determines the level of economic activity.
- The equality of savings and investment is brought about by variations in the level of output.

The corollary of the above two propositions is that since savings are not expenditures, it follows that investment determines savings and not the other way around. This is a completely new view that changed the conventional understanding of the way in which the economy operates. For example, the common idea that one finds even in the current economic literature on economic growth and development is that savings must increase in order to accelerate economic growth. Keynes would regard such a proposition as unfounded in modern economies with their sophisticated financial system. The idea of increasing savings is just impossible in Keynes's system, since the increase in savings can only decrease the value of the marginal propensity to consume (or to increase the marginal propensity to save) and decrease the price of the multiplier; consequently, lower the level of activity and push the economy into a greater slump.

[18]In parenthesis the year that the Nobel Prize in economics was awarded.

Thus savings, *i.e.*, the normal level of production which is not consumed decreases. This effect which is now quite usual in macroeconomic textbooks can make an excellent example of the 'fallacy of composition' and has been called the 'paradox of thrift'.

Once Keynes formulated his major theory which was completely independent of prices, he wanted to fill some of its gaps. First of all he needed a theory of investment determination and for that he used the idea of the MEC. Keynes's exegesis of the tendency of profitability to fall in the long run, which leads to economic crisis, has been largely misunderstood and its importance has been downplayed in the subsequent literature. However, Keynes following a long tradition of economists adopted the idea of the long run falling profitability, as he expressed it in the movement of the MEC. Moreover, he argued that a declining MEC is internally generated by an economic system, whose motion originates in expected profitability. This is the reason why Keynes was so much interested in the future of the system, and, most of all, the maintenance of its capitalist character. Keynes's concern is absolutely understood, if we think of the two alternative systems of his time, the national and the soviet type 'socialisms'. This is the historical context that we should place the exercise of caution with respect to the manipulation of investment, when he arrives at the conclusion that 'the duty of ordering the current volume of investment cannot safely be left in private hands' (*General Theory*, p. 320). His plea for substantial reforms, with 'a gradual disappearance of the rate of return on accumulated wealth' providing 'a sensible way of gradually getting rid of many of the objectionable features of capitalism [. . .]' (*General Theory*, p. 221), otherwise the 'socialist' alternative would prevail. Keynes's fairly radical conclusions, with today's (at least before the downturn that started in 2007 and continues at the present 2009) standards as well as the difficulty of his theoretical arguments for they were not cast in terms of the 'habitual modes of thinking', led many of his commentators to the relegation of the notion of the MEC and the business cycles associated with it to secondary importance. However, by ignoring the falling MEC from Keynes's overall theory of effective demand, we are left with an enormous lacuna, and, therefore, our understanding of the way in which the system operates, since profitability and its evolution shape both the present and the future of a system in continuous motion.

The MEC schedule however introduces prices and once prices are introduced in the analysis the next question is why prices and in the case of the MEC schedule the rate of interest does not adjust so as to lead investment to a level that would establish the full employment of labour and maintain it there. In order to deal with this issue Keynes introduced the theory of liquidity preference and the idea of money interest rate that 'rules the roost' and this money rate of interest forms a 'floor', or inflexibility in the downward direction. Many questions are raised with the last two components of his theory (*i.e.*, investment and interest rate). Keynes himself was sceptical about the theories that he proposed, but he was content, with the idea that he did the best he could do (Keynes, 1937b). Some Keynesian economists such as Hicks thought that the inflexibility in the interest rate in the downward direction, hence the famous 'liquidity trap'

was the essence of Keynes's theory. Other Keynesian economists, Franco Modigliani for example thought the inflexibility of the money wage to the downward direction is what distinguishes Keynes from the neoclassical economists. With both issues we deal with in the next chapter on the neoclassical synthesis.

In retrospect, we see that many decades after the *General Theory* was written, there has not been any substantial progress in these two areas of economic analysis, in both theoretical and empirical grounds. The liquidity trap idea or the inflexibility of the rate of interest, which is the same, that was forgotten for decades, resurfaced in the recent years. The difference that we find is that in the Hicksian case the interest rate happened to be simply low (minimum interest rate) nowadays the interest rate associated with the liquidity trap is near zero. The issue of money wages has been resurfaced and its flexibility to a downward direction is achieved through flexible forms of employment. Certainly, these are efforts that are within the neoclassical synthesis and are doubtful whether they are in Keynes's spirit. What is certain, though, is Keynes's macroeconomic look and the system of national accounts associated with it. In the *General Theory*, Keynes defined a research agenda for the generations to come and it is true that there has not been much theoretical progress since.

Questions for Thought and Discussion

1. What are the effects of public works on employment? Contrast the Treasury's with Kahn's views.
2. What exactly is Keynes's theory of 'effective demand'? How does it differ from Smith's and Malthus's 'effectual demand'?
3. In both Keynes and neoclassical economics the level of demand determines the supply. Discuss.
4. Present and critically evaluate Keynes's argument about the falling profitability.
5. Discuss the notion of liquidity preference in Keynes's *General Theory*.
6. How is the money interest rate determined in the *General Theory*?
7. Would a fall in money wages restore full employment of labour? What about the rate of interest? Explain.

Notes on Further Readings

The *General Theory* has been described as a difficult book and only a few economists have really studied it and if someone is interested in Keynes's ideas then the suggestion is the secondary literature. On closer examination, however, one discovers that the ideas contained in the book are much easier than it is usually thought and that the secondary literature more often than not distorts Keynes's ideas. This is a problem that is usually met with the secondary literature. If there is a problem with the book, it is that it was written in a period of time that economists were not accustomed to macroeconomics and there was not any tradition, everything had to be created from

(continued)

the beginning. We suggest that the serious reader must refer at least to the chapters that are mentioned in our presentation of the *General Theory*.

A comprehensive presentation of the *General Theory* is contained in Dillard's (1948) book which, to our view, remains, despite the passage of time, the best review of Keynes's ideas. Whereas the books by Hansen (1953) and Klein (1947) despite their popularity remain trapped within the IS–LM apparatus, which we discuss in the next chapter. The book by Eatwell and Milgate (1983), which contains a collection of macroeconomic articles of mainly Sraffian economists on Keynes. Our analysis of Keynes is based to a great extent on the articles by Garegnani (Chaps. 1 and 2) and Eatwell (Chap. 6). The hallmark of these articles is their persistent effort to base their conclusions on Keynes's own writings, an effort that we do not usually find in mainstream approaches. Hansen (1953), Samuelson (1948) are more in the spirit of the IS–LM approach that we discuss in the next chapter and less to the *General Theory* and Keynes's writings in general. The article by Don Patinkin (1990), whose point of reference is Keynes's text, remained faithful to his ideas and interpretation of Keynes.

With respect to the theory of effective demand, we think that the references in the above paragraph are a must. The interested reader must at some point see the controversies on the question of finance and on that the article by Assimakopoulos (1983) is the best way to start and the follow up discussion in the pages of the *Cambridge Journal of Economics*. On empirical grounds, the article by Pollin and Justice (1994) explores the relationship between finance–investment–interest rates, and on this issue refer also to the article by Apergis and Tsoulfidis (1997).

With respect to the ranking of investment in descending order, we recommend Alchian (1955), who noted that the ranking according to the internal rate of return might not be the same with ranking according to the maximisation of the present value. Garegnani (1977) argues that a normal MEC schedule is impossible in the case that there is substantial unemployment. For example, let us suppose that investment increases, we know that there is going to be a multiple increase in income and so expected income rises, and the MEC schedule shifts to the right. This means higher investment and so the MEC shifts even further to the right raising investment and so forth. In this sense, the final amount of investment is indeterminate or that all investment projects will be finally undertaken 'since the profitability of projects is itself a function of aggregate demand and thus endogenous to the problem'. Eatwell (1987) extends the analysis of (1983) and presents a thorough review of the MEC idea alongside with Fisher's theory of investment decisions.

As for Keynes's theory of liquidity preference (Chaps. 13–17) perhaps the most difficult of Keynes's writings we suggest the articles by Panico (1987) and Milgate (1977). Garegnani (1978–1979) argues that the interest rate is determined by the supply of and the demand for money and is not related in

(continued)

any direct way with savings and investment. The equality of the latter is determined through variations in income. Furthermore, Keynes's theory left the determination of the interest rate 'hanging in the air' or 'hanging by its own bootstraps'. On these general issues the reader is encouraged to see Minsky's book (1983). For Keynes's views after the publication of the *General Theory* we recommend his two articles published the following year (1937a, b). In the first article Keynes summarises the core theoretical propositions of the *General Theory*, whereas the second article refers mainly to Keynes's theory of the rate of interest and essentially claims that the theory of the rate of interest was complementary to the theory of effective demand and Keynes notes that he did what he could from what was available.

Chapter 11
The Neoclassical Synthesis

Those who are strongly wedded to what I shall call "the classical theory", will fluctuate, I expect, between a belief that I am quite wrong and a belief that I am saying nothing new. It is for others to determine if either of these or the third alternative is right.

(Keynes, *General Theory*, p. v)

It is usually considered as one of the most important achievements of the Keynesian theory that it explains the consistency of economic equilibrium with the presence of involuntary unemployment. It is, however, not sufficiently recognized that, except in a limiting case to be considered later, this result is due entirely to the assumption of "rigid wages" and not to the Keynesian liquidity preference.

(Modigliani 1944, p. 65)

11.1 Introduction

Many economists, soon after the publication of the *General Theory* (1936), set out to formulate and, at the same time, to clarify the difficult and often confusing content of the book. Among the first models that were specified was that of John Hicks (1937, 1983), which was to constitute the backbone of what today came to be known as macroeconomics.[1] In his article, Hicks sought to express the central propositions of the *General Theory* in terms of equations and graphs in the effort to illuminate the relation between the theory of effective demand and liquidity preference. Furthermore, Hicks clarified these relations with the aid of two curves the *SI* and the *LL*, which later became known as the *IS–LM* curves. Hicks's model became particularly popular in the US through the work of Paul Samuelson (1948) initially and subsequently through Alvin Hansen (1953). These two economists

[1]John Hicks (Nobel 1972) presented his model for the first time in a symposium of the newly instituted econometric society, which took place in Oxford in September of 1936 and the next year it was published in *Econometrica* (Hicks 1937). It is interesting to note that in the same symposium Roy Harrod (1937) and James Meade (1937), also presented similar models of systems of simultaneous equations.

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_11, © Springer-Verlag Berlin Heidelberg 2010

contributed more than anybody else to the popularisation of the Keynesian analysis and way of thought. The *IS–LM* conceptual apparatus has displayed remarkable longevity and resilience to various critiques and, since the late 1950s or early 1960s, continues to be part of the formal education of economists. At the same time, the *IS–LM* model plays a significant role by virtue of the fact that macroeconomic analyses, regardless of the approach, are cast to a great extent, in terms of the *IS–LM* representation of the economy. This is not to imply that the *IS–LM* model is without its problems; on the contrary, many economists expressed scepticism on the validity of the *IS–LM* as a representation of the *General Theory* and the way in which the economy works.[2]

In what follows in this chapter, we present and evaluate the Hicksian *IS–LM* model and continue with Keynes's reaction to the Hicksian restatement of the *General Theory*. Next, we introduce Modigliani's version of the Keynesian model, and the chapter ends with some concluding remarks.

11.2 Hicks's Analysis of IS–LM

Hicks's analysis focuses on the relation between savings and investment and seeks to establish the simultaneous determination of income and the rate of interest in both the real and monetary economy. According to Keynes's analysis in *General Theory,* income constitutes the principal variable in his analysis; nevertheless, one would continue to be in the spirit of Keynes by considering the important role of the rate of interest. Thus, Hicks argued that investment (I) is a function of the rate of interest (i) and also income (Y). Formally, we write the following function:

$$I = I(i,\ Y)$$

Similarly, for the saving function (S), we have:

$$S = S(i, Y)$$

The equilibrium condition is:

$$I = S$$

From the above equality, we derive the following particular functional formalisation, which is called *IS,* and it is defined as the locus of points that determine a relation between the rate of interest and the level of income, when investment

[2]Post-Keynesians are among those that dispute the validity of the *IS–LM* representation of the *General Theory*. Joan Robinson (1975), for example, has described these efforts as "bastard Keynesianism".

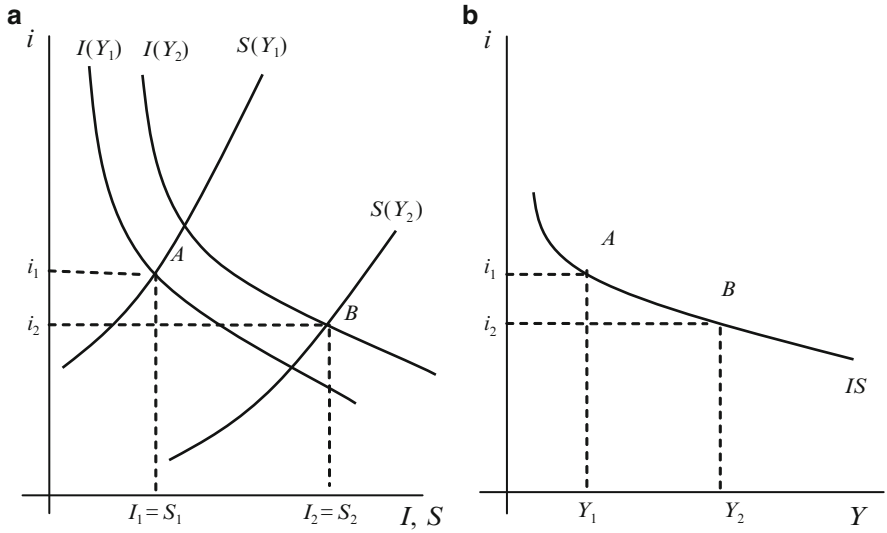

Fig. 11.1 Equilibrium in the goods market and the IS curve

and savings are equal to each other. The *IS* curve is formed in the way we show in Fig. 11.1 where we have the savings and the investment functions for each level of income.

Let us suppose that we are in an initial equilibrium point such as A and let us further suppose that income increases from Y_1 to Y_2. It follows that the savings and investment schedules – both have their positive first derivatives with respect to income – shift to the right and their intersection at point B determines the new equilibrium point. It is important to stress that the savings function is much more sensitive to variations in income, and, therefore, it shifts to the right by more than the investment function.[3] The two equilibrium points (i_1, Y_1) and (i_2, Y_2) are portrayed in Fig. 11.1b. In a similar fashion, we generate a series of such points, which when connected form the *IS* curve.[4]

Hicks furthermore incorporates in his analysis the money market, where the supply of money (M) is exogenously determined, that is, $M = M_0/P$, where M_0 is the exogenously given nominal money supply and P is the price level. The demand for money depends on income and the rate of interest, that is, $L = L(i, Y)$. By invoking the balancing condition $M = L$, we arrive at

$$M_0 = L(i, Y)$$

[3]The idea is that the investment decisions are much more complex and do not depend so much in income.

[4]One should not rule out the case of positive or other slopes for the *IS* curve which however complicate the analysis and raise questions about the stability of equilibrium. These issues however are beyond the scope of a book of history of economic thought.

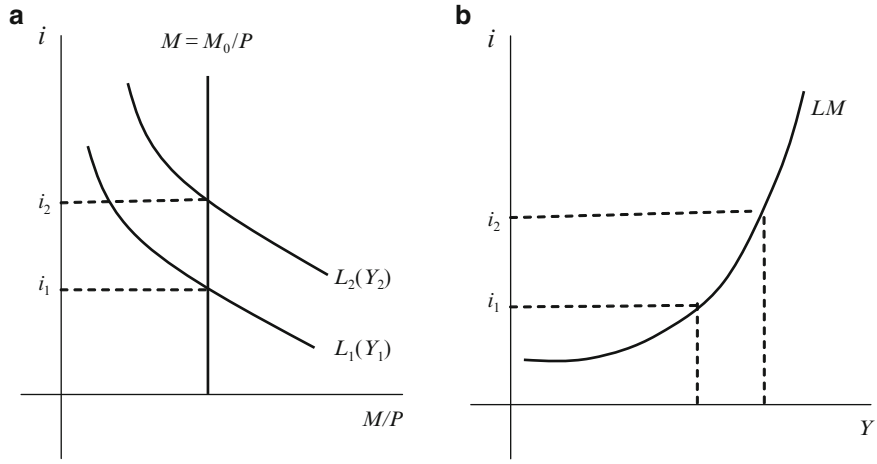

Fig. 11.2 Equilibrium in the money market and the LM curve

Figure 11.2 illustrates the equilibrium position in the money market, where the supply of money, for reasons of simplicity and clarity of presentation, is depicted with a vertical line indicating its exogenous character.[5] The demand for money, as we know, is inversely related to the rate of interest, a relation whose details have been analysed in the previous chapter. When income increases it follows that much more liquidity is required for the needs of transactions, and, therefore, the interest rate will increase for any given level of money supply.

We observe that with the supply of money given the demand for money for transaction purposes is directly related to income. The crucial question here is that while we refer to the money market the discussion is in terms of the bond market. In particular, we know that the excess demand for any good leads to an increase in its price until excess demand becomes zero and thus we get the equilibrium point. Since in the case of money market the equilibrium interest rate is derived in the market for bonds (see Chap. 9), then how can the same interest rate equilibrate the money market? In Keynes's analysis it seems that there is an implicit portfolio stock exchange constraint, which can be written as follows:

$$(L - M) + (B^d - B^s) = 0$$

where B symbolises the bond market, while the superscripts d and s symbolise the demand for and the supply of bonds, respectively. Consequently, we have the total demand for wealth ($L + B^d$) equal to its supply ($M + B^s$). If we, further, suppose Walras's Law, then the above equality necessarily holds and if the rate of interest

[5]The exogeneity or endogeneity of money supply is an old issue that continues to attract the interest of economists and policy makers.

brings equilibrium in the market for bonds then on the basis of Walras's Law we conclude that equilibrium will be also established in the money market, that is, $L = M$. As a consequence, we can follow Keynes, who argued that interest rates are determined in the money market. Based on Walras's Law, equilibrium in the bond market and equilibrium in the money market is one and the same. If, for example, $i > i^*$, then $B^d > B^s$ and because of the stock constraint we get $L < M$, that is there is an excess supply of money in the economy.

Returning to the above equilibrium relations, we end up with a system of four equations and four unknowns: Y, i, I, S. The equations IS and LM represent the reduced form of the above system of simultaneous equations, whose solution gives the equilibrium income together with the equilibrium interest rate. In the same figure, we present the interest rate that corresponds to the liquidity trap (i_{LT}), where the demand for money is infinitely elastic. Consequently, the LM curve is essentially the solid line.

The intersection of the two curves at point B determines the equilibrium pair of interest rate and income. Any point above the IS curve indicates excess supply of goods and every point below the IS curve indicates excess demand for goods. As for the LM curve, every point to the right indicates excess demand for money and every point left to the LM curve indicates excess supply of money. The intersection of the two curves defines four quadrants, which are portrayed in Fig. 11.3 above, and in each quadrant we indicate the excess demand or supply in the goods and money markets. The mechanism that establishes equilibrium in the economy works as follows: let us suppose that for some reason the economy is out of equilibrium at a point on the quadrant II. In such a case, savings exceed investment and thus income has a tendency to decrease, while the demand for money is greater than the supply

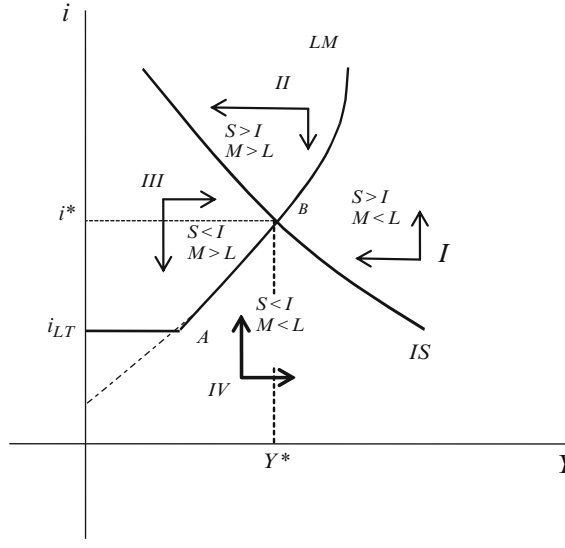

Fig. 11.3 Equilibrium in the market for goods and money

and the interest rate tends to increase. The changes are expected to lead the economy towards equilibrium at point B. In an analogous way, we can describe the mechanism that restores equilibrium at points in the other quadrants and this is left as an exercise.

11.3 Hicks and Keynes

Hicks's article was published in 1937, 8 months after the publication of the *General Theory*. Keynes already knew the content of the article since he was among the first that the article was given to for comments before its publication to the *Economic Journal*. Keynes never disapproved directly and explicitly the presentation in terms of the IS–LM apparatus. Don Patinkin (1922–1995) in a series of articles argues the fact that Keynes never said anything negative for the formalisation of his theory by Hicks and that this *ipso facto* implies an adoption of this presentation on his part (Patinkin 1990). If Keynes disagreed then he would have every reason to emphatically express his disagreement. After all Hicks's presentation in a sense was provocative, since Keynes's *General Theory* in it was viewed as a special case of the neoclassical true general theory.

Post Keynesian economists claim that the fact that Keynes did not exercise a negative critique can be attributed to his idiosyncrasy that would not pay attention to anyone's writings which might concern his *General Theory*. On the other hand, Keynes did not have any reason, to express, at least in the beginning, his strong disagreement to Hicks's presentation. It is possible that he did not think that Hicks's article would meet the success that it finally met.[6] It is certain that he disagreed with Hicks's view as this can be judged by a careful reading of his correspondence with Hicks and from the article that he wrote in the *Quarterly Journal of Economics* (1937), where he summarised his views. Specifically, he placed special emphasis, once again, on the fact that economies are characterised by uncertainty.

Hicks's approach is characteristically different from that of Keynes's. We know from Pasinetti (1974) that Keynes followed a sequential analysis starting from the marginal efficiency of capital, and then to the interest rate, to investment and through the investment multiplier to the equilibrium level of income. By contrast, in Hicks, all of the above take place simultaneously, as we show in Fig. 11.3. Furthermore, Hicks in his formulation of the demand for money refers to a single interest rate. In the *General Theory*, however, we know that Keynes refers to two interest rates, the current and the expected in the long run. Consequently, Keynes's

[6]Skidelsky claims that Keynes does not seem to have held a high opinion of Hicks and this was the reason why he did not pay much attention to the *IS–LM* presentation of his theory. After all Keynes at that time could not know that this article was going to become such a success (Skidelski 1992).

analysis is in sharp contrast to Hicks's and on top of all we have the issue of uncertainty that permeates the *General Theory* and is completely absent in Hicks's presentation.

Another important difference is that Hicks does not refer to the problem of unemployment equilibrium which is so central in Keynes – and really differentiates him from the classics –. Instead, Hicks locates the difference between Keynes and the classics to the interest rate and the issue of whether it increases with investment or not (Barens and Caspari 1999, p. 219). According to Hicks, in periods of stagnation the interest rate is particularly low and under these circumstances speculators are not willing to hold non-liquid assets; consequently, their demand for money is so high that it absorbs whatever quantity of money is available. Thus, every increase in the supply of money is counterbalanced by a corresponding increase in the demand for money and the rate of interest remains constant. Monetary policy therefore is completely ineffective and it cannot restore the economy to full employment equilibrium. Hicks notes,

> there are conditions in which the interest-mechanism will not work. The special form in which this appears in the *General Theory* is the doctrine of a floor to the rate of interest – [the liquidity trap] as Sir Dennis Robertson has called it. (Hicks 1957, p. 287)

If we suppose that the economy is in the liquidity trap,[7] then a monetary policy, regardless of how active it might be, cannot shift the economy beyond the initial equilibrium point. In terms of Fig. 11.4, if the economy is in equilibrium at point *A*, an expansionary monetary policy will shift the *LM* curve, for example to the position *LM′*, with no consequence what so ever for the initial equilibrium position.

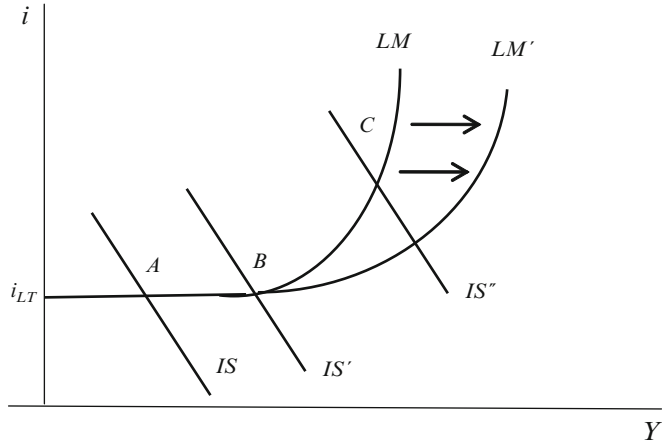

Fig. 11.4 Equilibrium in the markets for goods and money

[7]According to Boianovsky (2004) the notion of liquidity trap was introduced by Robertson in 1936.

Consequently, Hicks in his model claims that the *General Theory* is not so general as Mr. Keynes thought, but rather a special case of the neoclassical theory, where the liquidity trap has a prominent position. The truth, however, is that the idea of the liquidity trap is very hard to pin point in the writings of Keynes; of course, there are some sporadic hints in the *General Theory*, as for example is the following:

> There is the possibility, for reasons discussed above, that, after the interest rate has fallen to a certain level, liquidity-preference may become virtually absolute in the sense that almost everyone prefers cash to holding a debt which yields so low a rate of interest. In this event the monetary authority would have lost effective control over the rate of interest. But whilst this limiting case might become practically important in future, I know of no example of it hitherto. (Keynes 1936, p. 207)[8]

However, Keynes does not discuss this case in any detail so as to claim that this is the hallmark of his theory. What is certain, however, is that the liquidity trap is more Hicks's and subsequently Hansen's (1953, pp. 122–123) idea rather than Keynes's.[9] Consequently, the view that the liquidity trap is the essence of Keynes's theory is due to the influence that the Hicksian model exerted on macroeconomics and much less to Keynes and his writings.

Suppose, now, that for some reason investment increases, and then the increase in the rate of interest follows suit, a result which is consistent with neoclassical theory and with Hicks's argument. It is true, that in Keynes the arrow of causality is different from that in neoclassical economics. However, it continues to be true that, under normal conditions, the interest rate increases when investment increases except for the case of the liquidity trap, where only income changes in every change in investment. The trouble, however, with Hicks's view is that for Keynes the rate of interest is determined by monetary forces, while in the *IS–LM* framework the interest rate is determined by real forces. This is an issue that Keynes pointed out in his letter to Hicks. For example we read:

> From my point of view it is important to insist that my remark is to the effect that an increase in the inducement to invest need not raise the rate of interest. I should agree that, unless the monetary policy is appropriate, it is quite likely to. In this respect I consider that the difference between myself and the classicals lies in the fact that they regard the rate of interest as a no-monetary phenomenon, so that an increase in the inducement to invest would raise the rate of interest irrespective of monetary policy. (Keynes 1973, p. 80)

A final point relates to the inclusion of current income in the investment function. Keynes objected to this idea for the reason that income was already included in the definition of the marginal efficiency of capital through the prospective yields. The following quotation from his letter to Hicks, shows that Keynes was not only acquainted with the *IS–LM* apparatus but also as a modern econometrician

[8] A similar argument is in *General Theory* (Chap. 13, p. 172).

[9] In the liquidity trap, according to Hicks, the rate of interest is so low that the public considers it as being capable of increasing only. Under these conditions, a fall in the price of securities is expected and consequently the public prefers cash to other assets.

argued against the inclusion in the same specification of both income and interest rate. Specifically, Keynes notes:

> At one time I tried the equations, as you have done, with *I* in all of them. The objection to this is that it overemphasizes current income. In the case of the inducement to invest, expected income for the period of investment is the relevant variable. This I have attempted to take into account of in the definition of the marginal efficiency of capital. As soon as the prospective yields have been determined, account has been taken of income, actual and expected. But, whilst it may be true that entrepreneurs are over-influenced by present income, far too much stress is laid on the psychological influence, if present income is brought into such prominence. It is of course, all matter of degree. (Keynes 1973, pp. 80–81).

Barens and Caspari (1999) in their discussion of Hicks and Keynes note that while Hicks accepted all of Keynes's points he nevertheless insisted in his own formulation for merely pedagogical reasons.

11.4 Modigliani's Synthesis

Hicks's model does not refer explicitly to the labour market; it is simply confined to demonstrating that there is equilibrium in only two markets that is the market for goods and the market for money. In his model, Hicks explicitly argues that the money wage as well as the general price level is exogenously given. Franco Modigliani (1944) extended Hicks's model by including the labour market and the production function. Modigliani argued that the assumption of equilibrium with unemployment cannot be supported on the basis of the liquidity preference theory except for the particular case of the liquidity trap. In general, however, the Keynesian hypothesis can be supported on the assumption of the rigidity in the money wage. For Modigliani, the equilibrium in terms of the *IS–LM* model implies a pair of interest rate and money income that clears simultaneously the money and good markets. Consequently, we must take into account that the money income (Y) is equal to the price level (P) times the level of the real income (X). As a result, we may write:

$$Y = PX$$

The level of the real income (or output) is a function of the level of employment of labour (N). Consequently, we have:

$$X = F(N)$$

The level of employment in turn is determined at the point, where the marginal product of labour is equal to wage. Consequently, we have:

$$w = PF^{-1}(N)$$

Up until now we have a system of seven equations (the three equations above together with the system of four simultaneous equations of the *IS–LM*) with eight unknowns, that is I, S, i, Y, X, W, P. More specifically, we have the four equations of Hicks's model:

$$I = I(i,\ Y)$$

$$S = S(i,\ Y)$$

$$I = S$$

$$M = L(i,\ Y)$$

And the three new equations suggested by Modigliani:

$$Y = PX$$

$$\mathrm{X} = F(N)$$

$$w = PF^{-1}(N)$$

The system is over determined by one equation, the missing equation is the supply of labour. Modigliani in his article invokes Keynes's assumption of the given money wage. More specifically, the money wage is given if, and only if, the economy is at a level of output less than full employment. We know that in the neoclassical analysis the supply of labour is a function of the real wage $N = F(w/P)$ so the money wage can be written as $w = F^{-1}(N)P$ Formally, Modigliani stated his condition in the labour market in the following way:

$$w = aw_o + bPF^{-1}(\mathrm{N})$$

where, $a = 1, b = 0$ if $N < N_f$ and $a = 0, b = 1$ if $N = N_f$

The last equation indicates that if the current employment in the economy is smaller than full employment (N_f), then Keynes's view for the rigidity of money wage holds indeed, that is we have ($a = 1$ and $b = 0$). Money wage is viewed as a "datum, a result of history or of economic policy or of both" (Modigliani 1944, p. 47). If, however, the economy is at full employment, then the money wage becomes flexible ($a = 0$ and $b = 1$) and the last equation becomes an ordinary supply of labour function. Consequently, the money wage will be determined from the supply of labour at the point of full employment.

In Modigliani's presentation we find that the central assumption is the rigidity of the money wage, an assumption which, as with the liquidity trap does not really find any justification in the *General Theory*, where the nominal wage is being used simply to determine the price level. By contrast, in Modigliani's presentation the

nominal wage has another important role to play. This is revealed if we express Modigliani's system of simultaneous equations in terms of wage units or alternatively in terms of labour commanded.[10] Thus, we have:

$$\frac{I}{w} = \frac{I}{w}\left(i, \frac{P}{w}X\right) \text{ investment is given in terms of labour commanded}$$

$$\frac{S}{w} = \frac{S}{w}\left(i, \frac{P}{w}X\right) \text{ savings is given in terms of labour commanded}$$

$$\frac{I}{w} = \frac{S}{w} \text{ equilibrium in the goods market}$$

$$M_o = L\left(i, w\frac{P}{w}X\right) \text{ equilibrium in the money market}$$

$$\frac{Y}{w} = \frac{P}{w}X \text{ income given in terms of labour commanded}$$

$$X = F(N) \text{ the production function, which is by definition in real terms}$$

$$\frac{w}{P} = F'(N) \text{ the real wage is equal to the marginal product of labour}$$

$$\frac{w}{P} = F^{-1}(N) \text{ the supply of labour}$$

Hence, we have a system of eight equations and eight unknowns (I/w, S/w, i, Y/w, X, N, w, P). If, for a moment, we disregard the fourth equation and focus our attention on the remaining seven equations, we observe that these can determine all the variables but one, that is the money wage. The result is that the supply of money determines the money wage; since this is the only variable that remains to close the system. Such a determination is due to the quantity theory of money. Consequently, Modigliani's system of equations is dichotomised into the real economy – which includes all the equations except the fourth one – and the money economy, that is the equation of equilibrium in the money market. The real economy gives solutions in real terms (seven equations with seven unknowns, that is I/w, S/w, i, Y/w, X, N, w/P) while the money supply:

$$M_o = L\left(i, w\frac{P}{w}X\right)$$

[10]Hence, we ascertain one useful application of Smith's labour command theory of value.

determines the nominal wage, since the other variables are determined in the real economy. Consequently, the money supply determines the nominal wage and through the real wage it also determines the general price level. Thus, monetary policy may affect real magnitudes in the Keynesian model, contrary to Hicks's reasoning according to which the money supply does affect the real economy.

Modigliani's analysis leads to the conclusion that flexibility in prices and money wages establishes full employment in the economy. The mechanism that restores full employment works as follows: the existence of unemployment drives down nominal wages, and, therefore, incomes fall. The demand for money for transaction purposes, being directly related to income, falls as well, and with a given supply of money the rate of interest falls as well. From there on investment increases and the economy moves toward the full employment level of output. Modigliani managed to formalise Keynes's argument about the results of the flexibility in money wages. It is important to stress that the pivotal variable in this formalisation of the theory of employment is the idea of inflexibility of money wage. A corollary of this theory is that the role of money is not neutral. For example, the increase in the supply of money affects the price level and reduces the interest rate and thus output and employment are increased. If the nominal wage were perfectly flexible, then money's role would be neutral since it does not influence the interest rate i, or the liquidity preference L and output remains the same. Consequently, under conditions of a fully flexible nominal wage the increase in the supply of money leads only to an increase in the general price level. Consequently, Modigliani concludes that Keynes's theory works only in case of inflexibility of the nominal wage. If, however, the money wage is flexible then we derive the usual neoclassical results, where the real economy determines the level of output and employment and the money economy determines the nominal variables of the economy. This does not imply a rejection of Keynes's theory; on the contrary, economists accept the idea of inflexibility of the money wage as a stylised fact of modern economies and thus, Keynesian policy is viewed as both theoretically valid and necessary. The problem, however, relates to the theoretical consistency of the Keynesian system that once again became a special case of the general neoclassical model according to which the economy exhibits a sufficient flexibility in prices of commodities and the factors of production.

11.5 Summary and Conclusions

In an overall evaluation of the two models we see that they both represent aspects or partial arguments of the *General Theory*. Nevertheless their major problem in terms of the *General Theory* is the simultaneity issue and also the treatment of uncertainty. In Hicks's article we find an explanation of unemployment and recession as a result of the liquidity trap, which differentiates Keynes's theory from the (neo) classical one. In Modigliani, by contrast, the recession is the result of the inflexibility of money wage and not of the lack of effective demand. Commenting on this kind of

revision of the *General Theory* Paul Samuelson in the third edition of his popular text *Economics*, notes:

> In recent years 90 per cent of American economists have stopped being "Keynesian economists" or "anti-Keynesian economists". Instead they have worked towards a *synthesis* of whatever is valuable in older economics and in modern theories of income determination. The result might be called neoclassical economics and is accepted in its broad outlines by all but about 5 per cent of extreme left-wing and right-wing writers". (Samuelson 1955, p. 212).

These efforts to cast Keynesian theory in terms of *IS–LM*, Samuelson called *neoclassical synthesis*, since it puts together the neoclassical analysis of investment and savings and the market for labour with the analysis of Keynes about the interaction between the money market and the real level of economic activity. The neoclassical synthesis became the dominant presentation of the *General Theory*. According to this view, when there is unemployment, then we have Modigliani's supply of labour function with $a = 1$ and $b = 0$, and as a result of the exercise of appropriate monetary and fiscal policy the economy approaches the level of full employment. When the economy approaches the level of output that corresponds to full employment, then once again the neoclassical theory becomes relevant.

> If our central controls succeed in establishing an aggregate volume of output corresponding to full employment as nearly as is practicable, the classical theory comes into its own again from this point onwards. (Keynes 1936, p. 378)

In general, economists of the neoclassical synthesis argue that although the economy returns to full employment through the price mechanism, nevertheless this is a long run process. Consequently, for immediate results active fiscal and monetary policies are necessary.

Modigliani's ideas, which became the foundation of the neoclassical synthesis, and which essentially constitute a Marshallian partial equilibrium approach, became the object of criticism from Walrasian authors. They posited the following question: how is it possible to have equilibrium in all the markets but one? The protagonists of this critique of the neoclassical synthesis are Alex Lejonhufvund and Robert Clower, whose contributions we discuss in the next chapter.

Other criticisms included the phenomena of unemployment and later of the stagflation in the late 1960s or 1970s. Some economists, the monetarists for example, tried to fix the weaknesses of the model and others such as the New Classical economists claimed that the premises on which the *IS–LM* framework is based are dubious, while New Keynesian economists in the 1980s revived the old Keynesian models by injecting realism and by basing them on microeconomic foundations which simply were not used in the initial models. Whatever happens to the current macroeconomic debates and the various criticisms launched against the *IS–LM* models, one thing is certain, that these will continue to be part of the formal education of future generations of economists.

Questions for Discussion and Thought

1. Write down the Hicksian system of equations.
2. Draw a graph with the IS–LM system of equations and assuming a disequilibrium situation describe the dynamics of attaining equilibrium.
3. To what extent does Hicks's model represent Keynes's *General Theory*?
4. What was Keynes's reaction to Hicks's *IS–LM* representation of the *General Theory*?
5. Discuss Modigliani's Neoclassical Synthesis. To what extent does his model differ from Keynes's?
6. What are the major similarities and differences between Hicks's and Modigliani's models?
7. Critically evaluate the following statement: "I will assume all markets with the sole exception of the labour market that are in equilibrium".

Notes on Further Reading

Hicks's (1936) article is easy to follow, but the exposition of ideas (not "visions") is really dry. The reader discerns an effort on the part of Hicks to express Keynes's ideas in terms of equations and graphs without, however, the proper textual documentation. As we have mentioned, at the time that Hicks presented his article in the econometric society meeting at Oxford, two other related papers were presented in the same meeting by Meade and Harrod. Darity and Young (1995) present the details of these three articles and claim that Hicks had already read both of the other's papers prior to presenting his own. Darity and Young (1995) also note "there are grounds for believing that he also had read a third related paper by David Champernowne (1936)". From our point of view what is certain is that all four papers were systematic efforts to formalise the difficult and often confusing content of the *General Theory* and make it accessible to other economists. Whether or not these four papers were written independently of each other the fact is that all of them were the product of a period that many researchers were trying to offer answers to the same question. Consequently, it comes as no surprise that there was overlap in the models, since the authors knew each other well and were meeting regularly to present results of their research.

Hicks's model had the advantage over the others not only for his graphic presentation that made it accessible to the public "but also that he could use a single apparatus to draw a series of pictures that represented the classical and Keynes – whether accurately or not – as special cases of a more general model" (Darity and Young 1995, pp. 11–12). Patinkin (1990) claimed that Keynes approved of Hicks's presentation as this can be judged from his correspondence. In his letter with regard to Hicks's paper (1936) he notes that "I find it very interesting and really have next to nothing to say by way of criticism". However, we know that in the same letter Keynes went on to comment on several points in detail. Barens and Caspari (1999) also discuss

(continued)

this letter and identify four important criticisms that Keynes makes to Hicks's article. It is curious how major authors such as Patinkin fail to appreciate the importance of these criticisms and opt to cling instead on the seemingly approving opening remark of Keynes's letter. Keynes's clearly disapproved of Hicks's formulation and the various misinterpretations of his *General Theory* led him to his *Quarterly Journal of Economics* article, where Keynes (1937), once again, restated his theoretical differences from the classics and so he distanced himself from *IS–LM* formulations.

De Vroey (2000) focuses on the relation between Hicks's and Modigliani's papers and he concludes that Hicks original models have less in common with the subsequent developments in the *IS–LM* literature which are based more on Modigliani's paper. In fact, Modigliani's paper is included in Klein's (1963) influential book and also in Hansen's (1953) book that popularised Keynesian theory. Blanchard (1989) discusses the evolution of the neoclassical synthesis which sometime in the late 1960s displayed "anomalies", in Kuhn's sense of the term, that is there appear inconsistencies with the fundamental propositions of the approach that question its overall validity. The monetarist efforts, kept the *IS–LM* apparatus alive, but in the 1970s the New Classical economists (Lucas, Sargent, *inter alia*) discarded such representations of the economy "for their lack of sound microfoundations consistent with new classical standards" (Blanchard 1989, p. 634). In the 1980s the *IS–LM* models were revived through the work of New Keynesian economists, who by taking into account the new classical criticisms injected realism in the models. For the applications of *IS–LM* models we also recommend Vercelli's (1999) article.

Keynes's idea about current and expected interest rate is spirited away in both Hicks's and Modigliani's presentation and there is no uncertainty in the Keynesian sense, that is, as the purely incalculable risk. Only Tobin (1958) with his asset holding approach and the choice of risky monetary assets and money individually is in the spirit of Keynes.

Chapter 12
Disequilibrium Macroeconomics: From Its Brief Ascent to the Rapid Decline

[M]oney buys goods and goods buy money but goods do not buy goods.
(Clower 1969, pp. 207–208)

This unnatural use of language clouds the whole argument, but with care and patience the reader can work out a commonsense interpretation of what all means.
(Kahn 1977)

12.1 Introduction

The neoclassical synthesis prevailed for decades as the authentic interpretation of Keynes. This dominance is attributed, on the one hand, to the simplicity of the presentation of Keynesian ideas and, on the other hand, to the provision of the necessary rationale for the contemplation of fiscal and monetary policy. The neoclassical synthesis (of Hicks, Samuelson and Modigliani) was considered the true general theory, as it could be applied to any stage of the business cycle and not only during depressions, as the *General Theory* of Keynes. The depressions are caused either by rigid money wages, or because the rate of interest is in the liquidity trap region, or, perhaps even worse, a combination of rigidity of money wages and liquidity trap-interest rate. But once the economy is at its full employment stage, through appropriate government intervention, then the system operates in accordance with the (neo)classical theory.[1]

Robert Clower was among the first economists who criticised the usual formulation of the Keynesian model of the neoclassical synthesis. In particular, Clower (1965) argued that there is a logical inconsistency between the neoclassical synthesis and neoclassical economics, because in the neoclassical synthesis, we have equilibrium in all markets except the labour market, as a result of the rigidity of

[1]Keynes himself notes this case: 'If our central controls succeed in establishing an aggregate volume of output corresponding to full employment as nearly as is practicable, the classical theory comes into its own again from this point onwards' (Keynes 1936, p. 378).

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_12, © Springer-Verlag Berlin Heidelberg 2010

nominal wages. Neoclassical economics, however, posits through the Walras Law that equilibrium in all markets but one is impossible. A few years later, Axel Leijonhufvud extended the ideas of Clower, who was his teacher, in his book The *Keynesian Economics and the Economics of Keynes* (1967), a title that signifies the divergence between the Keynesian economics (of the neoclassical synthesis) and the economics of Keynes (*i.e.*, the economics of the *General Theory*). The investigation of Clower and Leijonhufvud was declared disequilibrium Keynesianism, since their model includes both unemployment and the lack of effective demand as a result of insufficient information and lack of coordination of economic agents. Subsequently, the works by Edmond Malinvaud, around which a whole literature has developed, gave new impetus to the approach by emphasising the existence of disequilibrium rather than equilibrium and by constituting an effort in general for a more realistic macroeconomic approach.

In this chapter, we are dealing with the overall approach of Clower and Leijonhufvud, where the crucial distinction is between Walrasian and Keynesian models. Furthermore, we explain the difference between notional and effective demand and the importance of this difference for the development of macroeconomic models of partial and general disequilibrium. The microeconomic foundations of macroeconomics and Malinvaud's contributions towards this direction together with the ensuing economic policy proposals conclude the chapter.

12.2 Walrasian Vs. Keynesian Models

Clower argued that the neoclassical synthesis failed to combine the price theory with the theory of income determination. In other words, the macroeconomic analysis of the neoclassical synthesis lacks sufficient microeconomic foundations. One wonders how it could be possible to develop such solid micro foundations for the macroeconomic theory of the determination of the level of output and to deduce from these foundations the disequilibria phenomena of unemployment and inflation. According to Clower, microeconomic foundations can be placed starting with the rejection of the notion of general equilibrium and of the idea that the economy works its way through the auctioneer towards exchange taking place at equilibrium prices. The neoclassical synthesis failed to raise this very fundamental difference of Keynes from Walras, that is, the economy (in Keynes) operates almost always out of equilibrium.

In order to illustrate this difference, that is, the neoclassical synthesis Keynesianism, and the version of the economics of Keynes presented by Clower and Leijonhufvud, it is necessary to review some of the important aspects of the way in which general equilibrium functions in a Walrasian system, where the auctioneer holds centre stage, and whose role is to announce prices and then to take into account the reactions of economic agents. The auctioneer keeps track of the differences between supply and demand for each price that he announces to potential traders and ensures that for the next set of prices which he will announce

these differences will get smaller and so forth. The auctioneer concludes this procedure whereby individuals grope towards the equilibrium prices, a process known as tâtonnement, when the difference between supply and demand for each good is zero. From this point onwards transactions take place at these (and never off) equilibrium prices. In addition, the agents of the economy are price takers and decide only for the quantities of the given stock of resources that they will offer or will demand.[2] Finally, in the Walrasian conceptualisation of the economy, from the point of attainment of equilibrium prices onwards, all goods have the same degree of liquidity, because anyone can seek to sell and buy any quantity of goods. Therefore, all goods function both as goods and as money (medium of exchange), which is equivalent to saying that all goods are readily accepted in exchange, their only difference with money proper being that they are consumed directly (Leijonhufvud 1967, p. 405).

The analysis shows that the Walrasian conceptualisation of the functioning of the economy renders superfluous the need for money, as a means of exchange or precaution. Transactions always take place at equilibrium prices, and, therefore, there is no excess demand or excess supply in any market. The auctioneer however is merely an analytical device since actual transactions do not take place in equilibrium prices. Consequently, barter is impossible[3] in an advanced society and thus the owner of good A would not prefer to exchange it against good B, of which he needs only a portion, regardless of whether the remaining portion of B, which he does not need, may be exchanged against another good which is desired. The individual would prefer to exchange good A with a good that is generally accepted in all markets; in short, money would be the preferred means of exchange. As a consequence, Clower and Leijonhufvud conclude that money is necessary in an economy, where transactions are conducted in non-equilibrium prices and that Keynes's analysis is in terms of historical (real) time and disequilibrium conditions.

12.3 Effective Demand and Notional Demand

Clower and Leijonhuvfud contend that Keynes sought to show the difficulties of coordination of a system based on the blind interplay of market forces. In contrast, the neoclassical synthesis is an attempt to incorporate the *General Theory* in the neoclassical paradigm, where Walras's law holds. However, Keynes's analysis does not require the auctioneer, and, therefore, transactions take place at disequilibrium. In such an analysis, Clower and Leijonhuvfud argue that the short end of

[2]Walras realised that exchange at prices out of equilibrium would change the endowments and so the economy would not be led to the theoretical equilibrium, so he had to assume that no exchange takes place at disequilibrium prices.

[3]Barter is at best an analytical concept rather than a reality in historical time. Exchanges in all societies take place through the use of money.

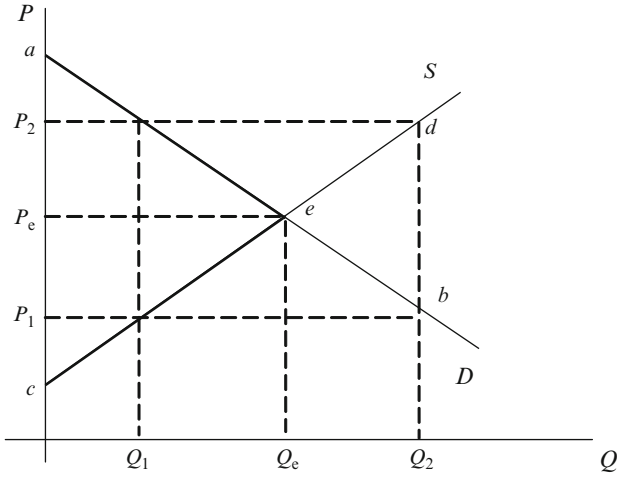

Fig. 12.1 Notional vs. effective demand and supply

the market represents real exchanges, while the long end of the market shows the rate of rationing (see Fig. 12.1).[4]

According to Clower at price P_2, producers would like to offer the quantity Q_2, while consumers would be willing and able to buy only the quantity Q_1. Therefore at price P_2 consumers would like to buy quantity Q_1, which is closer to the vertical axis and so the demand side, and, therefore, the buyers dominate the market. By contrast, at price P_1, the demand is equal to Q_2 and Q_1 is the quantity supplied, this time, the quantity supplied is closer to the vertical axis, and, therefore, the suppliers dominate the market. The quantity Q_2-Q_1 shows the degree of exclusion, when the market is out of equilibrium. In such a case, we should distinguish between effective demand and supply and notional demand and supply. The effective demand represents desires limited by the ability to buy or sell, whereas the notional demand refers to the unlimited desires. The curve *aec* represents the effective exchange curve and the curve *deb* represents the notional exchange curve.

Only in the Walrasian description of the economy do the notional and effective curves coincide in point *e*. Clower (1976) calls this case, the unitary decision hypothesis, according to which the decisions of economic agents regarding demand and supply coincide. In the neoclassical synthesis, for instance, the decisions on labour supply are perceived automatically as demand for goods, while in reality the decisions for sales are not automatically translated into decisions for purchases, since the expected consumers expenditures depend on current income. Thus, the description of the economy in Keynes, in the general case, the effective demand falls short of the notional demand. In such an economy the decision of each

[4]As it may happen with gas in case of scarcity or with markets functioning through rationing systems in abnormal times such as occupations, wars, and the like.

individual to sell is not necessarily transformed into a decision to buy, because he must sell first and then buy. This case Clower (1976) characterised as the dual decision hypothesis because individuals can choose from a range of market prices and decide, first for their notional demand (and supply) and second for their effective demand (and supply), because they depend on the quantities that can be offered from the given stock of resources. Thus, the notional and the effective demands do not coincide due to the fact that these include the volume of given resources that can be sold.

As a consequence, the decisions of individuals are divided into two categories: Firstly, the notional decisions, where individuals maximise their utility and decide on the quantities of goods that they purchase and the quantity of labour that they offer; and the effective decisions according to which the quantities purchased by the individuals are limited by their ability to sell. For example, the individual may not offer as much labour as he would wish, and, therefore, his demand for goods will be lower than that he would actually like. Overall we would say that the transactions that take place at disequilibrium prices fall short of those that take place at equilibrium prices. As a consequence, in an economy the output produced and employment in general will be less than their full employment levels.

The analysis is based on the specific price that is being selected and the question that arises is to what extent, if any, a reduction in the nominal wage would lead to equilibrium. Clower and Leijonhufvud would rule out this possibility on the basis of the difference between effective and notional (or planned) demand, a difference which is the very nature of the current money economies. More specifically, Clower and Leijonhufvud assume a case of unemployment, where workers would like to buy goods, if firms were willing to employ them. On the other hand, entrepreneurs would be willing to sell their goods provided that they knew there is an effective demand for them and without any doubt they would increase their employment in order to produce the extra goods for which they had secured the demand. In the case of the real economy, workers would exchange their work for the produced goods. But in a monetary economy, workers are paid money wages and not a portion of the output produced. The entrepreneur can respond to the payment of money wage only if and when there is demand for the produced goods. But there is no way for workers to communicate and above all to convince entrepreneurs of their intention to buy the extra goods produced. Entrepreneurs facing the lack of adequate demand are reluctant to increase their employment; on the other hand workers seek employment, in vain, though, because of the monetary character of modern economies!

Thus the presence of money complicates the supply and demand relations in the Walrasian models. The fall in wages does not mean necessarily an increase in employment, because employers are simply interested to sell their goods, and if they cannot be assured that there is a demand for their goods, they have no incentive to increase their employment. Therefore, the problem in the economy is not the lack of price flexibility but rather the failure of communicating notional quantities as well as the lack of sufficient coordination (there is no Walrasian auctioneer to announce prices and to correct the imbalances of the real economy) between economic agents.

Although it is not immediately obvious, the preceding analysis has interesting implications for the formulation of economic policy that focuses on guiding the aggregate demand. Theoretically at least, the lack of information about the notional quantities demanded and supplied is 'solved' by limiting the number of participants, in an effort to facilitate communication among them. Obviously this cannot be achieved at the macroeconomic level, and, therefore, the only possible solution is the management of demand in desired directions. The idea is that a temporary increase in demand, past a point, feeds on itself; in as much as, the restriction of the labour market is lifted and the economy is directed towards full employment. At the point where the real and notional quantities coincide, at the same time Walras's law holds.

There is no doubt that Clower and Leijonhufvud tried to synthesise a realistic theory and that the criticisms that they levelled against the neoclassical synthesis are meaningful and on target. But the problem is that in their theoretical construction there is no textual evidence based on the *General Theory*. Clower and Leijonhufvud argued that Keynes was trying to escape from Walras since Walras's analysis is conducted in logical and not in real time. The truth is that Keynes, in the *General Theory*, is not so interested in Walras, mentioning him only once (*General Theory*, pp. 176–177). Clearly, Keynes, methodologically speaking, is a follower of his teacher Alfred Marshall and of the other classical economists (as defined in the *General Theory*) at Cambridge. As a consequence, Clower's and Leijonhufvud's criticisms may be meaningful to the extent that they are levelled against the neoclassical synthesis, but are far from a reconstruction of what Keynes actually said. In other words, the reconstruction of Keynes by the two authors is neither historical (in the sense of what Keynes actually said), or logical (in the sense of what Keynes should have said had he followed his assumptions strictly).[5]

12.4 The Microeconomic Foundations of Disequilibrium Macroeconomics

The analysis of Clower and Leijonhufvud, if anything, made clear that the neoclassical synthesis (of Hicks, Samuelson and Modigliani) was without microeconomic foundations based on the optimisation behaviour of economic agents. More specifically, the theorisation of micro–macro linkages between price expectations and quantitative adjustments became necessary. This research programme was adopted by macroeconomists of all approaches (Keynesian, New Classical economists, *etc*.) and certainly continues until today. Despite the fact that both Clower and Leijonhufvud as well as modern neoclassical macroeconomics stress the need for providing microfoundations to macroeconomic analysis, it is important to

[5]For the distinction between historical and rational reconstruction, see Blaug (1999) and also Chap. 6.

emphasise the distinguishing feature between them, that is that modern neoclassical macroeconomics theorise the economy in terms of equilibrium and not disequilibrium. One weakness of disequilibrium macroeconomics is that they do not explain in a satisfactory way the observed price rigidity, a weakness that Edmond Malinvaud in his *The Theory of Unemployment Reconsidered* (1977) sought to deal with.

Malinvaud's central position is that disequilibrium is not restricted to the goods and labour markets, as in the case of Clower and Leijonhufvud, but that disequilibrium extends to all markets. Malinvaud (1977) *inter alia* sought to examine the problem of unemployment by means of a rationing model and of a new approach to general macroeconomic disequilibrium and presents a typology of different equilibrium positions of an economy, characterised by rationing and two markets: consumer goods and labour. Malinvaud argues that the theory of unemployment is very closely linked to the theory of rationing and this happens because the presence of involuntary unemployment indicates that workers are constrained, given that labour supply exceeds labour demand. Furthermore, Malinvaud contends that such an analysis is only possible under conditions of general equilibrium since any rationing, for instance, in the labour market has an effect on the goods market, and *vice versa*. As a consequence, a partial equilibrium analysis would be meaningless. Malinvaud adopts the Hicksian distinction between a flexprice system which includes the prices of agricultural products, raw materials, *etc*. and a fixprice system which includes industrial products. Malinvaud argues that in modern economies and with the passage of time the fixprice system tends to become dominant (1977, p. 9).

It is important to note that the rigidity or flexibility in pricing makes sense in the short run, and thus it is logical for Malinvaud to assume the wages and prices as given, which is equivalent to saying that only quantities can change. Based on the analysis of Clower and Leijonhufvud the short end of the market determines the quantity exchanged and the long end of the market the quantity rationed. With this in mind, Malinvaud concludes that once we have fixed the prices and hence only quantities change, supply and demand cannot be equal (1977, p. 13). The different cases of market equilibrium with rationing are conveniently shown in Table 12.1.

In the case of two markets (goods and labour) we have four cases depicted in the four cells formed in Table 12.1 (Malinvaud 1977, p. 30). In the first cell we have the

Table 12.1 Typology of macroeconomic equilibrium

		Goods market	
		Buyers (*demand for goods*)	Sellers (*supply of goods*)
Labour market	Buyers (demand for labour)	Keynesian Unemployment $E_L < 0, E_G < 0$	Classical Unemployment $E_L < 0, E_G > 0$
	Sellers (*supply of labour*)	Underconsumptionist Unemployment $E_L > 0, E_G < 0$	Repressed Inflation $E_L > 0, E_G > 0$

case of Keynesian unemployment, which excludes sales in both markets. In other words, we have excess supply of both labour and goods or what is the same thing negative excess demand for labour (EL) and for goods (EG). In this case employment is rationed for workers and the rationing for business is on the sales of goods. The classical unemployment is illustrated in the adjacent cell, where buyers are rationed in the purchase of goods and sellers are rationed in the labour market. The high wages deter firms from employing more labour or selling more goods. Malinvaud cites the case of repressed inflation that appears in the third cell, where the buyers are excluded in both markets. In other words, there is excess demand in both markets; therefore the quantities actually exchanged are those of effective supply. Malinvaud calls this case repressed inflation because the price level does not respond to the given market and remains stable. In the underconsumption cell, we have the exclusion of buyers of labour and of the sellers of goods. This means that we have excess demand in the labour market and excess supply on the market of goods. As a consequence, the goods have been already produced, but because of unemployment there are not enough incomes (wages) to buy these goods.

These four cases are illustrated in Fig. 12.2, where the two markets (labour and goods) form the plane on the vertical axis of which we set the wage (w) while on its horizontal axis we set the price (p) of goods (Malinvaud 1977; Muellbauer and Portes 1978). We have excess demand for labour when the economy is right and below the equilibrium curve in the labour market and when is left and above the equilibrium line we have excess supply of labour. As for the equilibrium curve in the goods market, if the economy lies above and right of this curve we have excess supply of goods, while on the left and below it we have excess demand for goods. If prices were volatile, the economy would be at the Walrasian equilibrium point WE.

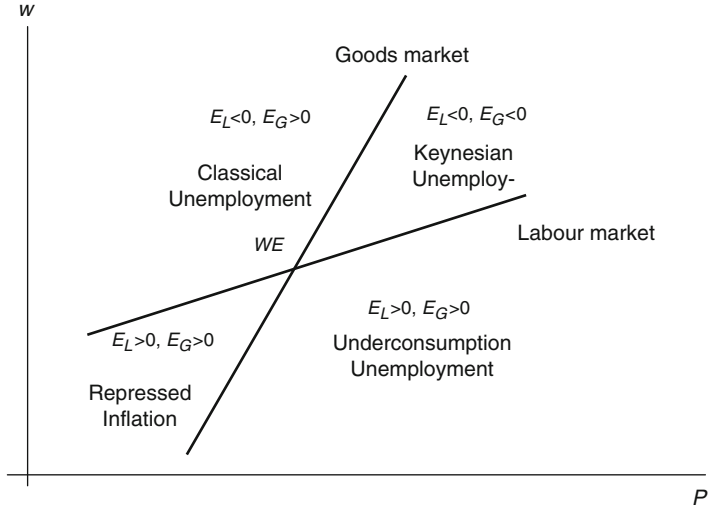

Fig. 12.2 Disequilibrium markets in the p, w plane

For example, in the case of Keynesian unemployment, we have excess supply of labour and goods. If prices of goods and the wage rate were flexible and could, for example, decrease, then the economy would reach the point WE. The same could have happened in the case of classical unemployment, where the attainment of equilibrium, that is, point WE would require a fall in wages and a rise in the prices of goods. In reality, however, because prices are not sufficiently flexible, it is possible to have different scenarios in different markets. Thus, it is possible to have cases where Keynesian unemployment coexists with classical unemployment. Therefore, one should be open to alternative explanations of the causes of unemployment. From this description it is interesting that prices and wages are considered almost stable in the short run and that quantities are those that redress the balance. As a consequence, in such an environment, disequilibrium is the rule and equilibrium the exception. In other words, Keynesian disequilibrium macroeconomics has shown that in the neoclassical equilibrium, the point WE, is only one possible case of 'equilibrium' among a number of other possible equilibrium points that the economy could find itself. In this way, the economists of disequilibrium macroeconomics have restricted the neoclassical synthesis into a special case of a broader Keynesian analysis conducted in real time.

From the four possible areas that the economy can find itself, Malinvaud pays more attention to the Keynesian and to the classical, and much less to the case of repressed inflation, while he regards the case of underconsumption unemployment as presenting only theoretical interest.[6] The case which is almost impossible to prevail is that of the Walrasian equilibrium, and this is because prices are characterised by rigidity due to full cost pricing practises followed by firms in the industrial sector of the economy. Therefore, full cost pricing combined with the rigidity in wages lead to a wage-price pair that falls into one of the three cases since the case of underconsumption unemployment can be ignored without problems.

12.5 The Effectiveness of Economic Policy

The above analysis leads to some interesting policy conclusions. By recognising the different situations that an economy can find itself, it follows that there will be different policy prescriptions to deal with each particular case. If, for instance, the economy is in classical unemployment, an expansionary fiscal policy can only displace the economy even further away from equilibrium. The same is not true, however, in the Keynesian unemployment, where an expansionary fiscal policy can have positive effects. In particular, the equilibrium curve in the goods market increases and thus shifts upward and to the right. Meanwhile, the labour market

[6]Malinvaud (1977, p. 93) refers to this case briefly and does not pay as much attention as to the other three cases of unemployment.

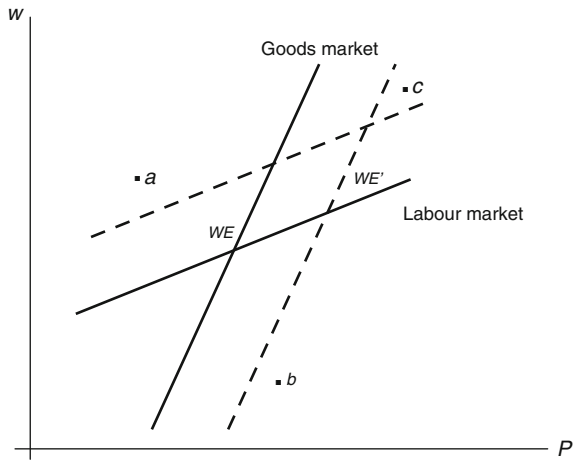

Fig. 12.3 Effects of an expansionary fiscal policy

equilibrium curve increases as well and thus shifts upward and to the left. The idea is that the expansionary policy increases both wages and prices, as businesses and households compete with the public sector in the labour market and in the goods market. As a consequence, an expansionary economic policy moves the economy towards the northeast direction and so the economy improves its position, that is, it gets closer to the equilibrium position. By contrast, if the economy is in the classical area or that of repressed inflation the position of the economy worsens.

The new position is described by the intersection of the dotted curves (Fig. 12.3) where we see that the new equilibrium position is WE'. Therefore, if the economy suffered from classical unemployment at the point *a*, or of repressed inflation at point b, an expansionary fiscal policy can only aggravate these two situations since the new equilibrium point is *WE*' which is even further away from points a and b. The same is not true, however, if the economy is in the Keynesian region where point c gets closer to the new equilibrium position, thus demonstrating the possibility of pursuing a successful economic policy with social and tangible results.

In this approach, and unlike the Keynesian approach of the neoclassical synthesis, prices and wages can also become policy tools with the aid of which the economy might move towards its equilibrium position. If, for example, the economy is in the classical unemployment region, a policy that leads to rising prices and falling wages or a combination thereof can become very effective in improving the position of the economy. Thus, for a range of money wages where prices are very low, we arrive at the classical unemployment area. Then we have also a range of nominal wages in which the economy is led to the region of repressed inflation and finally, for another range of money wages we have the case of Keynesian unemployment. As a consequence, we have three cases for the given wage and given price levels. When all these three cases coincide, we have what Malinvaud calls Walrasian equilibrium. Malinvaud argues that in modern economies, prices are characterised by rigidity. In many industry studies it has been observed that

entrepreneurs follow a pricing practise of simply adding a mark-up on the unit cost of production. Therefore, policies aimed at flexibility in prices and wages, such as the antitrust laws and the liberalisation of markets by removing various obstacles to their operation may have a positive effect.

12.5.1 The Critique of Richard Kahn

Kahn (1977) in his evaluation of Malinvaud's model formulated an empirical critique on the hypothesis of price rigidity. In particular, he shares Malinvaud's view of price rigidity as realistic to the extent that it refers to industrial products, where the supply of goods is variable, and, therefore, is under the control of business. However, a feature which Malinvaud took no account of is that industrial production is characterised by constant or increasing returns to scale. According to Okun's law the increased level of activity leads to higher production and lower average production costs. Assuming increasing returns to scale then, it is reasonable for the supply curve to display a negative (and not as Malinvaud assumes positive) slope, which means that output is limited only by demand. In other words, unemployment is always Keynesian when we have rising (or even constant) returns to scale production. As unemployment cannot be classical it follows that the reduction in the real wage not only does not increase employment, but it may even worsen rather than improve the economic situation.

If we assume decreasing returns to scale then we have cost curves with positive slope. This case (which Malinvaud does not address) is common in agricultural production, where prices are flexible. Therefore, in the case of Malinvaud we can have either a supply curve with a positive slope and price flexibility or a supply curve with negative slope but rigid price. According to Kahn, we are faced with a dilemma the only solution to which is the outright rejection of Malinvaud's approach.

12.6 Summary and Conclusions

The analysis of Clower and Leijonhufvud presents the following paradox: we know, from Adam Smith, that monetary economies allow and facilitate the further division of labour which in turn leads to an increase in labour productivity, *etc*. Therefore, the more monetised the economy, the greater the division of labour and the higher the productivity. The analysis of Clower and Leijonhufvud suggests exactly the opposite and the logical implication of the theory is that money makes an economy less efficient than barter economy does!

In an effort to investigate further the logical implications of the theory, let us suppose that the entrepreneurs are assured that every euro that they will pay in wages will be spent on output. Even in this case, entrepreneurs would be reluctant to increase their hiring, simply because wages may not be sufficient enough to buy

the output produced. This applies even in the case where all profits are spent. The idea is that there is always some portion of wages that can be saved and thus to have goods produced but not sold. So, the problem as Clower and Leijonhufvud argue is not the lack of adequate coordination, because even in the case of coordination disequilibrium persists. Therefore, better coordination does not really solve the problem.

The typical model of Clower and Leijonhufvud is that of equilibrium in the goods market and disequilibrium in the labour market, which seems to contradict Walras law. But in terms of notional demand or supply, there will be excess demand for goods and excess supply of labour. All these show that, given the marginal propensity to consume, we cannot increase employment, because of the increased production of consumer goods. Although the employment in the consumer goods sector increases, nevertheless total wages are less than total income (output). Therefore, if the marginal propensity to consume and invest is given, then the employment cannot be increased despite the increase in the production of consumer goods. With regard to investment the question is why it does not increase so as to lead the economy to full employment? Clower and Leijonhufvud argued that the high interest rate is the reason for this failure; such a view makes their model essentially similar to that of the neoclassical synthesis.

In Malinvaud's approach, disequilibrium is not limited to one market, but expands to include all markets; furthermore, we have for the first time systematic efforts to bridge the schism between micro and macroeconomics. A schism, which to our opinion, did essentially appear "in the years of high theory" to recall the characterisation of Shackle (1967), for the period beginning in 1926 and continuing until the Second World War. During this time period, we had the monopolistic competition revolution which attempted to form a theory of the firm under partial equilibrium and thus without the need for a macroeconomic analysis, while in the middle of that period we had the macroeconomic revolution, where the macroeconomic analysis is conducted without the necessary microfoundations. There is no doubt that the research of the general disequilibrium macroeconomists contributed more than anything else to the shift of interest in the development of macroeconomic analysis based on microeconomic foundations. This switch in the research agenda of many economists is the most significant accomplishment of this approach. Thus, the disequilibrium macroeconomics may be characterised by inventiveness and unusual vocabulary; nevertheless, it managed to incorporate its research questions in the analyses of the New Classical economists and of the economists of the real business cycles that we deal with in the next chapters, while their persistence on rigidities in prices and wages was adopted by the New Keynesian economists. Today we can say that the approach of economists of the disequilibrium macroeconomics such as Malinvaud is neglected by the vast majority of economists; nevertheless, the current macroeconomic research has been fed with new arguments and research questions of the disequilibrium macroeconomics. It seems that this approach has dropped to the stage of a degenerate research programme, however, the issues raised by the economists of disequilibrium macroeconomics and their insistence on the analysis of disequilibrium is, in our opinion, evidence of a realistic approach to macroeconomics

and it comes as no surprise that aspects of this approach have been integrated in other macroeconomic schools of economic thought.

Questions for Thought and Discussion

1. With the help of a graph show the distinction between notional and effectual demand and supply.
2. What exactly is the dual decision hypothesis?
3. In what sense if any, does Clower's and Leijonhufvud's approach differ from the neoclassical synthesis?
4. What are the main differences between the views of Clower and Leijonhufvud and those of Malinvaud?
5. Through a graph discuss the different cases of unemployment according to Malinvaud and rank them in order of significance.
6. Discuss the effectiveness of fiscal policy according to Malinvaud.

Notes for Further Study

In Clower's (1965 and 1967) model there is the distinction of the dual decision hypothesis whereby the factors of the economy should first obtain liquidity (*i.e.*, purchasing power) before they activate their notional demands and supplies. It is particularly interesting to note that Clower (1967), although he presents no textual evidence from the *General Theory*; nevertheless, he does not hesitate to claim that Keynes 'either had at the back of his head the case of dual decision or most of the General Theory is theoretical nonsense'. For Clower it is clear that Keynes's economics are inconsistent with the spirit of the Walrasian general equilibrium and that only by escaping from this is there a possibility of developing innovative theories.

Leijonhufvud (1968) extends and further clarifies the approach of his teacher by arguing that the disequilibrium in full employment is hampered by deficiencies in liquidity, real capital, adjusted expectations and rigidities in prices and wages. For a critical evaluation of the views of Clower and Leijonhufvud we suggest the article by Coddington (1976), who argues that the Clower and Leijonhufvud interpretation is not about the economics of Keynes, but rather about what in their opinion the economics of Keynes should be.

Barro and Grossman (1976) initially, but much more thoroughly Malinvaud (1977) later, extended the analysis of Clower and Leijonhufvud to include all markets. Barro (1979) criticised these approaches, including his own on the basis of rational expectations. For a severe criticism of Malinvaud we recommend the article by Kahn (1977). In this article Kahn (the student of Keynes) believes that the general disequilibrium approach ignores some central aspects of the analysis of Keynes. These include: the liquidity preference function, the role of uncertainty and the demand for investment, which is regarded as given. Kahn (1977) nevertheless recognises that this approach is

(continued)

susceptible to improvements in the three core issues mentioned in Sect. 12.4. Until these improvements take place Kahn regards this approach as a set of commonplace ideas, and the book by Malinvaud (1977) 'as a sad example of the effect that the study of orthodox economics can have upon a powerful mind' (Kahn 1977 [1983], p. 225). In the context of neoclassical economics Hahn (1977) objected to adopting the rigidity of prices and argued that even though the observed price rigidity is an economic reality, nevertheless it should be theorised and not simply taken for granted. Eatwell (1979) considers that neither Malinvaud nor Hahn take into account the relationship between savings and investment, an issue so central in the analysis of Keynes.

Schefold (1977) makes an evaluation of the general disequilibrium models in the same spirit as that of Kahn. In his article Schefold on the one hand clarifies some aspects of the argument in Kahn, and on the other hand suggests theorisations of the markup either in the spirit of Malinvaud or of the classical theory. Classical economists did not view the mark up as arbitrary, and, therefore, as evidence of monopoly power, but as a result of the competitive process in which fluctuations in demand are reflected in changes in the degree of capacity utilisation rather than in deviations of market prices from equilibrium prices. As a consequence, if we accept Kahn's criticism, then we either have to turn to the largely unknown routes proposed by Malinvaud (1977) or to return to the classical analysis. Schefold, of course, would opt for the classical analysis.

Chapter 13
The Rise and Fall of Monetarism

Inflation is always and everywhere a monetary problem in the sense that it is and can be produced only by a more rapid increase in the quantity of money than in output.
(Milton Friedman 1973, p. 28)

No country succeeded in stemming inflation without adopting measures directed at restraining the growth of the stock of money.
(Milton Friedman 1959, p. 2)

I would like to say to Milton [Friedman] and Anna [Schwartz]: Regarding the Great Depression. You're right, we did it. We're very sorry. But thanks to you, we won't do it again.
(Ben Bernanke, the Federal Reserve chairman, apologised to Friedman on his institution's behalf in November 2002)

The theoretical foundations of monetarism [. . .] do not lie in the realm of monetary theory, but in theories of the determination of output.
(John Eatwell 1983b, p. 203)

13.1 Introduction

The purpose of this chapter is to present and critically evaluate the major propositions of the monetarist school of economic thought, that is, the school of economic thought according to which the quantity of money is the utmost important economic variable whose changes affect the behaviour of the entire economic system. Although the characterisation Monetarism was coined by Karl Brunner in 1968 to describe a school of economic thought that includes, besides Milton Friedman (1912–2006) and Anna Schwartz (1915–), Bruner himself and Allan Metzler among others, it is in fact a very old school of economic thought since its traces can be found in early nineteenth century. The University of Chicago, where monetarism was developed, clings to a free market tradition that restricts government's intervention should be kept to a minimum and seeks to explain the major economic phenomena through a single variable, the supply of money.

The emergence of monetarism and its establishment in the late 1960s and early 1970s required the fulfilment of a number of preconditions, the more important of

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_13, © Springer-Verlag Berlin Heidelberg 2010

which was the total or partial failure of the established Keynesian orthodoxy to provide satisfactory answers to the simultaneous coexistence of inflation and unemployment, a phenomenon which came to be known as stagflation and which led to the collapse of the central idea that was associated with Keynesian economics of the neoclassical synthesis; that is, the inverse relationship between inflation and unemployment as is portrayed in the famous Phillips curve. The latter from the time of its appearance was to become a major topic of modern macroeconomics and contributed as no other empirical relation to the emergence of monetarism and the decay of Keynesian economics. In what follows, we are dealing with the Phillips curve in its initial version and the subsequent augmentation with expectations by Friedman. Next, we discuss the quantity theory of money in both its traditional as well as its modern monetarist version, as a theory of the demand for money. We also discuss the international aspects of the economy focusing on the balance of trade and the determination of exchange rates. We conclude with the reactions of Keynesians to the monetarist claims and with a summary of the major ideas of monetarists and the effectiveness of economic policies.

13.2 The Phillips Curve

The Phillips curve refers to the inverse relationship between inflation and unemployment, a relationship that was supposed to form one of the cornerstones of post-war macroeconomic analysis. Alban W. Phillips (1914–1975) in his famous 1958 article examined time series data of the rate of unemployment and the growth rate of the money wage in the UK for a period spanning almost a century (from 1861 to 1957). The idea was 'to see whether statistical evidence supports the hypothesis that the rate of change of money wage rates in the United Kingdom can be explained by the level of unemployment and the rate of change of unemployment' (Phillips 1958, p. 284). The intuition here is that low unemployment strengthens workers' bargaining position for higher wages and *vice versa*. The data showed an inverse quasi-linear relation (curvilinear) between the two variables of the form $\hat{w} = 9.638u^{-1.394} - 0.900$ which is shown in Fig. 13.1

Where $\hat{w}$ is the growth rate of the money wage (which later was replaced by inflation), and u is the unemployment rate. The higher wages imply higher cost of production, so firms increase their prices leading to an inflationary process. It is important to stress at this point that the Phillips curve in the beginning was viewed as a long-run relation, but soon this perception changed and the Phillips curve was transformed to a short-run concept according to which, given the unemployment rate, one could estimate the growth rate in the money wage (or inflation) according to the above formula. The Phillips curve in the 1960s was already used for policy purposes showing the impossibility of achieving simultaneously price stability and reduction of unemployment. The two variables were found to be inversely related. For example, a 5.5% rate of unemployment is associated with a nearly zero growth rate of the money wage whereas when unemployment is reduced to 2.5%, the growth rate of the money wage increases to 2%, as shown in Fig. 13.1.

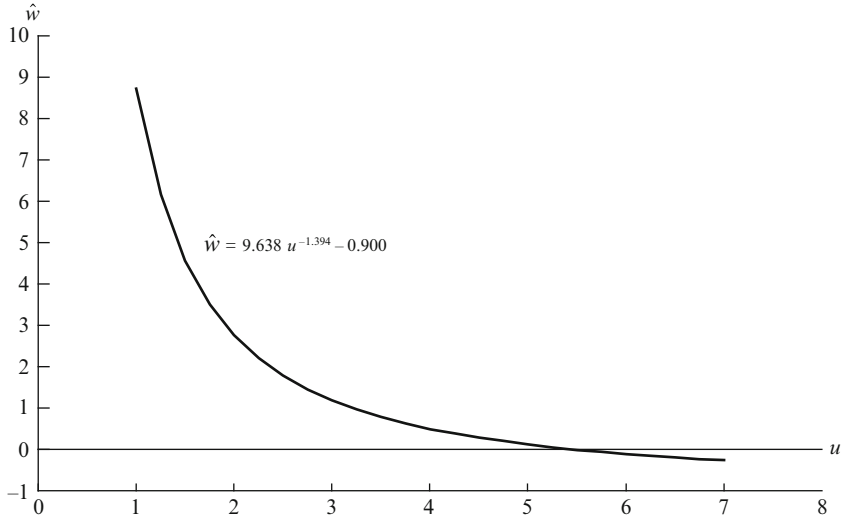

Fig. 13.1 The Phillips curve

The relationship that Phillips extracted from the time series data of the UK was purely statistical and by no means theoretical. The data, in other words, do not allow statements of the type that the fall in the rate of unemployment leads to increases in money wages, and *vice versa*. We know that the statistical relationships do not necessarily show causality, and, therefore, do not constitute theoretical relationships. This weakness, however, did not prevent Keynesians of the neoclassical synthesis to adopt a purely empirical relationship between unemployment and the growth rate of money wages and to convert it to a basic, for their argument, theoretical relationship between the two variables for short-run policies. More specifically, Samuelson and Solow (1960) replaced the growth rate of money wage by inflation. This idea was adopted from the majority of economists during the decade of 1960s, as an interpretation of the relationship between inflation and unemployment. In the end of the decade of 1960s, the statistical data of the USA and UK, however, showed that inflation and unemployment move in the same direction. This means that the Phillips curve does not necessarily have the usual negative slope, and thus there is no trade off relation between the two variables. The simultaneous occurrence of inflation and unemployment (stagflation) essentially appeared in the mid-to-late 1960s, and, after the oil shock of 1973–1974, the phenomenon became simply even more pronounced. This was the period that monetarism really was at its heydays as a modern school of economic thought.

13.2.1 Short-Run and Long-Run Phillips Curve

It follows from the above analysis that inflation and unemployment rates are either not related or they are related in a complex and not in a simple way as the Keynesian

economists thought. If we therefore assume that the two variables are independent of each other, then how does one explain the simultaneous rise of both inflation and unemployment in the late 1960s and the 1970s? The monetarists' answer is on the basis of the distinction between the short-run and long-run Phillips curve associated with the monetarist concept of natural unemployment rate. According to this analysis, equilibrium in the labour market comes through changes in the real wage. The nominal wage is characterised by inflexibility at least in the short run, in the long run; however, the nominal wage is flexible enough so as to lead the economy to the full employment level of output. In this perspective, the full employment level of labour does not correspond to zero unemployment but to a level of unemployment, which is called natural, and it is always higher than zero. More specifically, Milton Friedman (1912–2006) defined the natural level of unemployment as

> [...] the level [of unemployment] that would be ground out by the Walrasian system of general equilibrium equations, provided there is embedded in them the actual structural characteristics of the labour and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labour availabilities, and so on (Friedman 1968, p. 8).

The natural unemployment will be always greater than zero for a number of reasons. Among them, the major ones are:

1. There are always groups of people that are in the process of changing jobs and they are voluntarily registered as unemployed, so as to have all the time that they need to search for a job suitable to their skills.
2. There is another group of people that is voluntarily registered as unemployed, so as to take advantage of some possibilities that are offered to them by the modern welfare state, for example, motherhood, fatherhood, unemployment insurance and the like.

Consequently, there is always a set of people who, even though their wage is flexible, are not willing to supply their labour services and so are registered as unemployed, while at the same time the labour market is in equilibrium. This kind of unemployment is called natural and it is a long-run concept. What happens, however, in the short run? Friedman's answer is that if the equilibrium is stable, then the distinction between short-run and long-run equilibrium disappears. According to Friedman, in a stable equilibrium, the price level increases at a growth rate that can be predicted. As a consequence, workers take this into account when collectively bargaining for wage increases in their contracts.

If we now suppose that inflation for some reason accelerates and if this acceleration is unpredictable, then workers see their money wages increase and they perceive this initially, at least, as an increase in their real wage and so they are willing to increase their supply of labour. In Friedman's analysis, workers are assumed to suffer from money illusion in this case since they fail to realise that the acceleration of inflation decreased their real wage. Meanwhile, firms, since real wages have fallen, increase their demand for labour and workers react by increasing their supply of labour and thus unemployment falls. With the passage of time, workers

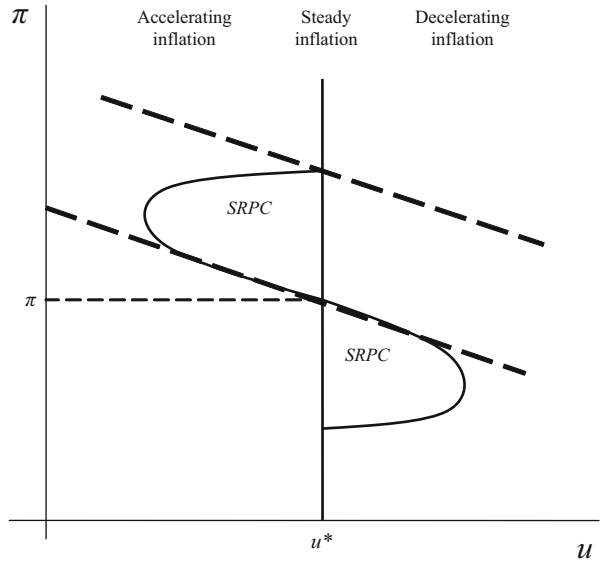

Fig. 13.2 The short-run and long-run Phillips curve

realise the reduction in their real wage and they ask for higher money wages in order to make up for the losses in their purchasing power. Entrepreneurs in turn respond to these increases by reducing employment and thus the economy returns to its starting point, that is, the point of the natural unemployment rate with the difference that this time the natural rate is accompanied by a higher inflation rate. This process is portrayed in Fig. 13.2

In this figure, inflation (π) is measured along the vertical axis and unemployment (u) is measured along the horizontal axis. We suppose initially an acceleration of inflation, which of course is not predicted; as a result, the real wage falls, firms are willing to increase employment and the economy moves left of u^* and unemployment falls. Workers eventually realise the fall in their real wage and ask for wage increases to make up for the losses due to inflation. The supply of labour decreases, real wages increase, employment falls and the economy returns to its natural unemployment rate; with the difference that now inflation has increased. If we now suppose a fall in inflation, workers because of their short-run money illusion perceive the fall in prices as a fall in their real wage and the supply of labour falls and unemployment rises, $u > u^*$. With the passage of time, workers realise that what really happened is an increase in their real wage and so they increase their labour supply returning to the natural rate of unemployment, which, this time, is associated with lower inflation. By juxtaposing the two cases, we form an S-shaped curve (see Fig. 13.2). Of course, one could draw a usual short-run Phillips curve (SRPC) as the one indicated by the dashed line, which conveys the idea that in the short run there is a trade off between inflation and unemployment. In the long run, however, there is no such trade off and thus the Phillips curve becomes completely

inelastic, as this is portrayed by the vertical line called the long-run Phillips curve (LRPC) which shows that there is no relation between unemployment and inflation. Once we return to the long-run Phillips curve, the whole process can start again.

13.2.2 Expectations-Augmented Phillips Curve

The interaction between short-run and long-run Phillips curve has been introduced by Friedman (1968) through his concept of the expectations-augmented Phillips curve, which can be formally stated as follows:

$$\pi_t - \pi_t^e = -\beta(u_t - u^*) \text{ with } \quad \beta > 0$$

where the left hand side of the above equation represents the unexpected inflation rate, and the right hand side the size of the cyclical unemployment, β is a positive reaction coefficient which stands for the speed of the adjustment process. If we suppose now that the expected rate of inflation (π_t^e) is approximated by last year inflation rate (π_{t-1}), then the expectations-augmented Phillips curve can be rewritten as follows:

$$\pi - \pi_{t-1} = -\beta(u - u^*)$$

which is interpreted to mean that when the actual unemployment is greater than the natural, then the change in the inflation rate is negative, and the change in the inflation rate is positive when the actual unemployment rate falls short of its natural rate.

The above equation also gives us another way of looking at the natural unemployment rate, defined as the rate that keeps inflation constant. This is the reason why the natural rate of unemployment is also called the non-accelerating inflation rate of unemployment or NAIRU, which has been an integral part of the modern macroeconomic literature. The acronym 'NAIRU' is attributed to two Keynesian economists Modigliani and Papademos (1975) and even today continues to be the cornerstone of modern macroeconomics (Espinosa-Vega and Russell 1997). Furthermore, many economists share the view that the NAIRU must be used as an additional tool in the design of monetary policy and also to define the unemployment rate. For example, if the current unemployment rate falls short of the natural then monetary authorities are advised to exercise tight monetary policy, and *vice versa*.[1] When inflation rate is stable this is an indication that the actual

[1]The difference between the natural unemployment rate and NAIRU is that the NAIRU is a variable unemployment rate as opposed to the natural unemployment rate in which the assumption is that there is a fixed socially optimum unemployment rate. The difference in terms of policies is that in the case of NAIRU the priority is to stabilise prices regardless of the rate of unemployment, whereas in the natural rate of unemployment the priority is the rate of unemployment which must come close to its natural level regardless of inflation rate.

unemployment is near the natural unemployment regardless of the rate of unemployment. What counts is the inflation rate and not the percentage of unemployment per se, and we know that the natural unemployment rate varies from country to country and time period to time period, thus in the 1970s and 1980s was much higher than in 1960s or earlier. For example, in the US economy from 4.4% in 1960s increased to 6.2% in the 1970s and in the 1980s exceeded the 7%, while it started to fall in the 1990s and increased in the recent years.

13.3 Quantity Theory of Money and Monetarism

During the 1960s, Keynesian economists saw in the inverse relationship between inflation and unemployment a set of alternatives from which they could choose the desired combination. This is the main reason why the Phillips curve became such an integral part of the Keynesian economics of the neoclassical synthesis. As a consequence, when the combination of stagnation and inflation plagued the western economies, this led to the rejection of the usual Phillips curve, and was registered as a blow against Keynesian economics facilitating along the way the reappearance of monetarism and its establishment as a school of economic thought with significant appeal.

The central idea of this school of economic thought is the quantity theory of money (QTM) as a theory of inflation. The QTM in its modern expression is based on two propositions. The first states that full employment equilibrium in the labour market occurs when unemployment is at its natural rate. This idea seems to have prevailed in the milieu of modern macroeconomics where it is argued that full employment does not mean zero unemployment, but a certain (natural) percentage of unemployment. The natural rate of unemployment exists because many unemployed are either in the stage of searching for better employment or they think that they require some more time in order to obtain the necessary skills needed, and so forth. Consequently, a significant unemployment rate is completely natural and expected. In short, unemployment in an economy is viewed as a real problem only in case where it exceeds the natural rate of unemployment.

Monetarism, as we know from the early version of the quantity theory of money, has existed as a theory explaining inflation prior to Keynesianism. Thus, we know from the following identity:

$$MV \equiv PQ$$

that if we assume the real output (Q), and the velocity of transactions (V) as given magnitudes, then the above identity is transformed to a theoretical relationship where the changes in the quantity of money (M) are imparted directly to the general price level (P). This version of the quantity theory of money is known from David Hume (1711–1776) and the classical economists, whereby the justification for the given velocity of money lies in the idea that V is an institutionally determined

variable and thus not easily changeable, while output is assumed given because of the acceptance of Say's law (see Chap. 6).

Modern monetarists express the above relationship in terms of growth rates, which means that they consider, in the beginning at least, the velocity of circulation of money as given and thus move on to argue that the growth rate of money supply influences the growth rate of nominal output identified with the nominal GDP, that is, the product of the real GDP times the general price level. Later on, when Friedman introduced the notion of natural unemployment, it could be argued by the monetarists that in the long run, at least, the real GDP would be equal to the level of real GDP that corresponded to natural unemployment and thus the growth rate of GDP can be safely assumed as known in the long run. Consequently, in the long run the growth rate of the supply of money – to the extent that it exceeds the growth rate of the real GDP – increases the growth rate of the price level, that is, the rate of inflation. The fundamental idea of the monetarists can be summarised in the following passage:

> [...] changes in the behaviour of the money stock have been closely associated with changes in economic activity, money income, and prices [...] The interrelation between monetary and economic change has been highly stable. (Friedman and Schwartz 1963, p. 676)

According to Keynesian economists, the velocity of money is a variable characterised by high volatility; consequently, changes in the supply of money is possible to be absorbed from changes in the velocity of money with negligible effects either on output or on the price level. A similar view is shared by the economists of the neoclassical synthesis, especially in the case where the economy is in the liquidity trap region; whereby, regardless of the changes in the supply of money, the real economy is not affected at all. Changes in the supply of money are absorbed by the corresponding changes in the velocity of money. The discussion between monetarists and Keynesians was restricted to the empirical level, where the differences in the velocity of money, however defined, that is with respect to M_1, M_2, M_3 and in the recent years even M_4, became the epicentre of controversy.

The empirical evidence with respect to the effects of the money supply on the price level so far has been mixed and depends on the definitions of the money supply (narrow or broad) and the time period. As a consequence, the velocity of the narrow money supply, $V_1 = GDP/M_1$, for the US economy has displayed a rising trend during the period 1920–1929, a falling trend during the period 1929–1946, an upward trend in the period 1947–1981, erratic behaviour along a falling trend during the period 1981–1991, and an upward trend since then. The erratic behaviour of the 1980s has been attributed to the deregulation of the banking industry and the appearance of new checkable accounts. Clearly, the overall movement of V_1 is associated with the long-run upward or downward stage of the economy. The results with respect to the U.S. data prove somewhat better for the monetarist argument with regard to the velocity $V_2 = GDP/M_2$ (see Fig. 13.3). A closer look at V_1 or V_2 in monthly or quarterly data reveals substantial fluctuations in the short run. The variability of the velocity of circulation has been attributed, among other

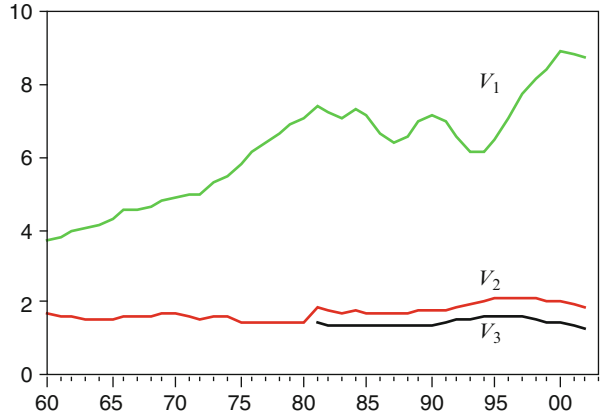

Fig. 13.3 Velocity of money in the US economy

things, to the frequency of payments, the efficiency of the banking system, the interest rate and the expected inflation rate. From the above, it follows that the causal relationship between money supply and price level – that is, the issue of exogeneity versus endogeneity – is not settled yet, and, therefore, continues to attract the attention of economists. There is no doubt that the discussion will continue in the future, as economists try to understand better the interrelations between monetary and real variables.

Monetarists in the beginning had many successful predictions, especially in the decade of 1970s using super simple econometric models, such as the St. Louis model which in its initial version of 1970 consisted of 8 (eight !) equations and 11 variables. This model is in sharp contrast with the usual macroeconometric models of a few hundred equations and variables. In short, the produced output in the St. Louis model depends linearly on the supply of money and the government expenditures with various time lags. In the regressions that were run, the monetary variables were found to have superior explanatory power to the fiscal ones. It is important to point out that the performance of the St. Louis model in the beginning at least gained a lot of support for its parsimony in terms of variables and data and for the accuracy of its predictions. Naturally, the St. Louis model was used extensively for economic policy purposes.

The continuously rising price level of the 1970s, however, led many economists to a re-examination of the idea of a stable increase in the supply of money. Although the rise in inflation was attributed to external facts, such as the Vietnam War, oil crises and so forth, nevertheless the data revealed a parallel movement in both inflation and the supply of money. From the mid-1970s, the Fed adopted the targeting of monetary variables, such as the money supply in the narrow sense, that is, the M_1. Soon, however, it was realised that what the Fed targeted was the Federal funds rate and not the supply of money. Only in October of 1979 did the Fed announce that its primary target will be the control of money supply and not the interest rate, that is, inflation targeting rather than interest rate targeting. Meanwhile,

the US economy and the world economy entered in their first post-WWII economic depression which was predictable within monetarism; the slowdown in the growth rate of money supply and the rise in the interest rates were responsible for the fall in the inflation rate, a fall which is known as reflation, that is the fall in the growth rate of inflation. At the same time, there had been a long-lasting stagnation in the economy which was manifested in the fall of investment and the increase in unemployment.

Monetarists, however, did not agree with the monetary authorities and argued (Friedman 1984) that the control of money supply was not exactly of the kind they wanted. More specifically, monetarists on the one hand distanced themselves from the monetary authorities; while on the other hand, they promoted their own proposals for the proper application of monetary policy. Meanwhile, the reaction of the public to the monetary authorities led to the abandonment of the monetary policy. The worse for the monetary approach came later, when the Fed in 1980 adopted the deregulation of the financial system which led to the increasing volatility of monetary variables and thus weakened the connections between monetary variables and the real economy. The situation got worse with a number of financial innovations that were introduced in the banking system and affected the volatility in the velocity of money in the narrow sense (V_1) in the beginning of the 1980s and thus set in doubt the monetarist prediction.[2] The endogeneity of the money supply, which has always been a central topic among radical economists, started to attract the attention of many orthodox economists and especially those of the real business approach, who in an unexpected and to a great extent unholy alliance with the Post-Keynesian economists considered the money supply as an endogenous variable.

The constancy (with the necessary qualifications) of the velocity of money is a term *sine qua none* for monetarism. In general, the historical empirical relationships that have been discussed by Friedman and other monetarists did not last for long. Under these circumstances, the targeting of money supply (regardless of its narrow or broad definition) was no longer in the goals of monetary authorities, whose target, once again, became the rate of interest.

More specifically, Taylor (1993) suggests that the Fed increases interest rates in times of high inflation, or when GDP is above its potential (natural) level, and decreases interest rates in the opposite situations. This method of controlling interest rates has been fairly consistent with interest policy decisions, even though the Fed does not explicitly subscribe to the rule. Formally, the rule can be stated as follows:

$$i_T = \pi_t + r_t^* + a_\pi(\pi_t - \pi_t^*) + a_y(y_t - y_t^*)$$

where i the target interest rate, π the rate of inflation as measured by the GDP deflator y is the logarithm of real GDP, starred variables represent the desired or equilibrium variables. The idea behind the use of such a rule is to ensure price

[2]Hence, the famous Goodhart laws that we examine in the next chapter are ascertained.

stability and full employment while reducing uncertainty and increasing credibility of future actions by the central bank. According to the rule, both reaction coefficients, a_π and a_y, should be positive (as a rough rule of thumb, Taylor's (1993) paper proposed setting $a_\pi = a_y = 0.5$). According to Taylor rule, when inflation or output is above the desired (equilibrium) level in order to reduce the inflation pressure a relatively high interest rate (a 'tight' monetary policy) is recommended. The converse is true in case where inflation or output falls short of the equilibrium value ('easy' monetary policy). Sometimes, monetary policy goals may conflict, as in the case of stagflation, when inflation is above its target rate, which in the US economy is at 2%,[3] while output is below full employment. In such a situation, Taylor rule specifies the relative weights given to reducing inflation versus increasing output. Taylor showed that the above relation describes accurately enough the behaviour of the Fed historically.

13.3.1 Friedman's Demand for Money

Friedman (1956) restored the credibility of the quantity theory of money, a theory that was associated with the University of Chicago. Friedman (1956) argued that the quantity theory of money must be viewed as a theory of the demand for money and not as a theory for the determination of the general price level or income. More specifically, he argued that the demand for money is due to the flow of services that money offers to its holder. The demand for money according to Friedman depends on three factors:

1. The total wealth of the households or firms
2. The opportunity cost of holding money
3. The preferences of agents of the economy that possess wealth

Money in Friedman's analysis is any good that gives utility to its holder. Consequently, the utility of money must be balanced by the disutility that one suffers from the lack of some other form of wealth. This approach has similarities with Keynes's approach that we analysed in Chap. 9. Nevertheless, there are essential differences that relate to the determinants of the demand for money. In symbolic terms, we have:

$$L = f\left(y_p, \omega, , r_b, r_e, p^e, u\right)$$

where L is the demand for money; y_p is the permanent income of individuals and also serves as a proxy for wealth;[4] ω is the ratio of human to non-human capital;

[3]The desired or target rate of inflation may change but the whole exercise is only meaningful for low rates of inflation.

[4]The expected annual return that the individual will earn from all his assets.

r_b and r_e are the expected returns of bonds and stocks (equity), respectively; p^e the expected general price level and finally, u refers to the preferences of individuals. The holders of wealth behave rationally and maximise their utility by ranking their assets according to their marginal returns. The ranking of the assets constitutes the pivotal point for the understanding of the transmission mechanism of changes in the stock of money in the economy and the way with which it may affect the real sector of the economy.

Friedman under the term wealth includes not only cash and bonds but also real capital such as investment in plant and equipment, durable consumer goods and human capital, that is, capital inherited or acquired through education and training. First of all, the value of human capital will be equal to the present value of the money stream that one expects to obtain through his abilities that are inherited or acquired through education, in general. Since there is no market for human capital that would establish a reward for this (unless we imagine a slave society), it is hard to determine its reward in any precise way. Every individual can substitute one form of capital for another, for example, through education; however, this requires the passage of time. Consequently, in the short run ω is relatively stable. With the passage of time, as human needs increase, it is logical to increase the demand for money, consequently, M_1 and ω are inversely related.

In order to have a better understanding of this relationship, we suppose that the government applies open market policies in its effort to control the supply of money. We suppose that the economy is still in a state of equilibrium, where individuals distribute their wealth over different assets so as to equalise their returns. If we suppose that the government increases its supply of money through the purchase of bonds, it follows that the price of bonds increases and their returns diminish. The increased supply of money will be used for the purchase of consumer or investment goods, whose price increases until the point that the returns of all assets are equalised. This is the reason why monetarists argue that the supply of money directly affects the real economy, while for Keynesians the effect of the supply of money on real variables is only indirect and thus weaker than the monetarists think. This difference is the result of the idea that for monetarists there is no asset that is a close substitute to money.

The above analysis can be cast alternatively via Walras law (Fonseca and Ussher 2009). We remind the reader of the determination of the interest rate in the money market that takes place through the purchase of bonds that we examined in Chap. 10. Stipulating now the presence of the goods market, we will have:

$$\left(M^d - M^s\right) + \left(B^d - B^s\right) + \left(Y^d - Y^s\right) = 0$$

where the three terms in parenthesis represent the excess demand in the money, bond and goods markets, respectively. Starting from an equilibrium position across all markets, let us suppose an expansionary monetary policy which brings about a negative excess demand in the money market $(M^d - M^s) < 0$. In this case, monetarists do not claim that the bonds market should be characterised by excess

demand $(B^d - B^s) > 0$, simply because there is no reason for the individuals to dispose of this excess supply of money in the bonds market, and, therefore, this difference will totally appear as an excess demand for goods, that is $(Y^d - Y^s) > 0$. Consequently, inventories run down and output expands and so does employment. This is the reason why for monetarists the supply of money affects the economy directly through the increase in the demand of consumer goods and not indirectly through the interest rate and investment, as for example Keynesians would argue. This idea of the monetarists is logically sound (within the neoclassical paradigm) and complements the neoclassical synthesis of the Keynesian economists.

13.4 The Monetarist Approach to the Balance of Payments

We are already familiar with the classical quantity theory of money from the discussion of the principle of comparative advantage. We also know that in the usual quantity theory of money any variation in the quantity of money is manifested in the price level and through shortages or surpluses in the balance of payments leads to variations in the quantity of gold reserves of a country; thereby restoring equilibrium domestically and internationally. For example, if in country A the supply of money increases, it follows that the domestic price level increases and so exports fall and imports rise and the deficit in the balance of payments is sustained through the outflow of gold (or silver). As a consequence, the supply of money in country A falls together with its price level; the exact opposite process takes place in country B. These two adjustment processes are completed with the restoration of equilibrium in the two countries.

It goes without saying that such an adjustment mechanism belongs to the past and in modern economies with their advanced financial system, gold reserves do not play such an equilibrating role. The role of gold is assumed by the currency (or currencies) which are accepted in international exchanges (dollar exchange standard or euro to a lesser extent). If we now hypothesise that the level of output is at full employment (natural level) and that the law of one price holds in both the product and money markets and that furthermore the domestic price level together with the interest rates are determined in international rather than the domestic markets,[5] then we end up with a demand for money as a stable function of the

[5]In other words, the principle of purchasing power parity (PPP) holds, that is, the international money (say $100) purchases the same amount of (traded) goods everywhere. If not, an outflow of goods from the relatively low prices countries towards the relatively high prices countries will take place. The subsequent surplus and deficits in trade balances will lead to changes in the exchange rates so as to restore the equality of purchasing power of $100 across countries for the traded commodities. Empirically, we know that the PPP hardly holds even over the very long run (a time period of a century).

price level, the income and the rate of interest. More specifically, in the simplest case the demand for money can be expressed as

$$L = Pf(y, r)$$

where L is the demand for money, P is the domestic price level, y is the real income and r is the real interest rate. The money supply on the other hand will be

$$\mathrm{M} = D + R$$

where M is the supply of money which consists of domestic credit flows (loans), D, and the stocks of exchange reserves, R, which are held in the central bank. We further suppose that the system is initially in equilibrium, thus we have, $L = M$.

If we suppose that for some reason the domestic credit flows increase, then in this case, with the demand for money given, the system is out of equilibrium. The monetarists argue that since the additional credit flows (loans) cannot be spent in the domestic market (because the demand for money is given), it follows that they flow out in the international money markets. In this way, the economy at hand is led to a balance of payments deficit.

If we suppose a regime of fixed exchange rates, such as the one under Bretton Woods that prevailed from 1945 until 1971, then the monetary authorities are forced to sell foreign exchange (dollars) in order to buy the domestic currency so as to cover the trade deficit and maintain the stability of the exchange rate of the domestic currency against the dollar. The success of such a policy depends foremost on the size of the deficit and the stock of foreign reserves held in the central bank. The loss of foreign reserves would reduce the initial increase of the domestic money supply to the point that the supply of money would offset the trade deficit. The system returns to its initial equilibrium, when the supply of money returns to its initial level with the increase in the domestic credit flows which are compensated with the reduction of foreign exchange reserves. In short, any deviation between supply of and demand for money is manifested in the deficit (or surplus) balance of payments, which in turn activates the mechanism through which the initial equilibrium is restored.

The consequences of the above mechanisms depend on the size of the economy. If we suppose a small open economy with fixed exchange rates, the money supply obviously becomes endogenous, and, therefore, monetary policy, in such an economy, is an exercise in futility. The expansion of domestic credit, merely, reduces the international reserves of this country while its money supply remains the same. By contrast, the increase of domestic credit flows in a large economy such as the USA can affect the world monetary expansion and also the international price level. For example, the inflation of the 1960s was attributed by many economists to the War in Vietnam that led the USA in the increase of domestic credits and thus in the increase of the domestic but also the international money supply since the dollar was the international medium of exchange. Thus, under a fixed exchange regime which lasted up until 1971, the price level in the US economy had repercussions in

the rest of the world through the domestic supply of money in the USA, which had to rise in order to stabilise the current account deficit. In fact, the USA determined the monetary policies in the rest of the world. This situation could not go on and many surplus current account countries (mainly France and W. Germany) required their dollar reserves to be converted into gold in compliance with the Bretton Woods agreement. The result was that the USA declined such demands, which in effect led to the collapse of Bretton Woods agreements in August 1971.

Up until now, we have supposed a fixed exchange rate regime. We know, however, that after 1971 we entered the regime of freely fluctuating exchange rates. In such a regime, the increase of domestic credits, given the demand for money, results in a profitable absorption of the additional credit flows in the foreign market of goods and titles. As a consequence, there is an excess supply of money in the market for foreign exchange. In a free market setting the exchange rate of the currency which is in abundance, other things constant, is expected to fall. The devaluation of the domestic currency leads to a rise in the domestic price level and thus to an increase in the demand for money and the restoration of equilibrium. These adjustments are activated through variations in the exchange rates and not in the stocks of precious metals (mainly gold) and the domestic money supply. It is important to emphasise that such a freely fluctuating exchange rate regime has not existed for long in any country, because governments are very careful with the exchange rates and intervene accordingly. The form of their intervention might be overt or it might take the form of raising obstacles in an effort to discourage certain imports. Examples of such intervention might include subsidies in agriculture, quality controls in the imported goods, various other non-tariff prohibitions of imported goods, preferential treatment of countries, and so forth.

13.5 Economic Policy Conclusions

Monetarists view capitalism as a fundamentally healthy system that possesses the internal mechanisms which are capable of maintaining equilibrium and full employment. The system may be out of equilibrium because of external shocks but even in this case the system must be left to its own devices in order to ensure that equilibrium will be restored. A precondition for all of the above is the operation of a free market, while any long-lasting recessions are attributed exclusively to mistaken government intervention especially with respect to the supply of money. The role of the government, according to the monetarists, is to safeguard the free operation of the market forces. Government intervention, in the form of either fiscal or monetary policy, cannot achieve anything quite different with respect to the outcomes of the real economy. According to Friedman, fiscal policy is ineffective and what counts is only the supply of money. If government expenditures are financed through the creation of new money (*e.g.,* through money deposits), then the only thing that is achieved, in the long run, is the worsening of inflation. This case

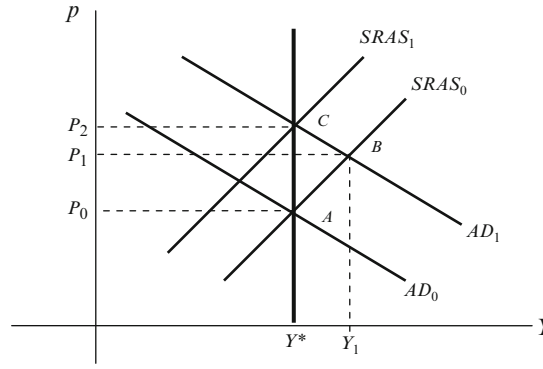

Fig. 13.4 The effectiveness of monetary policy

can be shown through the following graph of aggregate demand and aggregate supply curves.

In Fig. 13.4**,** we start with the economy at point *A* where all the three curves, that is, the short-run aggregate supply ($SRAS_0$), the aggregate demand (AD_0) and the long-run aggregate supply (*LRAS*) intersect and we further suppose that the government applies an expansionary monetary policy, which increases the aggregate demand as this is manifested in the outward shift of the aggregate demand curve from AD_0 to AD_1. This shift leads to an increase in the price level together with the money wages; we hypothesise that the increase in nominal wages is less than the general price level. Workers, according to monetarists, in the short run suffer from money illusion, perceive the increase in their money wage as an increase in their real wage and thus increase their supply of labour. Firms, on the other hand, since they observe the fall in the real wage, hire additional workers thereby reducing the unemployment rate. The result is that output expands from y^* to y_1 and with it the employment and the economy move from point *A* to point *B*, where AD_1 intersects the curve $SRAS_0$. This equilibrium, however, does not last for long; soon workers realise that the real wage has been in fact reduced, so they revise their plans and ask for higher money wages in order to make up for the losses caused by inflation. However, as their real wage increases the short-run aggregate supply curve shifts to the position $SRAS_1$, as firms are forced to reduce their output laying off workers and thus unemployment rises. The new equilibrium point is *C*, which implies that the economy returns to its initial position, with the difference that the price level is higher. Consequently, monetarists argue that an expansionary fiscal or monetary policy end up to be purely inflationary.[6]

With regard to the natural unemployment, monetarists argue that the aim of the government is to reduce it. In this regard, Monetarists favour the elimination of any

[6]The underlying mechanism for these adjustments is the same as with the case of the Phillips curve that we discussed above.

obstacles that do not allow the market to operate properly. Such obstacles may include:

- The lack of adequate information regarding the professions and the needs that are created.
- The reduction of the length of time that the welfare benefits last.

Consequently, monetarists argue that the aggregate supply curve shifts to the left; at the same time, however, in the economy, there is technological progress, which raises the marginal productivity of labour, and, therefore, increases the demand for labour. Consequently, micro- and not necessarily macroeconomic policies may be used for a solution to the problem of unemployment.

With regard to inflation, monetarists suggest that it might be controlled through macroeconomic policies and, in particular, they suggest the expansion of money supply at a fixed growth rate. For example, a certain percentage a year which accords with the long-run growth of the real GDP, which in the case of the US economy is about 3%. Some other monetarists suggest more flexible rules for the growth of the money supply.

13.6 Keynesian Responses

Two constituent components of the monetarist analysis of the phenomena were observed in the late 1960s and early 1970s: the notion of natural unemployment and the analysis of the short-run Phillips curve, whose negative slope has to do with the rise of government expenditures to reduce unemployment or the reduction in government expenditures to cure inflation. In both cases, the economy in the long-run returns to the level of output corresponding to full employment or what is the same thing to natural unemployment.

Friedman's analysis shows that in the long run the supply of money does not influence the real variables. In other words, in the long run, money is neutral. We cannot say the same thing about the short run, where the supply of money and thus government intervention can have some effects. Such a conclusion was agreed upon by many Keynesian economists of the neoclassical synthesis. One of them was Modigliani (1977), who as the president of the American Economic Association, argued in his annual address that in the long run the economy is led to a full employment level of output and that there are no important differences between monetarists and Keynesians. Tobin, on the other hand, argued that while the economy will tend to a full employment level of output in the long run, the problem of inflexibility of wages and prices of assets in a downward direction in the short run will remain. Friedman's (1973) response was that Tobin by focusing on the details of the operation of the market essentially builds a short-run macroeconomic theory. The worse thing with short-run analyses is to suggest economic policy prescriptions, since in these cases the policies that are applied are usually of the 'too little too late' type and end up worsening, rather than correcting, the economic problems

at hand. Friedman furthermore argued that it is preferable to construct a long-run theory and correct along the way those imperfections that do not let the system work properly rather than build a theory around imperfections and then try to correct them through government action. Government intervention, Friedman argued, would be much more effective if its actions are contemplated within a long-run framework since it is in this case that policy actions are more likely to secure full employment without inflation. Friedman, in his intervention, clarified his distinction between short-run and long-run arguments. More specifically, he pointed out the following characteristic example: the fact that a piece of paper does not fall to the ground with the speed that it is predicted by the law of gravity does not invalidate the law of gravity. In other words, friction and other 'imperfections' do not invalidate the operation of the law of gravity. What the Keynesian economists argue, continued Friedman, is to focus on the atmospheric pressure (that is to the short-run phenomena) rather than to the operation of the law of gravity (the long-run tendency).

The Keynesian economists (Modigliani, Tobin, *inter alia*) in their effort to defend their propositions justified their insistence on the short run by arguing that in the long run the economy finds itself on another short run. In other words, the long run consists of many short-run periods. It is true, Keynesians argued, that wages in the long run are flexible especially during periods of inflation; yet, there still exists a kind of imperfection that remains present even in the very long run and this is uncertainty. Consequently, despite whether analysis is of a long-run range, future (uncertainty) will always be a variable affecting investment decisions.

Another dimension of the controversy between monetarists and Keynesians relates to the stability of the demand function. The assumption of the stability of the demand for money has been questioned by Keynesian and especially Post-Keynesian economists (Kaldor 1970; Modigliani 1977, *inter alia*). For the monetarists, the stability of the demand for money was a precondition *sine qua non*; in other words, the demand for money is expected to be a stable function of the variables that determine it and furthermore it should be inelastic to the rate of interest. This is the reason why monetarists focused, from very early on, on the stability properties of the demand for money (Cagan 1956; Friedman 1959, 1966; Meltzer 1963). Monetarists begun their study about the stability of the demand for money using a partial adjustment model, known with the acronym PAM, where the demand for money depends on the level of income and a spectrum of interest rates. Furthermore, the demand for money, because of the adjustment cost, displays a hysteresis between the desired level and the realised one.[7] These functional relations, in the beginning, gave rise to some pretty good results regarding the 'explanation' of money demand, yet this

[7]The functional form of PAM is: $m_t - m_{t-1} = \delta\ (m_t^* - m_t)$, where m is the demand for money in constant prices ς, m^* is the desired demand for money, t is time and $0 < \delta < 1$ is the adjustment coefficient.

performance of the model did not continue under conditions of high inflation and stagnation of the economy from the mid-1970s onwards. More specifically, Goldfeld (1976) argued that the function of money demand collapsed as an idea because of the empirical research that took place in the 1970s. It is particularly ironic that while empirical results were poor during the 1970s and certainly not supportive to monetarism, central banks around the world assumed a more active position in targeting some monetary variables. As a consequence, the interest in the nature and the stability properties of money demand emanated from the central banks and in particular from the Fed. The idea of stability of the demand for money resurfaced in the 1980s for at least some monetary variables, but the results of empirical research were far from being satisfactory for monetarists.

The instability in the demand for money must be attributed, at least in part, to a number of institutional changes that took place especially during the 1980s. As a result, some monetarists modified the theory of the demand for money in order to interpret these puzzling empirical findings. Soon, the partial adjustment models were replaced by the shocks absorption models. These models were introduced for the first time by Carr and Darby (1981) and Laidler (1984). In these empirical investigations, the instability of the demand for money remained a disturbing feature for monetarists. The next step was to be taken many years later in the popular until today error correction models. The main characteristic of these models was to capture the long-run equilibrium through short-run adjustments. Finally, the co-integration models combine features such as the long-run equilibrium relation between the variables involved and, furthermore, they allow for the attainment of this equilibrium by including an error correction term and by allowing for short-run and long-run causality tests in the case of many variables. In this sense, they are superior when they are compared to simple models and causality tests of the past between pairs of variables.

The empirical research of many decades has shown beyond any doubt that the demand for money function is not that stable as monetarists would like to. As a matter of fact, monetarists managed to show that for some countries and for a specific period of time the demand for money is stable. However, this does not mean that the same result necessarily holds for a different country or for the same country at a different time period. Monetarists, for many decades, tried to sidestep these critiques either by redefining their variables (M_1, M_2, M_3 and lately M_4) and attributing the negative results to the quality of the data or the definitions of variables. It is certain, however, that the results of their analysis fall short of being a satisfactory proof of their theories and their view points.

At this point, it is worth mentioning the plea of Post Keynesian economists such as Kaldor (1970) and Minsky (1975) that relates to the issue of the exogeneity of money. We saw that this hypothesis cannot be supported even within the assumptions of monetarism, when we consider a small open economy, where the supply of money is certainly not within the control of the central bank. Monetarists claim that in the short run the central bank, by intervening in the exchange rate market, can reduce the exchange rate of its currency and increase its exports and, thus, increase the output produced and so forth. In the long run, however, because of the increase

of the price level, the economy is restored to the initial equilibrium position in terms of output produced. The trouble with this view is that if the economic agents have rational expectations, then we know from the New Classical Macroeconomics, which we study in the next chapter, that what is to happen in the long run will happen in the short run as well, provided that people know the policy of the central bank. The argument of the rational expectations is strengthened in the case of monetary policy, because then we are supposed to know not only the current monetary policy but also the future one.

According to Eatwell (1983a, b), the above criticisms are mainly empirical and in this sense inadequate for they do not question the logical consistency of monetarism. In particular, Eatwell focuses his analysis on the central feature of the quantity theory of money according to which the produced output is given, as a result of the operation of the market forces, that is, the forces of supply and demand. We know that in the neoclassical theory the quantities and prices are determined simultaneously from the forces of supply and demand, and, therefore, we cannot assume that we know the level of output without knowing the price level. Such an assumption is not in the spirit of Keynes's analysis, where the question at stake is the determination of the level of output something that monetarists take as given. This critique that targets the logical coherence of monetarism does not seem to have attracted the attention that it deserves. However, what the logical critique did not achieve was achieved by the lasting stagflation, the inadequacy of the monetarist explanation and the failure of the policies associated with it to come to terms with the phenomenon. Thus, the monetarist school of economic thought was put, once again, in the fridges of economic analysis mainly because of its inadequacy to interpret the lasting stagflation.

13.7 Summary and Conclusions

Monetarism as a school of thought culminated in the late 1960s and early 1970s, when the high rate of inflation induced many Keynesian economists to turn for answers to monetarism. In such an economic situation, monetarists could raise questions to which Keynesians could not give satisfactory answers and their approach was led to a state of decay. Only after the passage of at least a decade did the Keynesians manage to reorganise their ideas and make a strong come back. In their return, Keynesians not only managed to provide satisfactory answers to the questions that Monetarists raised but also to raise questions of their own that Monetarists could not answer. As a consequence, the Monetarist school of economic thought from its brief ascent found itself in the stage of decay and marginalisation, a situation that monetarists were familiar with during the previous decades. In short, the basic propositions of modern monetarists are the following:

1. Variations in the supply of money are responsible for the variations in nominal income.

2. The market forces are powerful enough to lead the economy to equilibrium except for sudden shocks in money supply. Yet even in this case the economy returns eventually to its initial equilibrium position characterised with natural unemployment.
3. The inverse relation between inflation and unemployment does not hold in the long run, where the Phillips curve is vertical to the point of natural unemployment.
4. Inflation is, always and everywhere, a monetary phenomenon.

The initial success of Monetarism is attributed to the fact that it was formulated in terms of the neoclassical synthesis and the IS–LM model and thus its theoretical underpinnings were not different from those of Keynesian economics. Furthermore, monetarism offered credible answers to the phenomenon of stagflation and at the same time placed its emphasis on the prominent role of the money supply and its control by the central bank.

More specifically, Keynesian economists of the neoclassical synthesis explain the fluctuations in the level of economic activity on the basis of the level of expenditures and in particular of investment, given the consumption function. By contrast, Friedman and the monetarists explain the fluctuations of the level of economic activity on the basis of money supply, with the demand for money as a stable function of a rather small number of variables. As a consequence, Keynesians (of the neoclassical synthesis) downplay the role of money and the effectiveness of monetary policy, while they attach special weight to fiscal policy. Monetarists, on the other hand, view fiscal policy as an impotent tool to affect the level of economic activity and regard its role in the long run as being neutral or even negative with respect to the level of output. The above do not mean that monetarists would be supporters of an active monetary policy; to the contrary, monetarists view the role of the government to be limited to some absolutely essential functions and argue that the size of the government should be small; otherwise, it might be an obstacle to the normal operation of the economy. As for monetary policy, Monetarists argue that it must be known in advance, that is to say, the monetary authorities should be allowed to increase the money supply at a constant rate which should be approximately equal to the long run growth rate of real GDP.

Such rigidity in the exercise of monetary policy created a lot of scepticism about the theoretical adequacy of the monetarist approach. Some monetarists proposed more flexible rules for the growth of money supply; nevertheless, partly as a result of weaknesses in the theoretical core of monetarism and partly because the fundamental propositions of monetarisms did not stand against empirical testing, especially after the mid-1970s, monetarism was again led to the fridges of theoretical economics although not to euthanasia. On our part, we can say that monetarism made some lasting contributions, which have been incorporated in the major premises of other schools of economic thought. The first of these contributions is the dismissal of the short-run Phillips curve, which was so much associated with Keynesianism that it became the reason for its collapse. More specifically, the assumption of Keynesians about the inflexibility in money wages was somehow

offset by the adoption of the Phillips curve that furthermore offered a trade off between inflation and unemployment, which was relevant in the exercise of fiscal policy. However, when the empirical facts cast doubt on the short-run Phillips curve and while at the same time the long-run Phillips curve was not amenable to any kind of active policy, then the collapse of Keynesianism of the neoclassical synthesis was essentially triggered off. The second contribution of monetarists was the argument for minimal government intervention, which has been accepted by all macroeconomic schools of economic thought at least up until the crisis that started in 2007 and continues as of this writing. The trouble with monetarism was that it placed too much emphasis on the role of money. What is certain, however, is that monetarism has managed to shift the attention of economists to the study of the way in which monetary and real variables are related to each other as well as on how and to what extent monetary policy can affect the real economy. The discussions about the independence of the central bank and its relation to economic growth have their roots to monetarism.

So what remains of monetarism? The answer is that most of its major propositions have been assimilated in the New Classical economics and the New Keynesian economics. Monetarism is currently a 'degenerate research programme', a state in which it had found itself for many years before its revival mainly by Milton Friedman. We cannot say with any degree of certainty what will happen in the future, but at present it seems that monetarism follows the example of the supply side economics approach, which once was very popular (in the early 1980s), but then fell into oblivion. Would monetarism have the same fate with supply side economics or with Keynesian economics? These are questions that only time will give a definitive answer.

Questions for Discussion and Thought

1. What were some of the major reasons that made monetarism particularly popular in the 1970s among economists and policy makers?
2. Write down the equation of the old quantity theory of money and explain each of its terms. In what sense does the modern version of the quantity theory of money differ from the traditional one?
3. In what sense, if any, is money neutral according to the Monetarists?
4. Suppose that the supply of money increases. How does this affect economic activity according to the Monetarists?
5. Friedman's analysis integrates the interaction of money and real variables in the economy. Do you agree?
6. What was the criticism launched by Tobin and Modigliani to Friedman's analysis of the Phillips curve?
7. Monetarism has been criticised on empirical and not on theoretical grounds. Do you agree?
8. What remains of monetarism?

Notes on Further Reading

The idea of a trade off between unemployment and inflation exists long before the appearance of Phillips's (1958) famous article. In fact, Irving Fisher in the 1920s noted such an inverse relationship between unemployment and the money wage. Thus, some contend that the Phillips curve should be called the 'Fisher curve' (Fisher 1973 [1926]).

Although Milton Friedman is considered to be the 'Father' of Monetarism, nevertheless he was a renowned economist from very early in his age because of his work initially on statistics and later on for his studies on the consumption function (1957), where he introduced the novel idea of the permanent and transitory income. Having shown that the marginal propensity out of transitory income is small and so is the multiplier and thus Friedman in effect showed the ineffectiveness of fiscal policy. The monetarist ideas started gaining popularity with the publication of a book of readings titled 'The Quantity Theory of Money' (1956) of which Friedman was the editor. It was the product of a money and banking workshop, in which he and his students worked out many of the ideas that have become part of Monetarism. Friedman in his own seminal article on the quantity theory of money: A Restatement argued that the demand for money and by implication the velocity of money are stable. However, the book by Friedman and Schwartz (1963) is the one that popularised the ideas of the role of the supply of money and that attracted attention. In this book, Friedman and Schwartz argued that the US historical record lends strong support to the view that the stock of money played a significant role in the ups and downs of the level of economic activity in the US economic history and that the Great Depression was caused by the contraction in the money supply, – an idea that caused a heated debate with the Keynesians (see, for example, Peter Temin 1973).

The other major contribution of the Monetarists is the expectations-augmented Phillips curve according to which the long-run Phillips curve is vertical, while money is neutral in the long run. This idea was absorbed into monetarist analysis in the mid-to-late 1960s, after the publication of Friedman's seminal articles in (1967) and (1968). Incidentally, this concept is now widely accepted, especially after Richard Lipsey's modification (1960) of Phillips's findings into a broader context, whereby the growth rate of the money wage was replaced by price inflation, and was considered along Keynesian lines, that is to say inflation was analysed as a demand-pull phenomenon. If we were to select what was so important in Friedman's argument, this is unquestionably the explicit introduction of expectations into macroeconomic theory.

It is important to stress that Friedman's contributions are known for their simplicity and clarity of presentation and gained popularity because of his lucid way of writing and his really revolutionary for some, counterrevolutionary for

(continued)

others, but certainly direct and to the point way of thinking. One of his books that popularised his ideas was 'Capitalism and Freedom' (1962) which was essentially restated with his wife Rosa Friedman (1980) under the title 'Free to Choose'. Friedman, in these books along with a series of articles written in newspapers and magazines, argued for several controversial propositions, which included the abolition of: minimum wage legislation, rent controls, farm supports, corporate income taxation, progressive income taxes, the fractional reserve system, public schools, *etc*. He was in favour though of negative income taxes a proposition that ironically finds strong support among many radicals today!

Chapter 14
New Classical Macroeconomics

Keynesian orthodoxy is in deep trouble, the deepest kind of trouble in which an applied body of theory can find itself. It appears to be giving wrong answers to the most basic questions of macroeconomic policy.

(Lucas 1981, p. 559)

We dwell on these halcyon days of Keynesian economics because, without conscious effort, they are difficult to recall today. In the 1970s, the US economy has undergone its first major depression since the 1930s, to the accompaniment of inflation rates in excess of 10 percent per annum. These events have been transmitted (by consent of the government involved) to other advanced countries and in many cases have been amplified. These events did not arise from a reactionary reversion to outmoded, "classical" principles of tight money and balanced budgets. On the contrary, they were accompanied by massive governmental budget deficits and high rates of monetary expansion: policies which, although bearing an admitted risk of inflation, promised according to modern Keynesian doctrine rapid real growth and low rates of unemployment.

(Lucas and Sargent Th 1978, p. 271)

My thesis in this lecture is that macroeconomics in this original sense has succeeded: Its central problem of depression-prevention has been solved, for all practical purposes, and has in fact been solved for many decades. There remain important gains in welfare from better fiscal policies, but I argue that these are gains from providing people with better incentives to work and to save, not from better fine tuning of spending flows.

(Robert Lucas, Presidential Address to the American Economic Association 2003)

14.1 Introduction

The stagflation that afflicted the US and the other economies after the late 1960s shook the economists' faith in Keynesian economics or Monetarism. The Keynesians could explain the rising price level, as a result of an expansionary fiscal policy; the trouble, however, was the continued and persistent recession (unemployment), which was inconsistent with their theory. Monetarists, on the other hand, could explain the recession through the tight monetary policy; however, this explanation was inconsistent with the rising price level. Thus, the inability of Keynesians

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_14, © Springer-Verlag Berlin Heidelberg 2010

and Monetarists to explain the key macroeconomic events of the late 1960s and early 1970s discredited their theories and created the need for alternative explanations.

It is under these circumstances that the process of succession of one economic theory by another starts. In this particular case, Monetarism tried to improve its predictive content by incorporating new tools and hypotheses resulting in what Tobin (1986) once characterised as "Monetarism Mark II", but, at the same time, an altogether new approach called New Classical macroeconomics emerged with Monetarism being just a constituent part of it, and it is in this sense that Mankiw (1998) characterised Monetarism as "the first wave of New Classical Economics". In spite of the fact that the New Classical (henceforth NC) Macroeconomics accepts many of the ideas of the Monetarists, its advocates have developed a set of ideas that form a separate approach. The word "classical" here refers to the restoration of the ideas that dominated before Keynes (1936). The ideas that were dominant – even during the Great Depression of the 1930s – included the notion of continuous equilibrium in all markets. Consequently, the efforts of government through fiscal and monetary policy to lead the economy towards the recovery stage were not only futile but also dangerous, since they were worsening an already bad situation. In the mid-1970s, the hypothesis of rational expectations, which we examine below, was gaining the support of an increasing number of economists who are known today as the New Classical (hence forth NC) economists[1] in the sense that their conclusions with regard to the effectiveness of economic policy were similar to those of the classical economists.[2] In the end of the 1970s, Keynesian economics was viewed by the majority of economists as the approach of the past, and the new majority was subscribing to the approach of NC economics, which marginalised even the Monetarists despite of the fact that it maintained its conservative orientation.

The founder of this new school of economic thought is Robert Lucas Jr. from the University of Chicago and a student of Milton Friedman. Lucas is recognised as exerting a significant influence on macroeconomics since the 1970s. There are three salient features in Lucas's approach:

1. The macroeconomic model must be based on solid microeconomic foundations, which is interpreted to mean that the behaviour of economic agents should be accounted for in the prices of macroeconomic variables.
2. The macroeconomic models obtain dynamic characteristics by accounting for the expectations of households and firms in the long run.
3. The macroeconomic model becomes stochastic, that is, it accounts for the uncertainty factor.

For his contributions to macroeconomic analysis, Lucas was awarded the Nobel Prize in economics in 1995. The major representatives of this school of economic thought include also Thomas Sargent of the New York University and Neil Wallace of the University of Minnesota.

[1]The name New Classical economists or economics was established in Sargent's (1979) article.

[2]By classical, they mean essentially the neoclassical economists before or at the time of Keynes.

In what follows in this chapter, we examine the basic propositions of the NC macroeconomics, that is, the rational expectations hypothesis, continuous market clearing (CMC) and the Lucas supply curve. We continue with an examination of business fluctuations and discuss the NC economists' subtle argument for their occurrence. The ineffectiveness of economic policy follows together with the famous Ricardian Equivalence Theorem. We also make some remarks on the large scale macroeconometric models in the context of the Lucas critique and end up with a summary and some concluding remarks.

14.2 The Rational Expectations Hypothesis

The Rational Expectations Hypothesis (henceforth REH) is attributed to John Muth (1930–2005). Mankiw (1990) in his overall critical evaluation of macroeconomic advances points out:

> The widespread acceptance of the axiom of rational expectations is perhaps the largest single change in macroeconomics in the past two decades. (Mankiw 1990, p. 1648)

The central hypothesis that was stated by Muth (1961) was that the rational expectations of the agents of the economy

> since they are informed predictions of future events, are essentially the same as the predictions of the relevant economic theory [...] we call such expectations rational. (Muth 1961, p. 315)

The relevant (or the prevailing) economic theory embodies the knowledge that is related to the economic process under study and economic agents use this knowledge. Notice that Muth did not hypothesise that all economic agents have the same expectations and that their predictions are the same as those that are derived from the economic theory or that their predictions are perfect. Muth simply supposed in statistical terms that expectations

> tend to be distributed, for the same information set, about the prediction of the theory (or the "objective" probability distribution of outcomes)". (Muth 1961, p. 316)

In order to simplify the presentation Muth (1961) assumed a one-good (*e.g.*, wheat) economy, where the decision to produce is taken at time t and the good is sold after a predetermined future period of time (*e.g.*, after a year). If we take the prices and quantities as deviations from their equilibrium points, then we are going to have for the demand of this good (q_t^d) at time t

$$q_t^d = -\beta p_t$$

where $\beta > 0$. The supply of the good depends upon the expected price and a random disturbance attributed for example to weather conditions. The supply function can be written

$$q_t^s = \gamma p_t^e + u_t$$

where $c > 0$. Furthermore, we assume that there is no autocorrelation between the disturbances and that $E(u_t) = 0$, where E stands for the expected value or expectations. Invoking the equilibrium condition ($q_t^d = q_t^s$) and solving for the price, we get

$$p_t = -(\gamma/\beta)p_t^e + (1/\beta)u_t$$

Since $Eu_t = 0$, we will have

$$Ep_t = -(\gamma/\beta)p_t^e$$

The model is solved once we assume something with regard to the behaviour of the expected price. The simplest hypothesis that can be made is:

$$p_t^e = p_{t-1}$$

that is to say, the expected price is equal to the price of the previous period. In other words, we adopt provisionally, at least, the adjusted expectations hypothesis according to which the prices of the previous years' guide the prices of the next. We substitute the above relation to the equilibrium price equation and we get,

$$Ep_t = -(\gamma/\beta)p_{t-1}$$

Up until now, we relied on the simplest possible assumption of the adjusted expectations. In the more complex case, we take more lags in prices subject to the constraint that the sum of the coefficients of the lags is equal to one. The advantage of the assumption of the adjusted expectations hypothesis (henceforth, AEH) is the recognition of the role that the past plays on current decisions. The assertion that today's situation does not differ significantly from that of yesterday is in accordance to common sense. Moreover, the mistakes of the past are corrected gradually and not automatically while the need to overcome the obstacles that decelerate the process of correction creates the necessary conditions that justify government intervention. In other words in the process of price formation there are various imperfections in the way in which expectations are formed. In this case there is fertile ground for the exercise of fiscal and monetary policy. This is a Keynesian argument that became the target of attack by NC economists.

A more careful look on the subject of the formation of expectations shows that the predictions that stem from the above relation do not agree with the way that expectations are formed in reality. While producers think that a high price will be followed by a still higher one the model builder knows that on an average a high price will be followed by a lower one. If the model is right, then someone could apply it profitably, for example, in the stock market. In reality, however, it is hard to

believe that the producers after such a long time did not learn their lesson. The hypothesis of adjusted expectations has been criticised for mainly two reasons: The first is that it is based on past observations of a single relevant variable. This is counterintuitive since economic agents in the formation of their expectations take into account all the available information. The second reason is that prices display a tendency, consequently, the differences in observations and predictions are autocorrelated. In a nutshell, the AEH leads to systematic forecasting errors.

Muth argued that if expectations are rational, that is, without systematic error, then they must accord to the predictions of the model, that is, we are going to have:

$$p_t{}^e = Ep_t = -(\gamma/\beta)p_t$$

If we rule out the case where $\gamma/\beta = -1$, then it is required that $P_t^e = 0$. The rational expectations require that the price of the next period will be stationary; that is, the one that restores the equilibrium in the market. If the producers use this prediction, the market price will fluctuate randomly around the stationary price. Consequently, the expectations are fulfilled on an average, minimising therefore the limits of speculation for those that happen to know the model. Consequently, if there are not any unpredictable shocks, the hypothesis of rational expectations gives results similar to those of perfect foresight.

14.3 Continuous Market Clearing

The second constituent component of the NCE refers to the hypothesis of CMC, including the labour market. Consequently, the full employment of labour is the rule rather than the exception. The hypothesis of CMC combines the old Walrasian general equilibrium approach and the more recent theory of efficient markets. The theory of efficient markets refers mainly to prices and to the equilibrium that prevails in the markets for securities and commodities. These markets concentrate many of the characteristics of auction markets. Walras's achievement was to develop a mathematical model in order to show that in an economy characterised by competitive markets general equilibrium prevails, whereby all prices are equilibrium prices and are determined simultaneously in a way that secures equilibrium between demand and supply across all markets (including the labour market). In order to interpret the way in which the economy approximates the state of equilibrium without the intervention of some centrally organising authority, Walras invented the parable of the auctioneer, who announces different prices in response to which buyers and sellers manifest their intentions with regard to the quantities that they would desire and at the same time would be willing and able to buy or sell. The auctioneer takes into account the resulting differences and announces new prices in an effort to eliminate the differences between supply and demand. This process of groping toward equilibrium prices and quantities is called tatônnement and continues for the time period required for the attainment of an equilibrium price for

each particular good. It is important to point out that in this process of trial and error; no actual transactions take place until the attainment of equilibrium prices. Hence, the Walrasian auctioneer clearly requires the concept of perfect competition, because no single participant knows anything more than anybody else and every participant is independent of the others. These conditions are satisfied when there are an infinitely large number of infinitesimally small participants. From a macroeconomic point of view the Walrasian system of general equilibrium implies that there are not any indisposed quantities of goods and services including those in the labour market. Consequently, the general equilibrium necessarily implies full employment in the labour market.

The assumption of the efficient market is an extension and further elaboration of the general equilibrium and investigates to a greater extent the nature of equilibrium prices. A market is efficient when the prevailing prices contain all the available information relative to goods and services that are being exchanged. An efficient market diffuses with infinite speed all the available information. This is the reason why the stock market is considered to be the most efficient market, since it makes known to the general public with a speed greater than any other market whatever useful information is available.

The problem with the hypothesis of efficient markets is that it is applied to markets other than the stock market, the foreign exchange market and the commodity (gold, oil and the like) markets and it is not applicable to the goods and services markets.[3] As a result, the NCE are forced to assume that all unemployment is virtually voluntary; implying that whoever really wants to work can find a job at the current equilibrium wage. Consequently, the unemployed must look for the causes of their situation in their unwillingness to work at the current wage.

14.4 The Lucas Supply Curve

One of central constituent components of the NCE is the aggregate supply curve which was developed in Lucas's articles and came to be known as the Lucas supply curve written as

$$y_t = a(p_t - p_t^e) + y^* + u_t$$

where y is the logarithm of total output, p is the logarithm of the general price level and p_t^e stands for the logarithm of the expectations formed on the basis of available information for the general price level, y^* is the logarithm of the output corresponding to the level output associated with the natural unemployment.

[3]The new Keynesian economists cite a series of arguments on the basis of which they claim that the markets do not attain the equilibrium situation as fast as it is required according to the NC economists.

The Lucas supply curve contains indirectly the Phillips curve in the sense that the output and the price level (relative to the expected price level) are positively correlated[4] and implies that current output y_t will increase *vis a vis* the full employment level of output y^* in a quantity proportional to the difference of the current price level (p_t) from the expected price level p_t^e. In the same equation, we add a supply random shock u_t, which is distributed over time with zero mean. This term is added to remind that the supply curve is influenced from a variety of unpredictable factors.

The Lucas supply curve can be interpreted in two ways. The first was proposed by Lucas and Rapping (1969) and refers to the labour market and the second was proposed a few years later by Lucas (1972, 1973) and refers to the goods market. Two are the major characteristics of the Lucas supply curve:

1. The decisions of the agents of the economy (firms, workers) in every case reflect optimising behaviour.
2. Workers and firms make their decisions according to relative prices; that is, the real wage and the price of the product, respectively.

With respect to the labour market Lucas and Rapping (1969) argue that the workers' decisions depend on the difference between the current real wage and the expected normal real wage. If the difference is positive then workers have an incentive to supply more work at the current period and less in the future period. If however, the real wage falls short of the future normal real wage then workers prefer to supply less labour today and more in the future. Only if the current real wage is equal to the future normal real wage is there equilibrium in the labour market.

With regard to the goods market, Lucas (1972, 1973) hypothesises that the firms only know with precision their own price, while the general price level is known with some degree of hysteresis. Thus, when the price of the product increases, the firm must identify the source of the change and if the change in price is caused by

1. the increase in the demand for its product, then the firm must increase its output, for its selling price is lower than the general price level.
2. the increase in demand in general, then there is no reason for the firm to increase its output, for the price of its product increases *pari pasu* with the other prices in the market.

Firms, therefore, in each price change, must decide whether the change is in relative prices, or in the general price level. It follows that in inflationary conditions firms are reluctant to expand or contract their production, thereby affecting the level of employment. The Lucas supply curve can be used to explain the short run cyclical fluctuations.

[4]We say that the Phillips curve is indirectly contained, since if inflation rises then nominal wages also rise, but at a lower rate; consequently, the real wage falls, and, therefore, profits increase and so does investment with the result that the unemployment rate is reduced.

14.5 The New Classical Economics and the Business Cycle

We know that in NC economics the natural state of the economy is that of equilibrium with full employment. Any deviations of the real from the normal magnitudes are considered temporary. Naturally, the question that comes to the fore is how does this theory explain the business cycle? The NC economists argue that if we have as a point of departure a usual supply function according to which the quantity supplied and the price of a product are positively related then every producer will be interested primarily for his own (relative) price and secondarily for the general price level. More specifically, Lucas (1972, 1973) claims that individuals pay attention to their own selling prices and not to the prices of the goods that they themselves buy. As a result, the agents of the economy confuse changes in the general price level with changes in their own relative prices. An unexpected inflation is regarded initially by the economic agents as an increase in their relative price and thus they increase their output and the supply of labour. Thus, if individuals could always distinguish relative price changes from absolute ones then there would never occur any departure from the equilibrium position. This behaviour of firms and their inability to distinguish clearly between relative prices and the general price level may lead to business fluctuations as we discuss below (Fig. 14.1).

Let us suppose the usual graphics where the real GDP fluctuates around its trend and let us further suppose that during a particular period, relative prices remain the same because there have been no changes in real economic conditions. At the same time, the price level rises by a larger-than-normal amount because there has been a larger than normal increase in aggregate demand (or in the money supply). Firms and workers will have no way of knowing that this increase in prices is general and not specific to their own output; consequently, they will proceed under the assumption that their relative price and wages have increased. As a result, they will

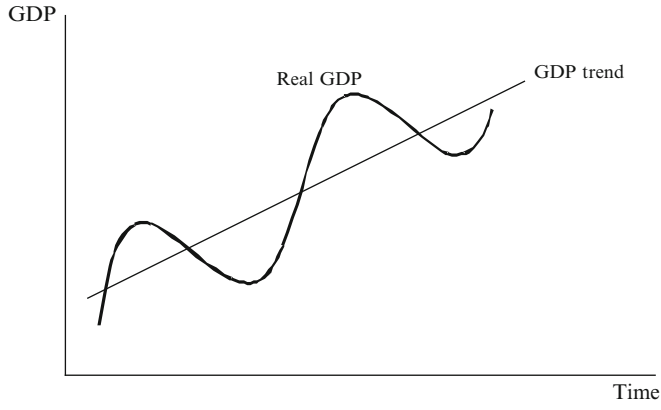

Fig. 14.1 Business fluctuations

increase their output and supply of labour. Firms and workers know precisely their own prices and wages, however, they do not know what happens to the rest of the economy, and, therefore, they increase their output and employment in their industry and if this behaviour is generalised to all industries, then we end up with an increase in the real GDP. This is how the NC economists explain the upturn of the business cycle. Consequently, businesses invest and workers decide to increase their supply of labour. From the moment that these decisions are taken it is not easy to be recalled. For example, in the case of investment, businesses are tied up for the economically useful life of the equipment while at the same time in the labour markets there are legal, moral and other constraints which are difficult to bypass. If, however, all knew a priori that this is an increase in the general price level and not just an increase in the demand for firm's product then, in such a case, no change would take place away from the natural level of (real) GDP. When, however, the public realises its mistake then it gradually revises its behaviour by reducing the quantity supplied and the real GDP returns to its equilibrium path. If there is a fall in the general price level, through a restrictive economic policy; for example, then the economy is led to a recession with deflation. In both cases Lucas's analysis is not based on asymmetric information, as is the case with Friedman's and the Monetarists' analysis according to which the workers are the only ones who suffer from money illusion. According to Lucas, both firms and workers commit expectational errors and respond positively to the variations in the price level, thinking, mistakenly, that the variations are in relative prices, thereby changing their supply decisions.[5]

The process is as follows: when the general price falls, entrepreneurs and workers of an industry think that only their own prices fall, and, therefore, reduce their supply of goods and labour, respectively. At some point, entrepreneurs and workers realise that the fall in prices was general and thus they revise their initial (mistaken) decisions for the price level, they set off the growth process, *etc*.

It seems that the NC economics succeeded in squaring the cycle. On the one hand they offered an interpretation of the business cycle, while on the other hand their interpretation did not contradict the three fundamental propositions of their theory. Nevertheless, a more careful examination of this theory shows that the major problem lies in which there is no evidence of this imperfection in the information of economic agents, and there is no reason why each and every one of the individuals should behave so naively. Consequently, the imperfection in the information which according to NC economics characterises the behaviour of individuals is both theoretically weak and at the same time contradicts our experience. As a result, a series of failures of the NC economics to interpret the business cycles in terms of the lack of adequate information led to the Real Business Cycles approach to macroeconomics with which we are dealing in the next chapter.

[5]In the final analysis, the money supply is what increases or decreases the output produced and this because individuals do not interpret correctly the price changes.

14.6 The Ineffectiveness of Government Intervention

Although Muth's article was published in 1961 nevertheless the hypothesis of rational expectations did not play any role in the economic theory until many years later when it was adopted by Lucas (1972) and later, by Sargent and Wallace (1975), that is from macroeconomists that were investigating the microeconomic foundations of macroeconomics. According to the rational expectations, hypothesis economic units do not adjust their expectations towards the equilibrium point in successive steps, but because economic units do know the result of the economic theory, they are directed towards the equilibrium position instantaneously. In other words, if we suppose that the economic units have rational expectations, then whatever is to happen in the long run will happen in the short run. Since, the effectiveness of economic policy depends on mistaken expectations, the rational expectation hypothesis leads to the conclusion that if the Keynesian economic policy is ineffective in the long run it will also be ineffective in the short run. Now, regarding the Monetarist view, in case that the public has rational expectations, then changes in the quantity of money do not affect the unemployment rate neither in the short run nor in the long run. These three approaches (Keynesian, Monetarist and the NC) can be depicted in terms of the usual graph of aggregate demand and aggregate supply as shown in Fig. 14.2

We initially suppose that the economy is in equilibrium at point A the point of intersection of aggregate demand (AD_0) and short run aggregate supply ($SRAS_0$) curves – and that the output produced corresponds to full employment.[6] According to the usual presentation of the Keynesian model, an increase in the aggregate demand is manifested with a parallel shift of the AD curve to the right. The new equilibrium is point B, that is to say the point of the intersection of the $SRAS_0$ and

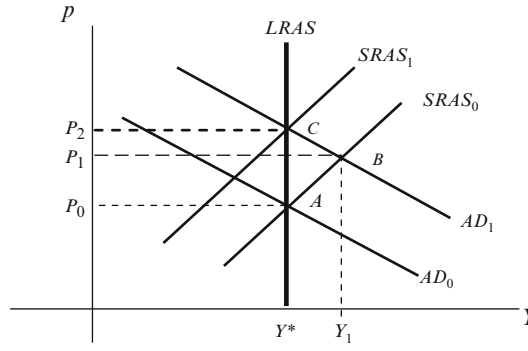

Fig. 14.2 Equilibrium in the market for goods

[6]There is only natural unemployment.

AD_1 curves. Total output increases from point Y^* to Y_1, but at the same time the general price level increases from P_0 to P_1.[7]

According to the monetarist approach the economy is led to point B because of mistaken expectations concerning the changes of the general price level. When the public revises its expectations the supply curve shifts to the left to the new position $SRAS_1$. Consequently, according to the Monetarists, an expansionary monetary policy in the long run can only increase the general price level with no effect on output. The real variables of the economy are affected only in the short run.

The critique of the rational expectation hypothesis to both the Keynesian and the Monetarist approaches is that the public is perceived to know the consequences of the change in the economic policy. As a result, when the government increases the supply of money Keynesian economists suppose that the public does not take into account the inflation, and thus is under money illusion. In other words, Keynesians suppose that for the public $p_t^e = p_m$, where p_m is the nominal price, while Monetarists argue that the public forms its expectations *post factum*, that is to say the agents in the economy form adjusted expectations $p_t^e = p_{t-1}$. In the rational expectations hypothesis the public is not under the influence of money illusion on the one hand and, on the other hand it reacts directly to every change in policy with the supply curve shifting automatically to the position $SR\varsigma S_1$, and the economy moving from equilibrium point A directly to point C skipping point B. Consequently, according to NCE the demand management policies are likely to fail in both the short run and the long run. If, for example, the monetary authorities decide to increase the supply of money at an annual growth rate of 4% the public will adjust its expectations taking into account the consequences of such an increase in the supply of money. If, however, the monetary authorities decide without warning to increase the supply of money at an annual rate of 6%, then the non-anticipated percentage of 2% would lead to an increase in output and employment above the normal level. Consequently, according to the NCE, only in the case that the change in the supply of money is sudden, and, therefore, unexpected, can we have in the short run an unstable equilibrium such as that at point B.

14.7 Barro's Ricardian Equivalence Theorem

The view about the ineffectiveness of government policy leads to some conclusions in regard to the financing of government expenditures. More specifically, it has been argued that taxation and public borrowing constitute essentially equivalent forms of financing of public expenditures. This idea has been expounded by Robert Barro and came to be known as the Ricardian Equivalence Theorem (Barro 1974, 1980) a term coined by Buckanan (1976), and then became a standard topic in public finance and macroeconomics. The rationale behind this view is that the

[7]In other words, the degree of capacity utilisation of the economy increases.

government is expected at some future time to redeem its debt. If we now suppose a closed economy then the repayment of debt will take place via increased future taxation, which means that on the basis of the rational expectations hypothesis individuals increase their savings buying the bonds that have been issued by the government. The amount of savings in other words matches the size of public deficit, and, therefore, the interest-rate remains the same, which means that there is crowding out effect of the private investment from the public expenditures and, therefore, the overall demand remains the same together with the other real variables of the economy. A similar operating mechanism is activated in the case of an open economy, where the redemption of public debt takes place via the sale of assets to international institutional agents. Such a possibility raises, once again, the question of limited future government income and, hence, the inevitable future increase of taxation. As a consequence, the final effects of deficit spending are similar regardless of the mode of the financing.

One criticism to the above argument is that individuals have a limited life time, and, therefore, do not care very much about the tax, since what they will pay at the end of their lives will be less than what they are called to pay once and for all. The individuals as a result of tax reductions engage themselves more in consumption spending rather than in saving. The counterargument here is that bequests must also be accounted for. If we therefore in our analysis use the household instead of the individual in effect we form an "institution" with an infinite life time (since the incomes from assets in general are inherited to the heirs as bequests) and thus the hypothesis of rising savings in the case of rising public expenditures continues to hold. As a consequence, the policy of cutting taxation through the issue of bonds in the effort to raise aggregate demand cannot but fail, because the economic agents anticipate the future increase in taxation for the payment of interest on the public debt and so they are prepared accordingly.

The Ricardian equivalence theorem was very popular especially in the 1980s because of its ineffectiveness of government intervention conclusions. The empirical evidence from various countries produced mixed results; however, it is clear that if we take the US economy then the theorem does not pass the test. We know that despite rhetorics about limited government in the US the public debt increased dramatically over the last decades in both absolute and relative to GDP terms. One would expect according to the Ricardian Equivalence Theorem that consumption expenditures would diminish and savings would increase in order to generate enough funds for the redemption of the public debt in the future. Only a cursory look at the data is sufficient enough to show that neither private consumption expenditures decreased nor private savings increased casting doubt to the validity of such a theorem.[8]

[8]In fact, saving as a percentage of the disposable income in the US dropped from about 10% in the 1980s to 7% in the 1990s and by the year 2000 fell to just 1%.

14.8 The Lucas Critique

Lucas (1976) argues that we cannot use the parameters of the usual functions of consumption, investment, demand for money, *etc*., that are derived from econometric techniques for the contemplation of economic policy. The parameters of such functions depend on the decisions that have been taken in the past with given the utility functions. In other words in these estimations of the above or similar functions, there is an optimising procedure prior to them, which is not accounted for in the macroeconometric estimations. If, in what follows, we suppose that on the basis of these functions we intend to exercise, for example, fiscal policy, then one of the results of this decision is that the utility functions of the consumers change. As a consequence, the consumption function changes with the result that the government expenditures multiplier will have a different value from the one we initially estimated and thus the final result of this policy will be different from the expected. Consequently, we cannot take directly the *IS–LM* or *AD–AS* models and use them in the exercise of fiscal and monetary policy, because the parameters of these models can be changed as the economic agents adjust their expectations and their behaviour in general with regard to the new political environment.

Lucas claims that the macroeconometric models must include the utility functions and the production functions which describe the behaviour of the economic agents and analyse how exactly these agents react in every change in policy, since these are the fundamental equations of the economic system. This was the reason that led to the development of economic models, which are called Real Business Cycles, the details of which we examine in the next chapter. The Lucas (1976) critique essentially constitutes an internal critique of NC economics and bears many similarities with the Goodhart laws, from the name of the former president of the Central Bank of England. According to these laws, if we observe various monetary variables, we may detect regularities in their behaviour. Charles Goodhart describes his findings as follows:

> any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes. (Goodhart 1975, pp. 322–323)

In other words, if we observe the monetary variables, over a period of time, we may discover strong statistical correlations and there is always the temptation to use these relations for the exercise of economic policy, however, Goodhart's laws say that such a policy is doomed to fail because when we target such correlated variables then a financial innovation is being created which contaminates the previously observedstrong statistical relationships between the variables.

There is no doubt that the Lucas critique is theoretically justified nevertheless one wonders about its significance at the empirical level. In other words, how sensitive are the parameters of the econometric models in the case of alternative economic policies? This is an interesting question that Taylor (1989) grappled with. In fact, Taylor, starting with the Lucas critique as being valid in the theoretical level, discovered that the parameters in the macro-econometric models are surprisingly

rigid in case of different economic policies. These results cast doubt to the Lucas Critique (Mankiw 1990, p. 1647) and thus the macroeconometric models can be used for policy purposes. This by no means invalidates Lucas Critique but it only emphasises that the practical significance of this critique is limited. The issue of stability of the parameters in econometric specifications remains open and in the future one can only expect positive results with respect to the ability of economists to understand the way in which the economy operates and the effectiveness of alternative economic policies. If there is something positive in the Lucas critique then this is the promotion of the idea of finding proper microfoundations of macroeconomics.

14.9 Concluding Remarks

The NC economics is in a similar position with the Monetarism that is a *degenerate research programme*. In fact, NC economics have contributed many ideas that have been assimilated from the Real Business Cycles and the New Keynesian approaches that we examine in the next chapters. In what follows, we try a critical evaluation of this school of economic thought. One of the weaknesses of this approach is the hypothesis of CMC in all markets. Another critique refers to the way in which the economy reacts in an anticipated change in economic policy (see Fig. 14.2) where in this case it is ascertained that the REH leads the agents of the economy automatically to point C without the intervention of point B. From the above we arrive at the conclusion that the REH essentially spirits away the time factor and the analysis essentially resembles the Walrasian one, where instead of the auctioneer we have the REH, with the result that in both cases there are no transactions in disequilibrium prices. Other critiques that have been launched against NCE are discussed in detail in Eatwell (1983) and concern a series of malfunctions (imperfections) of the market. The list of these criticisms includes the following:

1. Prices are determined in non-competitive markets as a mark-up on cost (Modigliani 1977; Tobin 1980). Consequently, the economic agents are anything but price-takers as they are supposed in the continuous market-clearing hypothesis.
2. The adjustment of wages and relative prices towards the equilibrium levels is slow (Tobin 1980).
3. Neither markets for labour nor economic agents are homogenised enough so as to give rise to the conclusions of the REH (Modigliani 1977).
4. It has been argued that the economy only accidentally can be found on the point of natural equilibrium (Modigliani 1977).
5. In econometrics autocorrelation is a frequent problem in times series data, consequently, the assumption of the REH that the error terms are not serially correlated does not have the required empirical content.

Eatwell (1983a) argues that these critiques simply discern weaknesses of the REH and the NC economics and if taken into account then they may improve this approach. Yet, in no way do they constitute a fundamental critique in the logical

foundations of the theory of NCE. The major problem according to Eatwell is that the rational expectation hypothesis accepts the orthodox model of the functioning of the market.

The empirical evidence was initially in favour of the NC approach, as was shown, for example, in the articles by Barro (1978, 1979). Nevertheless, some other studies such as that of Minskin (1982) and also by Gordon (1982) found that the level of output and employment are affected from the expected and also the unexpected economic policy. The so-called silent depression of the 1970s which was extended in the 1980s constituted another evidence against the REH. The opponents of the NC economics argued that it is impossible to have such a deep and simultaneously prolonged recession given that information on the supply of money, prices, *etc*., are readily available to anyone interested. Consequently, both the duration and the intensity of the recession could not be interpreted as a result of the lack of adequate information, as NC economists would argue.

This divergence between theory and empirical evidence led many NC economists to look for other theoretical explanations although they maintained the REH and the CMC hypothesis. These economists further argued that the public has perfect knowledge of the changes in monetary policy and that the business cycles come from the supply side of the economy, that is, from permanent shocks that come from real and not monetary variables.

There is no doubt that NC economics is an approach of the past and that some of its propositions have been assimilated to current approaches such as the Real Business Cycles approach and the New Keynesian economics. It is true that the REH is still being used in macrotheorization however more like a reminiscence of the past and less like a theory generating concept.

Questions for Discussion and Thought

1. Discuss the Lucas Critique and compare it with Goodhart's laws.
2. Is the Lucas critique theoretically and empirically justified?
3. Present, compare and evaluate the weak and strong version of the rational expectation hypothesis.
4. Discuss the policy prescriptions of the NC economics.
5. How do business cycles appear according to the NC economics?
6. According to Lucas workers suffer from a type of "money illusion". They supply additional labour in response to an expansionary monetary policy. Could the government exploit this money illusion?
7. What is the current status of NC macroeconomics?
8. Discuss and critically evaluate the Ricardian Equivalence Theorem

Notes for Further Reading

We know that Keynesian economics emerged during the period of the great depression of the 1930s and were established during the first post-war decades. The so called silent depression of the 1970s and 1980s was the one that

(continued)

cast doubts on the Keynesian orthodoxy and paved the way for new macroeconomic approaches. In the beginning, the Monetarists were those that offered their own explanations of the phenomenon of stagflation. Soon however it was realised that these explanations were at best inadequate to deal with the complexity of the phenomena that they were put to resolve. In this context, we had the emergence of the New Classical macroeconomics that was the result of the pioneering works of Robert Lucas Jr. of the University of Chicago. Lucas started as a Monetarist but quite early he turned the focus of his research on issues that came to form the backbone of a new stream of economic thought and at the same time to exert immense influence to macroeconomic thought in general. Among the works of Lucas two are those that exerted most of the influence, the first in 1972 and the second a year later. In these two articles, Lucas developed the fundamental concepts that formed the theoretical core of the new approach. In these concepts the first is the *Lucas Supply Curve* which as Lucas pointed out it should more accurately be called the *Lucas-Rapping Supply Curve* for it was discussed for the first time in a joint paper with Rapping (1968). In an overall appraisal of its significance in the modern macroeconomic analysis, the *Lucas-Rapping Supply Curve* would be put on par with Friedman's idea of permanent income and Modigliani's life cycle hypothesis in the theory of consumption, or the theorization of risk in Tobin's portfolio analysis.

It is interesting to note that the hypothesis of rational expectations was first presented in 1959 by Muth (a graduate student at that time) in a seminar and two years later it was published in *Econometrica*. Muth had written two more related articles which together with the first were reprinted in a book of readings edited by Lucas and Sargent (1981). There is no doubt that the REH is an entirely new idea that exerted a lasting influence on modern macroeconomics. Rarely, did so few pages have such an impact in the history of economics. Nevertheless, it is clear that without Lucas's work the idea of rational expectations and its author would go to oblivion.

The Ricardian Equivalence Theorem was presented for the first time by Barro (1974) but the characterization is attributed to Buchanan (1976). O'Driscol (1977), among many others, does not accept that Ricardo had really such an analysis of equivalence between taxation and public borrowing. In fact, according to O'Driscol's interpretation of Ricardo the introduction of taxation and the issuance of public bonds are indeed equivalent mechanisms for the collection of the necessary funds from the point of view of collection costs and not from the point of view of economic results. Barro (1989b) returning to this question argues that on the one hand there is textual evidence in Ricardo in favour of the theorem and on the other hand Barro invokes the so-called "Stigler's Law" according to which none of the important theoretical innovations is named after its originator. To our view the classical economists (Smith, Ricardo and J.S. Mill) did not support in

(continued)

general that the methods of financing of government expenditures are equivalent as we argued in Chaps. 2 and 3.

Lucas (1976) with his famous critique cast doubt on the large scale macroeconometric models that are based on the stability of parameters despite the change in economic policy. Theoretically speaking Lucas (1976) is right in his claim, however practically it was found that the parameters of these models do not change to such an extent so as to alter qualitatively the results of the analysis. For a critique of Lucas we recommend the articles by Blanchard (1984) and Taylor (1989). It is certain that the first economists working in the specification of their grand scale macroeconometric models (*e.g.*, Tinbergen, Klein, *inter alia*) were already aware of the possibility of the instability of the parameters from their own experience. Tinbergen had explicitly pointed out such a problem; however, economists before Lucas did not think that their results were affected in any significant way by the change in the parameters, which in any case did not have any clue how to estimate in the case of a regime change. A similar to Lucas's critique was advanced by Goodhart (1975) in the so called "Goodhart laws" which however is restricted to monetary variables. It might be remarked in passing that Goodhart published his law a year earlier than Lucas and that Lucas presented his critique in 1973 in a conference that took place in Carnegie–Rochester. There is no doubt that the two economists arrived at similar conclusions working independently of each other during a turbulent period of time with an economy hit by many and important shocks (*e.g.*, Vietnam War, oil crisis, collapse of Bretton Woods agreements, *etc.*). Naturally, one would be careful to claim the stability of the parameters in such an economic environment.

Lucas (1975) is also credited with the advancement of a business cycle model based on the general equilibrium, while his articles in (1969) (written jointly with Rapping) and (1972) focus on the supply curve that bears his name. It is important to note that during the 1980s Lucas dealt with economic growth and his contribution was adopted by other approaches such as the real business cycles and the New Keynesian economics that we deal with in the next two chapters.

Chapter 15
The Real Business Cycles Approach

The use of the term business cycle is unfortunate for two reasons. One is that it leads people to think in terms of a time series business cycle component which is to be explained independently of a growth component; our research has, instead, one unifying theory of both of these. The other reason that I do not like to use the expression is that it is not accurate; some systems of low-order linear stochastic difference equations with a non-oscillatory deterministic part, and, therefore, no cycle, display key business cycle features (see Slutzky 1927). I thus do not refer to business cycle phenomena, which are nothing more nor less than a certain set of statistical properties of a certain set of important aggregate time series.

(Prescott 1986, p. 10)

[P]eople got scared [...]. The press scared people. People running for office scared people. Bernanke scared people; Paulson scared people.... [P]eople began not to know what was going to happen. Then they stopped investing – by investing, I mean getting a new car or fixing up your house. And that led to the economy – it was depressed a bit that fourth quarter of last year... [With] benign neglect the economy would have come roaring back quite quickly [...].

(Prescott in an interview March 30, 2009)

15.1 Introduction

Up until now, we discussed the way in which the New Classical (henceforth, NC) economics restored and brought to centre stage the idea of cyclical fluctuations on the basis of the assumption of continuous market clearing in all markets and the rational expectation hypothesis that characterises the behaviour of the economic agents. According to NC economists, the lack of adequate information on the difference between relative prices and the general price level as well as the monetary shocks constitutes the major causes of fluctuations in the real GDP. This explanation of the source of business fluctuations, on the one hand, questioned the Keynesian orthodoxy and, on the other hand, paved the way for the emergence of the Real Business Cycles (henceforth, RBC) and the New Keynesian economics.

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_15, © Springer-Verlag Berlin Heidelberg 2010

This chapter begins with a history of the discussion of business cycles and continues with the major characteristics of the RBC approach, while the causes (real or monetary) of business cycles and the separation of cycles from growth (trend) follow. A short description of the RBC simulation models as well as the policy implications of the real business cycles is the next topic, and the chapter concludes with a summary and some critical remarks about the approach that has attracted a lot of attention at least up until the outbreak of the crisis of 2008.

15.2 Salient Features of the RBC Approach

The emergence of the RBC approach took place in the decade of the 1980s, when a small, but influential group of economists dissented from the NC economics and offered new interpretations of cyclical fluctuations.[1] According to this group of economists, business fluctuations are created from the supply and not from the demand side of the economy, as the majority of economists up until the early 1980s, at least, used to theorise.[2] In other words, economic fluctuations are attributed to stochastically appearing changes in real (and not in monetary variables) with these changes being diffused rapidly throughout the economy. Historically, theories that attribute the business fluctuations to changes in real variables can be looked either in their simple form, as is the case of sun spots and the associated changes in the weather (as for instance, the sunspot theory of business cycles of Jevons) or in the more sophisticated form of Schumpeter with the swarms of innovations. It is important to stress that in this approach the monetary variables continue to be important, nevertheless, their role is secondary to that of real variables.

Keynesian economists as well as monetarists argue that business cycles are mainly due to changes in aggregate demand. The NC economists claim that changes in the aggregate demand cannot exert lasting changes on real GDP. The flexibility in money wages and prices of the other inputs restore the real GDP to its normal level through direct and relatively fast changes in the short-run aggregate supply (see Fig. 3 in Chap. 7).[3] Historically, however, we know that modern economies have experienced recessions of long duration, and if the changes in aggregate demand do not constitute the cause for business fluctuations, then the causes of these fluctuations must be searched for in the supply side of the economy.

It is ironic that Tobin, a towering figure among Keynesian economists, provided the rationale for the development of the RBC approach! More specifically, Tobin

[1]Among the most important representatives of this approach are included: Edward Prescott (Minnesota), Finn Kyndland (Carnegie-Mellon), Charles Plosser, John Long, Robert King, Alan Stockman, Sergio Rebelo (all from the university of Rochester) and Robert Barro (Harvard).

[2]This approach cannot include changes in demand, such as change in preferences or government policies; the effects of these changes are considered to be strictly limited (Ploser 1989, p. 57).

[3]Robert J. Gordon characterised this possibility of presenting concisely many competing views in terms of slopes of two curves as “the paradox of convergence without agreement”.

(1980), in his critique of Lucas and the NC economics, observed that in the case of a model with perfect information the equilibrium position will be moving because of changes in the natural resources, technology and also in preferences. If these changes continue as random processes, then they are responsible for the generation of business cycles. Tobin's observation triggered the development of a whole new research programme which was initiated by the articles by Kydland and Prescott (1982) and also by Long and Plosser (1983). In these articles, it was argued that the cyclical fluctuations are due to disturbances in real variables such as:

1. Natural phenomena (earthquakes, floods, *etc*.) that influence mostly the agricultural production and then are transmitted to the rest of the economy.
2. Important changes in the prices of strategic inputs. A characteristic example is the price of oil in 1973 and in 1979 as well as in 2007. By contrast, in periods where the price of oil is low the economy is positively influenced.
3. Wars, political stalemate, general strikes, *etc*. The examples that are often referred to are the Vietnam War, dictatorships that succeed elected governments as well as sweeping political changes such as those that took place in the former Yugoslavia and earlier on in the former Soviet Union and Eastern Europe.
4. Government policies, which when unannounced may end up affecting real economic variables.
5. Changes in productivity which are created from improvements in the quality of capital and labour and are attributed to innovations of Schumpeterian type in products, but also in techniques of production and management.

From the list of the above variables, all have been proved important in different time periods; nevertheless, the hitherto research has shown that the changes in productivity are qualitatively different and are considered primarily responsible for both the economic growth and the business cycle. Technological change influences the long-run growth path in productivity but, at the same time, constitutes a disequilibrium factor. Within this framework, it is argued that cyclical fluctuations and economic growth are two intrinsically connected phenomena of the same process.

We know that the depression of the 1930s discredited the neoclassical approach and the associated idea that the market, if left to its own devices, possesses the ability to insulate the economy against cyclical fluctuations, which, in any case, have exogenous causes. The Keynesian economics had been established during the depression of the 1930s precisely because of the weakness of neoclassical economics to explain the long-run character of the recession as well as its intensity. Keynesian economists claim that the system is crisis prone because of its internal shortcomings; more specifically, the market system suffers from an endemic lack of adequate effective demand; consequently, Keynesian economists favoured fiscal and monetary policies that would lead to an increase of aggregate demand in the effort to restore the system to the full employment of all the factors of production including labour. The dominance of Keynesian economics was complete during the 1950s and early 1960s, when the major problem of economists was the optimal combination of fiscal and monetary policy in order to attain the desired goal in the best possible way. Yet, while the crisis of the 1930s led to the rise of Keynesian

economics, the crisis of the 1970s discredited the Keynesianism of the neoclassical synthesis and led to the rise and fall of various other approaches from which the most recent one is the RBC.

15.3 A Short Historical Excursion

The changes in real variables as the cause of business fluctuations had a prominent position in the texts of classical and the first neoclassical economists. The situation changed dramatically with the ideas of Marshall and Wicksell who argued that monetary variables obtain greater significance, although the idea of economic growth through cyclical fluctuations was fading away in the memory of most economists until the depression of 1930s. After the publication of Keynes's *General Theory,* a series of models of business fluctuations that focused on the interaction between accelerator and multiplier were developed (Samuelson 1939; Hicks 1950). These efforts can be essentially characterised as RBC models, since they attribute business fluctuations to changes in real variables, while the monetary variables play a secondary role. Keynesian economics of the neoclassical synthesis gave the impression that the economy is essentially manageable and through the appropriate mix of fiscal and monetary policy it is possible to achieve goals such as full employment, economic growth, stability in prices and balance in the foreign exchange rate. Soon, however, it was discovered that the simultaneous achievement of all the above goals was impossible, since the goals were competitive to each other and so attention shifted to optimisation. Thus, during the decade of the 1960s controversies between economists were about the optimal mix of fiscal and monetary policy the effectiveness of which was undisputed, excepting the cases of some radical economists. Thus, there was general satisfaction with the status of economic theory. The idea of serious fluctuations in the economy and moreover with some periodicity lost its popularity.

Burns (1960) was perhaps among the first economists who argued in favour of the idea of the end of the business cycle as we know it, at least for the economy of the USA, and this is because significant progress had been achieved in understanding the way in which the economic system is functioning. As a consequence, the right mix of fiscal and monetary policy can smooth out the effects of the business cycle and so ameliorate its negative effects. The problems that economies faced before WWII were not to appear again. Ten years later, Bronfenbrenner in his popular book titled: *Is the Business Cycle Obsolete*? (1969) argued that business cycles are phenomena of the past, when the economy was based almost exclusively on the operation of the blind forces of supply and demand; however, with the emergence of Keynesian economics and with the accompanying government intervention, Bronfenbrenner further argued, business cycles (at least the serious ones) were defeated for good. It is important to point out that such a viewpoint was receptive since the GDP in the USA and the other OECD countries, during the first

post-war decades, was growing at a very strong pace. Recessions, of course, were present, but they were short-lived and shallow.

The slowdown in the level of economic activity from the mid-to-late 1960s and to the 1980s made clear that, on the one hand, the growth rates of the advanced economies were much lower relative to those of the previous decades and, on the other hand, recessions were more frequent and deeper and were lasting longer than in the past decades. Many economists interpreted these recessions on the basis of Friedman's monetary theory. The NC economists, however, overturned this idea by arguing that the influence of monetary policy is neutral with regard to real variables except if the changes in the supply of money are not expected. But even in this extreme case, that is, the sudden change in monetary policy, the research of many economists of the RBC approach has shown that the relation between monetary and real magnitudes is much looser than what Monetarists and NC Economists thought (*e.g.,* Nelson and Ploser 1982).

The problems do not stop here, since even if we suppose that within the framework of NC economics there are monetary cycles then it must be demonstrated that the monetary shocks have a lasting influence on the real output and that on the other hand the economy is restored to its optimal growth rate after the initial shock. The NC economics, in order to justify the maintenance of deviations from the optimal growth rate, introduced in their analysis durable goods and stocks of finished goods. With these given together with the simultaneous failure of Keynesian economics, many macroeconomists directed their research efforts to the construction of macroeconomic theories, where the supply side is based on solid microeconomic foundations. Moreover, the research of Nelson and Plosser (1982) has shown that shocks in real variables are much more important than those of monetary variables in the interpretation of the long-run growth path of real GDP. Prescott (1986), on the other hand, argued that even though we assume away any monetary shocks in the economy after World War II, nevertheless we would have business cycles.

The fluctuations in economic activity according to the RBC approach are interpreted on the basis of changes in productivity and in particular of changes in total factor productivity, that is, the effectiveness with which capital and labour create output. Total factor productivity changes over time with the improvements in technology and workers skills. Furthermore, total factor productivity increases for a number of other reasons, which include the introduction of new goods and services, the reduction in the price of some basic inputs such as oil, *etc*. By contrast, total factor productivity may fall for reasons such as bad weather that affects agricultural production, the introduction of restrictive legislation, the rise of the price of some basic inputs, and so forth.

If the growth rate of total factor productivity exceeds the average, it follows that there are many investment opportunities and the same is true for employment opportunities. As a consequence, the economy is in its prosperity stage. Hence, we need to take into account that if total factor productivity exceeds its average growth rate and this difference persists for a few years then it follows that the rise in investment, employment and income tend also to persist. The converse will be true in periods of recession, that is, the period during which the growth rate of total

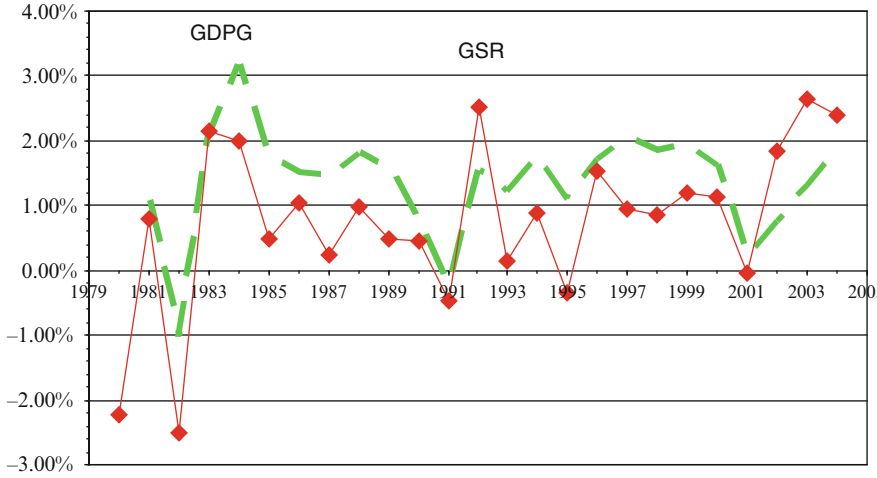

Fig. 15.1 Growth rate of GDP and the growth rate of total factor productivity[4]

factor productivity falls short of its long-run average growth rate. The evolution of total factor productivity, according to estimates provided by Kydland and Prescott (1991), explains the fluctuations in the GDP of the USA during the period 1954–1985 by about 70%. Because the epicentre of analysis of the RBC is the total factor productivity or what amounts to the same thing the growth of Solow's residual (GSR), we portray in the same figure the fluctuations in the growth of the real GDP (GDPG) of the US economy together with the fluctuations in the total factor productivity. Clearly, these two variables are too closely related (Fig. 15.1).[5]

According to the RBC argument, the agents of the economy form their expectations rationally as in the NC economics, but, moreover, are not subjected to mistaken information with regard to the path of the general price level which becomes publicly known. If there is a problem of information then this refers to the characterisation of shocks in productivity, that is, whether they are permanent or temporary. Other than these, the RBC and the NC economics share the following assumptions:

[4]Solow's residual is determined in an economy that functions in conditions of perfect competition and constant returns to scale by using the following form of production function $Y = AF(K, L)$, where A represents the exogenous technological change or the productivity shock. For the estimation of Solow's residual, we hypothesise a Cobb–Douglas production function $Y = AF(K, L) = AK^aL^{1-a}$. We assume constant income shares for the factors of production with the share of capital, $a = 0.3$. If we further suppose discrete time with $dt = \Delta t = 1$, then we arrive at the following relationship: $\Delta A/A = \Delta Y/Y - [a\Delta K/K - (1-a)\,\Delta L/L]$, where $\Delta A/A$ is the growth rate of the Solow's residual.

[5]The data for the growth of the real GDP in the US economy 1980–2004 and the total factor productivity come from The Groningen Growth and Development Centre (http://www.ggdc.net).

1. Continuous market clearing in all markets.
2. Fluctuations of the real GDP and employment stem from stochastic disturbances in technology.
3. The changes in employment are caused by the choices of households to allocate their time between work and leisure.
4. Monetary policy does not affect the real variables; in other words, money is neutral.

The distinction between short-run and long-run analysis is abandoned. In other words, the long-run tendency and the business fluctuations are considered intrinsically connected to common determining factors.

15.4 Economic Growth and Cycles

One of the most important issues in the literature on the evolution of various macroeconomic variables is the distinction between the long-run tendency of a variable such as the GDP and its fluctuations. According to the usual approach, we suppose that the real variable evolves around its long-run tendency as shown in Fig. 3 of Chap. 14. This view characterised almost all approaches to economic thought (classical, neoclassical, Keynesian, even that of the NC economics), at least until the decade of 1980s. The neoclassical economists viewed that these deviations from the trend are only transient and by and large not important, while Keynesians viewed these deviations as persistent and also as being serious enough so as to justify government intervention. Monetarists initially and the NC economists later viewed government intervention as unnecessary and argued that the market is endowed, with its own internal equilibrating devices that restore the economy along its long-run upward tendency. This view was questioned in the research of Nelson and Plosser, who claimed that:

> We conclude that macroeconomic models that focus on monetary disturbances as a source of purely transitory (stationary) fluctuations may never be successful in explaining a very large fraction of output fluctuations and that stochastic variation due to real factors is an essential element of any model of economic fluctuations. (Nelson and Plosser 1982, p. 141)

Nelson and Plosser were led to this significant conclusion (that is, that real variables are responsible for cyclical fluctuations) because in their research they discovered that they could not reject the hypothesis that the real GDP[6] evolves in such a way that it resembles a random walk. More specifically, Nelson and Plosser (1982) tried regressions with GDP data of the US economy using the following econometric specification:

$$y_t = a_t + by_{t-1} + e_t.$$

[6]Henceforth, when we refer to variables such as GDP and investment, we mean the real and not the nominal magnitudes.

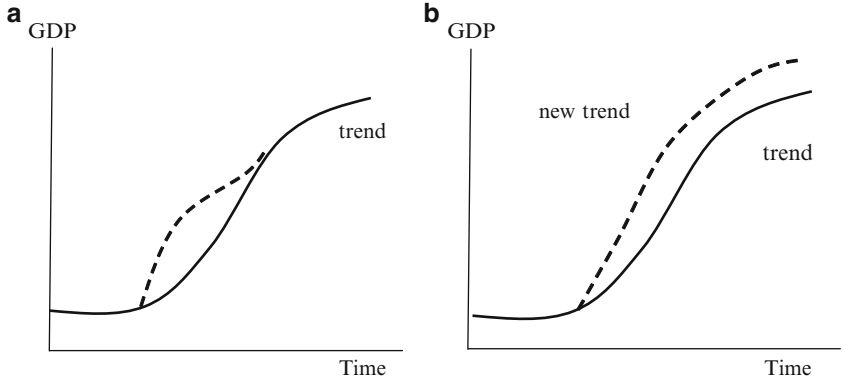

Fig. 15.2 Temporary and permanent changes in GDP

The authors ascertained that the hypothesis that the slope coefficient is equal to 1 ($b = 1$) cannot be rejected, which is equivalent to saying that the GDP follows a random walk with a_t showing the average growth rate of GDP, that is its long-run tendency, and b showing the drift of the long-run tendency of the GDP and the size of the shock that is exerted on the economy in year t.[7] In other words, according to the traditional view, if the economy suffers an external shock, then it deviates from its long-run equilibrium path and it returns to that only after the passage of some time proportional to the size of the initial shock. According to the traditional approach, the shocks have only a temporary character (Fig. 15.2a). Proponents of the RBC approach on the other hand argue that the shocks determine the new equilibrium path and that the economy continues on that path without necessarily returning to its previous path (Fig. 15.2b). Consequently, the shocks according to the RBC approach have a permanent and not a temporary character.

The logical consequence of the above findings of Nelson and Plosser is that if all the technological shocks are frequent and not stochastic, then the path of the GDP resembles a random walk and thus will display features that resemble a business cycle. Hence, some caution should be applied since the fluctuations in GDP essentially refer to fluctuations of the natural tendency of the real GDP and not to deviations from a regular and fully determined long-run tendency of the GDP. In other words, the observed fluctuations of the real GDP must not be interpreted to mean deviations of the observed GDP from some natural GDP, but rather as fluctuations of the natural GDP. These fluctuations might be caused by a multitude of permanent disturbances that determine an entirely new trajectory in the growth rate of the real GDP. We know that economists traditionally make the distinction between

[7]We say that the time series data variable y follows a random walk or it has a unit root if $b = 1$. In such a case, we know that the variable, with the passage of time, drifts away from the equilibrium position.

economic trend (growth) and cycle and construct relevant economic models. For the adherents of the RBC approach, the term 'business cycles' is not the proper one (Prescott 1986) since it indicates a phenomenon that is independent of the factors determining economic growth. More specifically, Nelson and Plosser in their own research ascertain that the determining factors of the natural tendency (growth) of the GDP do not differ from those of the cyclical fluctuations. Prescott (1986) notes: 'Indeed if the economy did not display business cycles phenomena, there would be a puzzle'. Since the monetary factors and the related monetary shocks do not exert permanent changes in the GDP because of the hypothesis of neutrality of money, it follows that only the real variables are capable of leading the economy to disequilibrium. The monetary shocks are considered to be of secondary importance in comparison to the real shocks, which exert changes in the real GDP, with long-run effects.

A corollary of Nelson's and Plosser's analysis is that the usual practices of the distinction between trend and business cycle are meaningless because business cycle and economic growth are two intrinsically connected phenomena. This is the direction of the research of many economists of the RBC approach, that is, they endeavour to synthesise a single theory of both economic growth and business cycle. The basic model of the RBC approach is based on the following three hypotheses:

1. Total output consists of a single good.
2. The analysis is carried out with the real variables without the mediation of money.
3. Households voluntarily supply their labour services according to the wage level.

In what follows, we examine each of these assumptions in the case of unexpected technological shocks and their effects on the economy.

15.5 Shocks in Technology

In Fig. 15.3, we illustrate the way in which output and employment change in a model of the RBC. In the upper part of Fig. 15.3, we present a production function, and in the lower part, the labour market. The RBC approach hypothesises technological shocks as opposed to technological progress to indicate the unexpected or random nature of technology and not the smooth, and, therefore, expected way of its introduction as is usually supposed. Assume now a technology shock described by an upward (leftward) rotation of the aggregate production function. The results of the new production function on employment are depicted in the lower part of Fig. 15.3 that portrays the labour market.

Technological change, according to the economists of the RBC approach, increases the demand for labour as well as the demand for output. The increase in employment depends on the elasticity of the supply of labour with respect to the wage, which according to the RBC approach is high enough, as shown in the lower part of Fig. 15.3. In such a case, the technology shock will lead to the increase of

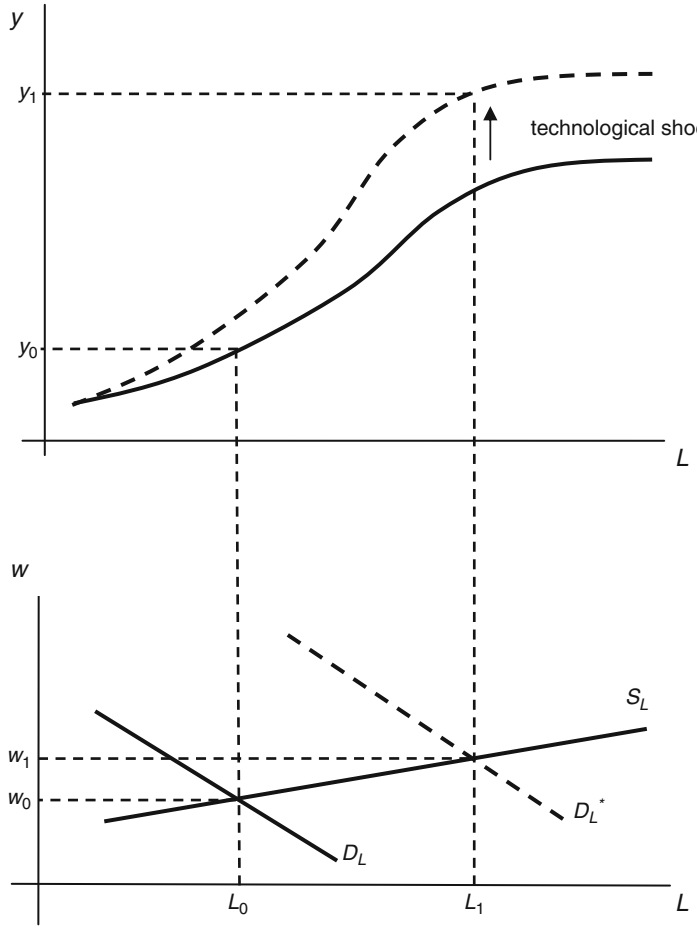

Fig. 15.3 Aggregate production function and the labour market

production from y_0 to y_1 and of employment from L_0 to L_1 with only a slight increase in the real wage. To what extent does an increase in the real wage lead to an increase in the supply of labour is mainly an empirical question. Theoretically, it could be argued both ways, for example, one could claim that the increase in the real wage may lead to a reduction in the labour supply in the case where workers have a target real wage, which once achieved, will induce workers to increase their leisure time and to decrease their working time (supply of labour). However, it is also possible that higher wages may stimulate an increase in the supply of labour and a decrease in the leisure time, a result also known as the substitution effect. One wonders which of these two effects is stronger, and the RBC economists argue that this depends on the type of technological progress, that is, if technological progress is transitory then the substitution effect is the stronger one, whereas if technological progress persists then the wealth effect will dominate.

It is important to point out that in the analysis of the RBC, the real interest rate also plays a significant role in the decisions of the economic agents. If, for example, the real interest rate increases, then it follows that the present value of future incomes decreases, as a consequence, workers have a reason to increase their hours of work and reduce their leisure time. Consequently, the supply of labour is increased, which in turn leads to an increase in both employment and output produced. However, the assumption of substitutability between labour and leisure in every change of wage is a question that is hard to be supported theoretically, while the empirical research is inconclusive and also casts doubt on the hypothesis of substitutability between leisure and labour time (Mankiw 1989, 1994).

Finally, positing technological change as a major factor of economic growth is plausible enough but nevertheless it becomes extremely difficult to interpret the notion of a negative technological change over a long-run period. In the short run, one could argue that this is due to natural phenomena which, however, have local and not necessarily global character.

15.6 Simulation Models

The RBC economists have developed a characteristically different way of testing their theory and of using it in the evaluations of the effectiveness of alternative policies. Thus, instead of subjecting their theory to econometric tests (as is usual for other theories) they resort to the use of simulation models, that is, models which are representative of their theory and then to the estimation of the parameters of their models so as to mimic the functioning of the real economy. This method of selecting values for the parameters of their system of equations that describe the operation of the economy is called calibration. More specifically, this method involves the following steps:

1. Construction of an *AS* and *AD* model.
2. Construction of a function with particular algebraic form that represents the decisions for production and consumption.
3. The model then is subjected to random shocks coming from technology.
4. The effects of these random shocks are reflected in the evolution of the major macroeconomic variables, and the results are compared with the evolution of the same variables in the real economy.

The simulation experiments show that when the model is subjected to a series of random shocks then it displays fluctuations, which resemble those of the respective variables in the real economy.[8] Prescott and Plosser (1982) claim that their research

[8]The detailed presentation of these simulation models is beyond the scope of this book whose main purpose is to point out the essential characteristics of each school of economic thought. The interested reader may consult the articles by Stadler (1994) or Hartley *et al*. (1998).

shows that the real and not the monetary shocks are responsible for the economic fluctuations.

In the RBC approach, we have an unusual form of testing the validity of the theory called 'calibration'. The idea is to use all the available econometric information to set the parameters of the model. The trouble with the calibration technique is that one does not estimate the parameters of the model, as in the econometric specifications. As a consequence, while with an econometric technique one can evaluate the explanatory content of a model, with calibration, by contrast, one does not easily falsify the model in question, even if there exists overwhelming evidence against the model being correct. Since RBC models explain data *ex post*, it follows that the falsification of these models is an extremely difficult task. Furthermore, it has been argued that RBC models are highly sample-specific and in this sense have little or no predictive content.

We point out that such simulation approaches have become a tradition since the decade of the 1970s and have been adopted in the case of computable general equilibrium models which continue to be used in the analyses of public finance and international trade. In these simulation models, various scenarios are being used; for example, one could examine the effects of substituting one kind of tax for another in order to derive results on relative prices and general welfare, while holding tax revenues and government expenditures constant. Similar models are being used in the theory of international trade (Shoven and Whaley 1992). Lately, these models have been adopted by heterodox economists, who by introducing real life characteristics such as monopoly power, class character of the economy and the policies pursued, *etc*., present an alternative to the orthodox computable general equilibrium model (Taylor 1990).

15.7 Economic Policy Implications

We have shown that the RBC approach has rejected three established principles which are shared by Keynesians, Monetarists and the NC economists alike. These principles are:

1. The fluctuations in total output are viewed as short-run deviations from some hypothetical trend growth rate.
2. Business fluctuations are socially undesired.
3. Monetary factors are important in the interpretation of business fluctuations.

As a consequence, the economists of the RBC approach are led to the conclusion that the government intervention in the economy through fiscal and monetary policy is meaningless. Furthermore, in the RBC approach growth and cycles are different aspects of the same process, while cyclical fluctuations in the GDP and employment are expected as an entirely normal reaction of rationally behaving economic agents to changes in the economic environment that they operate. Consequently, fluctuations in GDP and employment do not imply diminution of

social welfare. Moreover, some policies aimed at reducing the severity of business cycles are likely to give rise to more costs than benefits. Prescott, for example, points out:

> The policy implications of this research are that costly efforts at stabilisation are likely to be counterproductive. Economic fluctuations are optimal responses to uncertainty in the rate of technological change. However, this does not imply that the amount of technological change is optimal or invariant to policy. The average rate of technological change varies much both over time within a country and across national economies. What is needed is an understanding of the factors that determine the average rate at which technology advances. Such a theory surely will depend on the institutional arrangements societies adopt. If policies adopted to stabilise the economy reduce the average rate of technological change, then stabilisation policy is costly. (Prescott 1986, p. 21)

Since business fluctuations are caused mainly by technological and not by monetary shocks, it follows that the monetary policy cannot affect in any significant way the evolution of the real variables. In other words, for the RBC approach money is neutral in both the short and the long run, which is equivalent to saying that the so-called classical dichotomy between the real and monetary economy holds. Moreover, since workers essentially observe their real wage and on the basis of their real wage decide the supply of their labour services, it follows that the observed unemployment is voluntary. As for the fluctuations in the GDP, the RBC economists argue that these concern essentially the evolution of the equilibrium GDP, which by definition is accompanied by full employment of labour or what amounts to the same thing by zero involuntary unemployment. Consequently, it is meaningless for a government to set as a target the full employment of labour, since for the RBC approach full employment already exists; those that are 'unemployed' choose to be so for a number of reasons. Hence, one must be careful because it might be true that the government cannot affect the economy for the better, nevertheless it can affect the economy for the worse through inept fiscal and monetary policy.

The NC economics and, in particular, Lucas (1984) on the occasion of commemorating 30 years since the publication of the book by Friedman and Schwartz (1963) advanced the following critical evaluation of the RBC approach. In the beginning, he adopts the view according to which the US economy gives rise to results which are similar to those derived from perfect competition.[9] As a consequence, there is no need for government intervention in the economy. In fact, if there are any imperfections in the economy these are due to various restrictive legislations, fiscal and monetary policies, regulations of any kind, customs and traditions that do not match up with the perfect competition state. From all the

[9]One should not be surprised by this assumption since at the University of Chicago (where Lucas comes from) there is a long tradition according to which markets operate so efficiently that the performance of the economy is surprisingly close to that expected by perfect competition. This was the reason why Milton Friedman, George Stigler and Arnold Harberger (all from the University of Chicago) opposed the antimonopoly legislation and government regulation of the markets. The idea being that the antimonopoly legislation, in general, cannot achieve economic results quite different from the prevailing ones, which are not far from those predicted by the perfectly competitive model.

above, Lucas paradoxically enough selects monetary policy and argues, despite his adherence to monetary shocks, that changes in total factor productivity are associated with similar changes in the growth of output as well as a number of related variables (consumption, investment, *etc*.). These results, however, should not be considered as evidence of the predictive content of the RBC approach, which may approximate pretty well the real variables of the economy so long as monetary policy is conducted in the right way. From this perspective, the relative success of the RBC approach in the interpretation of post-war experience is absolutely consistent with the assumption that the post-war monetary policy was correctly carried out from the monetary authorities and led to optimal results, yet one should not deduce from this that monetary policy is without any impact on the real economy. The same is not true, however, with the monetary policy prior to Word War II period which was the focus of the Friedman and Schwartz (1963) book. In the post-war period, the monetary authorities were active and assumed a stabilisation role by applying an easy monetary policy during recessionary periods and stock market crises; by contrast, in expansionary periods of exchange rates devaluations they raised the interest rates through tight monetary policies in an effort to avert further devaluation of their currencies. They offered insurance to the deposits of the public up to a certain amount so as to secure the banking system from the possibility of panics. It has been argued that as a result of these interventions the fluctuations in economic variables were neither so deep nor as frequent as they were in the pre-WWII period. Of course, many economists would argue that the relative stability that was observed in the post-war period should be attributed, at least in part, to the progressive taxation, the social insurance and in general the welfare state.

Thus, for the RBC economists, any 'deviations' of the level of economic activity are not really deviations from some hypothetical equilibrium level of output with full employment, but rather constitute optimal reaction of the economic agents to changes in the production possibilities. In other words, the economy is continuously in full employment, and, therefore, there is no particular reason for government intervention, at least to correct for the short-run situations. The only significant role for the government is to contribute to the optimal growth rate of the economy. Required policy measures according to the RBC economists include a taxation policy that promotes economic growth and also the provision of public goods. In particular, government spending on health and education exerts a positive effect on the economic growth through the improvement of society's human capital which, in turn, may generate positive productivity shocks through innovative activity. The efforts of governments for continuing education and the acquisition of new skills by the labour force are certain to influence positively the real wage and level of productivity. It has been argued that the improvement of human capital solves the problem of competition between wages and profits because it is possible for both to increase and none to decrease.

If there is any government policy that the RBC economists would wholeheartedly support, then this would be the stabilisation of total factor productivity. The idea is that if business cycles are created from the ups and downs in the level of total factor productivity, then the stabilisation of the total factor productivity will

smooth out the evolution of the GDP. This stabilisation can be achieved in the case where the current growth rate of productivity exceeds its average growth rate and under these circumstances businesses are encouraged to slow down their investment. The converse is true in the case where the current growth rate of productivity falls short of the average. In conclusion, if private investment is directed in the above way, then it follows that the fluctuations in the output produced in the working hours supplied and in consumption will be smoothed out. Yet, it goes without saying that such policies are very easy to announce but extremely difficult to apply effectively.

15.8 Summary and Concluding Remarks

The RBC approach has attracted the attention of many economists because of its simplicity and the advancement of new techniques that opened new research areas. We point out that many of these techniques have been derived from the theoretical needs of the RBC approach. The RBC is not an altogether new approach and it is based on the same fundamental assumptions of NC economics. Thus, the assumption of continuous market clearing is not easily acceptable since reality is characterised by continuous disequilibrium in all markets and especially in the labour market. Hence, of course, the economists of the RBC as well as the NC have an interpretation based on the redefinition of unemployment, which is viewed as the result of optimal choice of workers, who, having to choose between work and leisure at the going low wage rate, choose leisure waiting for the wage rate to rise. Consequently, unemployment does not constitute a disequilibrium factor. Such an approach is very difficult to support if we think that the level of unemployment in the USA during the depression of 1930s was in the range of 25%. Unemployment of this sort cannot be explained neither with negative technological shocks nor with the intertemporal substitution of labour by leisure, that is, it is not plausible to argue that the unemployed chose leisure for a decade in order to be employed later when the wage and the interest rate would be at the right level. In the same vein, the observed unemployment during the decade of the 1980s cannot be interpreted by intertemporal substitution (Summers 1986).

Many economists find the view of the RBC that the technology shocks are mainly responsible for the cyclical fluctuations of the real GDP as an assumption that can hardly be accepted since these shocks are not so strong as to justify the observed fluctuations in the GDP (Lucas 1987, pp. 71–72 and Stadler 1994, p. 1751). The RBC model is reinforced by Schumpeter's (1936) view that innovations are introduced massively. Nevertheless the economists of the RBC approach cannot accept such a view since this automatically leads to the idea of long waves in economic activity, an idea which lies beyond their theoretical framework, which accepts fluctuations with duration of 3–5 years.

Moreover, the simple simulations of the RBC approach that are being used to mimic economic fluctuations, by and large, are not compared with the predictive

capacity of alternative competitive models so as to select the most appropriate. With regard to the hypothesis of substitutability, the fall in the wage may lead to an increase in the working hours supplied; however, as we have pointed out, if we suppose that the worker has a target wage rate, then what really happens in the economy (*i.e.* if and to what extent there is substitutability between work and leisure) is mainly an empirical question. The empirical research has shown that the wage exerts only a slight influence on employment; the fluctuations of employment that are observed are attributed to causes other than the changes in wage. On the other hand, the role of the interest rate is limited and thus does not affect the supply of labour in any significant way. Workers do not seem to take into account the level of the rate of interest rate in their decisions to supply their labour services.

Many economists view the changes in aggregate demand because of monetary policy as capable of exerting significant results because of the inflexibility in both prices and nominal wages. If markets do not clear fast enough and if the economy is characterised by changes in aggregate demand and aggregate supply, then the observed fluctuations consist of a stochastic trend around which orbits output because of shocks in aggregate demand. This new view is supported by Blanchard and Quah (1989), who interpret the fluctuations in GDP and unemployment on the basis of two shocks: those with permanent results on the GDP and those with temporary ones. The shocks with permanent results are attributed to the supply side and the shocks with temporary results to the side of demand. It is obvious that such a confirmation renders the role of stabilisation policy even more complex. For example, how can the government distinguish between shocks in supply and demand, especially when they are not independent of each other?

In an evaluation of the RBC approach, we observe that an effort is being made to interpret the economic phenomena beyond the established views. Thus, there is no *a priori* view which must be supported or be empirically verified. It is interesting to note that the results of the research decide on the credibility of the theory. In this context, the RBC approach views the supply of money as an endogenously determined variable and in so doing the RBC approach finds itself in a silent and, at the same time, unholy alliance not only with the Post-Keynesian economists but also with the Marxists. If we spot a difference, regarding this matter, between RBC economists and Post-keynesians or Marxists, then this would mean that for the RBC approach the endogeneity of the money supply is derived from the results of simulation models and not from a theoretical analysis. With regard to the business cycles, the RBC approach[10] cannot but take elements from Schumpeter, and from the economists that are systematically concerned with technology and its effects on the economy. Currently, the RBC approach is in a stage of decay;[11] however, it is

[10]It is ironic for a school of economic thought that endorses an outright rejection of business cycles to call itself the Real Business Cycles approach.

[11]In an early version of this chapter (Tsoulfidis 2004) we argued that the RBC will continue to be popular to the extent that the economies are expanding, if however, the economies fall in a serious recession then as the RBC approach would not have any serious economic policies to propose naturally one would expect its decay.

certain that it has contributed to the creation of a solid theoretical foundation with the aid of which interesting questions can be raised while at the same time techniques that have contributed to the research of macroeconomic questions have been developed. To the extent that this approach downgrades the social cost of the observed fluctuations is potentially dangerous in policy terms. This is so, because faced against a recession that lasts the RBC policy prescription would be to do nothing.

The RBC approach appeared in the middle of the 1980s and its popularity increased with the upturn of the 1990s. The crucial question is would this approach survive the current downturn in economic activity? We feel not and think that most likely it will be fused to a new approach that tries to reconcile some of the unresolved questions that the RBC approach fails to address.

Questions for Thought and Discussion

1. What are the major characteristics of the RBC approach?
2. The RBC approach does not explain the periodicity of business cycles. Discuss.
3. How does the RBC approach explain the phases of the business cycles?
4. Are recessions and unemployment socially undesired phenomena? What kind of policy measures do the RBC economists propose?
5. RBC economists neglect problems of income distribution because of the holistic nature of their approach. Discuss.
6. How does Lucas evaluate the RBC approach?

Notes for Further Reading

As was noted, the RBC economists are by and large former NCE, who in the early 1980s, were dissatisfied with the NCE attachment to a monetarist framework. The article by Kydland and Prescott (1982) is the first that summarises the ideas of these dissident economists and places them into an appropriate framework so as to form the RBC approach to economics. Since then, this school of economic thought has displayed significant progress. The article by Nelson and Plosser (1982) argues that the real shocks can be much more important than the monetary shocks in the interpretation of the evolution of total output. Nelson and Plosser arrived at this conclusion after the study of the evolution of the real GDP in the US economy, for which they could not falsify the assumption that the real GDP follows a random walk, a typical result that holds for many economies and not only for the USA. More specifically, if the shock is a real one, then the real GDP suffers a permanent change in its trend; in this case, we say that the track of the real GDP follows a statistical process which is known as a random walk something that is verified from the presence of a unit root. A year later, the articles by Long and Plosser (1983), and also the article by Barro (1984), came to ascertain that this approach is a promising one and that, at the same time, significant progress in the formation of the RBC approach had been achieved.

(continued)

The chief characteristic of this school of economic thought is that it presents the business cycles as optimal reactions of the economy in the case of exogenous shocks (Prescott 1986, p. 21). In this sense, business fluctuations are neither deviations from equilibrium nor are they caused by some sort of market failure. For example, recessions represent an unwanted but at the same time unavoidable displacement because of the constraints faced by the public. Given these constraints, markets react effectively and people achieve the best possible results. According to Barro (1982), the three basic constituent components of the new approach are: (a) the application of general equilibrium in all markets, (b) the adoption of the rational expectations hypothesis and (c) the application of game theory in the contemplation as well as in the conduct of economic policy. Barro, however, soon turned his research efforts on economic growth (Barro 1991). This course of action was opened up much earlier by Lucas, the economist that had attracted Barro from the macroeconomics of general disequilibrium to those of general equilibrium in the strand of NCE.

Monetarists (Friedman 1968) and NC economics (Lucas 1972 and 1973) view that the business cycle is caused by people misinformed about their real situation. For example, the economic upturn is caused because workers think that their real wages are higher than they actually are. As a result, workers are willing to offer more employment and output increases. The converse is true in the downturns, that is, workers consider their real wage lower than it actual is and thus they offer less employment, output falls and so forth. Friedman (1968) considers this case to last in the long run and to repeat itself any number of times because it seems that 'money illusion' is a chronic disease. By contrast, for Lucas the assumption of rational expectations led him to the conclusion that such situations can only be ephemeral. In both cases, however, if the flow of information were sufficient then there should not be any business cycles or persistent recessions. As a consequence, the weakness of both theories to provide adequate explanations for the stagnation of the 1970s and also of the early 1980s led to the development of competitive schools of economic thought such as the RBC economics and the resurgence of (new) Keynesian economics. Lucas's (1994) view about the RBC approach is that the shocks in total productivity can lead to changes in output of the same size as those observed in the US economy during the post-war period and so they can interpret the evolution of the other variables. What is important is Lucas's observation that the RBC approach constitutes an excellent test to check the extent to which monetary policy was conducted successfully. If one looks at the RBC from this point of view, then the success of this approach is strictly related to the successful conduct of monetary policy and not to the notion that money plays no role whatsoever in the economy.

It is interesting to note the view of the RBC economists regarding the development of a single theory that examines both economic growth and

(continued)

business fluctuations. This view is not new at all and Kydland and Prescott (1982) consider Slutsky (1929) as the precursor of this idea. On further consideration, however, we discover that this idea was developed originally in the Institute of Conjecture in Moscow with Nikolai Kondratiev (1892–1938) as its head. The interested reader can look in Kondratiev's (1999) work in order to ascertain the progress that was made in this research institute in the 1920s. Two young researchers at that time, Eugen Slutsky and Simon Kuznets, were also engaged in this area and were among the pioneers of the research on business cycles and economic growth. The idea of a single theory for both cycles and growth is also found in Schumpeter's (1942) work as well as in the work of his student at Harvard, Richard Goodwin (1968). The separation between cycle and growth must be attributed to the influence that was exerted by Solow's (1955 and 1956) article.

Chapter 16
The Return of (New) Keynesian Economics

Yes Keynes is dead; and so are Einstein and Newton.

(Samuelson, Paul 1988)

[T]he appeal of Keynesian economics stems from the evident unhappiness of workers and firms during recessions and depressions. Workers and firms do not act as if they were making a voluntary choice to cut production and hours worked.

(Gordon, Robert 1993)

16.1 Introduction

The period from the mid-1960s to the mid-1980s is characterised by a slowdown in the level of economic activity; however, government's stabilisation and welfare functions restrained the outbreak of a major depression and converted it to a long-lasting stagflation. This is the reason why this time period has been christened the "Silent Depression" and, as in the 1930s, it formed the fertile ground for the flourishing of new economic theories and for the withering away of the old ones. Dissatisfaction with the neoclassical synthesis as well as the dead end road of disequilibrium macroeconomics paved the way for the rise of Monetarism, which, however, soon lost its popularity, and already from the mid-1970s, the researches of New Classical economists with the works of Lucas and Sargent were leading the way. Many among the New Classical economists kept a critical stance towards basic propositions of their theory, and in the early 1980s, it led to the development of the real business cycles (RBC) approach. Meanwhile, Keynesian economists of the neoclassical synthesis (Tobin, Modigliani and Gordon) continued their critique which was initially levelled against Monetarism and subsequently against the New Classical economics and the RBC. The basic plea of New Keynesians is about the fundamental common hypothesis of all these approaches which is continuous market clearing. These new, so to speak, Keynesian economists argued that in actual economies there are many "obstacles" that hinder markets from complete clearing.

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_16,

The New Keynesians differ from the Keynesians of the neoclassical synthesis and disequilibrium macroeconomics, in that many of them during this long gestation period had accepted the hypothesis of rational expectations.[1] New Keynesians (not necessarily all of them) incorporate, in their analysis, microeconomic principles of rational behaviour of economic agents arguing, for instance, that the rational expectations hypothesis can be integrated profitably in the New Keynesian framework.[2] New Keynesians, like the Monetarists and the New Classicals and unlike the RBC approach, regard demand shocks as the major impulse mechanism that determines the actual stage of the economy. As a consequence, there is an active role for government. According to Mankiw and Romer (1993b), there are two basic propositions of this school of economic thought:

1. Money is not neutral in the short-run.
2. Economic fluctuations are caused by malfunctioning of markets.

The non-neutrality of money stems from the rigidities in prices, which originate from market imperfections.

In what follows, we discuss the major features of New Keynesian economics dealing first with the various rigidities (nominal or real) that prevent markets from clearing, and subsequently, we discuss two popular New Keynesian views of economic fluctuations. The chapter continues with some rather recent developments that attempt to integrate hitherto developments in macroeconomics to what came to be called "new consensus macroeconomics", and we conclude with some economic policy implications.

16.2 Nominal Rigidities

In Keynes and Keynesian economics, nominal wages are inflexible, at least in the downward direction, so Keynes, for example, uses the rigidity in money wages as a means to express all variables in terms of wage units. In Chap. 10, we saw that Modigliani (1944) also in his neoclassical synthesis uses the money wage to deflate the variables involved in his IS–LM system of equations. New Keynesians make an effort to theorise these rigidities in money wages and also in prices. It is interesting to note that the New Keynesians have, in common with the New Classicals, the idea that the economic agents display optimising behaviour (*e.g.*, households maximise their utility and firms maximise their profits). However, New Keynesians, contrary to New Classical economists, assume that many markets are characterised by some

[1]The characterisation "New" indicates recent advances in the Keynesian theory of the Neoclassical Synthesis. Among the New Keynesian are included (in parenthesis the University where they teach) Gregory Mankiw and Laurence Summers (Harvard), Olivier Blanchard and Stanley Fischer (MIT), Bruce Greenwald, Edmund Phelps, Joseph Stiglitz and Michael Woodford (Columbia).

[2]Blinder (1987) and Phelps (1992) are among the New Keynesians who question the hypothesis of rational expectations.

degree of price rigidity that permeates the entire economic system. In the labour market, for instance, money wages may easily increase but are more difficult to go in the downward direction, as hitherto historical evidence of many decades has demonstrated. This rigidity in nominal variables (wages and prices) shows that government intervention may affect the real variables of the economy, at least in the short run. There is no doubt that the hypothesis of rigidities in prices is the chief explanation of variations in the level of output, and, therefore, of businesses cycles. Nevertheless, New Keynesian economists, in the beginning, turned to the rigidities in money wages and subsequently to rigidity in prices in general.

The rigidity in money wages comes from contractual agreements signed between the parties involved either at the individual or at the collective level. Since money wages are characterised by rigidity, it follows that a flexible monetary policy could be effective with respect to the level of output and employment (Fisher 1977). The effectiveness of monetary policy, in the short-run, at least, can be shown in terms of the usual aggregate demand and aggregate supply graph (see Fig. 16.1).

Let us suppose that the economy is in equilibrium at point A, the point of intersection of aggregate demand, AD_0, the short run aggregate supply curve ($SRAS_0$), and long-run aggregate supply curve (LRAS). A negative demand shock (*e.g.*, a fall in consumers' confidence) hits the economy and thus shifts the aggregate demand curve to the left, from AD_0 to the new position AD_1. If prices are flexible but not the wage, the economy would move from point A to point B while output would fall from its full employment level Y^* to Y_1, where a percentage of the labour force is involuntarily unemployed. If there is full flexibility in both wages and prices, then the aggregate supply curve $SRAS_0$ shifts to the new position $SRAS_1$ (because of the fall in prices and the wage rate) and the economy returns to the full employment level of output, although at lower nominal wages and prices. If however money wages are rigid because of contracts that were signed prior to the

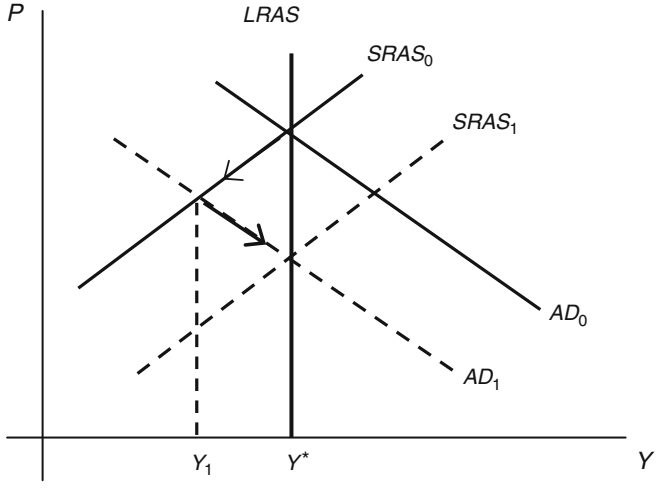

Fig. 16.1 Equilibrium in New Keynesian model

demand shock, the possibility of attaining point *C* is ruled out in the short run. In the long run if new contracts are signed then it is possible for wages to fall and attain point *C*. Once the economy finds itself on point *B*, then there is scope for an expansionary monetary policy which restores the full employment level of output and economy returns to the initial equilibrium point *A*, not withstanding that all the agents of the economy are assumed to act on the basis of rational expectations. As a consequence, government intervention can be effective in achieving its objectives, at least in the short run. Of course, in the absence of rigidities, and with rational expectations, the economy would move directly to point *C*, without the mediation of point *B*. It goes without saying that if the economy is in point *C* an expansionary policy will be inflationary.

The trouble with the above analysis is that real wage behaves countercyclically, something that does not agree with the empirical data which point towards the hypothesis of the procyclical character of real wage, as depicted in Fig. 16.2.[3]

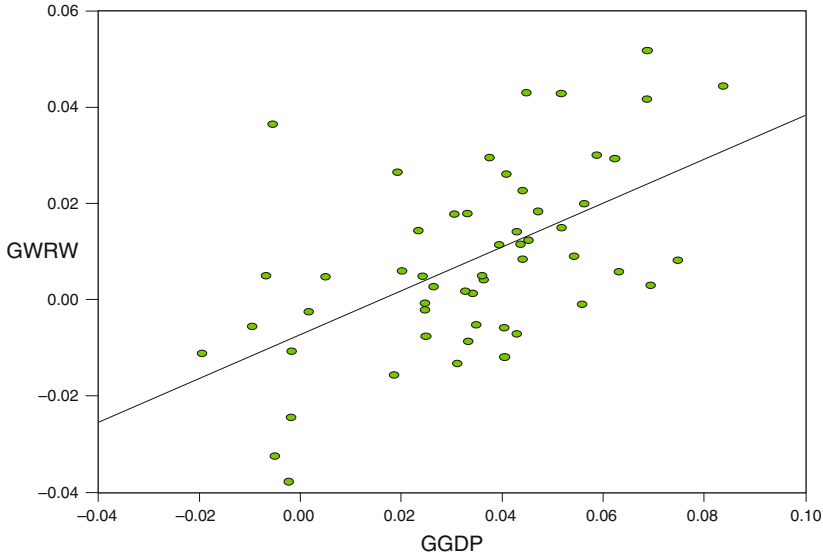

Fig. 16.2 Growth rate of the real weekly wage versus growth rate of the real GDP in the USA, 1948–2001

[3]The data are from the US economy (www.bea.gov) and refer to the growth rate of real weekly wages and the growth rate of the real GDP during the period 1948–2001. We also experimented with real hourly wages instead of weekly with no qualitative difference in the results. In a similar graph, Mankiw (1994) presents the rate of change of the real wage and the rate of change of GDP in the USA and he finds that the two variables are subjected to the same kind of fluctuations. Nevertheless, Gordon (1993) argues that the procyclical character of the real wage appears only sporadically as it happened during the decade of 1970s, while in the general case the real wage does not display significant cyclical fluctuations.

This observation led many New Keynesian economists to turn to interpretations that are based on nominal rigidities of prices. More specifically, they argue that in today's economies, imperfect competition dominates all markets, which means that firms are active price makers instead of being passive price takers, as is assumed in perfect competition.[4] Many New Keynesians (Akerlof and Yellen 1987; Parkin 1986; Rotemberg 1987) argue that the change in the pricing policy is by no means a simple process and that firms often hesitate to change their prices in response to small changes in some elements of cost. For example, in the case of industrial firms we know that they would rather react to a rise in cost by reducing their profits or the degree of their capacity utilisation instead of changing their prices (Rotemberg 1987; and also Chap. 8, Appendix). The small cost that must be paid to change prices has been called *menu cost* precisely because it was first observed in restaurants. It has been argued that in restaurants prices do not change with every change in cost because of the cost of preparation and printing of new menus and also because frequent changes in prices cause irritation to the (steady) customers of the restaurant. The expression *menu cost* is general enough, so that it includes all kinds of nominal rigidities that are observed in nominal prices. The above lend support to the view that if on top of rigidities in money wages, we add rigidities in prices, changes in aggregate demand give rise to an even higher instability in total output and employment. As a consequence, the scope of government intervention to stabilise the economy and attain a specific level of output is enhanced.

16.3 Real Rigidities

Many New Keynesians claim that nominal rigidities are not the only one aspect of overall rigidities in the economy but the other being real rigidities with their combination forming the major characteristic of the New Keynesian approach (Mankiw and Romer 1993b). According to Ball and Romer (1990) the real rigidities when combined with nominal ones strengthen the impact of the non-neutrality of money and the short-run effectiveness of government's stabilisation policy. Let us suppose that aggregate demand falls suddenly and there are menu costs involved, then it does not necessarily follow that prices will be affected in any significant way, though (other things equal) we expect fall in output and employment. If we suppose an inelastic supply of labour, it follows that the fall in the real wage becomes possible only if the money wage falls or the marginal product of labour rises, or if a combination of these two changes exists. The reason is that the fall in the real wage involves a fall in the marginal cost (MC) and the price of the product, provided demand remains the same (or does not increase by much). We know that

[4]Keynes distanced himself from these two polar conceptions of competition and, like his teacher Marshall, had a rather realistic view of competition akin to that of the classical economists (Shapiro 1997; Hayes 2008).

firms maximise their profits by equating their marginal revenue (MR) to their MC. The MR can be expressed in terms of elasticity of demand (e) and price of product (P) as follows:

$$\mathrm{MR} = P\left(1 - \frac{1}{e}\right).$$

We suppose for reasons of simplicity and clarity of presentation that the only variable cost of production is labour and hence, that the MC will be equal to the ratio of the real wage (w) to the marginal product of labour (MP_L),

$$\mathrm{MC} = \frac{w}{\mathrm{MP}_L}.$$

By invoking the equilibrium relation, that is, MR = MC we arrive at:

$$P = \frac{w}{\mathrm{MP}_L}\left(\frac{e}{e-1}\right),$$

where the term in the parenthesis depicts the mark-up on cost. We know that in monopolistic competition $e > 1$ and also that the elasticity of demand and mark-up on cost are inversely related. Consequently, it does not necessarily follow that a reduction in the real wage or an increase in the marginal product or some combination of these two variables will lead to a fall in the price of the product, because the price of the product also depends on changes in the mark-up. Therefore, it is possible for the MC to increase, but also for the markup $(e - 1)/e$ to move to the opposite direction. As a consequence, the change in price depends on the net effect of these two counteracting forces, which contribute to the relative stability of real prices. Thus, we have an empirical question and of course, there is no single answer. We know that in many industrial sectors the mark-ups move procyclically. In the general case, however, it has been argued that the movement of mark-ups is countercyclical thus contributing to the overall rigidity in real prices (Rotemberg and Woodford 1991). More specifically, in periods of expansion in the level of economic activity, competition is intensified and collusions between firms become all the more difficult to agree, let alone sticking to them. Meanwhile, competition reduces the mark-ups. The exact opposite tendencies take place in recessionary periods.

There are many sources of real rigidities in prices other than the mark-up. Among them, New Keynesians include the market research cost, because consumers usually do not have the time to investigate all possible alternatives, and so firms take into account this weakness of consumers in the pricing of their products. The same is true on the part of businesses, which in their effort to attract customers are obliged to advertise their products or services. Workers are reluctant to get involved in a process of search for employment, because this entails some cost (research in job market, preparation of their CVs, interviews, *etc*.). Similarly, businesses view hiring of workers as a costly and time-consuming process.

The rigidity in prices is a result of complex inter-industry connections; according to Gordon (1990) firms cannot easily change their prices in case of a fall in demand. The reason is that the same product is being sold by a number of other firms not only in domestic but also in international markets. If, for example, there is a fall in demand, it will be extremely difficult for a firm to reduce its price because the firm operates in an international environment and is subject to continuous supply shocks. Consequently, if the firm decides to cut its price then it will be exposed in case of an increase in cost. Because of a multitude of connections, it is almost impossible for each individual firm to know what precisely happens in the MC of its suppliers. These input–output interconnections between firms give rise to such an uncertainty that firms would be reluctant to change their prices, even when their cost of production changes. The analysis by Gordon offers an interpretation for both real as well as nominal rigidities.

Up until now, we discussed the real rigidities in prices, and we found that they are difficult to generalise theoretically. What really happens in the economy is an empirical question for which there is no single answer. New Keynesian economics offer explanations with respect to rigidities in real wages, which aim at interpreting the riddle of high unemployment rates in the US and other countries during the 1970s. According to New Keynesians, the economy may be in long run equilibrium and, at the same time, suffer from considerable unemployment as a result of real wage rigidities. There are three usual New Keynesian interpretations of involuntary unemployment, the implicit contracts, the efficient wage and the insiders outsiders, which we examine below.

16.3.1 Models of Implicit Contracts

It is important to point out that implicit contracts do not require the presence or the mediation of trade unions and so it comes as no surprise that they have been developed mainly in the USA, where the trade union factor is particularly weak and getting progressively weaker. The model of implicit contracts (Baily 1974; Gordon 1974; Azariadis 1975) seeks to explain the reason why firms and workers opt for long run agreements. There are three reasons that make these agreements possible:

1. Workers are protected against risks much less than firms are.
2. If at all, workers have limited access to capital markets.
3. Workers are in general reluctant to switch employment by moving between firms.

Firms are in a position to offer security of employment and, because of their easier access to financial markets, acquire the liquidity they want. It is obvious that workers do not have such possibilities of access to financial markets, and, therefore, they are exposed to the vagaries of economic activity. If workers leave the determination of their real wage to the free operation of market forces, it follows that in

the upturn in the level of economic activity, real wages rise while in the downturn phase they fall. Of course, the free operation of market forces by no means guarantees the continuation of their employment.

There are many weaknesses and unanswered questions in this theory; for example, during recessionary periods, unemployment rises, while the model predicts constant employment. Further, the model does not explain the reason why firms do not pay a lower wage to the newly hired. Questions of this kind are those that New Keynesians try to answer with the next two hypotheses.

16.3.2 The Efficient Wage Hypothesis

According to the efficient wage hypothesis, the observed unemployment is attributed to the reluctance of businesses to lower wages. The idea is that lower wages entail a smaller work effort and lack of sufficient discipline on the part of workers, wastefulness in materials, careless provision of services, *etc*. All these features lead to lower productivity, higher cost of production, and, therefore, lower profits (Yellen 1984; Gordon 1990). For these reasons, firms, in spite of being in a position to decrease wages, and, therefore, to increase employment, prefer nevertheless to maintain the real wage at a higher level than what would prevail under an unhindered operation of market forces. The objective, of course, is the increase in the productivity of labour while simultaneously maintaining the attractiveness of employment positions. The efficient wage model provides an interpretation of the coexistence of high unemployment rates with high real wages. The idea is that firms by maintaining the real wage at a level higher than that of the free market manage to keep the quality of their workforce at a high level and also enforce, in a "politically correct" and civilised way, the discipline and loyalty of their workforce.

This theory explains the high unemployment rates among the ranks of low productivity workers and also explains the phenomenon that in periods of recession the number of working hours is not distributed among more workers. The idea is that a fall in real wages implies a fall in productivity. The truth however is that for some countries such as the US, the real wage has been falling since 1972 and the argument that attributes high unemployment rates to rigidity of real wages does not really stand in the light of the movement of real wages in the last say 3–4 decades. This does not mean that the level of unemployment is not related at all to the level of real wage. To the contrary, unemployment and real wages are inversely related; however, one should think of possible lower limits in the fall in real wages as a fall past a certain point that may lead to social turmoil, altering thus the normal relations between the variables (see Keynes 1936, Chap. 17 and also our Chap. 10 on Keynes). One may envisage a level of real wage past which normal economic relationships no longer hold. This level of real wage may be called the "wage trap"- a concept which is symmetrical to the interest rate in the "liquidity trap". Although the liquidity trap has been discussed extensively in recent decades, especially by

New Keynesians, the "wage trap" has not received the attention that it deserves notwithstanding the fact that the level of real wages continues to be lower than that of the early 1970s in the US economy.

16.3.3 The Insider–Outsider Hypothesis

While in the model of efficient wages, firms determine the level of wages, in the insider–outsider model, the equilibrium wage is determined, at least partly, by the workers.[5] This model was developed by Lindbeck and Snower (1986) and explains why the unemployed do not accept lower wages to increase the demand for labour and to reduce the rate of unemployment. Insiders, that is, the workers who are already employed, have the power to impose the wage rate that they want and of course this wage is higher than the equilibrium wage, while the outsiders are the unemployed, that is, workers who are supposed to lack the strength to impose a lower wage and in doing so to change their employment status. The strength of the insiders stems from the high cost of business for firings and also for new hirings. These costs include expenses such as: research of job market, interviews, negotiations in the signing of new contracts, compensations of those fired, education and training cost of the newly hired and so forth. Furthermore, the insiders, that is, the senior workers make the hiring of the new personnel harder because they quite often see potential competitors in the newcomers; as a consequence, the insiders pose – usually tacitly and rarely overtly – obstacles and make it difficult for the newcomers to stay in the workplace once they are hired. This is the reason why the newcomers demand higher wages in order to cope with a difficult working environment. Firms are discouraged to continue with new hirings, because of their high cost (due to the behaviour of the insiders). Moreover, the conditions of insecurity of employment for the newly hired as well as the limited likelihood of advance in the hierarchy of the firm gives rise to working conditions that are not favourable at all for increase in productivity. The insiders have also the power to demand and usually get a share in the profits of the firm even though they are not unionised. It goes without saying that the insiders–outsiders model is strengthened with the presence of unions.

These models were mainly used for the interpretation of involuntary unemployment, but they simultaneously interpret the composition of unemployment also. Thus, according to this model, it is expected that unemployment will be high among young people, women and minorities.

[5]On further consideration, however, one concludes that the profit maximisation (or cost minimisation) is the decisive determining factor in business behaviour!

16.4 Economic Fluctuations

According to the New Keynesian economists the cause of fluctuations in the level of economic activity are shocks that stem either from the supply or the demand side of the economy. Between the two types of shocks, New Keynesians would consider the demand ones as being much more important, and further, they would claim that imperfections of various types enlarge the repercussions of these shocks, which are transformed into serious fluctuations in the level of output and employment. Contrary to the New Classical and RBC approaches, the New Keynesians are interested more in the repercussions of these shocks in output and employment and in the way in which the economy reacts to them and less with the identification of the source (supply or demand) of these shocks. There are two main strands, within the New Keynesian economics, for the interpretation of cyclical fluctuations, - the first that attributes business cycles to nominal rigidities and the second which identifies the source of business cycles in the presence of uncertainty.

16.4.1 Fluctuations Caused by Nominal Rigidities

The first interpretation based on nominal rigidities can be shown with the aid of Fig. 16.3, which includes a set of four diagrams. In the first diagram (a) we have the market of goods with the LRAS and with the SRAS in a horizontal position conveying the idea of rigidities in prices and in nominal wages. The two curves intersect at point A, which is also the point of intersection with the aggregate demand curve (AD_0). Point A is therefore the equilibrium point. If we suppose that the supply of money falls and, hence, the aggregate demand curve AD_0 shifts to the left to the new position AD_1, then the price level will remain unaltered at P_0, and the economy because of the rigidity of nominal variables, leads from the initial equilibrium point A to the new equilibrium point B and the equilibrium output decreases from Y^* to Y_1. In the second diagram (b) we draw the 45° line, whose role is simply to reflect the level of output from the horizontal to the vertical axis.[6] Thus, in diagram (c) we display a typical production function, where the fall in the level of output from Y^* in Y_1 is translated to a fall in the level of employment from L^* in L_1 in the diagram (d). With the price level and the money wage given, it follows that for the same wage W^* firms would wish to employ L^*, but they do not have the demand that they need to sell their additional output. The reduction in the supply of money leads to

[6]It is interesting to note that there are many New Keynesians placing emphasis on the importance of money supply, something that differentiates them from the real business cycles approach. This emphasis in monetary variables has made Mankiw and Romer (1993, p. 3) to regard the New Keynesian economics as a kind of New Monetarism. The difference often is restricted to various kinds of rigidities and so the monetary policy may have real results in the economy, as monetarists would argue.

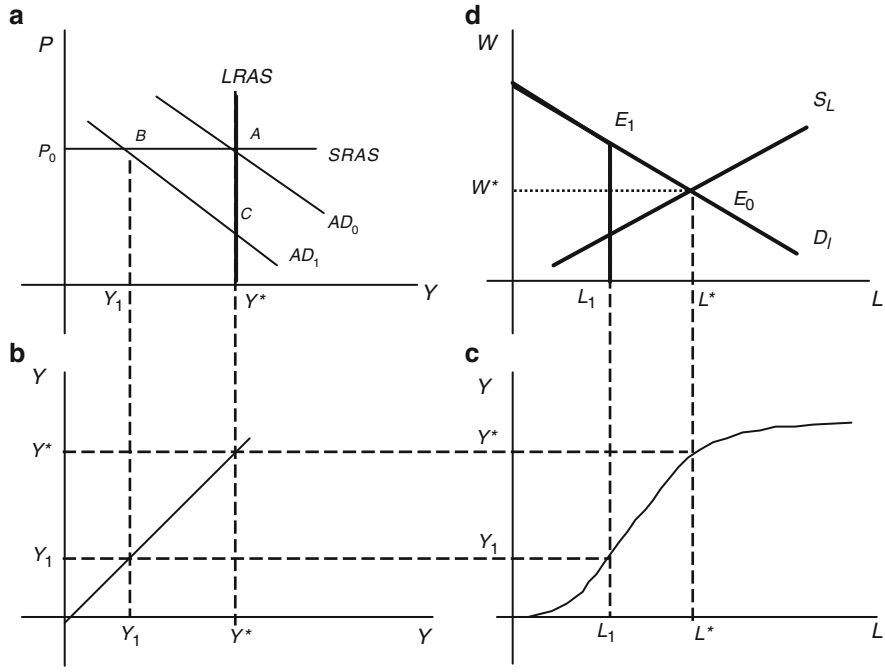

Fig. 16.3 Business fluctuations with nominal rigidities

a reduction in aggregate demand (from AD_0 to AD_1) and then to an increase in involuntary unemployment from L^* to L_1. This analysis leads New Keynesians to a set of policy proposals, which are focused on the increase in aggregate demand. It is important to note that in these policy proposals the money supply is no longer neutral in the short-run, although money is neutral in the long run, as is shown in the LRAS curve. From the above it follows that the unemployment and underutilisation of productive capacity is attributed to coordination failures of economic agents (Ball and Romer 1991). This weakness is explained through the argument that there are no benefits for each individual firm if it decided to undercut its price, whereas if all firms agreed and stuck to a certain pricing policy, they would all benefit.

New Keynesians unlike other macroeconomic approaches argue that if unemployment persists, the level of natural unemployment will tend to decrease. Therefore, the LRAS curve will tend to shift left towards point B. Consequently, for New Keynesians, the natural unemployment rate is not parametrically given, but is rather a variable that changes slowly over time, according to the following relation

$$U_t^* = U_{t-1}^* + \delta(U_{t-1} - U_{t-1}^*)$$

In other words, deviations of the actual unemployment rate (of the previous period) from the respective natural unemployment rate of the same time period,

$(U_{t-1} - U^*_{t-1})$ induce a proportional change in the current natural unemployment rate relative to the natural unemployment rate of the previous period $(U^*_t - U^*_{t-1})$. In this case, the reaction coefficient δ is expected to be positive. Phelps (1972) characterised this slow adjustment of natural unemployment towards current unemployment as path dependency hysteresis. If this hypothesis is correct, then the logical consequence is that efforts to reduce unemployment also have effects on the level of natural unemployment. Thus, it has been argued, that an expansive fiscal policy may reduce not only the actual but also the natural rate of unemployment, while the subsequent acceleration of inflation that accompanies this process can be controlled because the natural unemployment rate is closer to the actual. Of course, it could be argued that if the actual unemployment rate rises, this will increase the natural unemployment rate. Thus, in the US the natural unemployment rate of the 1970s was much lower than that of the 1980s and so were the actual unemployment rates in the same time periods.

16.4.2 Fluctuations Caused by Uncertainty

A second New Keynesian interpretation of economic fluctuations claims that the rigidity of prices not only cannot explain the economic fluctuations but also that the rigidities in nominal variables (prices and wage) limit the extent of cyclical fluctuations (Tobin 1993). Greenwald and Stiglitz (1993a, b) argue that fluctuations in the level of economic activity are caused by the risk averse behaviour of firms, when it comes to the expansion of their output and the financing of their investment activity. Firms are reluctant to borrow from capital markets because of the asymmetry in information, and, therefore, they cannot allocate their risk to the extent that they would like to. Consequently, their dependence on borrowing increases with the passage of time and so does their debt with the consequence that they are less protected in case of a recession. The result of this dependence is that firms in order to decrease the risk of their debt decide to decrease their output. The idea is that increase in output increases the dependence of firms on external sources of financing and, so increases the risk and the probability of their bankruptcy; in this respect, firms prefer to curtail their output in their effort to minimise the risk associated with a rising debt. During a recession, the marginal risk of bankruptcy increases the dependence of firms on external sources of financing and, thus, firms that are risk avert react by decreasing their output. Thus, a recession caused by a reduction of aggregate demand, (in terms of Fig. 16.4 the *AD* curve shifts to the left), leads also to a reduction of aggregate supply, (in terms of Fig. 16.2 the *AS* curve also shifts to the left). In such a case, the price level remains the same. If flexibility in prices exists then more uncertainty is created and the recession may become even worse.

The fall in the aggregate supply curve is caused by the increase in risk; as a consequence, the demand for labour falls. If the real wage remains the same, for

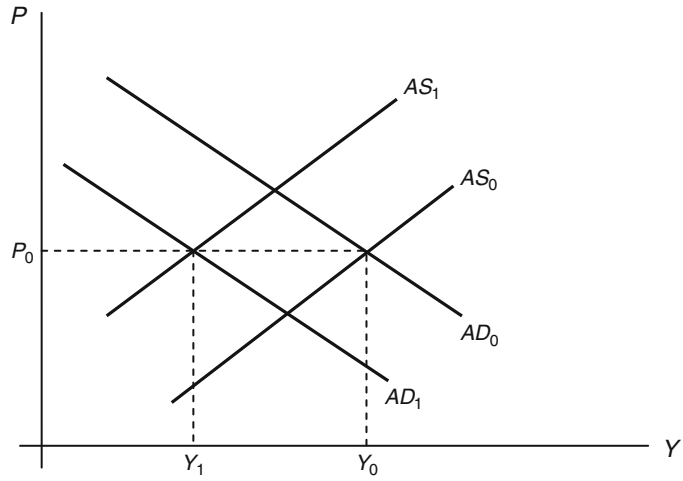

Fig. 16.4 The case of recession in conditions of uncertainty

example, because of the hypothesis of efficient wage, it follows that the involuntary unemployment is increased without changes in the real wage and the price level.

16.5 New Consensus Macroeconomics

In the past decade or so, we are witnessing an effort in macroeconomics to synthesise various ideas that were divisive in the past. Thus, a consensus has been achieved which naturally came to be known as new consensus macroeconomics. It is true that efforts were always there to synthesise various macroeconomic approaches to a single one; however, it is only in recent years that such efforts seem to have come to fruition. In what follows, we set to show the core propositions of this consensus as well as the viability of the entire enterprise. Below (Fontana 2009), we present a simplified version of the new consensus macroeconomics in a set of three equations describing the dynamics of change in a closed economy.[7]

The output gap:

$$(y - y^*)_t = a_0 + a_1(y - y^*)_{t-1} + a_2 E(y - y^*)_{t+1} + a_3(i_t - E(\pi)_{t+1}) + u_1.$$

The Phillips curve:

$$\pi_t = b_1\pi_{t-1} + b_2 E(\pi_{t+1}) + b_3(y - y^*)_t + u_2.$$

[7]See Arestis (2009) for a fuller description of the model with a foreign sector.

The interest rate policy (or Taylor) rule:

$$i_t = i^* + a_1(\pi - \pi^*) + c_2(y - y^*)_{t-1} + u_3.$$

All the parameters in the above three equations are positive with the exception of a_3; y is real output, y^* is the normal output and their difference $(y - y^*)$ is the output gap; a_0 is a constant designed to capture, among others things, the effects of fiscal variables on the output gap; E is the operator of expectations; i is the nominal interest rate, a policy instrument controlled by the central bank, π stands for inflation; u's are stochastic shocks and t is time.

The output gap equation bears certain similarities to the IS curve of the neoclassical synthesis in that it indicates how the current output gap is affected by the output gap lagged by one period, the expected output gap and also the real interest rate. The difference between the above IS curve and the traditional one is that it is derived from inter-temporal optimisation of a utility function and thus, unlike the old IS curve it is based on rigorous micro-foundations. Further, the above IS curve contains lagged and forward looking elements (rational expectations) that relate the real interest rate $(i_t - E(\pi)_{t+1})$ to the output gap.

The Phillips curve with inflation is determined by the current output gap, as well as the difference between past inflation rates and that expected in the future. In this equation all the parameters are positive and also $b_1 + b_2 = 1$. Such a difference in inflation rates is also an indirect indicator of the degree of commitment of the central bank to long run price stability. These two characteristics differentiate the new consensus Phillips curve from the traditional one, for the latter is based only on current output.

Finally, the third equation is a simple monetary policy rule (standard Taylor rule, see Chap. 13), with the nominal interest rate being explained by the current output gap, the deviation of current inflation from its target, π^* and the equilibrium real interest rate, i^*. The monetary rule equation is really an innovation whose role is to replace the conventional LM curve. Further, the equation suggests that monetary policy in the economy is derived from the output gap as well as the inflation rate.

The above three equations convey the idea that the appropriate economic policy is the monetary one. The central bank in the short run should target the stabilisation of output, whereas in the longer run the goal should be the maintenance of price stability. Both goals can be attained through the central bank's decision on the short-term interest rate. The mechanics are as follows: changes in the short term interest rate work themselves out via changes in the real interest rate, investment and consumption, that is changes in aggregate demand which affect the level of output, and, therefore, the rate of unemployment and the subsequent output gap change the inflation rate. Monetary policy can be modified appropriately so as to achieve the target inflation rate. The transmission mechanism can also work itself out through the difference between expected and current inflation rates. The adjustment process starts with interest rate changes which lead to differences between current inflation and target inflation rates which in turn affect the expected inflation rate, and so forth.

The problem in the above core equations are two strategic variables, that is, the equilibrium interest rate and the normal output, which are hard to define and therefore to devise methods for their estimation. Starting with the equilibrium interest rate Michael Woodford, one of the architects of this new consensus macroeconomics, defined it as the "equilibrium real rate of return in the case of fully flexible prices" (Woodford 2003, p. 248) and Fontana (2009) as the interest rate that prevails in the long-run when current output y is at its potential level y^*. It comes as no surprise that these definitions are far from adequate because we actually have no clue for the exact determination of the natural interest rate and output. In fact, this is not a new problem, and the answers that we get are on the negative side. And for that, we need not quote only Marx, who definitely ruled out the existence of such a natural interest rate that could be placed at par with the concept of natural prices, rate of profit or wages but also Keynes who, as we saw in Chap. 9, did not propose any method for its determination. It seems that he tilted towards Marx's and also Sraffa's position. Wicksell (1898) on the other hand differentiated between the current or observed rate of interest and the natural rate of interest, which he equated to the marginal product of capital – a non-observable variable. If the current rate of interest falls short of the natural rate of interest there is incentive to invest and consume and so prices tend to rise, and *vice versa*. Thus, one can indirectly determine the natural interest rate from the movement of the inflation rate. This analysis leads directly to the idea of the non-accelerating inflationary rate of interest (NAIRI) which has already been established in the relevant literature. As for the natural interest rate, it is hard to identify it theoretically and to approximate practically. The usually approximation of the natural interest rate is through the interest rate of long-term (5 or 10 years) government bonds. It is important to note though that in such an approximation a monetary variable is being selected when we know that in Wicksell (see also Chap. 8) the natural interest rate is not a monetary variable and that it corresponds to the payment for the services for the use of capital goods when all the factors of production are fully employed. Having defined the natural interest rate, Wicksell tried to proxy the natural interest rate in the real economy, for this purpose he picked the housing market and used the surrogate of the ratio of the flow of rent incomes to the value of stock of houses. If the so derived natural interest rate is higher than the market (money) interest rate, it follows that people tend to invest in houses thereby increasing the value of houses.[8] Clearly, if one cannot determine such a strategic variable as the natural (equilibrium) interest rate in a theoretically sound way, it follows that the whole exercise of monetary or rather interest rate policy is fraught with uncertainties as to its effectiveness. The uncertainties are multiplied, when we think of the second strategic variable which is the natural level of output, which is certainly more precisely defined; however, when it comes to its

[8]The recent housing bubble in the US and other economies might be interpreted along a Wicksellian cumulative process, where the whole process begins with central banks trying to keep the interest rate as low as possible.

measurement, we face a number of problems that cast doubts on the above outlined transmission mechanisms.[9]

Finally, this monetary policy exercise inspired by the new consensus macroeconomics may only be meaningful when the inflation and interest rates are at low levels. Further, the economy must be growing at a rather healthy rate, because in a sharp downturn, such as the one that began in 2007, it is very unlikely that a monetary policy of changing the interest rates by fractions of percentage points will have any perceptible effect on the aggregate economic variables. Fiscal policy would become imperative and with that the consensus in new consensus macroeconomics would be called to question. Paraphrasing Goodhart's law, we may say that "any observed macroeconomic concession will tend to collapse once pressure is placed on it for control purposes". For example, we know that in the mid-1960s the consensus was "we are all Keynesians now"[10] but a few years later macroeconomics became a deeply divided discipline.

16.6 Policy Implications

From the analysis so far it becomes clear that the New Keynesian economists restore the importance of the stabilisation role of government, which is absolutely necessary to correct certain weaknesses in the operation of the economy that result from rigidities in the markets (nominal or real), of the role of money which is not neutral, but in the short run affects the level of economic activity and, hence, employment and so forth. On the other hand, the model by Greenwald and Stiglitz (1993a, b) shows that the rigidity in prices is not the problem and that such rigidities may rather exert a positive effect on the economy (the idea is that higher flexibility may give rise to larger economic fluctuations). The government has a corrective role to play and this relates to the more effective co-ordination of agents in the economy. However, New Keynesians fully agree with New Classical economists that in the long run, monetary policy is neutral. The disagreement is about the short run, where New Keynesians argue that monetary or better interest rate policy may be effective in changing output and employment in the short run.

The New Keynesian economists, contrary to the Keynesians of the Neoclassical Synthesis, do not adopt any specific mix of policies which they believe will stabilise

[9]The issue of the estimation of capacity utilisation is certainly a controversial one with various methods being suggested which vary from questionnaires to econometrics (see Dergiades and Tsoulfidis 2007).

[10]This statement is attributed to Friedman by the *Time* magazine (December, 1965) although in the next issue of the magazine Friedman complained that he was misquoted and that what he really said was that "in one sense, we are all Keynesians now; in another, nobody is any longer a Keynesian". From this Delphic statement one thing is certain: the characterisation "Keynesian" was received favourably by the majority of economists at the time. The same characterisation would have not been so favourably received in the early 1980s.

the economy and also do not necessarily emphasise the importance of demand over the supply side of the economy (Gordon 1990, p. 1117). They argue, however, that there exist so many sources of disturbances that it is not by any means easy for a government to achieve its objectives. Certainly, in the case where the economy is found in deep crisis the New Keynesian economists argue for a more active role for the government. In normal periods they would support flexibility in the economic policy, because it is not possible to design long-term economic policies in such a fast changing economic environment (Stiglitz 1984). This is the most general principle that the majority of New Keynesians would share. However, the supporters of particular theories would propose very specific policies. For example, in the question of unemployment that is caused by the distinction of workers as insiders and outsiders, the New Keynesians would favour policies that aim at the weakening of insiders and strengthening of outsiders.

Although New Keynesian economists recognise that monetary policies may exert short run effects on the economy, they do not imply by that that they would support an active monetary policy just for short run gains in output and employment. The idea is that the result of such an easy monetary policy would give rise to inflationary expectations and thus defer problems for the future. They would however favour monetary policy for stabilisation purposes. For example, if the economy is hit by some unexpected external shock, such as a fall in consumer confidence, which tends to lower both output and inflation, lowering interest rates may help by increasing output, while stabilising inflation and inflationary expectations. From, the above it follows that the New Keynesians managed to give new meaning and direction to monetary policy which is really more an interest rate policy and much less a targeting of certain monetary aggregates. This becomes particularly pronounced in the new consensus macroeconomics which purports to unify macroeconomic theory.

16.7 Summary and Conclusions

In this chapter, we argued that the Keynesian theory, which many economists during the 1970s thought, would become, at best, a mere chapter in the history of economic thought books; nevertheless in the mid 1980s, (New) Keynesian economics made a triumphal come back. According to Mankiw (1993), one of the protagonists of this effort, Keynesian economics was resurrected. The hallmark of New Keynesian economists is their continuous efforts to inject realism in their analysis and for this reason they accept several sensible ideas of rival schools of economic thought. For example, New Keynesians adopted the Monetarist idea of natural unemployment and expectations augmented Phillips curve and adjusted it to the non-accelerating inflation rate of unemployment (NAIRU), – that is, the level of unemployment which is consistent with stable inflation. Another example of creative integration of a concept in New Keynesian economics is the rational expectations hypothesis. Once the rational expectations hypothesis is adopted in the New

Keynesian Models (with nominal or real rigidities in prices and wage), it gives rise to conclusions that render monetary and fiscal policies meaningful for at least short run results. Moreover, the emphasis of New Keynesians in the provision of micro-foundations has rendered their approach very promising and appealing to the majority of macroeconomists today.

Consequently, it is not accidental that many economists aspire to a synthesis between the efforts to inject realistic elements in the macroeconomic models of New Keynesian economists, which have not been tested sufficiently, with the simulation models of the RBC approach (in which there is no voluntary unemployment and the role money is absolutely neutral). In the future, this tendency is expected to be strengthened to lead to dynamic stochastic general equilibrium (DSGE) models and to simulation studies which will be more representative of reality and hence more suitable for the conduct of economic policy. These models are called dynamic because they evolve over time and are stochastic because they allow for random shocks that hit the economy. Further, as these models are based on micro-foundations they are not affected by the Lucas critique. The difference between the DSGE and the simulation models in RBC is the presence of monopolistic competition and various other rigidities (Woodford 2003, 2008).

In our view, New Keynesian economics will continue to attract the interest of economists in the years to come for two simple reasons: firstly, because of the recognition that unemployment is, to a great extent, involuntary and second because the operation of the market if left alone will not lead to socially desirable results. Consequently, New Keynesian economics provides the rationale for government intervention and also for the necessary guidance and tools. As a consequence, during a period of downturn as is the current one, the popularity of New Keynesian economics is expected to increase while the opposite will be true for the competing streams of economic thought that exemplify free operation of markets.

Questions for Thought and Discussion

1. Discuss the major characteristics of the New Keynesian approach.
2. What are the differences between New Keynesian and New Classical economics?
3. Discuss the model of unemployment with hysteresis of the New Keynesian economics.
4. Discuss the following concepts:
   - Menu cost.
   - Co-ordination failures.
   - Efficiency wage hypothesis.
   - Insiders–outsiders model.
5. How do nominal rigidities make government intervention in the economy effective?
6. How do New Keynesians interpret business fluctuations?
7. Can monetary policy affect the economy at least in the short run? (Discuss at least three approaches using the same model!)
8. “Where is the consensus in the new consensus macroeconomics”?

Notes for Further Reading

For a detailed presentation of New Keynesian economics we recommend readings that have been conveniently collected in a two volume book by Mankiw and Romer (1993a). Among these readings, the articles in parenthesis the year of the original publication by R Gordon (1990), Barro (1989a), Rotemberg (1987), Stiglitz (1984), Romer (1993) and Tobin (1993) are absolutely essential. A few of these articles are also included in the book of readings by Snowdown and Vane (1997) together with a summary view of the New Keynesian approach. What is striking about this approach is the pluralism in viewpoints expressed in very specific issues, for example, regarding profit margins and their variability in various stages of the business cycle. Thus, Bils (1987) as well as Rotemberg and Woodford (1991) in two frequently cited studies find that profit margins are countercyclical; by contrast, Domowitz *et al*. (1988) find that the profit margins are procyclical. Mixed results are found in the study by Kaskarelis and Tsoulfidis (1998) for specific branches of Greek manufacturing. The book by Snowdown et al. (1994) contains a chapter on New Keynesian economics, where there is a discussion of various strands of this school of economic thought and interviews with two major New Keynesians. Mankiw (in this interview) claims that the RBC approach has a bleak future whereas he is more positive towards Monetarism and New Classical economics. He thinks that the New Keynesian economics can use some of their ideas. More specifically, Mankiw disagrees with the neutral role of monetary policy in the RBC approach and sides with the Monetarists and the New Classical Economists in the idea that the effects of money in the short run may not be neutral. Of course the difference of New Keynesians from Monetarists and New Classicals is that monetary policy is not neutral in the short run because of the presence of various rigidities in markets. Thus, it comes as no surprise that Mankiw sees a convergence of various macroeconomic strands of thought to a New Synthesis or new consensus macroeconomics (Woodford 2008; Arestis 2009; Fontana 2008).

A quite different view is expressed by an old Keynesian and current founding "member" of New Keynesian economics, Edmund Phelps (1994), who in his interview (Snowdown *et al*. 1994) does not see any convergence process between the various macroeconomic schools of economic thought. Instead he sees what he calls the "Balkanisation" of macroeconomic schools of economic thought.

More recently, New Keynesians having accepted the RBC methodology of simulation models have begun to build DSGE macroeconomic models (for details see Woodford 2003). These models have in common with the RBC simulation models the microeconomic foundations and the rational expectations hypothesis but differ in that they accept market failures, that is, a number of nominal and real rigidities prevent the markets from clearing and so there could be involuntary unemployment. Unlike the RBC models the DSGE models find that monetary policy and fiscal policy may be effective and between the two New Keynesians tilt more towards monetary (interest rate) policy.

Chapter 17
Economic Theory in Historical Perspective

Thus [. . .] to explain the outstanding features of our actual experience; – namely, that we oscillate, avoiding the gravest extremes of fluctuation in employment and in prices in both directions, round an intermediate position appreciably below full employment and appreciably above the minimum employment a decline below which would endanger life.

(General Theory, p. 254)

[...] the ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from intellectual influences, are usually the slave of some defunct economist.

(General Theory, p. 383)

The truth is most of the individual mistakes boil down to just one: a belief that markets are self-adjusting and that the role of government should be minimal. Looking back at that belief during hearings this fall on Capitol Hill, Alan Greenspan said out loud, "I have found a flaw". Congressman Henry Waxman pushed him, responding, "In other words, you found that your view of the world, your ideology, was not right; it was not working". "Absolutely, precisely", Greenspan said. The embrace by America – and much of the rest of the world – of this flawed economic philosophy made it inevitable that we would eventually arrive at the place we are today

(J. Stiglitz, Capitalist Fools, 2009)

One would like to think that the market for economic theory is self correcting, but maybe, like the financial market, it is not.

(R. Solow 2010)

17.1 Introduction

In this chapter, we attempt a brief and at the same time critical review of the characteristics of each school of economic thought that withstood the test of time and is still fuelling with ideas and arguments of the conduct of economic policy. Furthermore, our interest focuses on the nature of economic theory, and its future inasmuch as we essentially accept the long period analysis, where the phases of

L. Tsoulfidis, *Competing Schools of Economic Thought*,
DOI 10.1007/978-3-540-92693-1_17, © Springer-Verlag Berlin Heidelberg 2010

economic growth are followed by economic downturn in a long wave-like evolutionary pattern. Economic history teaches that during times of depressions economic theories are being tested, and their failure to provide satisfactory answers to the problems at hand paves the way for the emergence of competitive economic theories, which establish themselves during the recovery stage.

In what follows, we deal with the conditions that contribute to the development of economic theories and continue with an examination of the concrete circumstances that gave rise to modern neoclassical macroeconomic theories. The chapter concludes with the claim that the current impasse in macroeconomics is indicative of the need for new directions in economic theory which will become imperative in case of a long economic downturn and concludes by suggesting the need for a synthesis between the classical analysis and the theory of effective demand.

17.2 Core Characteristics of Competing Economic Theories

In an effort to present and evaluate critically the various economic theories and schools of economic thought, we began with the Mercantilists and the Physiocrats, who were the first whose economic theorisation attained scientific status. The Physiocrats, in particular, saw economic life to be in deviation from an ideal (natural) condition, and considered that market forces, if left alone, would not succeed in leading economic life to the ideal model illustrated in the operation of the *Tableau Economique*. Therefore, the Physiocrats proposed economic policies to strengthen the economic forces at work, which were not developed enough to cope with the (feudal) obstacles that prevented the normal operation of the economy. We focused on the Physiocrats because the Mercantilists simply offered policy measures which, however, were neither consistent with the logic of a single school of economic thought nor did they possess the necessary theoretical cohesion. There is no doubt that the Physiocrats constitute the first school of economic thought with the necessary theoretical cohesion; nevertheless, the operation of endogenous economic forces in and of themselves was not strong enough to promote the economic progress that the Physiocrats would have liked to unless there was government intervention. Only through the active support of government, economic life would approximate the Physiocratic ideal society.

Adam Smith sought to discover the "natural order" of economic life, and for that, he was inspired by the movement of planets determined by the laws of gravity. Smith used this analogy to explain the movement of actual prices by the law of equal profitability which was supposed to hold in a particular set of hypothetical market prices which he called "natural prices". The term "natural" signifies the fact that economic phenomena have their own internal dynamics, just like natural phenomena, and operate, as François Quesnays observed, in a way that is "independent of men's will". A salient feature of Smith's and the other classical economists' argument was the determination of natural prices by the labour time content of commodities and the explanation of profit as a form of surplus, over and

above what is needed for the reproduction of workers capacity to work on the one hand and of the productively consumed means of production on the other. In this determination, some elements must be considered as moving slowly relatively to others, and, therefore, they can be treated parametrically; in this sense, the parameters or data of the classical analysis are the real wage, the output produced and the technology in use.

These ideas are continued by Ricardo, who fully accepts Smith's method as well as the object of analysis, which remains the determination of natural prices. In particular, David Ricardo sought to discover the "laws that determine the distribution of income" by correcting various aspects of Smith's labour theory of (exchange) value. We know that Smith abandoned the labour theory of value when he extended his analysis to a capitalist economy. Ricardo argued that while this labour theory of value for pre-capitalist societies deduces the movement of relative (equilibrium) prices from the respective relative labour times with accuracy, it, nevertheless, continues to hold in modern society as well albeit with modifications, which are designed to account for the presence of capital, the different durations required for the completion of the production process and the changes in distribution. Ricardo further argued that these modifications are predictable and of an acceptable order of magnitude and do not change fundamentally the originally correctly stated principle.

In this context we can also place Marx's mature work in *Capital*, where he sets out "to lay bare the law of motion of modern society" to which end the explanation of natural prices or prices of production was the first crucial step. By setting this analytical context classical economists established a scientific status for economics, as they were the first to argue that capitalism gives rise to economic phenomena that display regularities which can be subjected to abstract theorisation (Heilbroner 1983). Consequently, economics became an inquiry that could generalise and, therefore, theorise economic phenomena independently of ideology or religion.

Marx's analysis also accepts as data the real wage, the output and the technique in use. The differences are in the analyses. Marx introduced the two senses of socially necessary abstract labour time; the first refers to a weighted average of the labour time required in the production of commodities, while the second sense accounts for the deviation between the actual production and the social demand. Moreover, labour values must be transformed through the medium of money into direct prices (the first step prices in Max's analysis), since in *Capital* there is no barter economy and consequently the economy can only be monetary. Then, the direct prices are transformed into prices of production that form more immediate regulators of market prices than direct prices. It is interesting to note that the more concrete type of prices is the regulating direct prices and the corresponding regulating prices of production. Empirical research to date has shown that the use of average values (direct prices) and average prices of production has been proved to give rise to surprisingly accurate approximations to market prices. Should the research have been conducted in terms of regulating direct prices and regulating prices of production the results, to our opinion, would be expected to be even better for the labour theory of value.

The idea that commodities embody direct and indirect labour time enabled Marx to resolve the puzzle of the source of profits. In particular, Marx was the first to argue that in capitalism workers are exploited not because they are not paid their full wage, but because with the full wage they receive workers are able to pay only for the basket of goods required for the reproduction of their capacity to work (their labour power), which is acquired through what is only a portion of their total labour time. The difference between total labour time and that required to reproduce the workers' capacity to work is called surplus labour time and its monetary expression, the surplus value, is appropriated by the propertied classes (capitalists and landlords) and the state. The value of the non-labour inputs (indirect labour) is simply embodied once and for all in the value of the new commodity as, for instance, in the case with raw materials or bit-by-bit through depreciation as in the case of fixed capital. We argued that this theory is logically consistent and, at the same time, realistic in the sense that it explains both the sources of profits and the rate of profit as an endogenously determined variable.

The theory of value of the classical economists was a prerequisite for understanding the dynamics of the capitalist economy. Regarding this issue there has certainly been progress in the theorisation of the development of key economic variables, as well as their interdependence. These variables include the real wage, the mechanisation of the production process, as manifested in the rising capital–labour ratio and of course the rate of profit, whose fluctuations play the decisive role in the characterisation of the phase of the economy. There is no doubt that both Smith and Ricardo made important contributions to the understanding of the evolution of profitability; however, the analysis in Marx of the tendential fall in the rate of profit is shared by almost all major economists of the past, that is Smith, Ricardo, J.S. Mill, Walras, Jevons, Marshall, J.B. Clark, Veblen, Keynes and Schumpeter. The difference lies in both the rationale of the falling tendency in the rate of profit and also in the relationship between this fall and the phase of the economy. Marx argues that the rate of profit is falling as a result of a rising capital–labour ratio (*i.e.*, because of the mechanisation of the production process) while, at the same time the rate of surplus value is also rising but at a slower rate. In this scenario, Marx shows that the fall in the rate of profit is absolutely consistent with an increase in real wages and a simultaneous fall in the purchasing power of workers. The point is that whatever is the increase in the real wage it does not exceed the increase in productivity, as otherwise the real wage interferes with profitability and hence the smooth functioning of the economic system. Central in this analysis is the distinction between productive and non productive labour, which is one of the salient features of classical economic thought.

The neoclassical approach, which emerged in the last quarter of the nineteenth century, that is, during the great depression of 1873–1896, continued to utilise the long period method and the natural prices as the object of its inquiry. The difference from the classical approach lay in a theorisation that was based on an entirely different set of data comprising the preferences of individuals, the size of the endowment and its distribution among individuals as well as the state of technology. It took quite a long time for the neoclassical approach to become the established orthodoxy in economic

theory. Some historians of economic thought characterise this succession as a silent non-revolutionary process (Blaug 1983; Hollander 1985). No matter how long it took this process to fully unfold, its very purpose (stated or not) was to set aside the more realistic classical approach mainly because of its disturbing political implications-especially those emanating from the labour theory of value.

If the purpose of economic theory is to approximate the way in which the actual economy works, there is no doubt that the classical approach was more relevant than the neoclassical one to the practice of business people, and, therefore, to economic reality. For example, the distinction between productive and unproductive labour is part of established prudent business practices; the classical theory of competition is characterised by a realism that is not found in the neoclassical notion of perfect competition, which *prima facie* contradicts almost every aspect of real life competition (Shaikh 1980; Eatwell 1981).

The first neoclassical economists (*i.e.*, the triad Jevons, Menger and Walras initially and subsequently Marshall, J. B. Clark and Böhm-Bawerk) were affected but at the same time, also contributed to the creation of a new intellectual atmosphere in which the classical system was found to be unsatisfactory and its replacement by a theory that would legitimise property and emphasise the merits of an exchange economy became imperative although not necessarily urgent. It is important to point out that although these ideas were developed during the great depression of 1873–1896, they gained momentum in the "Victorian Era", which was a period of steady economic growth and so the demands for a realistic economic theory from policy makers were much more elastic. Whereas, in the period of depression 1873–1896, both classical and neoclassical theories were in agreement with respect to no government intervention.

In this context it is worth stressing that the concept of competition in Marshall had a lot in common with the competition of classical economists, as an evolutionary process of rivalry between firms rather than of perfect competition. In fact perfect competition was devised as a mathematical result in the work of Cournot and the latter was imposed by the requirements of the neoclassical theory and not because it was derived from the observation of how firms actually compete. This concept was not fully developed in the work of Marshall (1890) and took some time until its development to the present state. The criticism of Sraffa (Sraffa 1925, 1926, 1930) that followed attempted to demonstrate the logical inconsistencies of the Marshallian theory of the firm, with the U-shaped average cost curve that operates under perfect competition in the context of partial equilibrium. The criticism targeted the logical foundations of the theory and we saw that the subsequent monopolistic competition revolution emerged as a result of this criticism. Sraffa suggested the abandonment of the notion of perfect competition and the development of the more realistic idea of monopolistic competition. In doing so, it was implicitly admitted that actual economic life is somehow in deviation from what it ought to be. Gradually we had the return of the notion of perfect competition, which entered in economic analysis in the 1930s mainly through the antitrust legislation and the government regulation of industry. The idea is that if the government seeks to correct the operation of the market, there must be an ideal model that could be

used as a yardstick in the evaluation of actual competition and the steps needed towards its perfection. Perfect competition became the established orthodoxy in economic theory, gradually and unconsciously in the beginning of the monopolistic competition revolution, but then consciously and effectively by the economists of the University of Chicago.

Therefore, if perfect competition is an imaginary situation imposed by the requirements of the neoclassical theory, then we can say that the so called "monopolistic competition" revolution of the 1930s essentially led to the establishment of the unrealistic model of perfect competition not only for theoretical purposes but also to inform policy decisions. As a result the classical conception of competition, as a process of rivalry between firms over market shares was set to the fringes of economic analysis. It is only in recent years that we are witnessing the resurfacing of the notion of competition as a process in the works of Marxists, Schumpeterian and Austrian economists. It is important to stress that the classical conception of competition basing itself on realism gains more acceptance in business rather than in economic literature proper. The work of Porter (1990) as well as the resource advantage theory (Hunt 2000) has much in common with the concept of competition as a process of rivalry between firms in their battle to increase their market share by increasing productivity and undercutting prices, leading to a gradual displacement and subsequent absorption or simply the elimination of rival firms. It is obvious that this kind of competition is not the same as "competition" as a static situation, where companies have all the time that they need to decide on the amount of output to be produced based on a given price. The same argument holds true for other forms of competition theorisation, such as the monopolistic or oligopolistic competition, because these models are essentially attempts to inject realism to the static and unrealistic model of perfect competition. In our view, if one rejects the notion of perfect competition, then one should also reject the various derivative expressions of this "competition" which appear under the labels "monopolistic" or "oligopolistic" competition.

Parallel to monopolistic competition, and, therefore, the microeconomic revolution development in the 1930s, we had the Keynesian or macroeconomic revolution in its making. In fact, the distinction between micro and macroeconomics originates in the 1930s. We know that up until the 1930s, economic theory, regardless of the approach, was unified. Even in the neoclassical approach there was no distinction between micro and macroeconomics in the sense that results derived in microeconomics may be false in macroeconomics (the fallacy of composition). For example, the Walrasian general equilibrium could not be classified as microeconomic in the sense that it refers to the determination of equilibrium prices, because it encompasses the economy as a whole, and, therefore, the Walrasian general equilibrium is both micro and macroeconomic at the same time. The *General Theory* as we know was published in the middle of the Great Depression and contained ideas that could help in the contemplation of economic policies to deal with the depression. Keynes argued that the *raison d' être* of unemployment is not the malfunctioning of the labour market, but the scarcity of effective demand. The latter does not depend on prices or a lack of saving but on uncertainty and expectations, phenomena which are

very hard to theorise. The scarcity of effective demand and the inability of the market system to generate effective demand to the amount required for the establishment of full employment of labour, calls forth the activation of the political element, that is, of human intervention. Thus, the impact of the *General Theory* in both economic theory and the operation of the economy was immediate.

The very simple fact that human intervention or the visible hand is necessary for a solution to economic problems was argued for the first time in a theoretical and therefore convincing manner. The intellectual climate of the time was conducive to these ideas. In fact, in periods of depression, such as that of the 1930s, policy makers are not only prone to pay close attention to but also implement new ideas, especially if these ideas promise desired results and are presented in a logically coherent and practical way. Keynes's theory, even in the "rehabilitated" form that was given to it by Hicks and others in the so-called neoclassical synthesis, concentrated all these required characteristics and what was even more crucial, its implementation delivered results, as one may judge by the overall economic performance during the period of the "golden age of accumulation" that started after WWII and ended in the mid to late sixties. Furthermore, post-WWII Keynesian economics provided the theoretical justification, and, therefore, created the necessity for a number of institutions that promoted government spending that enhanced the process of capital accumulation. We observe that economic theory during this period became more practically based, in the sense that it helped solve real socioeconomic problems. Meanwhile, economics lost, at least partly, it's purely scientific character and allowed the political element to play an important role in the management of economic affairs through the appropriate combination of fiscal and monetary policies.

As a consequence, there was widespread satisfaction with economics as a science and the policies emanating from its theoretical conclusions, since this was a period characterised by high growth and low unemployment rates. By the mid-1960s, however, the situation gradually began to change, with growth rates slowing down, and unemployment levels starting to rise together with inflation. Both results were quite unexpected within the neoclassical synthesis version of the Keynesian paradigm based on the idea of trade-off between unemployment and inflation as this was exemplified in the Phillips curve. The slowdown in economic activity that started in the mid-1960s and continued through the 1970s and into the early 1980s led Keynesian economics into disrepute and decline. However, it is clear that this slowdown did not convert into a deep depression and so its results did not have the same destructive character as those of the depression of 1930s. It is a commonly held view that the causes of this difference were due to the presence of institutions that resisted to and in general ameliorated the harsh effects of this protracted slowdown.

Neoclassical economics, in its monetarist version, attributed the slowdown in the level of economic activity not to the internal defects of the market system but to external circumstances and in particular to government and its intervening role in the markets. These ideas found strong support by economists disappointed by the failure of Keynesians to explain the coexistence of high inflation and unemployment. By contrast, the monetarists found an answer to the puzzle of stagflation

through their idea of the expectations-augmented Phillips curve and its associated natural rate of unemployment. Actual unemployment is the result of households' choice insofar as they find the actual wage rate too low to motivate them to supply their labour services. This household decision, the monetarists argue, is further encouraged by governmental intervention in the labour markets. The solution to the problem of high unemployment is simply no government intervention and free operation of markets. However, the monetarist success in attaining popularity was only short-lived, since the continuation of stagflation cast doubt on the monetarists' capacity to understand the causes of the problems they sought to solve and because their policy proposals could not offer any socially acceptable solutions. The failures of monetarism to offer adequate explanations and also solutions to new phenomena made room for the emergence of the variant of "new classical economics", which became the orthodoxy of the 1970s.[1]

New classical economics essentially merges the hypothesis of rational expectations with monetarism. The idea behind the rational expectations hypothesis is that if we do not consider the element of surprise, then whatever is to happen in the long run will also happen in the short run. The rational expectations hypothesis postulates that economic agents, on average, know the outcomes of the true model of the economy and thus do not waste any of their time searching for the monetarist long-run equilibrium situation; they simply transfer themselves virtually instantly there. Consequently, systematic economic policy is ineffective in both the short and long run. In fact, there is no such dichotomy because the rational expectations hypothesis spirits away the notion of time. The attainment of equilibrium is the state to which an economy naturally and instantaneously is led to, unless there are exogenous shocks such as those emanating from technology, input prices, preferences or even from unanticipated government intervention.[2]

New classical economists, like the monetarists before them, could neither provide satisfactory explanations nor policy prescriptions for the lasting stagflation. Paradoxically, the new classical economics managed to theorise the way out of the slowdown, by propounding the encouragement of the working of market mechanisms and of non government intervention. The trough of this recent depression took place in the early 1980s and it became the starting point for the emergence of the real business cycle approach, where the exogenous technological change (and not necessarily the money supply) becomes responsible for the phase of the economy.

[1]It was hard, for policy makers, to accept the simplistic nature of monetarist "policy prescriptions" which called forth for no government intervention and growth of money supply at a level approximately equal to the long-term growth rate of GDP. But monetarism had an enormous influence on governments and set the tone for the era of the so-called neoliberalism that swept the globe since the 1980s.

[2]It is important to point out that the acceptance of a particular theory depends to a certain extent on the interests that it serves something that is at least partially independent of their ability to aid in prediction and the formulation of economic policy. One can place in this context the supply-side economics of the early 1980s, which sought to limit the role of the state. This was much more suited to the needs of capital in the age of globalisation, when domestic demand becomes less important as a source of growth.

In this analysis, taken to its extremes, the exact stage of the economy is due to the optimisation behaviour of the economic units, and consequently the recession or recovery phases are due to the optimal reaction of economic units to external shocks in the environment within which they operate. Recessions or recoveries are the results of voluntary reactions of households, which in the first case decide not to offer part of their labour services in order to offer them at more opportune times and in the second case find that their interests are served better by making their labour services available to economic activity. The results in both cases are optimal, which means that there is no reason for government to intervene because it cannot further improve the current situation (Prescott 1986, p. 21). In this perspective the economy is always at full capacity, therefore capital is always optimally used and so is labour. Equilibrium in the labour markets is determined by the behaviour of inflation. Hence the acronym the non-accelerating inflation rate of unemployment (NAIRU) can be seen as a characteristic example of the fetishism of economic categories within this approach, in the sense that the inflation rate "decides" the characterisation of actual unemployment as natural or as excess, since if the rate of inflation is steady, then the economy will find itself in its natural, that is, long run equilibrium position regardless of the actual number of unemployed.

From the above we may conclude that the monetarist, the new classical and especially the real business cycles approaches inevitably return to the tradition of neoclassical economics according to which economic theory is put on a par with the natural sciences and economic theory is, once again, deprived of its political element. The big problem with these views, and, especially, as of this writing, still popular real business cycles approach is that of a protracted recession, where they have no realistic policy proposals. Furthermore, the Keynesian safety networks – that were created during the "golden age of accumulation" – have been truncated in most countries during the previous "silent depression" of the 1980s and so the pressure for an active government role will be widespread. As the real business cycle approach does not theorise shocks of this dimension, naturally it does not propose any policies to, at least, ameliorate its consequences.

These considerations contributed to the increasing popularity of the New Keynesian approach according to which we should move to less restrictive economic models. A common theme of this approach is the continuous effort to provide microfoundations to macroeconomics. It is important to point out that the essential component of this microfoundationism is the neoclassical theory of competition in its monopolistic form together with various rigidities across markets which make government intervention especially in the field of monetary policy effective once again. This approach in its triumphant come back can also make use of rational expectations as an auxiliary hypothesis for faster attainment of the New Keynesian equilibrium. The idea is that if the New Keynesian model is the true model of the economy, then, naturally, economic agents will attain its results without wasting any of their time in non-equilibrium situations. Further, there is an ongoing discussion about the creation of a new neoclassical synthesis (or new consensus macroeconomics), which will combine all the progress that has been achieved hitherto in neoclassical macroeconomics. Even though such a hybrid

model is in the research agenda of economists today, we predict, however, that it will not last for long, because it is based on the notion of competition as an end-state rather than as a dynamic process of rivalry. The whole construction is therefore unstable and although efforts for improvements may continue, it is doomed either to disintegrate or mutate since the substance from which this synthesis is made of is problematic.

17.3 Elements for a New Direction

We argued that the purpose of economic theories is to lay bare the laws of motion of the economy and to determine the limits of variation of economic variables so as to provide guidelines for political intervention. We argued that the classical theory has solid theoretical foundations and in this sense is superior to neoclassical theory. Nevertheless as Paul Samuelson notes:

> [I]t is better to have a model with inexact foundations that gives you a good grip to handle reality than to wait for better foundations or to continue to use a model with good foundations that is not usefully relevant to explain the phenomena that we have to explain. (Samuelson 1988, p. 295)

With this in mind, the classical model concentrates the required properties of a theory that reveals the basic trends of the capitalist system and at the same time takes into account the specific institutional arrangements. This is clear for instance in Smith's or Ricardo's labour theory of value in a primitive society, which is modified to accommodate the concrete circumstances of capitalism. The same is true with Marx, whose "laws of motion" work more precisely in conditions of advanced capitalism. In this sense, the classical approach is preferred to the neoclassical one. This superiority, however, of a theoretical level must materialise in a concrete analysis that can interpret the phenomena and predict with relative accuracy the results of specific economic policies. Otherwise the classical approach remains superior only for its supporters, without having any further implications.

In this direction, classical analysis should be integrated with the theory of effective demand. The idea is that this theory leaves the question of integration of institutional elements with the classical economic theory open. Keynes was very conscious of these elements since his theory operates under the institutional set up of advanced capitalism with fully developed money and capital markets as well as governments capable of understanding the circumstances and intervening for desired effects. Neoclassical theory, by contrast, is a-historical in its approach for it conceptualises the market independently of institutions. In the perfect competition model, which constitutes the quintessence of the neoclassical structure, institutions are generally viewed as a kind of friction to the normal operation of the markets. The truth is that markets would simply perish without the existence of governments and the outer institutional shell.

If we can visualise a direction for the future of economic theory then this would lie in the creative synthesis of the classical research programme of economic

analysis with Keynes's principle of effective demand. There has been much discussion about the possibility of providing such microfoundations for the classical theory of value and distribution starting with the works of Kalecki already in the late thirties who made an effort on the one hand to dispense with the marginal productivity theory of income distribution and on the other hand to put the principle of effective demand in a classical perspective (Kalecki 1937). Subsequently, the works of post-Keynesian and especially neoRicardian economists made valiant efforts to integrate the classical theory of value with the theory of effective demand with much frustration and little progress (Eatwell 1983a). Such a synthesis, in our opinion, cannot be successful insofar as it merely juxtaposes the classical theory and the "principle of effective demand". In our view if a synthesis is to be fruitful it should intertwine the classical theory of value and the "principle of effective demand" in such a way, so as to expose the limitations of Keynes's ideas, where there is an exaggeration of the financial autonomy of capital and where there is brought in the analysis an essentially *deus ex machina* in the form of "animal spirits" or "expectations". The deficiency of effective demand must be seen more in a classical framework, where effective demand is cyclical and structural emanating from within the elemental process of capital accumulation. In such a conceptualisation one might be able to set up the boundaries within which the effective demand exerts its effects on the economy and use these boundaries to further develop the theory of capital accumulation.

The need today for such a synthesis has become particularly urgent. The reason is that so long as the economies are in their expansionary state, as they have been since the mid-1980s until the onset of the severe slowdown that started in 2007, the problem of lack of an adequate theory does not manifest itself all that seriously and we can afford the luxury of many competing paradigms in a never-ending contest. But as economies enter their long downward phase, the need for an adequate theory will become more and more urgent since in the new situation the safety networks that were in place during the downward phase of 1970s and early 1980s no longer exist to contain, as they did back then, to a great extent, the destructive effects of the depression. The need for such a theory certainly exists and as the problems of high unemployment, rising income inequalities and widespread poverty loom gravely on a global scale, the neoclassical theory and its various strands become more-and-more "scientific" and dismissive of the harsh reality by hypothesising perfect markets and perfect information, rational expectations and optimality, desired unemployment and inefficient government intervention. Under these circumstances, the need to explain these phenomena and their causes becomes imperative and classical economics will become particularly important in both understanding the causes of these phenomena and in proposing not only such policies that would strengthen the safety mechanisms that were in place in the 1970s and 1980s but also new policies that can be derived from such an economic analysis.

References

Allen, W. (1987). Mercantilism. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave: a dictionary of economic*. London: Macmillan.

Alchian, A. (1955). The rate of interest, Fisher's rate of return over cost, and Keynes' internal rate of return. *American Economic Review, 45*, 938–943.

Akerlof, G., & Yellen, J. (1987). Rational models of irrational behavior. *American Economic Review, 77*, 137–142.

Anderson, K. (2001). New Marx publications: a MEGA update. *Historical Materialism, 9*, 226–230.

Anderson, P. (1976). *Considerations on western Marxism*. London: New Left Review Books.

Anikin, A. (1975). *A science in its youth: pre-Marxian political economy*. Moscow: Progress Publishers.

Apergis, N., & Tsoulfidis, L. (1997). The relation between saving and finance in EU countries. *Research in Economics, 51*, 333–358.

Arestis, P. (2009). *New consensus macroeconomics and Keynesian critique. Mimeo*. New York: Levy Institute.

Asimakopoulos, A. (1983). Kalecki and Keynes on finance, investment and saving. *Cambridge Journal of Economics, 7*, 221–233.

Asimakopoulos, A. (1991). *Keynes's general theory and accumulation*. Cambridge: Cambridge University Press.

Aspromourgos, T. (1986). On the origins of the term neoclassical. *Cambridge Journal of Economics, 10*, 265–270.

Azariadis, C. (1975). Implicit contracts and unemployment equilibria. *Journal of Political Economy, 83*, 1183–1202.

Bacon, R., & Eltis, W. (1976). *Britain's economic problem: too few producers*. London: Macmillan.

Baily, M. (1974). Wages and employment under uncertain demand. *Review of Economic Studies, 41*, 37–50.

Ball, L., & Romer, D. (1990). Real rigidities and the non-neutrality of money. *Review of Economic Studies, 57*, 183–203.

Ball, L., & Romer, D. (1991). Sticky prices as coordination failure. *American Economic Review, 81*, 539–552.

Barens, I., & Caspari, V. (1999). Old views and new perspectives. On re-reading Hicks's "Mr. Keynes and the classics". *European Journal of the History of Economic Thought, 6*, 216–241.

Barro, R. (1974). Are government bonds net wealth? *Journal of Political Economy, 82*, 1095–1117.

Barro, R. (1978). Unanticipated money, output, and the price level in the United States. *Journal of Political Economy, 86*, 549–580.

Barro, R. (1979). Second thoughts on Keynesian economics. *American Economic Review, 69*, 54–59.
Barro, R. (1981). Intertemporal substitution and the business cycle. *Carnegie Rochester Series in Public Policy, 14*, 237–268.
Barro, R. (1984). Rational expectations and macroeconomics in 1984. *American Economic Review, 74*, 179–182.
Barro, R. (1989a). New classicals and Keynesians, or the good guys and the bad guys. Working Paper No 2982. MA: NBER.
Barro, R. (1989b). The Ricardian approach to budget deficits. *Journal of Economic Perspectives, 2*, 37–54.
Barro, R., & Grosman, H. (1976). *Money employment and inflation*. Cambridge, MA: Cambridge University Press.
Barro, R., & Gordon, D. B. (1983). Rules, discretion and reputation in a model of monetary policy. *Journal of Monetary Economics, 12*, 101–121.
Barro, R., & King, R. (1984). Time-separable preferences and intertemporal substitution models of business cycles. *Quarterly Journal of Economics, 99*, 817–839.
Baumol, W. (1962). *Economic dynamics: an introduction*. New York: Macmillan.
Baunol, W. (1967). Macroeconomics of unbalanced growth: the anatomy of urban crisis. *American Economic Review, 57*, 415–426.
Bellofiore, R. (Ed.). (1998). *Marxian economics: a reappraisal*. New York: St. Martin.
Berle, A., & Means, G. (1932). *The modern corporation and private property*. New York: Harcourt, Brace and World, 1967.
Berlin, I. (1978). *Karl Marx: his life and environment*. Oxford: Oxford University Press. Original work published 1939.
Bhaduri, A., & Harris, D. (1987). The complex dynamics of the simple Ricardian system. *Quarterly Journal of Economics, 102*, 893–901.
Bhagwati, J. (1982). Directly unproductive profit seeking activities. *Journal of Political Economy, 90*, 988–1002.
Bharadwai, K., & Schefold, B. (1990). *Essays on Pierro Sraffa. Critical perspectives on the revival of classical theory*. London: Unwin Hyman.
Bienenfeld, M. (1988). Regularities in price as an effect of changes in distribution. *Cambridge Journal of Economics, 12*, 247–255.
Bilginsoy, C. (1994). Quesnay's tableau economique: analytics and policy implications. *Oxford Economic Papers, 46*, 519–533.
Bils, M. (1987). The cyclical behavior of marginal cost and price. *American Economic Review, 77*, 838–855.
Bina, C. (1985). *The economics of the oil crisis*. New York: St. Martin.
Bishop, R. (1963). Monopolistic competition after thirty years: the impact on the general theory. *American Economic Review, Papers and Proceedings, 53*, 33–43.
Blackhouse, R. (1994). *A history of modern economic analysis*. New York: Basil Blackwell.
Blanchard, O. (1984). The Lucas critique and the Volcker deflation. *American Economic Review, 74*, 211–215.
Blanchard, O. (1987). Neoclassical synthesis. In J. Eatwell, M. Milgate, P. Newman (Eds.), *The New Palgrave: A dictionary of economics*. New York: Macmillan
Blanchard, O. (1989). A traditional interpretation of macroeconomic fluctuations. *American Economic Review, 79*, 1146–1163.
Blanchard, O., & Fisher, S. (1989). *Lectures on macroeconomics*. Cambridge, MA: MIT.
Blanchard, O., & Kiyotaki, N. (1987). Monopolistic competition and the effects of aggregate demand. *American Economic Review, 77*, 647–666.
Blanchard, O., & Quah, D. (1989). The dynamic effects of aggregate demand and aggregate supply disturbances. *American Economic Review, 79*, 655–673.
Blaug, M. (1980). *The methodology of economics*. Cambridge: Cambridge University Press.
Blaug, M. (1983). *Economic theory in retrospect*. Cambridge: Cambridge University Press.

Blaug, M. (1986). *Economic history and the history of economics*. New York: New York University Press.
Blaug, M. (1995). Competition as an end state and competition as a process. In C. Eaton & R. Harris (Eds.), *Essays in honour of Richard Lipsey*. Aldershot: Edward Elgar.
Blaug, M. (1999). Misunderstanding classical economics: the Sraffian interpretation of the surplus approach. *History of Political Economy, 31*, 213–236.
Bleaney, M. (1976). *Underconsumption theories: a history and critical analysis*. New York: International.
Bleaney, M. (1985). *The rise and fall of Keynesian economics*. London: McMillan.
Blinder, A. (1987). Keynes, Lucas and scientific progress. *American Economic Review, 77*, 130–136.
Blinder, A. (1988). The fall and rise of Keynesian economics. *Economic Record, 64*, 278–294.
Blitch, C. (1983). Alyn A. Young: a curious case of professional neglect. *History of Political Economy, 15*, 1–24.
Böhm-Bawek, E. (1891). *Positive theory of capital*. London: Macmillan.
Boianovsky, M. (2004). The IS–LM model and the liquidity trap concept: from Hicks to Krugman. *History of Political Economy, 36*, 92–126.
Bortkiewicz, L. (1907). Value and price in the Marxian system. *International Economic Papers, 2*, 5–61.
Botwinik, H. (1994). *Persistent inequalities*. Princeton: Princeton University Press.
Bronfrenbrener, M. (1969). *Is the business cycle obsolete?* New York: Wiley.
Brunner, K. (1968). The role of money and monetary policy. *Federal Reserve Bank of St Louis Review, 50*, 8–24.
Buckanan, J. (1976). Barro on the Ricardian equivalence theorem. *Journal of Political Economy, 84*, 337–342.
Burns, A. (1957). *Prosperity without inflation*. New York: Fordham University.
Cagan, P. H. (1956). The monetary dynamics of hyperinflation. In M. Friedman (Ed.), *Studies in the quantity theory of money*. Chicago: Chicago University Press.
Carr, J., & Ahiakpor, J. (1982). Ricardo on the non-neutrality of money in a world with taxes. *History of Political Economy, 14*, 147–165.
Carr, J., & Darby, M. (1981). The role of money supply shocks in the short-run demand for money. *Journal of Monetary Economics, 8*, 183–199.
Carter, A. (1981). *Structural change in the American economy*. Cambridge, MA: Harvard University Press.
Cassel, G. (1932). *Theory of social economy*. New York: Harcourt, Brace. Original work published 1918.
Chamberlin, E. (1962). *The theory of monopolistic competition* (8th ed.). Cambridge, MA: Harvard University Press. Original work published 1933.
Chamberlin, E. (1982). *Towards a more general theory of value*. Westport: Greenwood.
Champernowne, D. (1936). Unemployment, basic and monetary: the classical analysis and the Keynesian. *Review of Economic Studies, 3*, 201–216.
Chandler, A. D. (1977). *The visible hand*. Cambridge, MA: The Belknap Press.
Chatterjee, S. (1999). *Real business cycles: a legacy of countercyclical policies* (pp. 17–27). Philadelphia: Business Review, Federal Bank of Philadelphia.
Chatterjee, S. (2000). *From cycles to shocks: progress in business cycles theory* (pp. 1–11). Philadelphia: Business review, Federal Bank of Philadelphia.
Chilcote, E. (1997). Inter industry structure, relative prices, and productivity: an input–output study of the US and OECD countries, Doctoral dissertation, New School for Social Research, New York.
Christodoulopoulos, G. (1996). International competition and industrial rates of return. Unpublished doctoral dissertation, New School for Social Research, New York.
Clair, St O. (1965). *A key to Ricardo*. New York: Augustus M. Kelley.
Clapham, J. (1922). Of empty economic boxes. *Economic Journal, 32*, 305–314.

Clark, J. B. (1908). *The distribution of wealth. A theory of wages, interest and profits*. London: MacMillan.
Clifton, J. (1977). Competition and the evolution of the capitalist mode of production. *Cambridge Journal of Economics, 1*, 137–151.
Clower, R. (1965). The Keynesian counter-revolution: a theoretical appraisal. In F. Hahn & F. Brechling (Eds.), *The theory of interest rates*. London: Macmillan.
Cockshott, P., & Cottrell, A. (1997). Labour time versus alternative value bases: a research note. *Cambridge Journal of Economics, 21*, 545–549.
Cockshott, P., & Cottrell, A. (2005). Robust correlations between sectoral prices and labour values: a comment. *Cambridge Journal of Economics, 29*, 309–316.
Coddington, A. (1976). Keynesian economics: the search for first principles. *Journal of Economic Literature, 14*, 1258–1273.
Cohen, A. (1983). The laws of returns under competitive conditions: progress in microeconomics since Sraffa (1926)? *Eastern Economic Journal, 9*, 213–220.
Cohen, A., & Harcour, T. G. (2003). Whatever happened to the Cambridge capital theory controversies? *Journal of Economic Perspectives, 17*, 199–214.
Coontz, S. (1966). *Productive labour and effective demand*. New York: Augustus M. Kelley.
Corry, B. (1959). Malthus and Keynes – a reconsideration. *Economic Journal, 69*, 717–724.
Cournot, A. (1838). *Researches into the mathematical principles of the theory of wealth*. New York: AM Kelley.
Daal, J., & Jolink, A. (1993). *The equilibrium economics of Léon Walras*. London: Routledge.
Darity, W., & Young, W. (1995). IS–LM: an inquest. *History of Political Economy, 27*, 1–41.
Da Silva, E. (1992). Prices, wages, and profits in Brazil: an input–output analysis, 1975. In F. Moseley & E. Wolff (Eds.), *International perspectives on profitability*. Aldershot: Edward Elgar.
Dergiades, T. H., & Tsoulfidis, L. (2007). Estimating capacity utilization using a SVAR model: an application to the US and Canadian economies. *Economics Bulletin, 5*(4), 1–12.
De Vroey, M. (2000). IS–LM a la Hicks versus IS–LM a la Modigliani. *History of Political Economy, 32*, 293–316.
Diaz, E., & Osuna, R. (2006). Can we trust cross-sectional price-value correlation measures? Some evidence from the case of Spain. *Journal of Post Keynesian Economics, 28*, 345–363.
Dilorenzo, T. H., & High, J. (1988). Antitrust and competition, historically considered. *Economic Inquiry, 6*, 423–435.
Dobb, M. (1973). *Theories of value and distribution since Adam Smith*. Cambridge: Cambridge University Press.
Dome, T. (1992). Ricardo's theory of tax incidence: a Sraffian re-interpretation. *Cambridge Journal of Economics, 16*, 43–53.
Dome, T. (1994). *History of economic theory: a critical introduction*. Aldershot: Edward Elgar.
Domowitz, I. Hubbard, G., & Petersen, B. (1988). Market structure and cyclical fluctuations in United States manufacturing. *Review of Economics and Statistics*, *70,* 55–66.
Duménil, G. (1983). Beyond the transformation riddle: a labor theory of value. *Science and Society, 47*, 427–450.
Duménil, G., & Lévy, D. (1993). *The economics of the profit rate*. Aldershot: Edward Elgar.
Dupuit, J. (1969). On the measurement of the utility of public works (trans: K. Arrow, & T. Scitovsky) (Eds.), Readings in welfare economics. Homewood, IL: Richard D. Irwin. (Original work published 1844).
Eagly, R. (1984). *The structure of classical economic theory*. New York: Oxford University Press.
Eatwell, J. (1977). The irrelevance of returns to scale in Sraffa's analysis. *Journal of Economic Literature, 15*, 61–68.
Eatwell, J. (1980). A simple framework or the analysis of taxation, distribution and effective demand. In E. Nell (Ed.), *Growth, profits and property*. Cambridge: Cambridge University Press.
Eatwell, J. (1982). Competition. In I. Bradley & M. Howard (Eds.), *Classical and Marxian political economy: essays in memory of Ronald Meek*. London: Macmillan.

Eatwell, J. (1983a). Theories of value, output and employment. In J. Eatwell, & M. Milgate (Eds.), *Keynes's economics and the theory of value and distribution*. New York: Oxford University Press.

Eatwell, J. (1983b). Analytical foundations of monetarism. In J. Eatwell & M. Milgate (Eds.), *Keynes's economics and the theory of value and distribution*. New York: Oxford University Press.

Eatwell, J. (1987). Marginal efficiency of capital. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave: A dictionary of economics*. New York: Macmillan.

Eatwell, J. (1990). Comment. In K. Baradwaj & B. Schefold (Eds.), *Essays on Piero Sraffa*. London: Unwin-Hyman.

Eatwell, J. (2008). Returns to scale. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave dictionary of economics* (2nd ed.). London: Palgrave Macmillan.

Eatwell, J., Milgate, M., & Newman, P. (Eds.). (1987). *The New Palgrave: a dictionary of economics*. New York: Macmillan.

Eatwell, J., Milgate, M., & Newman, P. (Eds.). (1990). *The New Palgrave capital theory*. London: Macmillan.

Eatwell, J., & Panico, C. (2008). Sraffa, Piero (1898–1983). The New Palgrave dictionary of economics (2nd ed.) In S. N. Durlauf, & L. E. Blume (Eds.). London: Palgrave Macmillan.

Edgeworth, F. (1881). *Mathematical physics*. London: Kegan Paul & Co.

Eichenbaums, M. (1991). Real business cycles theory: wisdom or whimsy? *Journal of Economic Dynamics and Control, 15*, 607–626.

Eichenbaum, M., & Singleton, K. (1986). Do equilibrium real business cycle theories explain post-war U.S. business cycles? NBER Working Papers No 1932.

Einaudi, L. (1967). The physiocratic theory of taxation. In *Economic essays in honour of Gustav Cassell*. New York: Augustus M. Kelley.

Ekelund, R., & Hebert, R. (1983). *A history of economic theory*. New York: MacGraw-Hill.

Eltis, W. (1975). Adam Smith's theory of economic growth. In A. Skinner & T. Wilson (Eds.), *Essays on Adam Smith*. Oxford: Clarendon.

Eltis, W. (1984). *The classical theory of economic growth*. London: Macmillan.

Eltis, W. (1987). Classical economics. In J. Eatwell, M. Milgate, P. Newman (Eds.), *The New Palgrave: A dictionary of economics*. New York: Macmillan.

Erreygers, G. (1995). Ricardo's theory of tax incidence: a comment. *Cambridge Journal of Economics, 19,* 819–825.

Espinosa-Vega, M., & Russell, S. (1997). History and theory of the NAIRU: a critical review, economic review. *Federal Reserve Bank of Atlanta, 82*(2), 4–25.

Felipe, J., & Fisher, F. M. (2003). Aggregation in production functions: what applied economists should know. *Metroeconomica, 54*(2), 208–262.

Felipe, J., & McCombie, J. S. L. (2001). The CES production function, the accounting identity, and Occam's razor. *Applied Economics, 33*(10), 1221–1232.

Felipe, J., & McCombie, J. S. L. (2005). How sound are the foundations of the aggregate production function. *Eastern Economic Journal, 31*(3), 467–488.

Felipe, J., & McCombie, J. S. L. (2006). The tyranny of the identity: growth accounting revisited. *International Review of Applied Economics, 20*(3), 283–299.

Ferguson, C., & Gould, J. (1975). *Microeconomic theory*. Homewood, IL: R.D. Irwin.

Fisher, I. (1973). I discovered the Phillips curve: a statistical relation between unemployment and price changes. *Journal of Political Economy, 81*(2), 496–502. Reprinted from 1926 edition of International Labour Review.

Fisher, S. (1977). Long-term contracts, rational expectations, and the optimal money supply rule. *Journal of Political Economy, 85*, 191–206.

Flaschel, P., & Semmler, W. (1990). On composite classical and Keynesian microdynamic adjustment processes. In N. Christodoulakis (Ed.), *Dynamic modelling and control of national economies*. Oxford: Pergamon.

Foley, D. (1986). *Understanding capital, Marx's economic theory*. Cambridge, MA: Harvard University Press.
Foley, D. (2000). *Notes on the theoretical foundations of political economy*. Mimeo, New York: New School University.
Foley, D. (1982). The value of money, the value of labor power and the transformation problem. *Review of Radical Political Economics, 14*(2), 37–47.
Fonseca, G., & Ussher, L. (2009). *The history of economic through website*. New York: New School for Social Research.
Fontana, G. (2009). Whither new consensus macroeconomics? The role of government in fiscal policy in modern macroeconomics. The Levy Economics Institute, Working Paper No 563.
Freeman, A., & Carchedi, G. (1996). *Marx and non-equilibrium economics*. Cheltenham: Edward Elgar.
Friedman, M. (1953). *Essays on positive economics*. Chicago: Chicago University Press.
Friedman, M. (1956). The quantity theory of money, a restatement. In M. Friedman (Ed.), *The quantity theory of money*. Chicago: Chicago University Press.
Friedman, M. (1957). *A theory of the consumption function*. Princeton: Princeton University Press.
Friedman, M. (1959). *A program for monetary stability*. New York: Fordham University Press.
Friedman, M. (1962). *Capitalism and freedom*. Chicago: University of Chicago Press.
Friedman, M. (1968). The role of monetary policy. *American Economic Review, 58*, 1–17.
Friedman, M. (1973). *Money and economic development*. New York: Praeger.
Friedman, M. (Ed.). (1976). *The quantity theory of money*. Chicago: Chicago University Press.
Friedman, M. (1984). Monetary policy for the 1980s. In J. Moore (Ed.), *To promote prosperity: U.S. domestic policy in the mid-1980s*. Glenn Campbell, Stanford: Hoover Institution Press.
Friedman, M., & Friedman, R. (1980). *Free to choose*. New York: Harcourt Brace and Jovanovicz.
Friedman, M., & Schwartz, A. (1963). *A monetary history of the United States, 1867–1960*. Princeton: Princeton University Press.
Friedman, M., & Schwartz, A. (1982). *Monetary trends in the United States and the United Kingdom*. Chicago: Chicago University Press.
Garegnani, P. (1976). On a change in the notion of equilibrium in recent work on value and distribution. In M. Brown, K. Sato, & P. Zarembka (Eds.), *Essays in modern capital theory*. Amsterdam: North Holland.
Garegnani, P. (1978–1979). Notes on consumption, investment and effective demand, I & II. Cambridge Journal of Economics, 2, 335–353; 3, 63–82. Reprinted in Eatwell, & Milgate (Eds.) (1983).
Garegnani, P. (1983). The classical theory of wages and the role of demand schedules in the determination of relative prices. *American Economic Review, 73*, 309–313.
Garegnani, P. (1984). Value and distribution in classical economists and Marx. *Oxford Economic Papers, 36*, 193–226.
Garegnani, P. (1990). Quantity of capital. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave series: capital theory* (pp. 1–78). London: Macmillan.
Garegnani, P. (1998). Sraffa: the theoretical world of the old classical economists. *The European Journal of the History of Economic Thought, 5*, 415–429.
Gehrke, C., & Kurz, H. (2006). Sraffa on von Bortkiewicz: reconstructing the classical theory of value and distribution. *History of Political Economy, 38*(1), 91–149.
George, H. (1879). *Progress and Poverty: An Inquiry into the Cause of Industrial Depressions and of Increase of Want with Increase of Wealth*. New York: Robert Schalkenbach Foundation.
Giacomin, A. (1995). *Il Mercato e il Potere*. Bologna: CLUEB.
Glick, M. (1985). Monopoly or competition in the US economy. *Review of Radical Political Economics, 17*, 121–127.
Glyn, A., Hughes, A. Lipietz, A., & Sinh, A. (1990). The rise and fall of the golden age. In Maglin, S., & Schor, J. (eds.), *The Golden Age of Capitalism*. New York: Oxford Universiry Press.
Goldfeld, S. (1976). The case of the missing money. *Brookings Papers on Economic Activity, 3*, 683–730.

Goodhart, C. (1975). Problems of monetary management: the U.K. experience. In A. Courakis (Ed.) (1981), *Inflation, depression and economic policy in the West*. Totowa, NJ: Barnes and Nobles.
Gordon, D. (1974). A neoclassical theory of Keynesian unemployment. *Economic Inquiry, 12*, 431–459.
Gordon, R. J. (1972). Wage-price controls and the shifting Phillips curve. *Brooking Papers on Economic Activity, 3*, 385–421.
Gordon, R. J. (1982). Inflation, flexible exchange rates and the natural rate of unemployment. In M. Baily (Ed.), *Workers, jobs and inflation*. Washington, DC: The Brookings Institution.
Gordon, R. J. (1990). What is new-Keynesian economics? *Journal of Economic Literature, 28*(3), 1115–1171.
Gordon, R. J. (1993). Measuring the aggregate price level: implications for economic performance and policy. In K. Shigehara (Ed.), *Price stabilization in the 1990s*. London: Macmillan.
Green, R. (1982). Money output and inflation in classical economics. *Contributions to Political Economy, 1*, 59–85.
Green, R. (1992). *Classical theories of money, output and inflation: a study in historical economics*. London: Macmillan.
Greenwald, A., & Stiglitz, J. (1993a). Financial market imperfections and business cycles. *The Quarterly Journal of Economics, 108*, 77–113.
Greenwald, A., & Stiglitz, J. (1993b). New and old Keynesians. *Journal of Economic Perspectives, 7*, 23–44.
Hahn, F. (1977). Keynesian economics and general equilibrium theory. In G. Harcourt (Ed.), *Reflections on the current debates in the microfoundations of macroeconomics*. London: MacMillan.
Hall, R. (1988). The relation between price and marginal cost in U.S. industry. *Journal of Political Economy, 96*, 921–947.
Hall, R. (1993). Market structure and macroeconomic fluctuations. In G. Mankiw, & D. Romer (Eds.), *New Keynesian economics*. Cambridge, MA: MIT University Press
Hall, R., & Hitch, C. (1939). Price theory and business behaviour. *Oxford Economic Papers, 2*, 12–45.
Han, J., & Schefold, B. (2006). An empirical investigation of paradoxes: reswitching and reverse capital deepening in capital theory. *Cambridge Journal of Economics, 30*(5), 737–765.
Hansen, A. (1953). *A guide to Keynes*. New York: McGraw Hill.
Harberger, A. (1954). Monopoly and resource allocation. *American Economic Review, 44*, 77–87.
Harcourt, G. (1972). *Some Cambridge controversies in the theory of capital*. Cambridge: Cambridge University Press.
Harris, D. (1988). On the classical theory of competition. *Cambridge Journal of Economics, 12,* 139–167.
Hart, N. (2003). Marshall's dilemma: equilibrium versus evolution. *Journal of Economic Issues, 37*, 1139–1160.
Hartley, J., Hoover, K., & Salyer, K. (1998). *Real business cycles: a reader*. London: Routledge.
Harvey, J. (1996). Orthodox approaches to exchange rate determination: a survey. *Journal of Post Keynesian Economics, 18*, 567–583.
Hayes, M. (2008). Keynes's degree of competition. *European Journal of the History of Economic Thought, 15*, 275–291.
Heckscher, E. (1955). *Mercantilism*. London: Allen & Unwin. Original work published 1931.
Heilbroner, R. (1975). The paradox of progress: decline and decay in the wealth of nations. In A. Skinner & T. Wilson (Eds.), *Essays on Adam Smith*. Oxford: Oxford University Press.
Heilbroner, R. (1980). *Marxism for and Against*. New York: Norton.
Heilbroner, R. (1981). *The worldly philosophers*. New York: Simon and Schuster.
Heilbroner, R. (1983). The problem of value in the constitution of economic thought. *Social Research, 50*, 253–277.
Heilbroner, R. (1985). *The Nature and Logic of Capitalism.* New York: W. W. Norton.

Heilbroner, R., & Milberg, W. (1996). *The crisis of vision in modern economic thought*. Cambridge, MA: Cambridge University Press.
Hicks, J. (1937). Mr Keynes and the classics. *Econometrica, 5*, 147–159.
Hicks, J. (1939). *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*. Oxford: Clarendon Press, 1946.
Hicks, J. (1957). A rehabilitation of 'classical' economics? *Economic Journal, 67*, 278–289.
Hicks, J. (1983). IS–LM: an explanation. In J. P. Fitoussi (Ed.), *Modern macroeconomic theory*. New York: Barnes and Noble.
Hollander, S. (1971). The development of Ricardo's position on machinery. *History of Political Economy, 1*, 105–135.
Hollander, S. (1973). *The economics of Adam Smith*. Toronto: Toronto University Press.
Hollander, S. (1979). *The economics of David Ricardo*. Toronto: Toronto University Press.
Hollander, S. (1985). On the substantive identity of the Ricardian and neoclassical conceptions of economic organization: the French connection in British classicism. In G. Caravale (Ed.), *The legacy of Ricardo*. New York: Basil Blackwell.
Hollander, S. (1989). On P. Mirowski's "Physics and the marginal revolution". *Cambridge Journal of Economics, 13*, 459–470.
Hollander, S. (1991). Marx and the falling rate of profit. In G. Caravale (Ed.), *Marx and modern economic analysis*. Aldershot: Edward Elgar.
Howitt, P. (1987). Macroeconomics: relations with microeconomics. In J. Eatwell, M. Milgate, & P., Newman (Eds.), *The New Palgrave: a dictionary of economics*. New York: Macmillan. (Vol. 3, pp. 273–276).
Hunt, E. (1992). *History of economic thought: a critical perspective*. New York: Harper and Collins.
Hunt, S. (2000). *A general theory of competition. Resources, competences, productivity, economic growth*. London: Sage.
Jevons, S. (1911). *The theory of political economy*. London: Macmillan. Original work published 1871.
Judd, J., & Scadding, J. (1982). The search for a stable money demand function: a survey of the post-1973 literature. *Journal of Economic Literature, 20*, 993–1023.
Kahn, R. (1989). *The economics of the short period*. London: Macmillan. Original work published 1929.
Kahn, R. (1931). The relation of home investment to unemployment. *Economic Journal, 41*, 173–198.
Kahn, R. (1977). Malinvaud on Keynes. *Cambridge Journal of Economics, 1*, 375–388.
Kaldor, N. (1935). Market imperfection and excess capacity. Economica, 2, 35–50. In N. Kaldor (Ed.) (1980) Economic essays, I, Essays on value and distribution. New York: Holmes & Meier.
Kaldor, N. (1956). Alternative theories of distribution. *Review of Economic Studies, 23*, 83–100.
Kaldor, N. (1970). The new monetarism. *Lloyds Bank Review, 97*, 1–18.
Kalecki, M. (1937). A theory of the business cycle. *Review of Economic Studies, 4*, 77–97.
Kalecki, M. (1969). *Theory of economic dynamics: an essay on cyclical and long-run changes in capitalist economy*. New York: Augustus M. Kelley.
Kaskarelis, I., & Tsoulfidis, L. (1999). Inflation, demand, imports competition and markups in greek manufacturing industries. *Economia Internazionale, 52,* 151–171.
Keynes, J. (1971). Thomas Robert Malthus: the first of the Cambridge economists. In D.E. Moggridge (ed.), *The collected writings of John Maynard Keynes*. London: Macmillan. (Vol. 10, pp. 71–108). (Original work published 1933).
Keynes, J. (1936). *The general theory of employment, interest, and money*. New York: Harcourt, Brace and Jovanovich.
Keynes, J. (1937a). The general theory of employment. *Quarterly Journal of Economics, 51*, 209–223.
Keynes, J. (1937b). Alternative theories of the rate of interest. *Economic Journal, 47*, 241–252.

Keynes, J. (1973). The general theory of employment. In *The collected writings of John Maynard Keynes, Vol. 14: the general theory and after, part II: defence and development*. London: Royal Economic Society.
Keynes, J., & Henderson, H. (1929). *Can Lloyd George do it?* (9th ed.). D.E. Moggridge (ed.), Reprinted in *The Collected writings of John Maynard Keynes*. London: Macmillan.
Kenway, P. (1980). Marx, Keynes and the possibility of crisis. *Cambridge Journal of Economics, 4*, 23–36.
King, R., & Plosser, C. (1984). Money, credit and prices in a real business cycle model. *American Economic Review, 74*, 363–380.
Kirzner, I. (1987). Competition: Austrian conceptions. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave*. London: Macmillan.
Klein, L. (1963). *The Keynesian revolution*. New York: Macmillan.
Kliman, A., & McGlone, T. (1988). The transformation non-problem and the non-transformation problem. *Capital and Class, 35*, 56–83.
Kliman, A. (2002). The law of value and laws of statistics: sectoral values and prices in the US economy, 1977–1997. *Cambridge Journal of Economics, 26*, 299–311.
Kliman, A. (2004). Spurious value–price correlations: some additional evidence and arguments. *Research in Political Economy, 21*, 223–239.
Knight, F. (1921). *Risk uncertainty and profits*. Boston: Houghton Mifflin.
Kondratiev, N. (1999). *The works of N. Kondratiev*. London: Pickering & Chato.
Krelle, W. (1976). Basic facts in capital theory. Some lessons for the controversy in capital theory. *Review d' Economie Politique, 87*, 283–329.
Kreps, D. (1990). *A course in microeconomic theory*. Princeton: Princeton University Press.
Kuczynski, M., & Meek, R. (1972). *Quesnay's Tableau Economique*. London: Macmillan.
Kurz, H. (1990a). Accumulation, distribution and the 'Keynesian hypothesis'. In K. Baradwaj & B. Schefold (Eds.), *Essays on Piero Sraffa*. London: Unwin-Hyman.
Kurz, H. (1990b). Debates in capital theory. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Capital theory, the New Palgrave*. New York: W.W. Norton.
Kurz, H. (1998). Marx on Technological Change: The Ricardian Heritage. In R. Bellofiore (ed.), *Marxian Economics. A Reappraisal, vol. 2. Houndmills,* Basingstoke and London: Macmillan, 119–138.
Kurz, H. (2007). Capital. In *International encyclopedia of social sciences* (2nd ed., pp. 424–439). London: MacMillan.
Kurz, H. (2010). Technical Progress, Capital Accumulation and Income Distribution in Classical Economics: Adam Smith, David Ricardo and Karl Marx. *European Journal of the History of Economic Thought* (forthcoming).
Kurz, H., & Salvadori, N. (1992). Morishima on Ricardo: A Review Article. *Cambridge Journal of Economics. 16,* 227–247.
Kurz, H., & Salvadori, N. (1995). *Theory of production*. Cambridge: Cambridge University Press.
Kurz, H., & Salvadori, N. (2002). Mark Blaug on the Sraffian interpretation of the surplus approach. *History of Political Economy, 34*, 237–240.
Kyndland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica, 50*, 1345–1370.
Laidler, D. (1984). The buffer stock notion in monetary economics. *Economic Journal, 94*, 17–33.
Landreth, H., & Colander, D. (1993). *History of economic theory*. Boston: Houghton-Mifflin.
Layard, R., & Walters, A. (1978). *Microeconomic theory*. New York: MacGrow Hill.
Lavoie, M. (2006). *Introduction to post-Keynesian economics*. New York: Palgrave Macmillan.
Leijonhufvud, A. (1968). *On Keynesian economics and the economics of Keynes*. Oxford: Oxford University Press.
Leontief, W. (1939). The structure of the American Economy, *1919–1939*. New York: Oxford University Press (2d ed. p. 1951).
Leontief, W. (1986). *Input–output economics*. Oxford: Oxford University Press.

Levhari, D., & Samuelson, P. (1966). The nonswitching theorem is false. *Quarterly Journal of Economics, 80*, 518–519.
Lindbeck, A., & Snower, D. (1986). Wage setting, unemployment, and insider-outsider relations. *American Economic Review, 76*, 235–239.
Lipsey, R. (1960). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862–1957: a further analysis. *Economica, 27*, 1–31.
Litterman, B., & Weiss, L. (1985). Money, real interest rates and output: a reinterpretation of postwar U.S. data. *Econometrica, 53*, 129–155.
Long, J., & Plosser, C. (1983). Real business cycles. *Journal of Political Economy, 91*, 1345–1370.
Löwe, A. (1975). Adam Smith's system of equilibrium growth. In A. Skinner & T. Wilson (Eds.), *Essays on Adam Smith*. Oxford: Oxford University Press.
Lucas, R. (1972). Expectations and the neutrality of money. *Journal of Economic Theory, 4*, 103–124.
Lucas, R. (1973). Some international evidence on output-inflation trade offs. *American Economic Review, 63*, 326–334.
Lucas, R. (1975). An equilibrium model of the business cycle. *Journal of Political Economy, 83*, 1113–1144.
Lucas, R. (1976). Econometric policy evaluation: a critique. *Journal of Monetary Economics, 1*, 19–46.
Lucas, R. (1980). The death of Keynesian economics: issues and ideas, 18–19, University of Chicago, Winter.
Lucas, R. (1981). *Studies in business-cycle theory*. Cambridge, MA: MIT.
Lucas, R. (1987). *Models of business cycles*. Oxford: Basil Blackwell.
Lucas, R. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 12*, 3–42.
Lucas, R. (1994). Review of Milton Friedman and Anna Schwartz's 'a monetary history of the United States, 1876–1960'. *Journal of Monetary Economics, 34*, 5–16.
Lucas, R., & Rapping, L. (1969). Price expectations and the Phillips curve. *American Economic Review, 59*, 342–350.
Lucas, R., & Rapping, L. (1972). Unemployment in the great depression: is there a full explanation? *Journal of Political Economy, 80*, 6–91.
Lucas, R., & Sargent, T. (1978). *After the Phillips curve: Persistence of high inflation and high unemployment*. Boston, MA: Federal Resrve Bank of Boston.
Lucas, R., & Sargent, T. (Eds.). (1981). *Rational expectations and econometric practice*. New York: George Allen and Unwin.
Lynn, A. (1977). Adam Smith's fiscal ideas: an eclectic revisited. *National Tax Journal, 29*, 369–378.
Maital, S. (1972). The Tableau Economique as a simple Leontief model: an amendment. *Quarterly Journal of Economics, 86*, 504–507.
Makridakis, S., Wheelwright, S., & Hyndman, R. (1998). *Forecasting: methods and applications*. New York: Wiley.
Malanos, G. (1946). The evolution of the general theory, Doctoral dissertation, Harvard University, MA.
Malinvaud, E. (1977). *The theory of unemployment reconsidered*. Oxford: Blackwell.
Malthus, T. (1827). Definitions in political economy, London. Reprinted in E. Wrigley, & D. Souden (Eds.) (1986) The works of Thomas Robert Malthus (Vol. VIII). London: Pickering & Chatto.
Mankiw, G. (1990). A quick refresher course in macroeconomics. *Journal of Economic Literature, 28*, 1645–1660.
Mankiw, G. (1997). *Macroeconomics*. New York: Worth.
Mankiw, G. (1998). Recent developments in macroeconomics. A very quick refresher course. *Journal of Money, Credit and Banking, 20*, 436–449.

Mankiw, G., & Romer, D. (Eds.). (1993a). *New Keynesian economics*. Cambridge, MA: MIT University Press.
Mankiw, G., & Romer, D. (1993b). Introduction. In G. Mankiw, & D. Romer (Eds.), *New Keynesian economics*. Cambridge, MA: MIT University Press.
Marcuzzo, M. C. (2001). Sraffa and Cambridge economics, 1928–1931. In T. Cozzi & R. Marchionatti (Eds.), *P. Sraffa's political economy. A centenary estimate* (pp. 81–99). London: Routledge.
Marchionatti, R. (2003). On the methodological foundations of modern microeconomics: Frank Knight and the cost controversy in the 1920s. *History of Political Economy, 35*, 49–75.
Mariolis, T. H. (2004). Pure joint production and international trade: a note. *Cambridge Journal of Economics, 28*, 449–456.
Mariolis, T. H., & Tsoulfidis, L. (2009). Decomposing the changes in production prices into 'capital-intensity' and 'price' effects: theory and evidence from the Chinese economy. *Contributions to Political Economy, 28*, 1–22.
Marshall, A. (1890). *Principles of economics*. London: MacMillan.
Marx, K. (1847). *Wage labour and capital*. Moscow: International.
Marx, K. (1859). *A contribution to the critique of political economy*. New York: International, 1981.
Marx, K. (1857). *Grundrisse, foundations of the critique of political economy* (p. 1973). New York: Vintage.
Marx, K. (1867). *Capital* (1st ed., p. 1977). New York: International.
Marx, K. (1885). *Capital* (2nd ed., p. 1977). New York: International.
Marx, K. (1894). *Capital* (3rd ed., p. 1977). New York: International.
Marx, K. (1969). *Theories of surplus value, parts I, II and III*. Moscow: Progress.
Marx, K., & Engels, F. (1955). *Selected correspondence*. Moscow: Progress.
Mattick, P. (1969). *Marx and Keynes. The limits of the mixed economy*. Boston: Horizon.
McCloskey, D. (1985). *The rhetoric of economics*. Madison: The Winsconsin University Press.
McLellan, D. (1973). *Karl Marx: his life and thought*. London: MacMillan.
McLellan, D. (1978). *Friedrich Engels*. London: Penguin.
McNulty, P. (1967). Economic theory and the meaning of competition. *Quarterly Journal of Economics, 82*, 639–656.
Meacci, F. (1985). Ricardo's chapter on machinery and the theory of capital. In G. Caravale (Ed.), *The legacy of Ricardo*. Oxford: Basil Blackwell.
Meade, J. (1937). A simplified model of Keynes's system. *Review of Economic Studies, 4*, 98–107.
Meek, R. (1962). *The economics of physiocracy: essays and translations*. London: Allen & Unwin.
Meek, R. (1973). *Studies in the labour theory of value*. London: Lawrence and Wishart.
Mehring, F. (1936). Karl Marx: the story of his life (trans: Fitzgerald, E.). London: George Allen & Unwin.
Meltzer, A. (1963). The demand for money: the evidence from time series. *Journal of Political Economy, 71*, 219–246.
Milgate, M. (1977). Keynes on the 'classical' theory of interest. *Cambridge Journal of Economics, 1*, 307–315.
Mill, J. S. (1848). *Principles of political economy*. Fairfield, New Jersey: Augustus M. Kelley.
Miller, R., & Blair, P. (1985). *Input–output analysis: foundations and extensions*. New Jersey: Prentice Hall.
Miller, R. (1957). A note on some Walrasian geometry. *Zeitschrift für Nationalökonomie, XIX*, 315–317.
Minskin, F. (1982). Does anticipated monetary policy matter? An econometric investigation. *Journal of Political Economy, 90*, 22–51.
Minsky, H. (1975). *John Maynard Keynes*. New York: Columbia University Press.
Mirowski, P. (1984). Physics and the Marshallian revolution. *Cambridge Journal of Economics, 8*, 361–379.

Mirowski, P. (1989). On Hollander's 'substantive identity' of classical and neoclassical economics: Mirowski on Hollander on Mirowski. *Cambridge Journal of Economics, 13*, 471–477.
Modigliani, F. (1944). Liquidity preference and the theory of interest and money. *Econometrica, 12*, 45–88.
Modigliani, F. (1977). The monetarist controversy: or should we forsake stabilization policies? *American Economic Review, 67*, 1–19.
Modigliani, F., & Papademos, L. (1975). Targets for monetary policy in the coming years. *The Brookings Papers on Economic Activity, 1*, 141–156.
Moggridge, D. (1992). *Maynard Keynes. An economist's biography*. London: Routledge.
Mohun, S. (1994). A re(in)statement of the labor theory of value. *Cambridge Journal of Economics, 18*, 391–412.
Mohun, S. (2004). The labour theory of value as foundation for empirical investigations. *Metroeconomica, 55*, 65–95.
Morishima, M. (1973). *Marx's economics: a dual theory of value and growth*. Cambridge: Cambridge University Press.
Morishima, M. (1977). *Walras's economics*. Cambridge: Cambridge University Press.
Morishima, M. (1989). *Ricardo's economics: a general equilibrium theory of distribution and growth*. Cambridge: Cambridge University Press.
Morishima, M., & Catephores, G. (1978). *Value, exploitation and growth: Marx in the light of modern economic theory*. London: McGraw-Hill.
Moseley, F. (2000). The "new solution" to the transformation problem: a sympathetic critique. *Review of Radical Economics, 32*, 282–316.
Moss, S. (1983). The history of the theory of the firm from Marshall to Robinson and Chamberlin: the source of positivism in economics. *Economica, 51*, 307–318.
Muellbauer, J., & Portes, R. (1978). Macroeconomic models with quantity rationing. *Economic Journal, 88*, 788–821.
Muth, J. (1961). Rational expectations and the theory of price movement. *Econometrica, 29*, 315–335.
Niehans, J. (1990). *A history of economic theory: classic contributions, 1720–1890*. Baltimore: Johns Hopkins University Press.
Nelson, C., & Plosser, C. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics, 10*, 139–162.
Ochoa, E. (1989). Values, prices and wage-profit curves in the U.S. economy. *Cambridge Journal of Economics, 13*, 413–430.
Ochoa, E., & Glick, M. (1992). Competing microeconomic theories of industrial profits: an empirical approach. In W. Milberg (Ed.), *The megacorp and macrodynamics: essays in memory of Alfred Eichner*. New York: Sharpe.
O'Driscoll, G. (1977). The Ricardian nonequivalence theorem. *Journal of Political Economy, 85*, 207–210.
Okishio, N. (1961). Technical change and the rate of profit. *Kobe University Economic Review, 7*, 86–96.
Pagan, A. (2003). *Report on modelling and forecasting of the Bank of England*. Bank of England: Mimeo.
Paglin, M. (1973). *Malthus and Lauderdale: the anti-Ricardian tradition*. New Jersey: A.M. Kelley.
Paitaridis, D., & Tsoulfidis, L. (2009) Smith's theory of the falling rate of profit. Revisited. Paper presented at the European society for the history of economic thought 13th conference in Thessaloniki.
Pajis, Ph. Van (1980). The falling rate of profit theory of crisis: A rational reconstruction by way of obituary. *Review of Radical Political Economy,* 12, 1–16.
Papageorgiou, A., & Tsoulfidis, L. (2006). Kondratiev, Marx and the Long cycle. *Indian Development Review*. Reprinted in Mariolis, & Tsoulfidis (Eds.), Distribution, development *and prices, critical perspectives*. New Delhi: Serials.

Palley, T. H. (1998). Walras' law and Keynesian macroeconomics. *Australian Economic Papers, 37*, 330–340.
Panico. (1987). Interest and profit. In J. Eatwell, M. Milgate, P. Newman (Eds.) *The New Palgrave: A dictionary of economics*. New York: Macmillan.
Papadimitriou, D., & Wray, R. (1998). The economic contributions of Hyman Minsky. *Review of Political Economy, 10*, 199–226.
Parkin, M. (1986). The output-inflation trade off when prices are costly to change. *Journal of Political Economy, 94*, 200–224.
Pasinetti, L. (1966). Changes in the rate of profit and switches of techniques. *Quarterly Journal of Economics, 80*, 503–517.
Pasinetti, L. (1974). *Growth and Income Distribution*. Cambridge: Cambridge University Press.
Pasinetti, L. (1977). *Lectures on the theory of production*. New York: Columbia University Press.
Pasinetti, L. (2003). *Continuity and change in Sraffa's thought: an archival excurses*. USA: Mimeo.
Patinkin, D. (1990). On different interpretations of the general theory. *Journal of Monetary Economics, 26*, 205–243.
Peach, T. (1993). *Interpreting Ricardo*. Cambridge: Cambridge University Press.
Penrose, E. (1959). *The theory of the growth of the firm*. New York: Wiley.
Petrovic, P. (1987). The deviation of production prices from labor values: some methodological and empirical evidence. *Cambridge Journal of Economics, 11*, 197–210.
Phelps, E. (1972). *Inflation policy and unemployment theory*. London: MacMillan.
Phelps, E. (1992). A review of unemployment. *Journal of Economic Literature, 30*, 1476–1490.
Phillips, A. (1955). The Tableau Economique as a simple Leontief model. *Quarterly Journal of Economics, 69*, 137–144.
Phillips, A. (1958). The relation between unemployment and the rate of change in money wage rates in the United Kingdom, 1861–1957. *Econometrica, 25*, 283–299.
Pollin, R., & Justice, C. (1994). Savings, finance, and interest rates: an empirical consideration of some basic Keynesian propositions. In G. Dymski & R. Pollin (Eds.), *New perspectives in monetary macroeconomics*. Michigan: University of Michigan.
Porter, M. (1990). *The competitive advantage of nations*. New York: Free.
Plosser, C. (1989). Understanding real business cycles. *Journal of Economic Perspective, 3*, 51–77.
Prasch, R., & Sheth, F. (1999). The economics and ethics of minimum wage legislation. *Review of Social Economy, 57*, 488–506.
Prescott, E. (1986). Theory ahead of business cycles measurement. *Federal Reserve Bank of Minneapolis, Quarterly Review, 10*, 9–21.
Pressman, S. (1994). *Quesnay's Tableau Economique: a critique and reconstruction*. New York: Kelley.
Reid, G. (1989). *Classical economic growth. An analysis in the tradition of Adam Smith*. London: Basil Blackwell.
Reinwald, T. (1977). The genesis of Chamberlin's monopolistic competition theory. *History of Political Economy, 9*, 522–534.
Ricardo, D. (1951a). On the principles of political economy and taxation. In P. Sraffa & M. Dobb (Eds.), *The works and correspondence of David Ricardo* (1st ed.). Cambridge: Cambridge University Press.
Ricardo, D. (1951b). The works and correspondence of David Ricardo. In P. Sraffa, & M. Dobb (Eds.) (Vol. IV). Cambridge: Cambridge University Press.
Ricardo, D. (1951c). The works and correspondence of David Ricardo. In P. Sraffa, & M. Dobb (Eds.) (Vol. IX). Cambridge: Cambridge University Press.
Robertson, D. (1930). The trees and the forest. *Economic Journal, 40*, 80–89.
Robinson, J. (1933). *The economics of imperfect competition*. London: Macmillan.
Robinson, J. (1953). The production function and the theory of capital. *Review of Economic Studies, 21*, 81–106.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature, 34*, 647–668.
Roncaglia, A. (1983). Piero Sraffa and the reconstruction of political economy. *Banca Nazionale del Lavoro Quarterly Review, 36*, 337–350.
Rosdolsky, R. (1977). *The making of Marx's capital*. London: Pluto.
Rotemberg, J. (1987).The new Keynesian microfoundations. NBER Macroeconomics Annual, 69–104.
Rotemberg, J., & Woodford, M. (1991). Markup and the business cycle. NBER Macro Annual, 63–128.
Rubin, I. (1928). *A history of economic thought* (p. 1979). London: Ink Links.
Rubin, I. (1972). *Essays on Marx's theory of value*. Detroit: Black and Red.
Samuelson, P. (1948). *Foundations of economic analysis*. New York: McGraw Hill.
Samuelson, P. (1955). *Economics* (3rd ed.). New York: McGraw-Hill.
Samuelson, P. (1962). Parable and realism in capital theory: the surrogate production function. *Review of Economic Studies, 29*, 193–206.
Samuelson, P. (1966). A summing-up. *Quarterly Journal of Economics, 80*, 568–583.
Samuelson, P. (1978). The canonical classical model of political economy. *Journal of Economic Literature, 16,* 1415–1434.
Samuelson, P. (1982). Quesnays Tableau Economique as a theorist would formulate it today. In I. Bradley & H. Michael (Eds.), *Classical and Marxian political economy: essays in honour of Ronald L. Meek*. London: Macmillan.
Samuelson, P. (1986). The monopolistic competition revolution. In R. Merton (Ed.), *Collected scientific papers of Paul A. Samuelson* (3rd ed.). Cambridge, MA: MIT. Original work published 1967.
Samuelson, P. (1990). Revisionist findings on Sraffa. In K. Baradwaj & B. Schefold (Eds.), *Essays on Piero Sraffa*. London: Unwin-Hyman.
Samuelson, P., & Solow, R. (1960). Analytical aspects of anti-inflationary policy. *American Economic Revie, Papers and Proceedings, 50*, 177–194.
Sargent, T. (1979). A note on maximum likelihood estimation of the rational expectations model of the term structure. *Journal of Monetary Economics, 5*, 133.
Sargent, T., & Wallace, N. (1975). Rational expectations, the optimal monetary instrument and the optimal money supply rule. *Journal of Political Economy, 83*, 241–254.
Scherer, M., & Ross, D. (1990). *Industrial market structure and economic performance*. Boston: Houghton and Mifflin.
Schefold, B. (1976). Different forms of technical progress. *Economic Journal, 86*, 806–809.
Schefold, B. (1978). Kahn on Malinvaud. *Cambridge Journal of Economics, 2*, 407–408.
Schefold, B. (2008). Families of strongly curved and of nearly linear wage curves: a contribution to the debate about the surrogate production function. *Bulletin of Political Economy, 2*, 1.
Schumpeter, J. (1942). *Capitalism socialism and democracy*. New York: Harper and Row.
Schumpeter, J. (1954). *History of economic analysis*. New York: Oxford University Press.
Schwartzman, D. (1960). The burden of monopoly. *The Journal of Political Economy, 68*, 627–630.
Seligman, E. (1969). *Essays in taxation.* New York: Agustus Kelley.
Screpanti, E., & Zamagni, S. (2002). *An outline of the history of economic thought*. New York: Oxford University Press.
Semmler, W. (1984). *Competition monopoly and differential profit rates: on the relevance of the classical and Marxian theories of production prices for modern industrial and corporate pricing*. New York: Columbia University Press.
Shackle, G. (1967). *Years of high theory*. Cambridge: Cambridge University Press.
Shaikh, A. (1973). *Theories of value and theories of distribution.* Ph.D. Dissertation. New York: Columbia University.
Shaikh, A. (1974). Laws of algebra and laws of production: the humbug production function. *Review of Economics and Statistics, 56*, 115–120.

Shaikh, A. (1977). Marx's theory of value and the transformation problems. In J. Schwartz (Ed.), *The subtle anatomy of capitalism*. Santa Monica: Goodyear.
Shaikh, A. (1978). Political economy and capitalism: notes on Dobb's theory of crisis. *Cambridge Journal of Economics, 2*, 233–251.
Shaikh, A. (1980a). The laws of international exchange. In E. Nell (Ed.), *Growth, profits and property*. Cambridge: Cambridge University Press.
Shaikh, A. (1980b). Marxian competition versus perfect competition: further comments on the so-called choice of technique. *Cambridge Journal of Economics, 4*, 75–83.
Shaikh, A. (1982). Neoricardian economics: a wealth of algebra a poverty of theory. *Review of Radical Political Economics, 14*, 67–83.
Shaikh, A. (1984). The transformation from Marx to Sraffa. In A. Freeman & E. Mandel (Eds.), *Ricardo, Marx and Sraffa*. London: Verso.
Shaikh, A. (1990). The humbug production function. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave series: capital theory*. London: Macmillan.
Shaikh, A. (1992). The falling rate of profit as the cause of long waves: Theory and empirical evidence. In Kleinknecht, A., Mandel, E., & Wallerstein, I. (Eds.), *New Findings in Long Wave Research*. London: Macmillan.
Shaikh, A. (1995). The stock market and the corporate sector: a profit based approach. Working Paper No 146. New York: Jerome Levy Institute.
Shaikh, A. (1996). Free trade, unemployment and economic policy. In J. Eatwell (Ed.), *Global unemployment: loss of Jobs in the 90s*. Armonk, New York: M.E. Sharpe.
Shaikh, A. (1998). The empirical strength of the labour theory of value. In R. Bellofiore (Ed.), *Marxian economics: a reappraisal* (Vol. 2, pp. 225–251). New York: St. Martin.
Shaikh, A. (1999). Explaining the global economic crisis: a critique of Brenner. *Historical Materialism, 5*, 103–144.
Shaikh, A. (2003). *Nonlinear dynamics and pseudo-production functions*. Mimeo, New York: New School for Social Research.
Shaikh, A. (2008). Competition and industrial rates of return. In Arestis, & Eatwell. (Eds.). Issues in finance and industry essays in honour of Ajit Singh. Houndmills: Palgrave Macmillan.
Shaikh, A., & Antonopoulos, R. (1999). Explaining long-term exchange rate behavior in the United States and Japan. Working Paper No 250. The Levy Economics Institute.
Shaikh, A., & Tonak, A. (1994). *Measuring the wealth of nations. The political economy of national accounts*. Cambridge: Cambridge University Press.
Shapiro, N. (1997). Imperfect competition and Keynes. In G. C. Harcourt & P. A. Riach (Eds.), *A 'second edition' of the general theory* (Vol. 1, pp. 83–92). London: Routledge.
Shoup, C. (1960). *Ricardo on taxation*. New York: Columbia University Press.
Shoven, J., & Whalley, J. (1992). *Applying general equilibrium*. New York: Cambridge University Press.
Silberberg, E. (1978). *The structure of economics: a mathematical analysis*. New York: MacGraw Hill.
Skidelsky, R. (1992). *John Maynard Keynes: Vol. 2 – the economist as saviour, 1920–1937*. London: Macmillan.
Slutsky, E. (1937). The summation of random causes as the source of cyclical processes. *Econometrica, 5*, 19–60.
Smith, A. (1937). The wealth of nations. In E. Cannan (Ed.). New York: Random House. (Original work published 1776).
Snowdown, B., & Vane, H. (Eds.). (1997). *A macroeconomic reader*. London: Routledge.
Snowdown, B., & Vane, H. (1998). Transforming macroeconomics: an interview with Robert E. Lucas Jr. *Journal of Economic Methodology, 5*, 115–146.
Snowdown, B., Vane, H., & Wynarczyk, P. (1994). *A modern guide to macroeconomics: an introduction to competing schools of economic thought*. Aldershot: Edward Elgar.
Solow, R. (1956). A contribution to economic growth. *Quarterly Journal of Economics, 70*, 65–94.

Solow, R. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics, 39*, 312–320.
Solow, R. (1974). Laws of production and laws of algebra: the humbug production function. *Review of Economics and Statistics, 56*, 115–120.
Solow, R. (2010). Stories on technological change. *European Journal of the History of Economic Thought* (forthcoming).
Sraffa, P. (1925). Sulle Relazioni fra Costo e Quantita Prodotta. *Annali di Economia, 2*, 277–328 (trans: A. Roncaglia, & J. Eatwell).
Sraffa, P. (1926). The laws of returns under competitive conditions. *Economic Journal, 36*, 535–550.
Sraffa, P. (1930). A criticism – a rejoinder. *Economic Journal, 40*, 89–93.
Sraffa, P. (1960). *Production of commodities by means of commodities: a prelude to economic theory*. Cambridge: Cambridge University Press.
Stadler, G. (1994). Real business cycles. *Journal of Economic Literature, 32*, 1750–1783.
Steedman, I. (1977). *Marx after Sraffa*. London: New Left.
Steedman, I. (1982). Marx on Ricardo. In I. Bradley & M. Howard (Eds.), *Classical and Marxian political economy*. London: Macmillan.
Steedman, I., & Tomkins, J. (1998). On measuring the deviation of prices from values. *Cambridge Journal of Economics, 22*, 379–385.
Steenge, A. (2000). The rents problem in the Tableau Economique: revisiting the Phillips model. *Economic Systems Research, 12*, 181–197.
Stigler, G. (1937). A generalization of the theory of imperfect competition. *Journal of Farm Economics, 19*, 707–717.
Stigler, G. (1949). *Five Lectures on Economic Problems*. New York: Longmans, Green Co.
Stigler, G. (1957). Perfect competition historically contemplated. *Journal of Political Economy, 65*, 1–16.
Stigler, G. (1958). Ricardo and the 93% labor theory of value. *American Economic Review, 48*, 357–367.
Stigler, G. (1966). *The theory of price*. New York: Macmillan. Original work published 1942.
Stigler, G. (1983). Monopolistic competition in retrospect. Reprinted in The organization of industry. Chicago: Chicago University Press. (Original work published 1949).
Stiglitz, J. (1984). Price rigidities and market structure. *American Economic Review, 74,* 350–355.
Summers, L. (1986). Some sceptical observations on real business cycle theory. *Federal Reserve Bank of Minneapolis, Quarterly Review, 10*, 23–27.
Sweezy, P. (1939). Demand under conditions of oligopoly. *Journal of Political Economy, 47*(4), 568–573.
Sweezy, P. (1942). *The theory of capitalist development: principles of Marxian political economy*. New York: Monthly Review.
Sylos-Labini, P. (1990). Sraffa's critique of the Marshallian theory of prices. In K. Bharadwai, & B. Schefold (Eds.), *Essays on Pierro Sraffa. Critical perspectives on the revival of classical theory*. London: Unwin Hyman.
Taylor, J. (1979). Staggered wage setting in a macromodel. *American Economic Review, 69*, 100–13. Reprinted in Mankiw and Romer (eds.), Vol. 1.
Taylor, J. (1989). Monetary policy and the stability of macroeconomic relationships. *Journal of Applied Econometrics, 4*, 161–178.
Taylor, L. (1990). *Relevant policy analysis structuralist computable general equilibrium models for the developing world*. Cambridge, MA: MIT.
Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies, 25*, 15–29.
Tobin, J. (1970). Friedman's theoretical framework. In R. J. Gordon (Ed.), *Milton Friedman's monetary framework*. Chicago: Chicago University Press.
Tobin, J. (1980). Are new classical models plausible enough to guide policy? *Journal of Money Credit and Banking, 12*, 788–799.

Tobin, J. (1993). Price flexibility and output stability: an old Keynesian view. *Journal of Economic Perspectives, 7*, 45–65.
Tsaliki, P., & Tsoulfidis, L. (1998). Alternative theories of competition: evidence from the Greek manufacturing industries. *International Review of Applied Economics, 12*, 187–204.
Tsoulfidis, L. (1989). The physiocratic theory of tax incidence. *Scottish Journal of Political Economy, 36*, 301–310.
Tsoulfidis, L. (1993). On Ricardo's theory of taxation. *Spoudai, 43*, 111–127.
Tsoulfidis, L. (1998). Ricardo's theory of value and Marx's critique. *History of Economic Ideas, 2*, 69–88.
Tsoulfidis, L. (2000). Corn laws. In *International encyclopedia of political economy*. London: Routledge.
Tsoulfidis, L. (2004) A *history of economic theory and policy (in Greek)*. Thessaloniki: University of Macedonia Press.
Tsoulfidis, L. (2007). Classical economists and public debt. *International Review of Economics, 54*, 1–12.
Tsoulfidis, L. (2008a). Price–value deviations: further evidence from input–output data of Japan. *International Review of Applied Economics, 22*, 707–724.
Tsoulfidis, L. (2008b). Capacity output. In *International encyclopaedia of social sciences* (2nd ed.). London: Macmillan.
Tsoulfidis, L. (2008c). Capital controversy. In *International encyclopaedia of social sciences* (2nd ed.). London: Macmillan.
Tsoulfidis, L. (2009). The rise and fall of monopolistic competition revolution. *International Review of Economics, 56*(1), 29–45.
Tsoulfidis, L., & Maniatis, T. H. (2002). Values, prices of production and market prices: some more evidence from the Greek economy. *Cambridge Journal of Economics, 26*, 359–369.
Tsoulfidis, L., & Tsaliki, P. (2005). Marxian Theory of Competition and the Concept of Regulating Capital: Evidence from the Greek Manufacturing. *Review of Radical Political Economics, 37,* 5–22.
Tsoulfidis, L., & Rieu, D.-M. (2006). Labor values, prices of production, and wage-profit rate frontiers of the Korean economy. *Seoul Journal of Economics, 19*, 275–295.
Tsoulfidis, L., & Mariolis, T. H. (2007). Labour values, prices of production and the effects of income distribution. Evidence from the Greek economy. *Economic Systems Research, 19*, 425–437.
Tsoulfidis, L., & Paitaridis, D. (2009). On the labor theory value: statistical artefacts or regularities? *Research in Political Economy, 21*, 223–239.
Tullio, G. (1987). Smith and Ricardo on the long run effects of the growth of government expenditure, taxation, and debt: is relevant today? *History of Political Economy, 21*, 723–737.
Van Parijs, Ph. (1980). The falling rate of profit theory of crisis. A rational reconstruction by way of obituary. *Review of Radical Economics, 12*, 1–16.
Von Neumann, J. (1945). A model of general economic equilibrium. *Review of Economic Studies, 13*, 1–9.
Veneziani, R. (2004). The temporal single-system interpretation of Marx's economics: a critical evaluation. *Metroeconomica, 55*, 96–114.
Vercelli, A. (1999). The evolution of IS–LM models: empirical evidence and theoretical presuppositions. *Journal of Economic Methodology, 6*, 199–219.
Viner, J. (1931). Cost curves and supply curves. Zeitschrift für Nationalökonomie. Reprinted in Stigler, G., & Boulding, K. (Eds.) (1952), *Readings in price theory*. Chicago: Richard D. Irwin.
Viner, J. (1968). Mercantilist thought. In Sills. (Ed.), International encyclopedia of social sciences New York: Macmillan.
Walras, L. (1954). *Elements of pure economics*. Homewood, IL: Irwin. Original work published 1874.
Warming, J. (1932). International difficulties arising out of the financing of public works during depression. *Economic Journal, 42*, 211–224.
Weeks, J. (1981). *Capital and exploitation*. Princeton: Princeton University Press.

Wheen, F. (2001). *Karl Marx. A life*. New York: Norton.
Wicksell, K. (1934–1935). *Lectures on political economy* (I, IIth ed.). London: Routledge and Kegan Paul.
Wicksteed, P. (1914). The scope and method of political economy in the light of the "Marginal theory of value and distribution". *Economic Journal, 24*, 1–23.
Woodford, M. (2003). *Interest and prices: foundations of a theory of monetary policy*. Princeton: Princeton University Press.
Woodford, M. (2008). *Convergence in macroeconomics: elements of the new synthesis*. New York: Mimeo.
Yellen, J. (1984). Efficiency wage models of unemployment. *American Economic Review, 74*, 200–205.
Young, A. (1928). Increasing returns and economic progress. *Economic Journal, 38*, 527–542.

Index