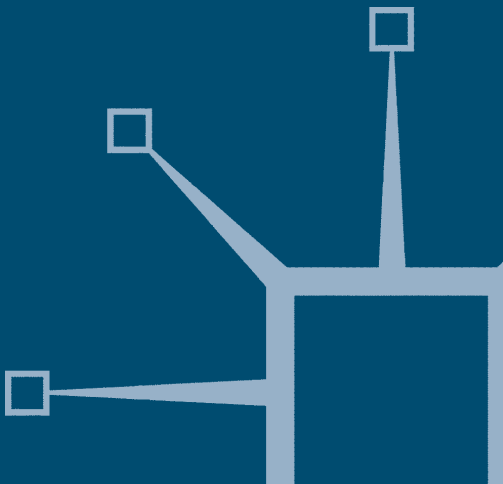
palgrave
macmillan

Modelling Non-Stationary Time Series

A Multivariate Approach

Simon P. Burke and John Hunter



Modelling Non-Stationary Time Series

Palgrave Texts in Econometrics

Series Editor: **Kerry Patterson**

Titles include:

Simon P. Burke and John Hunter
MODELLING NON-STATIONARY TIME SERIES

Michael P. Clements
EVALUATING ECONOMETRIC FORECASTS OF ECONOMIC AND FINANCIAL VARIABLES

Terence C.Mills
MODELLING TRENDS AND CYCLES IN ECONOMIC TIME SERIES

Kerry Patterson
UNIT ROOTS IN ECONOMIC TIME SERIES

Jan Podivinsky
MODELLING VOLATILITY

Palgrave Texts in Econometrics
Series Standing Order ISBN 1–4039–0172–4 Hardcover
Series Standing Order ISBN 1–4039–0173–2 Paperback
(*outside North America only*)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBN quoted above.

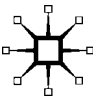
Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England.

Modelling Non-Stationary Time Series:

A Multivariate Approach

Simon P. Burke and John Hunter

palgrave
macmillan



First published 2005 by
PALGRAVE MACMILLAN
Houndmills, Basingstoke, Hampshire RG21 6XS and
175 Fifth Avenue, New York, N. Y. 10010
Companies and representatives throughout the world

PALGRAVE MACMILLAN is the global academic imprint of the Palgrave Macmillan division of St. Martin's Press, LLC and of Palgrave Macmillan Ltd. Macmillan® is a registered trademark in the United States, United Kingdom and other countries. Palgrave is a registered trademark in the European Union and other countries.

ISBN-10: 1–4039–0202–X hardback
ISBN-13: 978–1–4039–0202–3 hardback
ISBN-10: 1–4039–0203–8 paperback
ISBN-13: 978–1–4039–0203–0 paperback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Burke, Simon P.
Modelling non-stationary economic time series : a multivariate approach / by Simon P. Burke and John Hunter.
p. cm. – (Palgrave texts in econometrics)
Includes bibliographical references and index.
ISBN 1–4039–0202–X (cloth)–ISBN 1–4039–0203–8 (pbk.)
1. Econometric models. 2. Time-series analysis. I. Title. II. Series.

HB141.B866 2004
330¢.01¢51955–dc22 2004056896

10 9 8 7 6 5 4 3 2 1
14 13 12 11 10 09 08 07 06 05

Printed and bound in Great Britain by
Antony Rowe Ltd, Chippenham and Eastbourne

Contents

Preface

This book deals with an analysis of non-stationary time series that has been very influential in applied research in econometrics, economics and finance. The notion that series are non-stationary alters the way in which series are grouped and may even prove to be relevant to some aspects of regulation and competition policy when the definition of market becomes an economic issue. The later might also apply to any discussion of the nature of globalized financial markets. In terms of econometric and statistical theory an enormous literature has grown up to handle the behaviour of the different forms of persistence and non-stationary behaviour that economic and financial data might exhibit. This is emphasized by the Nobel Prize that has been presented to Clive Granger and Robert Engle in relation to their extension of our understanding of the way in which non-stationary series behave. However, the requirement to analyze non-stationary behaviour has spawned a wide range of approaches that relate to and interrelate with the notion that series are non-stationary and/or cointegrated.

It has been our privilege to in some part be involved in these developments and to have learned much from our colleagues and teachers alike. We must acknowledge our debt of gratitude to those who taught us and supervised us over the years. We would also like to thank participants at various Econometrics Society Conferences, EC^2, Econometrics Study Group Conferences held at the Burwells campus of Bristol University and participants in the Econometrics workshop for their incites, comments and stimulating research. Dr. Lindsey Anne Gillan also provided us with some guidance through the potential minefied that is academic publishing. However, all errors are our own.

Simon P. Burke
John Hunter

1
Introduction: Cointegration, Economic Equilibrium and the Long Run

The econometrician or statistician might be viewed as a forensic scientist, trying to detect from the splatter (of blood), a line through space from which it may be determined, how and by whom a crime was committed. The tools available to calculate and describe this evidence are estimators and tests, and then – conditional on the model selected – identification of the cause or the perpetrator of the crime.

At the very core of econometrics lies measurement, the quality of measurement and the existence of the measure. When a measure is considered then there is the practical question of whether measurement is feasible or not. Conventional statistical measurement and inference considered the behaviour of processes that are associated with distributions that are generally viewed as being fixed across the sample. When economists started to apply statistical measurement to economic data then the notion that the data were identically and independently distributed (IID) had to be rejected. Regression was used to measure the heterogeneity by estimating a mean conditional on exogenous information while the assumption that the data are independently and identically distributed (IID), was used to give structure to the unknown error in the model. Essentially some form of least squares regression became the method generally applied to explain economic phenomena, but in the early literature it is hard to find reference to the notion of non-stationarity. One exception is the book written by Herman Wold with Lars Jureen on the subject of demand analysis, which does consider the behaviour of stationary economic time series. However, Wold and Jureen (1953) analyzed data for the inter-war years, a period when price series fell in relative terms and growth of output was relatively stagnant. Hence, any question of how demand models might be derived when time series are non-stationary was apart from some exceptions ignored. It is of interest to note that, in a study of the demand for food, James Tobin estimated both a logarithmic inverse demand curve and in an attempt to remove serial correlation the same relationship in differences. The latter

equation became the basis of the Rotterdam model developed by Theil (1965) and Barten (1969). In the early 1970s, Box and Jenkins wrote a book that became highly influential in the statistical analysis of time series data. Box and Jenkins set out a methodology for building time series models, that firstly considers the appropriate degree of differencing required to render a series stationary, and then discusses the type of alternative models autoregressive (AR) or moving average (MA), or ARMA that might be used to estimate univariate time series and then considered the method of estimation. Fama (1970) suggests that the observation that financial time series follow random walks is consistent with the idea that markets were efficient. The random walk model implies that financial time series are non-stationary and, following Box and Jenkins, need to be differenced to make them stationary. The difference in the log of the share price approximates a return and when the financial market is efficient then returns are not supposed to be predictable.

The structure of time series models pre-dates Box and Jenkins. Yule (1927) first estimated AR processes and in 1929 Kolmogorov considered the behaviour of sums of independent random variables (see the discussion in Wold and Jureen (1953)). In the regression context, Sargan (1964) applied an MA error structure to a dynamic model of UK wage inflation. The Sargan model became the basis of most of the UK wage equations used in the large macroeconomic models (Wallis et al. 1984). In demand analysis, approximation rather than non-stationarity was behind differencing and developments in economic theory related to the structure of demand equations was more interested in issues of aggregation as compared with the possible time series structure of the data (Deaton and Muellbauer 1980). To difference time series became common practice in modelling univariate time series and this approach was also applied in finance where it was common to consider returns of different assets rather than share prices. The market model relates the return on a share to the return on the market. There was now a discrepancy between the methods applied in statistics and finance to time series data and the approach predominantly used by economists.

However, the first oil shock precipitated a crisis in macroeconomic model building. Most of the world's large macroeconomic models were unable to resolve many of the problems that ensued from this shock. Forecasts and policy simulations that provide the governments' predictions of the future and a practical tool for understanding the impact of policy on the economy were unable to explain what had happened and what policies might remedy the situation (Wallis et al. 1984). The UK Treasury's inability to forecast the balance of payments position led to the ludicrous situation of a developed economy being forced to borrow from the IMF – a remedy that would not have been sought had reasonable estimates been available of the true payments position. The whole approach to the econometric modelling of economic time series was in doubt.

Econometric modelling was criticized on three grounds – the specification of the models used, their forecast accuracy and their existence. The model building approach adopted at the London School of Economics (LSE) built on the methodology developed by Sargan (1964). The Sargan approach attempted to combine the lessons of conventional time series modelling by applying the difference operator to the dependent variable with the practical requirement of the economist that the model could be solved back to reveal relationships from which the levels of the data might be forecast. The LSE approach implied that economic time series were dynamic and best modelled as regressions that included an appropriate description of the dynamic process underlying the data. The approach reinforced the proposition that a valid regression was required to satisfy the Gauss–Markov conditions (Patterson 2000) and that any regression models estimated ought to be well specified. This became what has been called the Hendry methodology and in the UK and Europe this approach has provided a potent mechanism to generate reasonable approximations to many aggregate economic time series. In particular, the articles by Davidson et al. (1978) and Hendry and Mizon (1978) expound a single equation modelling methodology for consumption and money. Davidson et al. (1978) emphasize that correct specification follows from estimating general autoregressive distributed lag (ADL) models, states that the dynamic model explains the short-run behaviour of the stationary form of the data in differences, that any levels variables explain the long run and that the long run is associated with conventional economic theory. Hendry and Richard (1982, 1983) elaborated on these ideas further by explaining what an adequate approximation of the data is and how systems models are sequentially reduced into valid sub-models. The final important development that came out of this approach was the categorization of exogeneity into strict, weak, strong and super. As far as inference and the estimation of single equation regression models is concerned, weak exogeneity justified the use of contemporaneous variables such as income in consumption and money equations.

The LSE approach provided model builders with a methodology for estimating single equations by regression. Poor forecast performance was viewed as a sign of a poorly performing model and was viewed then as, correctable by valid model selection. In the US the failure of econometric model building was viewed as a failure of economic theory. Forecasts based on large macro models broke down because the postwar Keynesian consensus had broken down and the basis of failure was neoclassical monetary neutrality combined with hyper-rational agent behaviour. The Lucas critique suggested that the conventional macro models were unable to capture changes in agent responses to government policy, the deep parameters of the economic system. Models based on classical assumptions purported to show that monetary policy was not effective, while the notion that macroeconomic time series followed random walks was embedded in the article by Robert Hall (1978),

which showed that consumption followed a random walk.[1] In 1978 Sargent derived dynamic models based on rational expectations, which impose theoretical propositions about the underlying behaviour of agents on the short-run behaviour of the data. However, Sargent explicitly requires that the series are stationary for the solution to exist.[2] The literature derived from the neoclassical rational expectations solution to macro modelling has adopted two approaches to the problem of model specification. The first is to build dynamic models with data that are differenced and then to solve the expectations problem or to estimate the models using an unrestricted vector autoregressive (VAR) model. The former approach often uses Generalized Method of Moments to estimate the Euler equation via the errors in variables approach best explained by Wickens (1982). While Sims (1980) proposed the VAR methodology, first differencing the data to render it stationary and then estimating economic behaviour by systems of autoregressive models, suggesting that all the variables modelled are endogenous. Policy invariance is tested by looking at impulse responses and causal structure, rather than by deriving structural models.[3]

The LSE methodology assumed that long-run relationships existed and that conventional inference was valid irrespective of whether series are stationary or not. The rational expectations literature that transformed the data into differences risked the possibility that there may be over-differencing. Both approaches understood that time series modelling required dynamic models, the former assuming that conventional economic theory can be detected in terms of long-run relationships from the data, the latter approach that it cannot be. The idea that a correlation is not valid is best explained in Yule (1926) who considers a number of correlations that can only be viewed as nonsense. In particular, Yule found that the fall in Church of England marriages was positively correlated with the fall in the death rate between 1861 and 1913. This idea of nonsense correlation along with many of the problems associated with econometric modelling, including the appropriate measurement of expectations, was discussed by Keynes (1939).[4] Keynes emphasizes the role of economics in statistical model building and explains that economists need to be looking at true causes as compared with correlations that derive from the dependence of variables on an underlying primary cause. In 1974 Granger and Newbold presented simulation results for nonsense regressions – relationships that are observed to be correlated, but cannot be. Granger and Newbold (1986) describe how univariate and multivariate economic time series ought to be modelled. Simulations presented in Granger and Newbold (1974, 1986) show that it is possible to run regressions on unrelated data and find significant relationships where there should be none. The 1974 article suggests that the discovery of an R^2 that exceeds the Durbin Watson (DW)

statistic ought to be indicative of the problem as then the DW statistic has to be less than one and as a result the model must suffer from significant serial correlation. The article appears to emphasize that badly misspecified models should be viewed with deep suspicion, because they may reveal relationships that are spurious. It is apparent that the econometrics profession had adopted this research agenda by building on one side of the Atlantic ADL models and on the other VARs in differences. However, the results associated with Granger and Newbold (1986) were somewhat subtler, in that when the data were generated via random walks with MA errors, spurious regressions could be observed with DW statistics in excess of one. Hence, the question of what determines a true regression relationship is further complicated by the existence of more complex explanations of individual time series.

This book considers methods by which it can be determined whether time series are stationary or non-stationary in differences, difference stationary or trend stationary or rendered stationary by subtracting from the non-stationary series some part of another series. The latter case is the cointegration case, which occurs when two or more series combine to produce stationary variables and a conventional regression equation between these variables has economic meaning in a long-run sense. This notion of cointegration is then developed in the context of multiple time series. A conclusion for the VAR methodology in differences is that when long-run behaviour exists, in terms of combinations of stationary variables in levels, the VAR is fundamentally misspecified. However, the generalization of the ADL to a system, can under the restrictions associated with cointegration provide a short-run explanation of the data, with long-run behaviour explained by restrictions on the levels in each equation.

In chapter 2, the characteristics of economic and financial time series are considered. The properties of the variance, covariance and autocovariance of stationary and non-stationarity time series are defined, in addition to the alternative definitions of stationarity. Time series models are defined for both their stationary and non-stationary representations. The statistical properties of the error are defined in terms of white noise residuals and the Wold decomposition. Non-invertibility, random walks and alternative notions of persistence are dealt with, as, before time series are modelled, they ought to be stationary. The proposition that a series is stationary needs to be tested and the data transformed to take account of non-stationarity or persistence. Having decided on the stationary form of the data, a time series model can be identified and estimated. Much of the existing literature handles persistence by first or second differencing data. The former is often appropriate for real variables such as output or employment, while second differences might often be required for nominal variables in economic models, GDP, sales and retail

prices or in finance, share prices, stock indices and dividends. Otherwise, fractional differencing might be required, with the resulting models being special cases of the autoregressive fractionally integrated moving average (ARFIMA) model.

In chapter 3, modelling non-stationary time series is handled in a single equation framework. When more than one series is analyzed, differencing might be more than is required. This occurs when series in combination are stationary (cointegration). Non-integer differencing is often required, in the case of series such as interest rates. Single equation models, which incorporate some different right-hand side variables in levels, are classified as error correction models. When the original data or their logarithms are non-stationary, cointegration may be observed when linear combinations of two or more levels variables are stationary. Then cointegration is valid when the relationships are bivariate or there is one cointegrating relationships in a system. When the regressors are exogenous, in a univariate time series context, the regressions can be viewed as ARMAX or ARMA models with exogenous variables.

In chapter 4, the multivariate time series model is developed from a stationary representation of the data that is known always to exist, the vector or VMA model in differences. The book explains the nature of multivariate time series under stationarity and then extends this to the cointegration case. We then explain how the VMA in differences can be transformed into an error correction model using the Granger representation theorem and the Smith–McMillan form developed by Yoo (1986). Cointegration is then described in terms of error correcting VARs or VECMs. A procedure for determining the existence of the VAR is described along with the Johansen approach to estimation and inference. The book explains the asymptotic theory that lies behind the Johansen test statistic. An application is developed based on the models of the UK effective exchange rate estimated by Hunter (1992), Johansen and Juselius (1992) and Hunter and Simpson (1995). Finally a number of alternative representations are developed and the question of multi-cointegration discussed.

In chapter 5, the exogeneity of variables in the VAR and the identification of long-run parameters are considered. Exogeneity is discussed in terms of the restrictions required for weak, strict and cointegrating exogeneity in the long run. Then alternative forms of exogeneity and causality are considered and the results associated with Hunter (1992) and Hunter and Simpson (1995) are presented. Identification is discussed in terms of conventional systems with I(0) series, this approach is extended to show when the parameters can be identified via imposing the restrictions and solving out for the long-run parameters and their loadings. Identification is then discussed in terms of the results derived by Bauwens and Hunter (2000), Johansen (1995) and Boswijk

(1996). All three approaches are applied to the model estimated by Hunter (1992).

In chapter 6, more advanced topics are considered in some detail. Firstly, the I(2) case, firstly using an extention to the Sargan–Bézout approach adopted by Hunter (1994), then in terms of the representation and test due to Johansen (1992) and Paruolo (1996), and finally the test procedures due to Johansen and Paruolo are applied to the exchange rate data in Hunter (1992). Fractional cointegration is briefly discussed in terms of the estimator due to Robinson and Marinucci (1998) and the test due to Robinson and Yajima (2002). Secondly, forecasting of non-stationary and stationary components is considered. The results produced by Lin and Tsay (1996) and Clements and Hendry (1995, 1998) are presented with a graphical analysis of the performance of the simulations developed by Lin and Tsay (1996). Finally, models with short-run structural equations are discussed – in particular, models with unit roots in the endogenous and exogenous processes. It is shown how to estimate models where the unit roots relate to the endogenous variables and then to the case associated with the exogenous variables.

In chapter 7, the reader is guided to further issues in the literature. Firstly, a plethora of articles on testing stationarity and non-stationarity has developed; the reader is directed where appropriate to the book by Patterson (2005). A condensed discussion of structural breaks is provided along with direction to appropriate references.

2
Properties of Univariate Time Series

2.1 Introduction

This chapter introduces a number of concepts in the analysis of univariate time series that are important for an understanding of non-stationarity in the multivariate case. The fundamental building block is the autocorrelation structure of a time series. This describes the way in which current and past values of a time series are related to one another. Capturing the main characteristics of these relationships can be thought of as the primary task of a time series model: to provide theoretical structures the properties of which closely approximate those of observed time series, and to provide estimates of such models using specific time series that can be used to draw inferences about other aspects of behaviour.

Linear models designed to capture the leading properties of autocorrelation structures, namely the autoregressive and moving average models, define a set of structures for which generic concepts, especially non-stationarity, have very specific but simply stated implications. The discussion below begins by discussing autocorrelation and non-stationarity in fairly general terms. It moves on to describe how these properties can be reasonably approximated by univariate autoregressive moving average models, and lastly to how, they can be used to test for a limited form of non-stationarity. The treatment throughout is univariate.

2.2 Non-stationarity

2.2.1 Time series structure: autocorrelation

There are various aspects to the idea of stationarity and so to non-stationarity. A general definition may be very difficult to exploit in practice. A practical definition has to be precise, but will be more prescriptive, dealing with a limited set of situations relevant to the problem at hand. The characteristics of the set of problems dealt with in this book relate to the fact that the data are

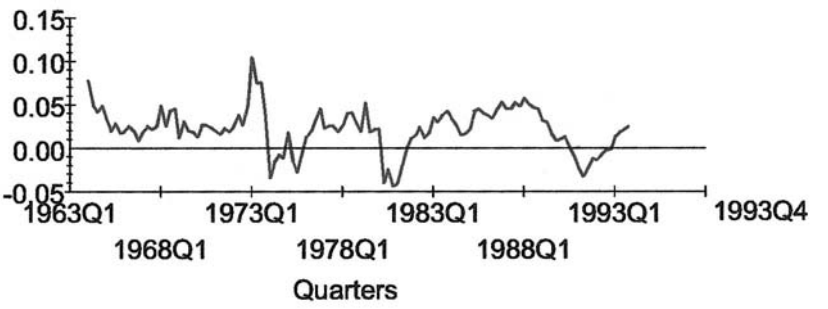


Figure 2.1 UK annual rate of growth of real output, quarterly, 1963Q1–1993Q4, T = 84

time series and that it is the temporal dependence between elements of these series that is of concern. Furthermore, the dependence will be considered at a relatively simple level: that of covariance. This last point does not matter if the distribution being used is the normal (or Gaussian) distribution, since this distribution is characterized entirely by its mean and variance and covariance.

Consider Figure 2.1. This shows the time series plot of the annual rate of growth of UK real output from 1963 to 1993. Its characteristics are that it varies around a more or less fixed level, that it does not drift away from this level for any great length of time, and that higher values at some point in time tend to be followed by other high values, or at least that changes from the high values or are often smooth. The same applies for low values, followed by low values or changing relatively smoothly.[1] The controlled variability around a fixed level is a manifestation of stationarity. The relationship between neighbouring values can be described by autocorrelation – literally, the quantification of the correlation between values in the time series separated by fixed periods of time. A type of stationarity can be defined in terms of the autocorrelation and mean of a time series. This is a restricted but very useful and practical definition.

In theory, the individual observations comprising the time series are thought of as realizations of underlying random variables. The autocorrelation of a time series is defined in terms of these underlying random variables as follows. Let X_t $t = 1, 2, \ldots$ be a sequence of scalar random variables, one for each equally spaced point in time, t, but otherwise referring to the same random variable, X. Such a sequence may (loosely) be called a stochastic process.[2] Let $E(.)$ be the expectation operator.

2.2.1.1 Autocovariance and autocorrelation

The autocovariance between two random variables at different points in time is their covariance, and is given by

$$\gamma_x(j) = E[X_t - E(X_t))(X_{t-j} - E(X_{t-j}))]. \qquad (2.1)$$
$$j = \ldots, -2, -1, 0, 1, 2, \ldots$$

The autocorrelation is the correlation between the two random variables. Noting that the variance of the process is given by

$$Var\,(X_t) = E[X_t - E(X_t))^2] = \gamma_x(0)$$

the autocorrelation is given by

$$\rho_x(j) = \frac{\gamma_x(j)}{\gamma_x(0)}. \qquad (2.2)$$
$$j = \ldots,\ -2,\ -1,\ 0,\ 1,\ 2,\ \ldots$$

Being a correlation, it follows that

$$-1 \le \rho_x(j) \le 1$$

making it a useful basis on which to compare time series.

The sequence of autocovariances and autocorrelations obtained as j, the time gap between random variables changes, are often referred to as functions. That is, (2.1) is called the autocovariance function and (2.2) the autocorrelation function (abbreviated to ACF).

2.2.2 Stationarity

The definitions of autocovariance and autocorrelation have been written to indicate that they depend only on the time gap, not the point in time. That is, for example, considering two different points in time, t and $t - j$,

$$E[(X_t - E(X_t))(X_{t-j} - E(X_{t-j}))] = \gamma_x(j)$$

and

$$E[(X_\tau - E(X_\tau))(X_{\tau-j} - E_{\tau-j}))] = \gamma_x(j)$$

even though $t \neq \tau$. But the time gap, j, is the same so they have the same autocovariance. This is an assumption consisting of two components.

It is assumed that the expected value, or mean of the time series does not change over time, so that for any $t \neq \tau$.

$$E(X_t) = E(X_\tau). \qquad (2.3a)$$

It is also assumed that, given the mean is constant, the autocovariance between equally separated random variables does not change. As a special case of this last assumption, the variance does not alter over time so that

$$\gamma_x(0) = E[(X_t - E(X_t))^2] = E[X_\tau - E(X_\tau))^2] \qquad (2.3b)$$

from which it follows that the autocorrelations depend only on the time gap, not on the time itself. The assumption that these quantities remain fixed over time is a fundamental aspect of stationarity, and goes most of the way to

providing a practical definition of stationarity for the purposes of time series analysis.

The expectation of the process, such as in equation (2.3a), is referred to as the first moment, the (co)variance as the second moment (about the mean). Thus, in equations (2.3a) and (2.3b), it has been assumed that the first two moments of the process are constant over time. Some definitions stop at this point and use the constancy of these moments to define covariance stationarity.[3] The definition used here will add one clarification: that these moments must be finite. (This is only a clarification because the definition obviously requires the moments to exist, but if infinite, they do not exist.) This is a common addition, see for example Banerjee et al. (1993, p. 11). The definition of covariance stationarity used in this book is at the beginning of the next section.

2.2.2.1 Covariance stationarity[4]

The sequence of random variables X_t, $t = 1, 2, \ldots$ is said to be covariance stationary if, for all t,

$$E(X_t) = \mu, \ |\mu| < \infty,$$
$$Var(X_t) = \sigma^2 < \infty,$$
$$E[(X_t - E(X_t))(X_{t-j} - E(X_{t-j}))] = \gamma_x(j), \ |\gamma_x(j)| < \infty.$$

Under the assumption of covariance stationarity it is meaningful to estimate the autocorrelation function in the following way. Let x_t, $t = 1, 2, \ldots, T$ be the observations on a time series. Then the autocovariance function may be estimated as

$$\hat{\gamma}_x(j) = \frac{\sum_{t=j+1}^{T}(x_t - \bar{x})(x_{t-j} - \bar{x})}{T - j}$$

and the ACF as

$$\hat{\rho}_x(j) = \frac{\hat{\gamma}_x(j)}{\hat{\gamma}_x(0)} \tag{2.4}$$

where $\bar{x}$ is the sample mean, $\bar{x} = \sum_{t=1}^{T} x_t/T$. Equation (2.4) is referred to as the sample ACF. The sample ACF for the UK output growth data is presented in Figure 2.2.

Figure 2.2 has two leading characteristics. The sample autocorrelations damp off over time, that is they decline towards zero as the time gap, or lag (j), gets larger. There is a degree of oscillation, so that the autocorrelations start off positive, then decline to zero, but go through zero before returning to

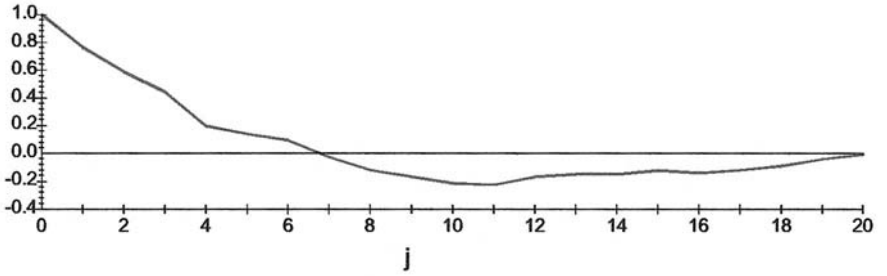


Figure 2.2 Sample ACF for rate of growth of UK real output, quarterly, 1963–1993

this level. Of course, zero indicates an absence of association between the values concerned.[5] So, in this case, after the gap between observations reaches 20 (5 years) there is no discernible relationship between values. If confidence limits can be added to the sample ACF, it may well turn out that any statistically significant relationship dies out earlier than 5 years.

The damping off of the sample ACF is an empirical characteristic of covariance stationary time series. In the case of Figure 2.1, it suggests that, while there is a relationship between temporally close values (say up to gaps of one year), values separated by a greater length of time are not much correlated. As a special case of this, the values of the time series observations are not dependent on the initial value, x_1, because if this was so, autocorrelations at very large time gaps would remain high.

Of course, the sample ACF can be calculated using equation (2.4) whether or not the conditions for covariance stationarity apply. Consider the \$/£ exchange rate data plotted in Figure 2.3. The characteristics of this time series plot are markedly different from those of Figure 2.1. Apart from the fact that

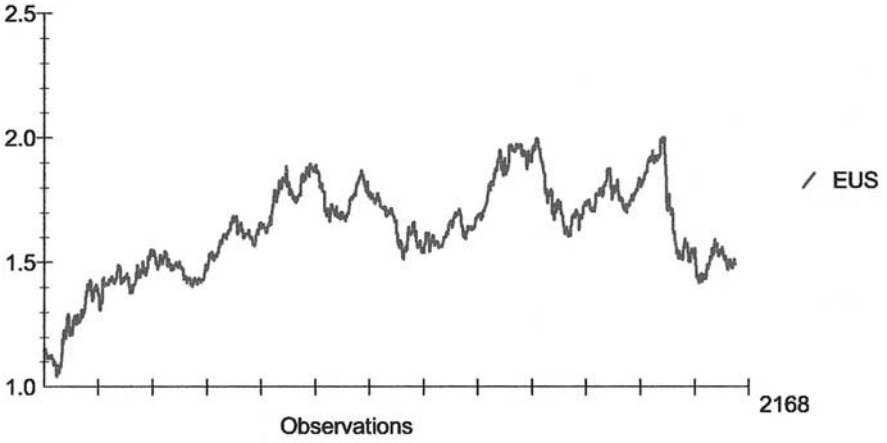


Figure 2.3 Daily \$/£ exchange rate, January 1985–July 1993, T = 2168

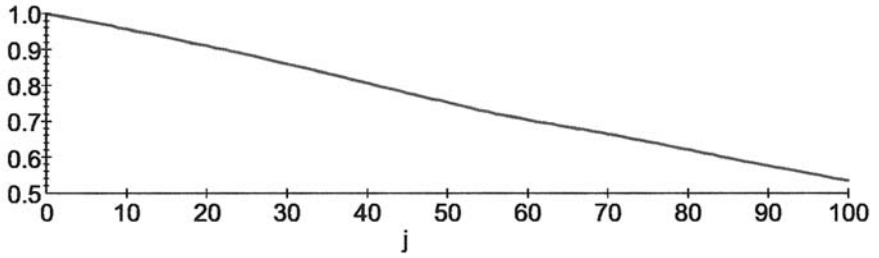


Figure 2.4 Sample of ACF of $/£ exchange rate data

the line is fuzzier, caused by the fact that a great many more observations are being plotted to the same real length of horizontal axis, which is a matter of scaling only, the series is seen to wander away from its starting point, to such an extent that it is difficult to argue that it appears to be varying around a fixed level. If it isn't varying about a fixed level (that is, there doesn't seem to be a fixed mean), then it is difficult to see how the variances or covariances might be behaving. It seems that they must also be varying with time, although, care should be taken since it is quite possible to imagine a series that varies to a constant degree around a mean that is changing.[6] However, in this case it is difficult to discern what that mean could be. The sample ACF for this series is given in Figure 2.4. In contrast to the ACF for the growth data, this declines linearly, and has not reached zero, even by the 100th lag ($\hat{\gamma}$ (100) =0.53521). This series appears to have very long memory, in terms of lags. Its sample ACF does not look like it is damping off at all. This is not consistent with the idea of covariance stationarity and suggests that the calculations may

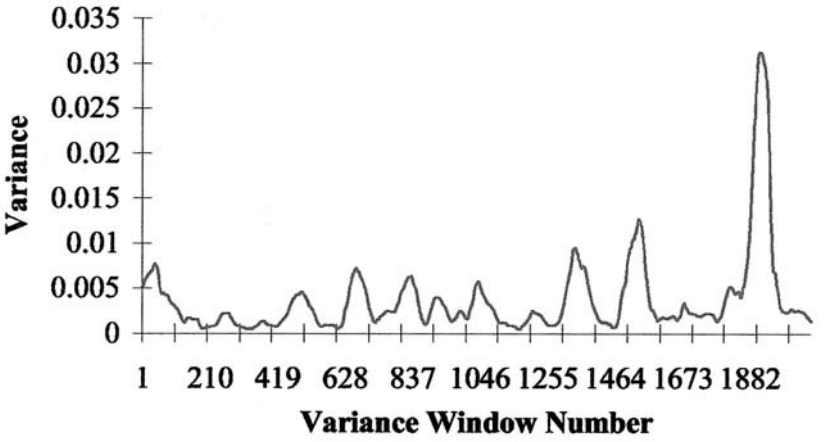


Figure 2.5 Moving window sample variance estimates of the $/£ exchange rate data, window length 100

indeed be meaningless. It seems likely that the exchange rate series is not covariance stationary.

To emphasize the point, Figure 2.5 plots the moving window sample variance estimates of the \$/£ exchange rate series, computing the sample variance for observations 1–100, followed by that for observations 2–101, and so on. From this it is clear that the variance around the mean does not remain constant even when the mean itself is allowed to vary across windows.

2.2.2.2 Strict stationarity

Covariance stationarity is a useful but rather specific version of stationarity. It is useful because it relates only to the first two moments, and because it can be defined precisely in terms of the parameters of the commonly used autoregressive-moving average (ARMA) time series models, as well as their multivariate counterpart, the vector autoregressive (VAR) model. Furthermore, if the distribution of the random variables is normal, then it is only necessary to consider the first two moments.

However, a more general definition is available, and is expressed in terms of the joint (entire) distribution of the set of random variables underlying the time series observations. Suppose there are T time series observations, x_t, $t = 1, 2, \ldots T$. Consider a subsample of n of these, x_t, $t = t + 1, t + 2, \ldots, t + n$. Each of these is thought of as a realization of an underlying random variable, X_t, $t = t + 1, t + 2, \ldots, t + n$. If the joint distribution of these n random variables remains unchanged through time, then the time series is said to be strictly stationary.[7]

2.2.3 Strict (joint distribution) stationarity

Let $F(\cdot)$ be the joint distribution function of, X_t, $t = t + 1, t + 2, \ldots, t + n$, written as $F(X_{\tau+1}, \ldots, X_{\tau+n})$. Then if

$$F(X_{\tau+1}, \ldots, X_{\tau+n}) = F(X_{\tau+h+1}, \ldots, X_{\tau+h+n}), \; |h| \geq 0 \tag{2.5}$$

then the process generating the time series observations is said to be strictly stationary.

Equation (2.5) simply states that the joint distribution of the sequence of random variables is unchanged when considering the distribution any number of periods earlier or later. In the case of covariance stationarity, it is not the distribution as a whole that is considered, but only its first two moments, the mean, and the (co)variances. This is clearly a weaker requirement.

It can be seen that strict stationarity, while having appeal from a philosophical point of view, is very demanding and so impracticable. In common with most textbook treatments, econometric research, applied and theoretical, this book will adopt covariance stationarity as its definition of stationarity, and, unless otherwise stated, stationarity will mean covariance stationarity. In addition, a common – though not universal – assumption of time series models is of normality, in which case the two definitions are coincident.

2.3 Univariate statistical time series models and non-stationarity

2.3.1 Describing covariance non-stationarity: parametric models

Covariance non-stationary is an observable feature of a time series, as seen from Figures 2.3–2.5. The failure of the sample autocorrelations of a time series to damp off over time suggests non-stationarity, or the wandering of a time series away from its starting (initial) value with a tendency not to return to it. These properties relate to the dynamic properties of the series rather than to the joint distribution of observations.

Such properties can be captured, by very simple models of the series, that relate the current value of a series to its past values and to the current and past values of a largely structureless stochastic component. That is, it is possible to invent theoretical models of the underlying random variables that would produce realizations whose sample properties approximate those observed in actual data.

It is important to realize at this early stage, that what is going on here is the approximation of the underlying process generating the data. This process is known as the data generating process (DGP). It can be expected to be highly complex, and incapable of exact description.[8] No model will be exact, not simply in terms of the parameter values chosen, but also in terms of the basic form of the model used. This having been said, models can often capture key features of data relevant for the purposes at hand. The key feature of interest in this case is covariance non-stationarity.

2.3.2 The white noise process

The building block of the time series models considered here is a stochastic process with simplest possible structure, having no temporal dependence and constant moments over time. Typically, the moments described are only the first two, thought there is no reason why this should not be extended to cover all moments. Put in terms of stationarity, the process is a zero mean stationary process. It is called a white noise process.

2.3.2.1 White noise

Let ϵ_t, $t = 1, 2, \ldots$ be a sequence of random variables. Then if

$$E(\epsilon_t) = 0 \tag{2.6a}$$

$$Var(\epsilon_t) = \sigma_\epsilon^2 \ \forall t \tag{2.6b}$$

$$\gamma_\epsilon(j) = 0 \ \ \forall j \neq 0 \tag{2.6c}$$

$$(E(\epsilon_t \epsilon_{t-j}) = 0, \ \forall j \neq 0) \tag{2.6d}$$

the sequence is said to be white noise[9] and the symbol $\forall$ means 'for all' or 'for any'.

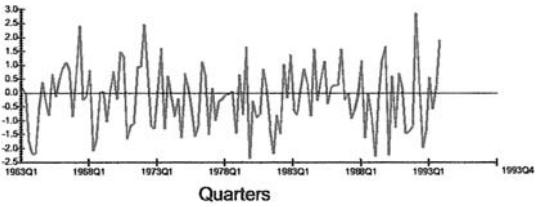


Figure 2.6 Realizations of a NIID(0,1) white noise sequence

Equations (2.6a–2.6d) state that the process has zero mean, constant variance, and that different members of the sequence are uncorrelated. In addition, there will often be a distributional assumption, which is that the random variables are normally distributed. Since, under normality, non-correlation is equivalent to independence, the sequence is then described as normally independently identically distributed (NIID) with a mean of zero and variance σ_ϵ^2 In short, $\epsilon_t \sim NIID\ (0, \sigma_\epsilon^2)$. Realizations of an NIID(0, 1) sequence are provided in Figure 2.6, with the time index labelled as though for the same time period and frequency of data as Figure 2.1 for the growth in output data.

2.3.3 The moving average process

It is easy to construct a correlated sequence from white noise by forming linear combinations of different members of the white noise sequence. For example, define

$$a_t = \epsilon_t - 1/2\epsilon_{t-1}.$$

Then clearly there is some temporal structure to the a_t, that is they are autocorrelated. Note that the mean of the process is given by

$$E(a_t) = E(\epsilon_t) - 1/2E(\epsilon_{t-1}) = 0; \tag{2.7a}$$

the variance by

$$Var(a_t) = Var(\epsilon_t) + 1/4E(\epsilon_{t-1}) - 1/2Cov(\epsilon_t,\ \epsilon_{t-1}), \tag{2.7b}$$

where, because ϵ_t is white noise, the covariance between ϵ_t and ϵ_{t-1}, $Cov(\epsilon_t, \epsilon_t - 1)$, is zero, this becomes

$$Var(a_t) = Var(\epsilon_t) + 1/4E(\epsilon_{t-1}) = 5/4\sigma_\epsilon^2 \tag{2.7c}$$

the autocovariance for $j \neq 0$ by

$$\begin{aligned}
\gamma_a(j) &= E(a_t a_{t-j}) \\
&= E((\epsilon_t - 1/2\epsilon_{t-1})(\epsilon_{t-j} - 1/2\epsilon_{t-j-1})) \\
&= E(\epsilon_t\epsilon_{t-j}) - 1/2E(\epsilon_{t-1}\epsilon_{t-j}) - 1/2E(\epsilon_t\epsilon_{t-j-1}) + 1/4E(\epsilon_{t-1}\epsilon_{t-j-1}) \\
&= \begin{cases} -\dfrac{1}{2}\sigma_\epsilon^2 & \text{for } j = 1 \\ 0 & \text{for } j > 1 \end{cases}
\end{aligned}$$

since the expectation terms in the last expression will be zero if the time index on the random variables is not the same because the white noise series is uncorrelated; if the index is the same then the expectation is the expectation of a square of a zero mean process, and so is its variance, σ_ϵ^2. So the process is autocorrelated as far as but not beyond the first lag and this is because it is a function of the current white noise term and its previous value.

It is possible to build more general models of this type. Let θ_i, $i = 1, 2, \ldots, q$ be constant coefficients and define

$$a_t = \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i,} \tag{2.8}$$

$$E(a_t) = 0. \tag{2.9a}$$

$$Var(a_t) = \left(1 + \sum_{i=1}^{q} q_i^2\right)\sigma_\epsilon^2, \tag{2.9b}$$

$$\gamma_a(j) = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \ldots + \theta_q\theta_{q-j})\sigma_\epsilon^2 & for\ j = 1,\ 2,\ \ldots,\ q \\ 0 & otherwise \end{cases} \tag{2.9c}$$

These equations show that the mean and variance are fixed, that the autocovariances depend not only on the time gap, but on the time itself, and that all moments are finite as long as the parameters are. The process is therefore stationary, but has an autocorrelation structure that cuts off after q lags. However, since there are q parameters in the model, these values may be chosen so as to reproduce any desired sequence of q autocovariances, and hence any ACF cutting off after lag q.

Equation (2.8) defines a moving average (MA) process (or model) of order q. It is important to note that all such processes are stationary, and that they are very flexible in terms of reproducing autocorrelation structure. To obtain a process whose autocorrelations last out to lag 15, a MA(15) model can be used. In theory, the model can extend to an infinite number of lags if the autocorrelations damp off asymptotically rather than all at once. Then it is necessary to place a restriction on the coefficients so that the variance (2.9b) exists, namely, that $\sum_{i=1}^{\infty} \theta_i^2 < \infty$. These properties also demonstrate the drawbacks of the MA model: it isn't practical to work with a very large number of lags; the model cannot capture non-stationary behaviour; and, finally, it is not easy to motivate in terms of the real life structures that might have given rise to data.

2.3.4 Wold's representation theorem

The approximation of autocorrelation structures of stationary processes by moving average models lies at the heart of one of the most important theories of time series analysis. Wold's representation theorem states that:

> 'any covariance stationary time series with only stochastic components[10] can be represented by an infinite order MA model with appropriately chosen coefficient values and white noise variance'.

The point is that, by extending the order of the MA far enough, it is always possible to provide a MA process whose ACF approximates that of any given ACF to whatever degree of accuracy is required, and that the approximation error goes to zero as the order of the MA increases. As long as the foregoing is understood, this may be abbreviated by stating that any (covariance) stationary time series with no deterministic components has an infinite order MA representation.

Thus, if x_t is a stationary time series with only stochastic components, it is always possible to represent it as

$$x_t = \epsilon_t - \sum_{i=1}^{\infty} \theta_i \epsilon_{t-q}$$

where ϵ_t is zero mean white noise with variance σ_ϵ^2, the only restriction on the parameters being that $\sum_{i=1}^{\infty} \theta_i^2 < \infty$. A more detailed account of this theorem may be found in Hamilton (1994, section 4.8) and a rigorous one in Brockwell and Davis (1991).[11]

2.3.5 The autoregressive process

A moving average process cannot capture non-stationarity. In addition, it cannot capture an autocorrelation structure that damps slowly off to zero other than in the case of an arbitrarily high-order (q) process. An alternative model relates the current value of a process to its past values plus a white noise disturbance term. This is the autoregressive process, and equations (2.10a) and (2.10b) below define an autoregressive process of order I and of order p respectively:

$$x_t = \phi x_{t-1} + \epsilon_t, \tag{2.10a}$$

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t, \tag{2.10b}$$

where ϕ and $\neq i = 1, 2, \ldots, p$ are constant coefficients.[12]

2.3.5.1 The ACF of an autoregressive process

It is straightforward to show that the autocovariance and autocorrelation functions of the AR(1) process are

$$\gamma_x(j) = \frac{\sigma_\epsilon^2 \phi^i}{1 - \phi^2} \tag{2.11a}$$

$$\rho_x(j) = \frac{\gamma_x(j)}{\gamma_x(0)} = \phi^j. \tag{2.11b}$$

In the case of an AR(p) process, the ACF is the solution to the difference equation $\rho_x(j) = \sum_{i=1}^{p} \phi_i \rho_x(j-i)$.[13]

The solution of this equation depends on the p solutions of the polynomial equation,

$$1 - \sum_{i=1}^{p} \phi_i z^i = 0, \tag{2.12}$$

where z is the argument of the function on the left-hand side of (2.12). Let these solutions be λ_i, $i = 1, 2, \ldots, p$. Then, in the case where the solutions are all distinct, the solution will be of the form[14]

$$\rho_x(j) = \sum_{i=1}^{p} A_i \lambda_i^{-j} \tag{2.13}$$

where the A_i, $i = 1, 2, \ldots, p$ are constant coefficients determined from the initial values of the ACF, $\rho_x(j)$, $j = 0, 1, \ldots, p-1$. A special case of particular interest in economics, is where the solutions occur in complex conjugate pairs.[15]

2.3.6 Lag polynomials and their roots

2.3.6.1 The lag operator and lag polynomials

The representation and analysis of the autoregressive and moving average models is made more succinct by the use of a functional representation of the lag structure involved. In the case of the AR(p) model, rewrite equation (2.10b) as

$$x_t - \sum_{i=1}^{p} \phi_i x_{t-i} = \epsilon_t. \tag{2.14}$$

Using the lag operator, L, defined such that

$$Lx_t = x_{t-1},$$
$$L^n x_t = x_{t-n},$$

equation (2.14) may be written in terms of x_t as

$$x_t - \sum_{i=1}^{p} \phi_i L^i x_t = \epsilon_t$$

or

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) x_t = \epsilon_t.$$

The term $\left(1-\sum_{i=1}^{p}\phi_i L^i\right)$ of this equation is a polynomial of degree p in the lag operator L (and so is itself an operator). That is, it is a polynomial function of L. It is therefore conveniently rewritten as

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i. \tag{2.15}$$

This function is called a lag polynomial operator (of order p). In general, the coefficient of $L^0 = 1$ does not have to be equal to 1 as it is here. This has arisen because the starting point was an autoregressive model.

Using (2.15), the AR(p) model of (2.10b) may be written

$$\phi(L)x_t = \epsilon_t.$$

Similarly, defining the q^{th} order lag polynomial,

$$\mathrm{q}(L) = 1 - \sum_{i=1}^{p} \mathrm{q}_i L^i, \tag{2.16}$$

the MA(q) model of (2.8) may be written as

$$a_t = \theta(L)\epsilon_t.$$

2.3.6.2 The roots of a lag polynomial

In obtaining the ACF of an AR(p) process, it was necessary to refer to the solutions to the equation

$$1 - \sum_{i=1}^{p} \phi_i z^i = 0, \tag{2.12 again}$$

But the left-hand side of this equation is the same function as (2.15) except that the lag operator has been replaced by the general complex argument, z. So, writing

$$\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$$

so that it can be seen that $\phi(z)$ is a polynomial function of z, equation (2.12) may be written,

$$\phi(z) = 0. \tag{2.17}$$

The values of z that satisfy (2.17) are called the roots of the polynomial $\phi(z)$. As a short hand, they are also referred to as the roots of the lag polynomial operator, $\phi(L)$, although, obviously, it is not correct in any sense to assign numerical values to an operator (the lag operator in this case).

2.3.6.3 Convenient short-hands for referring to functions of the coefficients of lag polynomials

Let $\pi(L)$ be a lag polynomial given by

$$\pi(L) = \pi_0 - \sum_{i=1}^{n} \pi_i L^i \tag{2.18}$$

where π_i, $i = 0, 1, \ldots, n$ are constant coefficients. By 'evaluating' the function at certain values of its argument, useful functions of the coefficients can result.

There are two important cases:

(i) Replace L by 0. Then (2.18) becomes

$$\pi(0) = \pi_0 - \sum_{i=1}^{n} \pi_i 0^i = \pi_0.$$

That is, $\pi(0)$ is the value of the coefficient of the zero lag term of $\pi(L)$.

(ii) Replace L by 1. Then (2.18) becomes

$$\pi(1) = \pi_0 - \sum_{i=1}^{n} \pi_i 1^i = \pi_0 - \sum_{i=1}^{n} \pi_i.$$

So $\pi(1)$ is the sum of the coefficients of $\pi(L)$.

2.3.6.4 Roots and the ACF of an autoregressive process

Having defined the lag polynomial and its roots, it is possible to refer very easily to an AR model and its ACF as follows. Consider the AR(p) model $\phi(L)\, x_t = \epsilon_t$, (where, it could be added for precision, but would normally be understood from the definition of an AR process, ϵ_t is white noise and $\phi(0) = 1$). Let λ_i $i = 1, 2, \ldots, p$ be the (distinct) roots of the lag polynomial. Then the ACF of x_t is given by $\rho_x(j) = \sum_{i=1}^{p} A_i \lambda_i^{-j}$, where the A_i depend on the parameters of the process, including the white noise variance. So, it is the roots of the autoregressive lag polynomial that determine the evolution of the ACF as a function of the time gap, or lag, j.

2.3.7 Non-stationarity and the autoregressive process

Equation (2.13) and its simplification in the first order case, equation (2.11b),[20] show that the pattern of the autocorrelations of an AR process depend on the roots, λ_i, $i = 1, 2, \ldots, n$ of the lag polynomial, $\phi(L)$. A necessary and sufficient condition that $\rho_x(j) \to 0$ as $j \to \infty$ is that $|\lambda^{-1}_i| < 1$ for all $i = 1, 2, \ldots, n$. In terms of the roots instead of their inverses, the condition is that all roots are such that $|\lambda_i| > 1$. This condition is referred to as 'all the roots lying outside the unit circle'. It applies to complex as well as to real roots. In the complex case, a root may be written $\lambda_c = a + ib$, where in this case $i = \sqrt{-1}$ (not an indexing subscript) and a and b are real coefficients. Then $|\lambda_c| = \sqrt{(a^2 + b^2)}$.

The condition that all the roots (of the autoregressive lag polynomial) lie outside the unit circle is the stationarity condition for autoregressive processes.

2.3.7.1 Stationarity of an autoregressive process

The AR (p) process $\phi(L)\, y_t = \epsilon_t$ is stationary if and only if λ_i such that

$$\phi(\lambda_i) = o \Rightarrow |\lambda_i| > 1 \; \forall i = 1,\ 2,\ \ldots,\ p.$$

2.3.8 The random walk and the unit root

Notice that, by definition, if the polynomial function evaluated at some number is equal to zero, then that number is a root of the polynomial. So, for example, if $\phi(1) = 0$ then 1 is a root, usually referred to as a unit root. Such a root would mean the autoregressive process with this lag polynomial was non-stationary because this root is not greater than 1 in modulus (i.e. $|1| = 1$), and hence the stationarity condition is contradicted.

2.3.8.1 The random walk process

The random walk is an AR(1) process with a unit root. It is therefore a non-stationary process. Equation (2.19) below defines a random walk,

$$x_t = x_{t-1} + \epsilon_t,$$

which can be written

$$\begin{aligned} &\phi(L)x_t = \epsilon_t, \\ &\phi(L) = 1 - L. \end{aligned} \tag{2.19}$$

To see that (2.19) has a unit root, and is therefore non-stationary, note that $\phi(z) = 0$ has the solution $z = 1$ from (2.19). That is, the lag polynomial of this model has a root of 1.

2.3.8.2 Differencing and stationarity

The period on period changes of a process are known as its first difference. Thus, $x_t - x_{t-1}$ is the first difference of x_t. It is denoted Δx_t. Clearly, D can be represented as the (autoregressive) lag operator $(1 - L)$. That is, it is a first order operator with a unit root. The random walk may thus be written $\Delta x_t = \epsilon_t$ which illustrates an important principle. Since ϵ_t is white noise it is stationary. Therefore Δx_t is stationary. But x_t itself is non-stationary. That is, taking the first difference of the non-stationary process has reduced it to stationarity. Such a process is said to be integrated of order 1.

The second difference of the random walk would be

$$\begin{aligned} &\Delta\Delta x_t = \Delta\epsilon_t \\ &\Rightarrow \Delta(x_t - x_{t-1}) = \epsilon_t - \epsilon_{t-1} \\ &\Rightarrow \Delta x_t - \Delta x_{t-1} = \epsilon_t - \epsilon_{t-1} \\ &\Rightarrow x_t - x_{t-1} - x_{t-1} + x_{t-2} = \epsilon_t - \epsilon_{t-1} \\ &\Rightarrow (1 - L)^2 x_t = \epsilon_t - \epsilon_{t-1} \\ &\Rightarrow \Delta^2 x_t = \epsilon_t - \epsilon_{t-1}. \end{aligned}$$

That is,

$$\Delta\Delta = \Delta^2$$

and in general, if the process is differenced n times, the operation can be represented as Δ^n, the lag operator representation of which can be calculated from $(1 - L)^n$. Although the first difference of the random walk is stationary, so is the second, because it is an MA(1) process, $\Delta^2 x_t = \epsilon_t - \epsilon_{t-1}$, and all MA processes are stationary. However, it has been over-differenced, meaning that in order to reduce the original (random walk) process to stationarity it was only necessary to difference once. This can be detected in the time series structure by observing that $\epsilon_t - \epsilon_{t-1}$ is a MA(1) process with a unit root. (It could be said that differencing the minimal number of times to reduce the series to stationarity removes the unit root altogether, while over-differencing moves it from the AR to the MA side of the equation.) Strictly speaking, it is the minimal number of times it is necessary to difference a non-stationary series to stationarity that defines its order of integration. This is made precise in section 2.3.11 below.

2.3.8.3 The random walk as stochastic trend

The idea of a trend is a process that increases by the same amount each time period.[17] Thus, a process defined as

$$y_t = a + bt \tag{2.20}$$

where a and b are constant coefficients, increases by an amount b each period since

$$y_t - y_{t-1} = (a + bt) - (a + b(t-1)) = b.$$

So, (2.20) could be written

$$y_t = y_{t-1} + b. \tag{2.21a}$$

However, (2.21a) does not tie down the value of the process at any point in time, whereas (2.20) does. In particular, considering the value of the process at $t = 0$, called the initial value, (2.20) gives

$$y_0 = a. \tag{2.21b}$$

The time trend model is fully described by equations (2.21a) and (2.21b). Because the amount added each time period is fixed, b this is known as a deterministic trend. If instead of adding a fixed amount, a white noise is added, the resultant process is still called a trend, but it is now termed a stochastic trend. Thus in place of (2.21a) write

$$y_t^* = y_{t-1}^* + \epsilon_t \tag{2.22}$$

Comparing (2.9a and 2.22), it is clear that y_t^* is a random walk. That is, the random walk and the stochastic trend are, when defined in this way, the same thing.

By the same argument as in the deterministic case, in order to tie the process down, it is necessary to provide some information about the process at some point in time. As before, it is most convenient and intuitively appealing to make this the initial value, y_0^*. The simplest case is $y_0^* = 0$.

It is possible to obtain the expression for the stochastic trend analogous to (2.20) in the deterministic case, that is to express the process in terms of its initial value and the accumulation of its increments. This is done by the process of back-substitution, which means using the original equation for y_t^* (2.22), lagging it one period to get an expression for y_{t-1}^*, and substituting this into (2.37). This generates an expression in y_{t-2}^* which can be substituted for in a similar manner. The process is repeated until the y^* variable in the equation is the initial value. These steps are:

$$\begin{aligned} y_t^* &= y_{t-1}^* + \epsilon \\ &= y_{t-2}^* + \epsilon_{t-1} + \epsilon_t \\ &= y_{t-3}^* + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &= \ldots = y_0^* + \sum_{j=1}^{t} \epsilon_j. \end{aligned} \tag{2.23}$$

The fact that (2.23) involves a simple (unweighted) sum of white noise terms leads to the general label of integrated for processes of this type, although the class is not restricted to pure random walks of the type illustrated here.[18]

Using equation (2.23), it is straightforward to show that both the variance and autocorrelation structure of the random walk are varying over time according to

$$\begin{aligned} Var(y_t^*) &= t\sigma_\epsilon^2, \\ Cov(y_t^*,\ y_{t-j}^*) &= (t-j)\sigma_\epsilon^2 \end{aligned}$$

and defining the correlation to be the covariance divided by the variance of the process at time t, the autocorrelation is

$$Cor(y_t^*,\ y_{t-j}^*) = \left(1 - \frac{i}{j}\right). \tag{2.24}$$

Alternatively, dividing by the square root of the product of the variances at t and $t - j$ would give the autocorrelation[19]

$$Cor(y_t^*,\ y_{t-j}^*) = \sqrt{1 - j/t}.$$

It is clear that the process is non-stationary since its moments are not constant over time. From this non-constancy, it also follows that the manipulations underlying the derivation of the difference equation for the ACF of an autoregressive process are not valid. So, in fact, equation (2.13) only applies in the stationary case.[20]

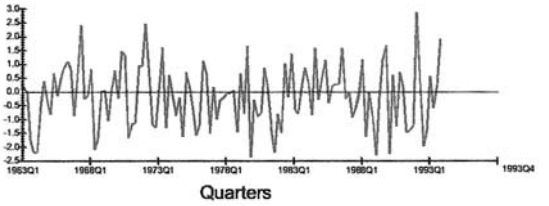


Figure 2.7 Random walk, 2,168 observations, initial value 0, NIID(0,1) white noise

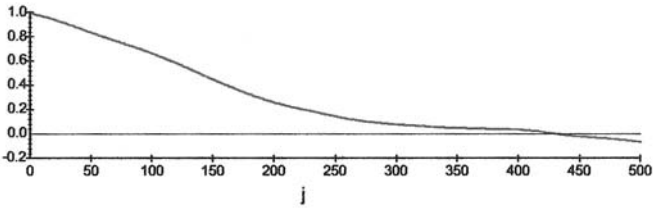


Figure 2.8 Sample ACF of random walk series plotted in Figure 2.7

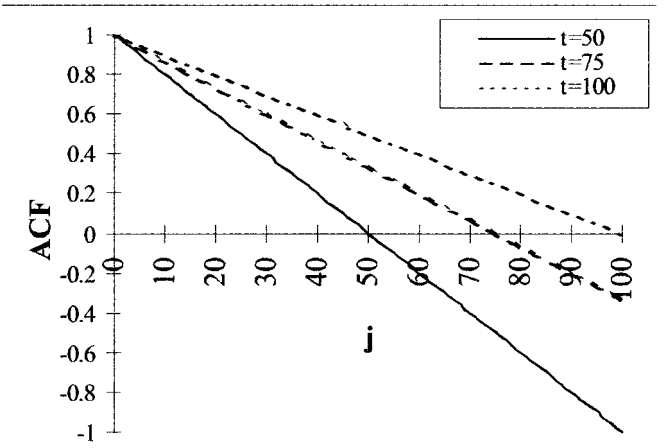


Figure 2.9 Theoretical ACF of a random walk at different points in time: t = 50, 75, 100

Figure 2.7 presents an artificially generated random walk sequence based on an NIID(0,1) white noise sequence, and Figure 2.8 is its sample ACF, although recall that this plot cannot have the meaning it possesses in the stationary case.

The theoretical ACF will vary with the time *t*, according to equation (2.24). Figure 2.9 provides a suite of such functions for three different points in time.

2.3.8.4 The random walk with drift

An important case for economic time series involves a generalization of the random walk so that the process consists of the sum of both a stochastic and a linear deterministic trend,

$$x_t = x_0 + \mu t + \sum_{j=1}^{t} \epsilon_j$$

x_0 being the initial value of the process. In this case,

$$\Delta x_t = x_t - x_{t-1} = x_0 + \mu t + \sum_{j=1}^{t} \epsilon_j \left(x_0 + \mu(t-1) + \sum_{j=1}^{t-1} \epsilon_j \right) = \mu + \epsilon_t$$

Such a process is called a random walk with drift and μ is called the drift parameter. There are now two aspects to the non-stationarity: not only is the variance growing over time (and the autocorrelation structure changing over time) but the mean of the process is also evolving since

$$E(x_t) = E(x_0) + E(\mu t) + E(\sum_{j=1}^{t} \epsilon_j) = x_0 + \mu t$$

assuming x_0 is fixed (non-stochastic).

2.3.9 The autoregressive moving average process and operator inversion

Autoregressive and moving average models can be combined to form a single model. This may be written

$$\phi(L)x_t = \theta(L)\epsilon_t \tag{2.25}$$

where ϵ_t is white noise, $\phi(L)$ a p^{th} order lag polynomial with ϕ (0) = 1 and θ (*L*) q^{th} order with θ (0) = 1 as defined by equations (2.15) and (2.16). The model is then autoregressive-moving average of order (*p*, *q*) or ARMA(*p*, *q*). Since θ (*L*) ϵ_t is a moving average process, it is stationary. Thus (2.25) is stationary if and only if the autoregressive contribution is stationary, that is if all the roots of ϕ (*L*) lie outside the unit circle. In the stationary case, models of this type give rise to ACFs which begin with *q* irregular autocorrelations determined by the MA coefficients, followed by a pattern of values generated to the solution of the difference equation arising from the autoregressive polynomial. For details see Box and Jenkins (1976).

ARMA models are also of practical importance since they provide a way of representing a relatively complex ACF with relatively few parameters: they are said to be parsimonious.

In the stationary case, both sides of (2.25) can be divided by ϕ (*L*) to give

$$x_t = \psi(L)\epsilon_t$$

where

$$\psi(L) = \theta(L)/\phi(L).$$

It is stronger to think of this as the inversion of the AR operator, and write instead

$$\psi(L) = \phi^{-1}(L)\theta(L).$$

where $\phi^{-1}(L)$ is such that $\phi^{-1}(L)\,\phi(L) = 1$.[21] The inverse operator does not exist unless all the roots of $\phi(L)$ lie outside the unit circle. The operator $\psi(L)$ is of infinite order, and so ARMA models can be thought of as a restricted way of obtaining an MA(∞) representation. That is, the ARMA model of finite orders provides an approximation to the infinite order MA representation of a stationary process.

2.3.9.1 Illustration of operator inversion

Let $\phi(L) = 1 - \phi L$, $|\phi| < 1$. Then a Taylor series expansion may be used to obtain

$$\phi^{-1}(L) = 1/(1-\phi L) = (1 + \phi L + \phi^2 L^2 + \ldots = \sum_{i=0}^{\infty} \phi^i L^i.$$

It is easily verified that $\phi^{-1}(L)\,\phi(L) = 1$ in this case. Then the ARMA(1,1) model

$$(1-\phi L)x_t = (1-\theta L)\epsilon_t \tag{2.26}$$

can be written

$$x_t = \left(\sum_{i=0}^{\infty} \phi^i L^i\right)(1-\theta L)\epsilon_t \tag{2.27}$$

Multiplying out the operators on the right hand side of (2.27) gives

$$\begin{aligned}\left(\sum_{i=0}^{\infty} \phi^i L^i\right)(1-\theta L) &= \sum_{i=0}^{\infty} \phi^i L^i - \theta \sum_{i=0}^{\infty} \phi^i L^{i+1} \\ &= 1 + \sum_{i=1}^{\infty} \phi^i L^i - \sum_{i=1}^{\infty} \theta\phi^{i-1} L^i \\ &= 1 + \sum_{i=1}^{\infty} (\phi^i - \theta\phi^{i-1}) L^i.\end{aligned}$$

This ARMA process (2.26) can therefore be represented as

$$x_t = \psi(L)\epsilon_t \tag{2.28a}$$

where $\psi(L) = 1 - \sum_{i=1}^{\infty} \psi_i L^i$, with

$$\psi_i = -(\phi^i - \theta\phi^{i-1}) = \phi^{i-1}(\theta - \phi). \tag{2.28b}$$

Equation (2.28b) also illustrates another point. If $\theta = \phi$ then $\psi_i = 0$, $\forall_i \geq 1$, that is $\psi(L) = 1$. Substituting this into (2.47a) gives $x_t = \epsilon_t$. Comparing this with

(2.26) it appears that the lag polynomials have cancelled. In the stationary case, where the common operator has its root outside the unit circle, this is a reasonable way to describe what has happened, since if $\theta = \phi$ the AR and MA operators are indeed the same. The situation is a little more complex in the non-stationary case where dependency on initial values is not negligible.

2.3.10 Factorizing polynomial lag operators

Any polynomial lag operator may be factorized in terms of its roots. Consider the nth order lag polynomial[22]

$$\pi(L) = 1 - \sum_{i=1}^{n} \pi_i L^i.$$

If the roots of $\pi(L)$ are λ_i, $i = 1, 2, \ldots, n$, then it may be written as the product of first order factors $(1 - \lambda^{-1}{}_i L)$,

$$\pi(L) = \prod_{i=1}^{n} (1 - \lambda_i^{-1} L).$$

This factorization does not depend on whether the roots are outside the unit circle. Thus, if in a stationary ARMA(p, q) model, there is a common factor (i.e. root) between the AR and the MA polynomial, this may be cancelled to give an ARMA($p - 1$, $q - 1$) model with exactly the same time series characteristics.

2.3.10.1 Invertibility

An AR(p) or ARMA(p, q) model is said to be invertible if the moving average operator has all roots outside the unit circle. That is, if

$$\phi(L) x_t = \theta(L) \epsilon_t$$

then, the process is invertible if and only if

$$\theta(z) = 0 \Rightarrow |z| > 1.$$

2.3.10.2 Identifiability and invertibility

Moving average models have the property that a given set of coefficients is not the only one that reproduces a specific autocorrelation structure. In particular, if any root of the moving average operator, λ^* is replaced by its inverse, then the autocorrelation structure is unchanged. In the simple MA(1) case, the processes

$$x_{1,t} = \epsilon_t - \theta \epsilon_{t-1},$$
$$x_{2,t} = \epsilon_t - 1/\theta \epsilon_{t-1},$$

have the same ACF,[23]

$$\rho_{x_2}(1) = \frac{-1/\theta}{1 + (-1/\theta)^2} = -\frac{1/\theta}{1 + 1/\theta^2} = -\frac{\theta}{1 + \theta^2} = \rho_{x_1}(1)$$

(Both are stationary because they are pure moving average processes.) In general, an ARMA(p, q) or MA(q) process will have 2^q different parameterizations that generate the same ACF, because any subset of the moving average roots may be replaced by their inverses.[24] The coefficients of the MA component may not therefore be uniquely identified from the ACF. However, if invertibility holds, there is a unique set of MA coefficients corresponding to the ACF.[25]

2.3.10.3 Comparison with stationarity

Stationarity and invertibility are the same mathematical condition applied to different operators. Stationarity ensures that an AR(p) process may be expressed as an MA (∞) (in other words that the inverse of the AR lag polynomial exists). Similarly, invertibility means that the MA operator may be inverted and so the model expressed in AR(∞) form.

2.3.11 Order of integration and autoregressive integrated moving average models

As already mentioned, there are many ways in which a time series can be covariance non-stationary. All that is required is that at least one of the mean, variance or covariance are changing over time. However, one particular way of capturing or describing non-stationarity relates to the stochastic trend. A stochastic trend (random walk) is non-stationary as has been seen. The difference of the process however, is stationary. The unit root associated with the differencing operator is in this case responsible for the non-stationarity. Such a series is described as integrated of order 1 because differencing it once removes the non-stationarity. A general definition of an integrated process in the ARMA context requires a condition to avoid over-differencing. This is invertibility.

2.3.11.1 Integration: Definition 1[26]

Let x_t be a non-stationary process. If

$$y_t = \Delta^d x_t \tag{2.29a}$$

has a stationary and invertible ARMA representation[27]

$$\phi(L)y_t = \theta(L)\epsilon_t \tag{2.29b}$$

then x_t is said to be integrated of order d, denoted I(d).

2.3.11.2 ARIMA models

Substituting (2.29a) into (2.29b) gives

$$\phi(L)\Delta^d x_t = \theta(L)\epsilon_t \qquad (2.30) \tag{2.30}$$

A time series having this representation is said to be autoregressive integrated moving average (ARIMA) of order (p, d, q), where $\phi(L)$ and $\theta(L)$ have all their roots outside the unit circle and are of order p and q respectively. The operators of the left-hand side of (2.30) can be expressed as a single operator, say

$$\pi(L) = \phi(L)\Delta^d.$$

For example, if $\phi(L) = 1 - 0.1L$ and $d = 1$ then

$$\pi(L) = (1-0.1L)\Delta = (1-0.1L)(1-L) = 1-1.1L+0.1L^2$$

where $\pi(L)$ has roots of 10 and 1, one stationary root and a unit root respectively. This suggests another way of thinking about the order of integration as being the number of unit roots in the autoregressive lag polynomial.

If x_t has an ARMA(m, q) representation $\pi(L)\, x_t = \theta(L)\, \epsilon_t$ which is invertible and where non-stationarities are due only to unit roots, then the order of integration is equal to the number of unit roots of $\pi(L)$. If $\pi(L)$ has $d \leq m$ unit roots, then it may be factorized as $\phi(L)\, \Delta^d x_t = \theta(L)\, \epsilon_t$, where $\phi(L)$ is of order $m - d$.

A time series with a positive order of integration is said to be integrated. Clearly, integrated time series are not the only type of non-stationary time series, but this is a very popular way of modelling non-stationarity, not least because it is simple and because a great deal of statistical theory has been developed to further this approach.

2.3.12 Trend and difference stationarity

The random walk with drift illustrated that non-stationarity can be due to deterministic or stochastic trends. In this example it was both. However, interest often focuses on the distinction between non-stationarity due to time trends and that due to stochastic trends. A time series that is non-stationary due to a linear time trend is called trend stationary, because it consists simply of stationary fluctuations around a trend. So, if the fluctuations about trend are white noise, this would be written,

$$x_t = a + bt + \epsilon_t. \tag{2.31}$$

If the trend is removed, the process is stationary since $x_t - a - bt = \epsilon_t$.

However, careful consideration of definition 1 of integration shows that this process is not an integrated process. To see this, first note that by (2.31) x_t is non-stationary as its expected values is the trend and so changing over time. Now consider differencing removes that non-stationarity:

$$\Delta x_t = \Delta a + \Delta bt + \Delta\epsilon_t = 0 + b\Delta t + \Delta\epsilon_t = b(t-(t-1)) + \Delta\epsilon_t = b + \Delta\epsilon_t. \tag{2.32}$$

The definition of integration is trivially generalized to include a non-zero mean for the difference process (here b),[28] but the real problem with (2.32) lies with the MA process $\Delta\epsilon_t$. This is first order but with a unit root and so is non-invertible. The definition requires the ARMA representation of the differenced

process to be invertible. Therefore x_t is not integrated of order 1 (or higher order as further differencing will just induce further MA unit roots). In contrast, a simple random walk is white noise after differencing by definition, and so clearly I(1). It is the quintessential I(1) process. It is said to be difference stationary.

But there is an uncomfortable wrinkle in this terminology: a difference stationary process can still have a deterministic trend. The simplest case is the random walk with drift, $x_t = x_{t-1} + b + \epsilon_t$, or $\Delta x_t = b + \epsilon_t$, which is stationary and invertible. So the process is I(1).[29]

2.3.13 Other models

2.3.13.1 Fractional integration

The order of integration of an ARIMA model need not be integer valued. If d is non-integer, the model is known as fractionally integrated and is abbreviated to ARFIMA(p, d, q). The process is stationary for $d < 0.5$, but the autocorrelations die down more slowly than those of a stationary AR process. For $d = 0.5$ the process is non-stationary.

The definition of the fractional differencing operator requires the gamma function, Γ (.), and is defined by

$$\Delta^d = 1 + \sum_{k=1}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k$$
$$= 1 - dL - \frac{1}{2d(1-d)} L^2 - \frac{1}{6d(1-d)(2-d)} L^3 \ldots + \gamma_k L^k \ldots$$

where $\gamma_k \propto k^{-(1+d)}$ for large k, and so die away slowly, such that a very high order autoregressive model would be needed to approximate the ACF reasonably well. The ARFIMA model was developed by Granger and Joyeux (1980) and Hosking (1981).

2.3.13.2 Structural models

Rather than embed time series properties in a single statement, as in the ARIMA class of models, components having identifiably different characteristics can be modelled separately so that each of its components can be interpreted directly. Such models are described extensively in Harvey (1989) and despite fairly persuasive arguments in their favour will not be dealt with in this book.[30]

2.4 Testing for non-stationarity in single series

2.4.1 Background

The form of non-stationarity that is commonly tested for is the unit root. The structure within which such tests are performed is the AR or ARMA model.

The idea is to obtain a parameterization of the model that allows the hypothesis to be tested to involve a single parameter. This subject is discussed in detail in Patterson (2005). However, we illustrate some of the structure of these tests briefly here for two reasons. Firstly, because multivariate generalizations form the basis of tests discussed in greater length in chapters 3 and 4;[31] and secondly because prior testing for non-stationarity is crucial to a great deal of the methodology of time series modelling used in economics and finance.

2.4.2 Reparameterizing the autoregressive model

Consider the AR (p) model,

$$\phi(L)x_t = \epsilon_t \tag{2.33}$$

where $\phi(L) = 1 - \sum_{i=1}^{p} \phi_i$. As discussed by many authors (e.g. Burke, 1996), and applied by Dickey and Fuller (1979, 1981) in their ground-breaking work on testing for unit roots, this polynomial can be rewritten as

$$\phi(L) = -\psi L + \phi^*(L)(1-L) \tag{2.34a}$$

where

$$\phi^*(L) = 1 - \sum_{i=1}^{p-1} \phi_i^* L^i,$$

$$\phi_i^* = -\sum_{i=i+1}^{p} \phi_i, i = 1,\ 2,\ \ldots,\ p-1$$

and, most relevantly,

$$\psi = -\phi(1). \tag{2.3b}$$

Equation (2.34b) shows that $\psi = 0$ if and only if $\phi(L)$ has a unit root. For convenience, define

$$\phi^*(L) = \sum_{i=0}^{p-1} \phi_i^* L^i = 1 + \sum_{i=1}^{p-1} \phi_i^* L^i.$$

Substituting this and (2.34a) into (2.33) and using $\Delta = (1 - L)$ gives

$$-(\psi L + \phi^*(L)\Delta)x_t = \epsilon_t$$

or

$$\phi^*(L)\Delta x_t = \psi x_{t-1} + \epsilon_t \tag{2.35}$$

where $\phi^*(L)$ is a $p - 1^{th}$ order lag polynomial with $\phi^*(0) = 1$. That is, the AR(p) model may be reparameterized as an AR($p - 1$) model in first differences ($\phi^*(L)$

Δx_t), together with a correction term in the lagged level (ψx_{t-1}). The unit root test is then a test of the null hypothesis

$$H_0: \psi = 0$$

in the model obtained by rearranging (2.35) in regression format as,

$$\Delta x_t = \psi x_{t-1} + \sum_{i=1}^{p-1} \phi_i^* \Delta x_{t-1} + \epsilon_t. \tag{2.36}$$

The summation term on the right-hand side of (2.36) does not appear if $p = 1$, and so can be thought of as the correction for autocorrelation beyond that which would be due to an AR(1) process. The alternative hypothesis can be one or two sided, according to whether the alternative of interest is stationarity ($\psi < 0$) or explosiveness ($\psi > 0$), or either. Typically the alternative of interest is stationarity, and so that used is

$$H_A: \psi < 0.$$

Unit root tests based on (2.36) are called augmented Dickey–Fuller (ADF) tests (see Patterson 2005).

2.4.3 Semi-parametric methods

A contrasting approach is based on the observation that it is only the parameter ψ that is of interest in (2.36). The others, while important because they correct for autocorrelation, are of no direct interest and are consequently known as nuisance parameters. Their specific values are certainly of no interest in this context. To see this, rewrite (2.36) so that the correction (or augmentation) terms do not appear explicitly, as

$$\Delta x_t = \psi x_{t-1} + u_t \tag{2.37}$$

where $u_t = \sum_{i=1}^{p-1} \phi_i^* \Delta x_{t-i} + \epsilon_t$, is an autocorrelated disturbance term, and no longer white noise. Tests based on (2.37) that assume the disturbances are white noise will not be valid and inferences from them could be seriously misleading. However, it is possible to correct the test statistics for the disturbance autocorrelation so that inferences are once again valid. This methodology is that developed by Phillips (1987) and Phillips and Perron (1988). These tests require calculation of a term that has become known as the long-term variance which is computed using a weighted average of autocorrelations in a way related to spectral estimation and heteroscedastic variance–covariance matrix (HAC) estimation (see Andrews 1991; Newey and West 1987; and White 1980). Again, more details may be found in Patterson (2005).

2.4.4 Other issues

A number of complicating features present themselves in testing for unit roots that have analogues in the multivariate tests discussed later in this book. Some important ones are listed below:

(i) The underlying model may not be AR but ARMA. In this case the AR approximation would be arbitrarily long and impractical in an empirical setting. Practically speaking, the optimal length of the pure AR approximation depends on the sample size, and longer models can only be entertained as more data becomes available. The relationship between the sample size and the AR order is critical to the (asymptotic) validity of the test. Ng and Perron (1995) discuss this problem and Hall (1989) offers a different approach. See also Galbraith and Zinde-Walsh (1999).

(ii) The number of unit roots may be greater than 1 in which case testing can become unreliable if performed in such a way that unit roots remain unparameterized in the model. Dickey and Pantula (1987) advise on this issue. It is relevant since economic time series, especially those recorded in nominal terms, can be integrated of higher order, especially I(2).

(iii) Economic time series are often subject to structural breaks. This is a portmanteau term to cover many possibilities, but relates simply to the assumption of constancy of parameters where this does not exist. This may affect the parameters of interest, so that, for example, a series may change from being I(1) to being stationary. Alternatively, a time series may in fact be stationary around a trend (or mean level) that is subject to jumps or sudden changes in slope. Since the tests themselves look at the stochastic behaviour around the trend, misspecification of this trend leads to unreliable inferences about the stochastic component of the series. This is a topic of current research, but established papers in the area are Perron (1989), Zivot and Andrews (1992) and Perron (1990).

(iv) As already observed, a unit root test is an examination of the stochastic component of a series, that is the random fluctuations about some deterministically determined level. This could be many things: zero, non-zero but fixed, or a trend of some polynomial degree. But misspecification of the deterministic component can lead to incorrect inference on the stochastic properties of the data. Dickey and Fuller (1979) address this to some extent, developing tests for the trend as well as the unit root. Patterson (2000, section 6.4) discusses a framework for joint determination of the stochastic and deterministic components of a univariate time series.

2.4.5 Other approaches

Unit root tests of the type outlined above have been criticized for a number of reasons. In empirical work, great importance is placed on the distinction

between unit root and stationary processes, so high precision is required of the tests. This is unfortunate because it has been demonstrated that even in the stationary case significant distortions can occur in the estimation and testing of autoregressive roots (Nankervis and Savin, 1985, 1988). The main attack is on the power of the tests: their ability to correctly reject the null of non-stationarity. To be useful, any test must, asymptotically, be able to reject a false null with certainty. Such tests are called consistent tests. The tests of the previous section satisfy this requirement under reasonable conditions. However, they are likely to perform less well if the root is stationary but close to unity, or if the process is fractionally integrated with an integration parameter close to one half.

The dual of the power problem can also be encountered, where size, the probability of rejecting a true null, is distorted from its nominal value. Of course, what this amounts to saying is that the appropriate tail of the null distribution is altered. This will occur, for example in an ARIMA(p,1,1), model,

$$\phi(L)\Delta x_t = (1-\theta L)\epsilon_t \tag{2.38}$$

where $|\theta| < 1$, and $\phi(L)$ has all its roots outside the unit circle, so the differencing operator is the only source of the unit root and $x_t \sim I(1)$. But as $\theta \to 1$, so $(1 - \theta L) \to (1 - L)$ and the MA operator will tend to cancel with the differencing operator. In the limit where this occurs, the process will be stationary and the null ought to be rejected. But in finite samples, this will be a smooth rather than a sudden transition, leading to a tendency for tests to reject the null of a unit root for θ close to unity, even though strictly speaking the process is still I(1). (See Blough, 1992, for a discussion of this issue.)[32]

This idea has formed the basis of a set of stationarity tests where the null is of stationarity. This amounts to a null hypothesis of

$$H_0: \theta = 1$$

in models such as (2.38). Naturally, this literature is closely related to that for testing for moving average unit roots, important contributions being Kwiatowski, Phillips, Schmidt and Shin (1992), and Leybourne and McCabe (1994). Of course, such tests suffer from finite sample power problems for θ close to unity, and size problems when an additional AR root tends to 1 (see also Lee and Schmidt, 1996, for behaviour in the presence of fractional integration). KPSS also suggest using both unit root and stationarity tests jointly in confirmatory data analysis. This was investigated in Burke (1994) and found to add little to the use of either test individually.

The power of unit root tests can be improved by the use of covariates (Hansen 1995). This has not yet become a popular approach. A method that is becoming as popular as the ADF test is that advanced by Elliott, Rothenberg and Stock (1996).

The Bayesian approach to unit root testing is now well developed and may be found more appealing since the impact of the unit root's presence or absence is not so crucial for the distribution theory (see Bauwens, Lubrano and Richard, 2000, chapter 6). Harvey's (1989) structural models relax the concentration on the ARIMA model that has taken such a firm hold in the analysis of non-stationary economic time series, and offer an alternative set of testing techniques. Lastly, in an alternative view of the uncertainty of structure, Leybourne, McCabe and Tremayne (1996) have developed tests for a stochastic unit root, where, rather than have a fixed value of 1, the root being tested is stochastic, having a distribution centred on unity under the null hypothesis.

2.4.6 Relevance of unit root testing to multivariate methods

There are a number of reasons why it is necessary to test for unit roots. Among these are that the presence of unit roots alters the statistical properties of estimators and test statistics used in the econometric analysis of the relationships between variables. Another is that the presence of a unit root in a group of series makes it possible to identify the presence of a long run relationship between the series.

Consider an $n \times 1$ vector of I(1) time series $X_t = (x_{1,t} \ldots x_{n,t})'$. Consider any function of these series,

$$\eta_t = f(X_t) \tag{2.39}$$

although a linear function is the easiest to work with, say,

$$f(X_t) = a_0 + \sum_{i=1}^{n} a_i x_{i,t} \tag{2.40}$$

where the a_i, $i = 0, 1, \ldots, n$ are constant coefficients.

If this combination results in a zero mean stationary process, then, substituting (2.40) into (2.39),

$$\eta_t = a_0 + \sum_{i=1}^{n} a_i x_{i,t}$$

is stationary. That is to say that the relationship

$$a_0 + \sum_{i=1}^{n} a_i x_{i,t} = 0 \tag{2.41}$$

holds with an error that has mean zero, constant variance, the ACF of which damps off quite quickly. Being stationary, it will not wander widely from its mean value of zero and will cross it frequently. That is, $a_0 + \sum_{i=1}^{n} a_i x_{i,t}$ will not depart from zero in any permanent way. So (2.41) holds in the long run –

never exactly, but without long periods of failure. This property is known as cointegration.

Unit root tests may feature in two ways in order to establish the existence of such a long-run relationship. First, it is necessary to test for the unit roots in the first place. Secondly, in order to establish that a long-run relationship exists, it is necessary to test if the function of the data is stationary – that is, to check that it does not contain a unit root. Thus one might perform a unit root or stationarity test on $f(X_t)$, although, importantly, if the parameters of this function are estimated then adjustments to the critical values of the tests are necessary due to the uncertainty of the estimates as representative values of the true parameter values.

2.5 Conclusion

This chapter has considered the characterization of non-stationarity for univariate time series. Testing for non-stationarity is mainly considered in the context of univariate models that are autoregressive (Dickey and Fuller 1979), and thus focuses on the presence of unit roots. The unit root is a very powerful property of a time series, and in cases where there are a reasonable number of observations, it is generally fairly straightforward to determine its presence or otherwise. Not only that, but its presence or absence has important structural implications since unit roots are associated with long-run behaviour.

A powerful tool of empirical analysis can be based on testing, whether such long-run behaviour is exhaustively shared by a set of time series. This is the notion of cointegration. To develop the idea, it is necessary to consider the relationship between series, rather than the properties of individual series alone. This is the subject of the next chapter.

3
Relationships Between Non-Stationary Time Series

3.1 Introduction

The previous chapter dealt with the properties of univariate time series, and in particular non-stationarity as characterized by the autoregressive unit root. This chapter develops the theme by looking at the way in which this type of non-stationarity can be modelled as a common feature such that the non-stationarity in one series is fully explained by that present in an appropriate combination of other series. It is natural to think of this in terms of a single regression equation.

The unit root corresponds to long-run behaviour of a series, that is to a component that has an arbitrarily low frequency. Thus, an equation which fully explains unit root behaviour can be thought of as fully describing the long-run relationship between the series concerned, or, in other words, it describes the underlying equilibrium behaviour. If the equation fails to capture all unit root behaviour it cannot be an equilibrium relationship.

These ideas are discussed below. The context is intuitively appealing, being that of a single equation with an unambiguous distinction between dependent variable and (weakly exogenous) regressors.[1] This does have limitations, however, among which is that only one equilibrium relationship can be considered. This is relaxed in later chapters.

3.2 Equilibrium and equilibrium correction

3.2.1 Long-run relationships

The idea of equilibrium is fundamental to the interrelationship of economic processes. In the time series econometric context, the idea is encapsulated in many ways. In general, the concept implies an underlying relationship about which the process or processes under examination vary, without deviating too far or for too long away from the values that would have to exist if the relationship held exactly at each period in time.

3.2.1.1 In a static model

A static model is one in which all the processes appear with the same time index so that only current values are concerned and there are no intertemporal links between them. Consider two scalar processes, z_t and y_t. Suppose there exists an exact linear relationship between them so that, at all points in time t

$$y_t = \mu + \beta_0 z_t. \tag{3.1}$$

There is never any deviation from this relationship. To emphasize that there is zero deviation, rewrite (3.1) as

$$y_t - \mu - \beta_0 z_t = 0. \tag{3.2}$$

However, rather than hold exactly, (3.2) might be subject to deviation. So, for a given value z_t, if the relationship (3.2) held exactly, the value of the y process would be

$$y_t^e = \mu + \beta_0 z_t. \tag{3.3}$$

But the y process is not equal to y_t^e but some other value, simply y_t. Denote the difference between these two, the extent to which the exact relationship does not hold, by

$$\eta_t = y_t - y_t^e. \tag{3.4}$$

If the η_t process varies about zero with a controlled size, then it is reasonable to regard the exact relationship (3.3) as the underlying relationship between the variables. Such a relationship is referred to as a long-run relationship. If, on the other hand, the deviations η_t seem to grow without bound, or become increasing dispersed about zero, then the exact relationship seems to be irrelevant. The stochastic property required of the deviations η_t is stationarity (and zero mean). Substituting (3.3) into (3.4) and rearranging gives

$$y_t = \mu + \beta_0 z_t + \eta_t. \tag{3.5}$$

Taking expectations of this gives

$$E(y_t) = \mu + \beta_0 E(z_t) + E(\eta_t). \tag{3.6}$$

Then, if $E(\eta_t) = 0$ and

$$E(y_t) = \bar{y},$$
$$E(z_t) = \bar{z},$$

equation (3.6) can be written

$$\bar{y} = \mu + \beta_0 \bar{z}. \tag{3.7}$$

This is the same functional relationship as (3.3), that is, it is the underlying or long-run relationship. The sequence of operations leading to (3.7) can be stylized as follows:

(i) Assume the processes z_t and y_t have settled to fixed values $\bar{z}$ and $\bar{y}$ respectively.
(ii) Assume also that there are no more deviations to the system, i.e. assume $\eta_t = 0$ (which can be regarded as its settled value).
(iii) Substitute these values into the complete relationship (3.5).

The resultant function relates the long-run static values of the variables. It is known as the static equilibrium. The condition that all variables have settled down in this way is known as the steady state.

The treatment here attempts to point out that while it is perfectly possible to make the above substitutions and obtain the long-run static solution in this way, this does not prove its existence. Rather it says that if z_t and y_t settle to fixed values and disturbances are stationary then the long-run solution to the model is given by (3.7). This discussion also indicates that the origin of these settled values should be the expected value of the processes.

3.2.1.2 In a dynamic model

Consider the model

$$y_t = \mu + \alpha_1 y_{t-1} + \beta_0 z_t + \beta_1 z_{t-1} + u_t. \tag{3.8}$$

where the u_t are disturbances to the dynamic relationship, but it remains to be seen how these relate to the deviations from equilibrium.[2] Since this relationship includes lags, it is said to be dynamic. Taking expected values of (3.8) treating all variables as stochastic, gives

$$E(y_t) = \mu + \alpha_1 E(y_{t-1}) + \beta_0 E(z_t) + \beta_1 E(z_{t-1}) + E(u_t). \tag{3.9}$$

If it is assumed that

$$E(y_t) = E(y_{t-1}) = \bar{y}, \tag{3.10a}$$
$$E(z_t) = E(z_{t-1}) = \bar{z}, \tag{3.10b}$$
$$E(u_t) = 0, \tag{3.10c}$$

then (3.9) can be used to derive a relationship between $\bar{z}$ and $\bar{y}$. Substituting equations (3.10a), (3.10b) and (3.10c) into (3.9) and rearranging gives

$$\bar{y} = \frac{\mu}{(1-\alpha_1)} + \frac{(\beta_0 + \beta_1)}{(1-\alpha_1)} \bar{z}. \tag{3.11}$$

3.2.2 Equilibrium and equilibrium error

Equation (3.11) allows the definition of deviations from equilibrium. It can be rearranged as

$$\bar{y} - \frac{\mu}{(1-\alpha_1)} - \frac{(\beta_0 + \beta_1)}{(1-\alpha_1)} \bar{z} = 0. \tag{3.12}$$

The left-hand side of (3.12) can be evaluated at a pair of actual values (z_t, y_t). If the system was in equilibrium, this should be zero. The extent to which it is not zero is the equilibrium error, which has been denoted η_t. That is,

$$y_t - \frac{\mu}{(1-\alpha_1)} - \frac{(\beta_0+\beta_1)}{(1-\alpha_1)} z_t = \eta_t \tag{3.13}$$

This defines the equilibrium error. If this process is non-stationary then it doesn't make much sense to regard the long-run solution as an equilibrium of course.

3.2.3 Equilibrium correction

The dynamic model (3.8) can be rewritten in terms of the equilibrium error (3.13). Subtracting y_{t-1} from both sides and adding $\beta_0 z_{t-1} - \beta_0 z_{t-1}$ to the right-hand side of (3.8) gives

$$\begin{aligned}\Delta y_t &= \mu + \beta_0 \Delta z_t + (\alpha_1 - 1) y_{t-1} + (\beta_0 + \beta_1) z_{t-1} + u_t \\ &= \mu + \beta_0 \Delta z_t - (1-\alpha_1) y_{t-1} + (\beta_0 + \beta_1) z_{t-1} + u_t\end{aligned}$$

which, on grouping terms in the lagged levels on the right-hand side, gives

$$\Delta y_t = \mu + \beta_0 \Delta z_t - (1-\alpha_1)\left(y_{t-1} - \frac{(\beta_0+\beta_1)}{(1-\alpha_1)} z_{t-1}\right) + u_t \tag{3.14}$$

From (3.13),

$$y_{t-1} - \frac{(\beta_0+\beta_1)}{(1-\alpha_1)} z_{t-1} = \eta_{t-1} + \frac{\mu}{(1-\alpha_1)},$$

which on substitution into (3.14) gives[3]

$$\Delta y_t = \beta_0 \Delta z_t - (1-\alpha_1)\eta_{t-1} + u_t. \tag{3.15}$$

This can also be called an equilibrium correction model.[4] Changes in y_t are seen to be due to changes in z_t and the extent to which the system was out of equilibrium in the past period, that is η_{t-1}. From (3.13), the equilibrium error is positive if the y process is of a value higher than is consistent with equilibrium ($\eta_t > 0 \Leftrightarrow y_t > \frac{\mu}{(1-\alpha_1)} + \frac{(\beta_0+\beta_1)}{(1-\alpha_1)} z_t$). This suggests that such an error should exert a downward pressure on y_t in the next period, in other words that there should be a negative pressure on the change. This means the coefficient on the lagged equilibrium error in (3.15) ought to be negative if this simple behavioural rule applies. Similarly, if the equilibrium error is negative, there should be upward pressure, and again the argument is for a negative coefficient on η_{t-1}. In this simple model this requires $\alpha_1 < 1$.

The speed of adjustment to equilibrium is measured by the size of the coefficient on the disequilibrium error ($1 - \alpha_1$). The larger this is in absolute

value (assuming it is of the "correct" negative sign), the quicker is the adjustment.

3.2.4 Equilibrium correction and autoregressive distributed lag models in general

Equation (3.8) is an example of an autoregressive distributed lag (ADL) model. There are two variables in the model, each appearing with a maximum lag of 1. The model is therefore referred to as an ADL(1, 1) model, the first number of the ordered pair referring to the maximum lag of the dependent variable. Note that (3.8) could be written in terms of lag polynomials. Define

$$\alpha(L) = 1 - \alpha_1 L$$
$$\beta(L) = \beta_0 + \beta_1 L$$

Then (3.8) can be written

$$\alpha(L) y_t = \mu + \beta(L) z_t + u_t. \tag{3.16}$$

The long-run solution and hence the equilibrium error can also be written in terms of these polynomials since

$$\alpha(1) = 1 - \alpha_1$$
$$\beta(1) = \beta_0 + \beta_1$$

which can be substituted into (3.13) to give

$$\eta_t = y_t - \frac{\mu}{\alpha(1)} - \frac{\beta(1)}{\alpha(1)} z_t. \tag{3.17}$$

Similarly, the ECM may be written

$$\Delta y_t = \beta_0 \Delta z_t - \alpha(1) z_{t-1} + u_t. \tag{3.18}$$

These results can be generalized for the ADL (m, n) model which is (3.16) with

$$\alpha(L) = 1 - \sum_{i=1}^{m} \alpha_i L^i, \tag{3.19a}$$

$$\beta(L) = \sum_{i=0}^{n} \beta_i L^i. \tag{3.19b}$$

Using a slightly generalized version of the reparameterization used in section 2.4.2, equation (2.34a), that is if

$$\gamma(L) = \gamma_0 - \sum_{i=1}^{p} \gamma_i L^i \tag{3.20a}$$

then it can be written

$$\gamma(L) = \gamma(1) L + \gamma^*(L)(1 - L) \tag{3.20b}$$

where

$$\gamma^*(L) = \gamma_0 - \sum_{i=1}^{p-1} \gamma_i^* L^i \tag{3.20c}$$

$$\gamma_i^* = -\sum_{i=i+1}^{p} \gamma_i, i = 1,\ 2,\ \ldots,\ p-1. \tag{3.20d}$$

Thus (3.16) becomes

$$(\alpha(1)L + \alpha^*(L)(1-L)y_t = \mu + \beta(1)L + (\beta^*(L)(1-L))z_t + u_t$$

This can be rearranged as

$$\alpha^*(L)\Delta y_t = \mu - \alpha(1)y_{t-1} + \beta(1)z_{t-1} + \beta^*(L)\Delta z_t + u_t$$

$$\Rightarrow$$

$$a^*(L)\Delta y_t = \mu + \beta^*(L)\Delta z_t - \alpha(1)\left(y_{t-1}\frac{\beta(1)}{\alpha(1)}z_{t-1}\right) + u_t$$

or

$$\alpha^*(L)\Delta y_t = \beta^*(L)\Delta z_t - \alpha(1)\left(y_{t-1} - \frac{\mu}{\alpha(1)} - \frac{\beta(1)}{\alpha(1)}z_{t-1}\right) + u_t \tag{3.21}$$

Equation (3.21) is the ECM in this general case. Noting that the lag of a fixed value is the same as the fixed value, $L^i\bar{y} = \bar{y}$ for $i = 0,\ 1,2\ \ldots$, and so

$$\alpha(L)\bar{y} = \left(1 - \sum_{i=1}^{p}\alpha_i L^i\right)\bar{y} = \bar{y} - \sum_{i=1}^{p}\alpha_i\bar{y} = \left(1 - \sum_{i=1}^{p}\alpha_i\right)\bar{y} = \alpha(1)\bar{y}$$

and similarly

$$\beta(L)\bar{z} = \beta(1)\bar{z}$$

the long-run static solution to the model can be written

$$\bar{y} - \frac{\mu}{\alpha(1)} - \frac{\beta(1)}{\alpha(1)}\bar{z} = 0$$

and hence the equilibrium error as (3.17) but using the operators (3.19a and 3.19b). The ECM may therefore be written in terms of the equilibrium error as

$$\alpha^*(L)\Delta y_t = \beta^*(L)\Delta z_t - \alpha(1)\eta_{t-1} + u_t.$$

Equation (3.19a) gives the form of $\alpha\ (L)$ and from (3.20a to 3.20d) it can be seen that $\alpha^*(0) = 1$ and $\alpha^*(L)$ is of order $m - 1$, that is

$$\alpha^*(L) = 1 - \sum_{i=1}^{m-1}\alpha_i^* L^i$$

Similarly

$$\beta^*(L) = \beta_0 + \sum_{i=1}^{n-1} \beta_i^* L^i$$

and so the ECM can be rewritten

$$\Delta y_t = \sum_{i=1}^{m-1} \alpha_i^* \Delta y_{t-i} + \sum_{i=0}^{n-1} \beta_i^* \Delta z_t - \alpha(1)\eta_{t-1} + u_t \tag{3.22}$$

equation (3.22) shows the ECM reparameterization of the ADL (*m*, *n*) model is in terms of the differences of the processes and the lagged equilibrium error, the maximum lag of the differences of each variable being one less than the maximum lag of the level in the ADL. Notice that the current value of z_t appears on the right-hand side of (3.22), that is the summation involving its lags begins at 0, while the summation involving the lags of y_t begins at 1 because the current value is on the left-hand side of the equation.

3.2.4.1 Solving the ECM for the long-run solution

The ECM reparameterization of the ADL is performed without any initial reference to the long-run solution of the model. That is

$$\alpha(L)y_t = \mu + \beta(L)z_t + u_t$$
$$\Leftrightarrow$$
$$a^*(L)\Delta y_t = \beta^*(L)\Delta z_t - \alpha(1)\left(y_{t-1} - \frac{\mu}{\alpha(1)} - \frac{\beta(1)}{\alpha(1)} z_{t-1}\right) + u_t.$$

The ECM provides an immediate calculation of the long-run solution. Under the assumption of a steady state without growth:[5]

$$y_t = \bar{y}, z_t = \bar{z}, u_t = 0, \forall t$$

and hence

$$\Delta y_t = y_t - y_{t-1} = \bar{y} - \bar{y} = 0,$$
$$\Delta z_t = z_t - z_{t-1} = \bar{z} - \bar{z} = 0.$$

Substituting these equations into the ECM gives

$$\alpha^*(L)0 = \beta^*(L)0 - \alpha(1)\left(\bar{y} - \frac{\mu}{\alpha(1)} - \frac{\beta(1)}{\alpha(1)}\bar{z}\right) + 0$$

which on rearrangement gives

$$\bar{y} = \frac{\mu}{\alpha(1)} + \frac{\beta(1)}{\alpha(1)}\bar{z}$$

as the long-run or steady-state solution to the model.[6]

3.2.4.2 Generalization to ADL models with more than two variables

The reparameterization generalizes in a straightforward manner to the situation where there are more than two variables in the equation. The general ADL may be written

$$\alpha(L)y_t = \mu + \sum_{j=1}^{r} \beta_j(L)z_{j,t} + u_t$$

where $\alpha(L)$ and $\beta_j(L)$, $j = 1, 2, \ldots, r$ are lag polynomials of possibly different order and $\alpha(0) = 1$, and $z_{j,t}$, $j = 1, 2, \ldots, r$ are a set of explanatory variables. Then, using the reparameterization on each of the polynomials and rearranging, the ECM is of the form

$$\alpha^*(L)\Delta y_t = \sum_{j=1}^{r} \beta_j^*(L)\Delta z_{j,t} - \alpha(1)\left(y_{t-1} - \frac{\mu}{\alpha(1)} - \sum_{j=1}^{r} \frac{\beta_j(1)}{\alpha(1)} z_{j,t-1} \right) = u_t. \qquad (3.23)$$

The lag polynomials $\alpha^*(L)$ and $\beta_j^*(L)$ are interpreted as before and are of order one less than $\alpha(L)$ and $\beta_j(L)$, respectively.

3.2.5 Unit roots and the ECM reparameterization

Equations (3.14),[7] (3.21), and (3.23) demonstrate the ECM reparameterization of ADL models of increasing generality. But they all have the same structure and (3.23) covers them all. It is clear that the value of $\alpha(1)$ is very important. It directly determines the speed of adjustment to equilibrium as it is the coefficient of the lagged equilibrium error term. It should also be negative if the equilibrium error interpretation is to make sense. Furthermore, in order for the long run (i.e. equilibrium) solution to exist, it must be non-zero because it appears as a divisor in the equilibrium error term.

The condition $\alpha(1) \neq 0$ means the lag polynomial $\alpha(L)$ must have no unit roots. Of course, as $\alpha(1) \to 0$, the speed of adjustment gets slower. In the limit there is no equilibrium to which to adjust. This point is returned to in the context of non-stationary processes. It is also clear that if any of the lag polynomials on the explanatory variables, that is any of the $\beta_j(L)$ have unit roots, then the corresponding variable disappears from the long-run solution since then $\beta_j(1) = 0$. If $\alpha(L)$ has a unit root, then one avenue to consider, if economically meaningful, would be to work instead with the differences of y_t right from the start, since under these circumstances $\alpha(L)$ could be factorized thus extracting the unit root as

$$\alpha(L) = \tilde{\alpha}(L)(1 - L) = \tilde{\alpha}(L)\Delta$$

and so

$$\alpha(L)y_t = \tilde{\alpha}(L)(1 - L)y_t = \tilde{\alpha}(L)\Delta y_t$$

The ADL would then be

$$\tilde{\alpha}(L)\Delta y_t = \mu + \sum_{j=1}^{r} \beta_j(L) z_{j,t} + u_t$$

and the long-run solution

$$\overline{\Delta y} = \frac{\mu}{\tilde{\alpha}(1)} + \sum_{j=1}^{r} \frac{\beta_j(1)}{\tilde{\alpha}(1)} \overline{z}_j$$

where $\overline{\Delta y}$ means that the y_t process is still changing, but by the same (steady state) amount every period. This long-run equilibrium relates the steady-state change in the y process to the steady-state levels of the z processes. If there are unit roots in any of the β_j (L) the same approach could be used here resulting in the replacement of the steady-state level of the corresponding z_j by its steady-state change. The possibly uncomfortable result is a long-run solution that mixes equilibrium levels, changes along a steady-state growth path and changes that might be best described as flows rather than stocks.

To give a slightly more concrete example, suppose α (L) does not have a unit root, and suppose there are two explanatory variables, the lag polynomial on the second of which has a unit root. The ADL is

$$\alpha(L)y_t = \mu + \beta_1(L)z_{1,t} + \beta_2(L)z_{2,t} + u_t,$$
$$\alpha(1), \beta_1(1) \neq 0, \quad \beta_2(1) = 0.$$

There are then two possible interpretations in terms of the long-run solution (also see the discussion of I(2) processes in chapter 6). Leaving $\beta_2(L)$ as it is, that is not factorizing out the unit root, gives the long-run solution

$$\overline{y} = \frac{\mu}{\alpha(1)} + \frac{\beta_1(1)}{\alpha(1)}\overline{z}_1 + \frac{\beta_2(1)}{\alpha(1)}\overline{z}_2 = \frac{\mu}{\alpha(1)} + \frac{\beta_1(1)}{\alpha(1)}\overline{z}_1$$

that is, x_2 drops out of the long-run solution. Alternatively, β_2 (L) could be factorized so as to draw out the unit root, β_2 (L) = $\tilde{\beta}_2$ (L) Δ. The ADL would then be

$$\alpha(L)y_t = \mu + \beta_1(L)z_{1,t} + \tilde{\beta}_2(L)\Delta z_{2,t} + u_t$$

and the long-run solution

$$\overline{y} = \frac{\mu}{\alpha(1)} + \frac{\beta_1(1)}{\alpha(1)}\overline{z}_1 + \frac{\beta_2(1)}{\alpha(1)}\overline{\Delta x_2} \tag{3.24}$$

Equation (3.24) relates the steady-state level of y to the steady-state level of x_1 and the steady-state change of x_2 The speed of adjustment to equilibrium is the same in both cases, but in the former, the first differences of $z_{2,t}$ will appear in the short-run dynamics, whereas in the latter it will be the second differences (Δ^2). The ECMs for the two cases are given below in equations (3.25a) and (3.25b) respectively.

$$\alpha^*(L)\Delta y_t = \beta_1^*(L)\Delta z_{1,t} + \beta_2^*(L)\Delta z_{2,t} - \alpha(1)\left(y_{t-1} - \frac{\mu}{\alpha(1)} - \frac{\beta_1(1)}{\alpha(1)} z_{1,t-1}\right) + u_t, \tag{3.25a}$$

$$\alpha^*(L)\Delta y_t = \beta_1^*(L)\Delta z_{1,t} + \tilde{\beta}_2^*(L)\Delta^2 z_{2,t} - \alpha(1)\left(y_{t-1} - \frac{\mu}{\alpha(1)} - \frac{\beta_1(1)}{\alpha(1)} z_{1,t-1} - \frac{\beta_2(1)}{\alpha(1)} \Delta z_{2,t-1}\right) + u_t, \tag{3.25b}$$

This type of choice over specification is not very comfortable as the two long-run equilibria are different, one including a steady state change, the other not. However, some clarification is often available either from the underlying economic theory or, in the case where some or all of the variables are non-stationary, their orders of integration.

3.3 Cointegration and equilibrium

3.3.1 Equilibrium error and static equilibrium

For purposes of illustration, consider again the two variable ADL (m, n) model,

$$\alpha(L)y_t = \mu + \beta(L)z_t + u_t \qquad \text{(3.16) again}$$

where the lag polynomials are as in equations (3.19a, 3.19b). Assuming it exists, this has a long-run static solution given by

$$\bar{y} = \frac{\mu}{\alpha(1)} + \frac{\beta(1)}{\alpha(1)}\bar{z}. \tag{3.26}$$

In any period t, the equilibrium error will be

$$\eta_t = y_t - \frac{\mu}{\alpha(1)} - \frac{\beta(1)}{\alpha(1)} z_t. \tag{3.27}$$

But in what sense is (3.26) an equilibrium rather than just a long-run solution that may or may not be relevant? It won't be relevant if, for example, the variables do not tend to steady-state values, that is $\bar{z}$ or $\bar{y}$ don't exist. This depends on the properties of the error sequence, η_t. Writing (3.27) as

$$y_t = \frac{\mu}{\alpha(1)} + \frac{\beta(1)}{\alpha(1)} z_t + \eta_t$$

shows that y_t is its putative equilibrium value, $\frac{\mu}{\alpha(1)} + \frac{\beta(1)}{\alpha(1)} z_t$, plus a deviation, η_t from this value. If these deviations display any form of permanence, then it is not sensible to regard (3.26) as the underlying relationship. One definition of a lack of permanence is stationarity, or, more precisely, the property of

being I(0). If η_t is I(1), then the idea that (3.26) represents an equilibrium is entirely unhelpful. Granger (1991) and Engle and Granger (1991) compare the properties of I(1) and I(0) variables. An I(0) series has a mean and there is a tendency for the series to return to this value frequently with deviations that are large on a relatively small number of occasions. An I(1) process will wander widely, only rarely return to an earlier value and its autocorrelations will remain close to one even at long lags. Theoretically, the expected time it takes for a random walk to return to some fixed value is infinite.[8] Clearly then, it makes no sense whatsoever for η_t to be I(1), but it does seem reasonable to require it to be I(0). There is the issue of its mean value as well. Clearly this should be zero, although this does not affect the time series properties of the variable, meaning its stationarity, variance, and autocorrelation structure.[9]

So the working definition of a static equilibrium will be as follows.

3.3.1.1 Static equilibrium

The relationship

$$\bar{y} = \delta_0 + \delta_1 \bar{z} \tag{3.28}$$

is a static equilibrium relationship for the processes z_t and y_t if

$$y_t - \delta_0 - \delta_1 z_t \sim I(0). \tag{3.29}$$

More generally, there may be an arbitrary number of variables and the function need not be linear. Engle and Granger (1991) use the term attractors to describe relationships such as (3.28) when (3.29) holds.

3.3.2 Static equilibrium with I(1) variables

3.3.2.1 Sums of ARMA processes

Suppose that z_t and y_z have ARMA(p_z, q_z) and ARMA(p_y, q_y) representations respectively. In particular, let

$$\begin{aligned} \phi_z(L) z_t &= \theta_z(L) \epsilon_{z,t} \\ \phi_y(L) y_t &= \theta_y(L) \epsilon_{y,t} \end{aligned} \tag{3.30}$$

where $\epsilon_{z,t}$ and $\epsilon_{y,t}$ are two white noise processes. These relationships can be used to obtain the ARMA representations for any linear combination of z_t and y_t Consider

$$w_t = z_t + y_t. \tag{3.31}$$

Equations in (3.30) indicate that it will be necessary to work with lag operators applied to both processes. To remove z_t from (3.31) multiply through by $\phi_z(L)$, to give

$$\phi_z(L) w_t = \phi_z(L) z_t + \phi_z(L) y_t = \mathrm{q}_z(L) \epsilon_{z,t} + \phi_z(L) y_t. \tag{3.32}$$

To substitute for y_t multiply (3.32) through by $\phi_y(L)$, to give

$$\begin{aligned}\phi_y(L)\phi_z(L)w_t &= \phi_y(L)\theta_z(L)\epsilon_{z,t} + \phi_y(L)\phi_z(L)y_t \\ &= \phi_y(L)\theta_z(L)\epsilon_{z,t} + \phi_z(L)\phi_y(L)y_t \\ &= \phi_y(L)\theta_z(L)\epsilon_{z,t} + \phi_z(L)\theta_y(L)\epsilon_{y,t}. \end{aligned} \tag{3.33}$$

The left-hand side of (3.33) is easily simplified by writing

$$\phi(L) = \phi_y(L)\phi_z(L) \tag{3.34}$$

which is a polynomial lag operator of order $p = p_z + p_y$. The last line of (3.33) is the sum of two MA processes, $\phi_y(L)\,\theta_z(L)\,\epsilon_{z,t}$ and $\phi_z(L)\,\theta_y(L)\,\epsilon_{y,t}$, of orders $p_y + q_z$ and $p_z + q_y$ respectively. As long as the white noise processes $\epsilon_{z,t}$ and $\epsilon_{y,t}$ are only contemporaneously correlated (i.e. $E(\epsilon_{z,t-i}, \epsilon_{y,t-i}) = 0$ if $i \neq j$), then the autocorrelations of the sum of these process will extend only as far as the larger of the two individual orders. That is, the sum will be a MA process whose order is the larger of $p_y + q_z$ and $p_z + q_y$. The variance of the new white noise driving sequence and the MA coefficients will depend on the variance–covariance matrix of $\epsilon_{y,t}$ and $\epsilon_{z,t}$ and the coefficient values of the original operators, $\phi_z(L)$, $\theta_z(L)$, $\phi_y(L)$ and $\theta_y(L)$.[10] Thus, the final line of (3.33) may be written

$$\phi_y(L)\theta_z(L)\epsilon_{z,t} + \phi_z(L)\theta_y(L)\epsilon_{y,t} = \theta(L)\epsilon_t \tag{3.35}$$

where $\theta(L)$ is a lag polynomial of order $q = max(p_y + q_z, p_z + q_y)$, and the variance of ϵ_t is chosen so that $\theta(0) = 1$. Thus the time series model for w_t is ARMA(p,q), were

$$\begin{aligned} p &= p_z + p_y \\ q &= max(p_y + q_z, p_z + q_y) \\ \phi(L)w_t &= \theta(L)\epsilon_t. \end{aligned}$$

From (3.34), the roots of $\phi(L)$ will be those of $\phi_z(L)$ and $\phi_y(L)$. Consider three important cases.

(i) If all the roots of these lie outside the unit circle, then all the roots of $\phi(L)$ lie outside the unit circle and so w_t is stationary. This means if z_t and y_t are stationary so is their sum, w_t.

(ii) Suppose z_t is I(1) and y_t I(0). Then $\phi(L)$ has one unit root, all the other lying outside the unit circle and all the roots of $\phi(L)$ lie outside the unit circle. Since $\phi(L) = \phi_y(L)\,\phi_z(L)$, it follows that $\phi(L)$ has one unit root, all other outside the unit circle. Thus, w_t is I(1). This means that the sum of an I(0) and an I(1) process is I(1).

(iii) Suppose that both z_t and y_t are I(1). This case is a little more complicated and it is necessary to go back to the working used when obtaining the ARMA structure for the sum. Consider equation (3.33), using the last equality on the right-hand side,

$$\phi_y(L)\phi_z(L)w_t = \phi_y(L)\theta_z(L)\epsilon_{z,t} + \phi_z(L)\theta_y(L)\epsilon_{y,t}.$$

Since z_t and y_t are I(1), the AR operators from their separate ARMA representations may be written in terms of a new AR operator consisting of all and only the stationary roots, and the differencing operator. So:

$$\phi_z(L) = \tilde{\phi}_z(L)\Delta$$
$$\phi_y(L) = \tilde{\phi}_y(L)\Delta$$

where $\tilde{\phi}_z(L)$ and $\tilde{\phi}_y(L)$ are of orders $p_z - 1$ and $p_y - 1$ respectively. Thus (3.33) may be written

$$\begin{aligned}\tilde{\phi}_y(L)\tilde{\phi}_z(L)\Delta^2 w_t &= \tilde{\phi}_y(L)\Delta\theta_z(L)\epsilon_{z,t} + \tilde{\phi}_z(L)\Delta\theta_y(L)\epsilon_{y,t} \\ &= \Delta(\tilde{\phi}_y(L)\theta_z(L)\epsilon_{z,t} + \tilde{\phi}_z(L)\theta_y(L)\epsilon_{y,t}).\end{aligned}$$

The common factor of Δ can now be cancelled on each side of the equation[11] to give

$$\tilde{\phi}_y(L)\tilde{\phi}_z(L)\Delta w_t = \tilde{\phi}_y(L)\theta_z(L)\epsilon_{z,t} + \tilde{\phi}_z(L)\theta_y(L)\epsilon_{y,t}. \tag{3.36}$$

Thus w_t has one (not two) unit roots and Δw_t has an ARMA(p^*, q^*) structure where

$$p^* = p_z + p_y - 2,$$
$$q^* = max(p_y + q_z + p_z + q_y) - 1.$$

Alternatively, in ARIMA terminology, w_t is ARIMA(p^*, 1, q^*). This important result shows that the sum of two I(1) processes is also I(1).

3.3.2.2 Linear functions of ARMA processes

It is also easily shown that multiplying or adding constants alters neither the ARMA orders nor their integration properties. Let

$$\phi(L)z_t = \theta(L)\epsilon_t, \tag{3.37}$$

and define

$$\tilde{z}_t = z_t + \delta \tag{3.38}$$

where δ is a constant. Multiplying both sides of (3.38) by $\phi(L)$ gives

$$\phi(L)\tilde{z}_t = \phi(L)z_t + \phi(L)\delta = \theta(L)\epsilon_t + \phi(1)\delta. \tag{3.39}$$

In the stationary case, $E(\tilde{z}_t) = \delta$, $Var(\tilde{z}_t) = Var(z_t)$, and the covariances are given by $\gamma_{\tilde{z}}(j) = E((\tilde{z}_t - \delta)(\tilde{z}_{t-j} - \delta)) = E(z_t z_{t-j})$, so are the same as those of the original process, z_t, and since the variance is also unchanged, so are the autocorrelations. Thus although a constant has to be added to the model, it is otherwise unchanged.

When z_t is I(1), it is the case that ϕ (1) = 0 since ϕ (L) has a unit root. Equation (3.39) can therefore be written as:

$$\phi(L)\tilde{z}_t = \theta(L)\epsilon_t$$

which is exactly the same process as the original. To show multiplying by a constant makes no difference, continue to define the z_t process by equation (3.37), and consider the transformed process,

$$\breve{z}_t = \lambda z_t \tag{3.40}$$

Multiplying both sides of (3.40) by ϕ (L) gives

$$\phi(L)\breve{z}_t = \phi(L)\lambda z_t = \lambda\phi(L)\lambda z_t = \lambda\theta(L)\epsilon_t = \theta(L)\lambda\epsilon_t.$$

Since ϵ_t is zero mean white noise, so is any scalar multiple, so $\lambda\epsilon_t$ is zero mean white noise. The structure is therefore unchanged as no new autocorrelation has been induced. To summarize, if z_t is ARMA (p, q) then so is $\nu_t = \delta + \lambda z_t$ and furthermore this process has the same autocorrelation structure, so that its AR and MA operators are unchanged. In particular, if z_t is I(d), then so is any linear transformation. Mathematically,

$$\begin{aligned} &\phi(L)z_t = \theta(L)\epsilon_t \\ &\Rightarrow \phi(L)\nu_t = \mu + \theta(L)\epsilon_t \end{aligned} \tag{3.41}$$

where

$$\begin{aligned} \nu_t &= \delta + \lambda z_t \\ \mu &= \phi(1)\delta. \end{aligned}$$

3.3.2.3 Linear combinations of ARIMA processes

Let z_t and y_t be the ARIMA(p_z, d_z, q_z) and ARIMA(p_y, d_y, q_y), where d_z, d_y = 0, 1. Then (3.1) states that $\nu_t = \delta + \lambda z_t$ is ARIMA(p_z, d_z, q_z), and, defining

$$\xi_t = y_t + \delta + \lambda z_t \tag{3.42}$$

it follows from the results for the sum of two ARIMA processes that ξ_t is ARIMA(p^*, d^*, q^*) where

$$\begin{aligned} p^* &= p_z + p_y, \\ q^* &= max(p_y + q_z, p_z + q_y), \\ d^* &= max(d_z, d_y). \end{aligned}$$

Equations (3.43) are easily generalized to show that any linear combination of an ARIMA(p_z, d_z, q_z) and an ARIMA(p_y, d_y, q_y) is ARIMA(p^*, d^*, q^*), with p^*, d^*, q^* as defined by equations (3.43). In particular:

(i) A linear combination of I(0) processes is I(0).
(ii) A linear combination of an I(0) and an I(1) process is I(1).
(iii) A linear combination of I(1) processes is I(1).

These results generalize trivially to the case of a linear combination of an arbitrary number of ARIMA processes as follows. If $x_{i,t}$ is ARIMA(p_i, d_i, q_i), $d_i = 0, 1$, $i = 1, 2, \ldots, n$, and

$$\xi_t = \delta_0 + \sum_{i=1}^{n} \delta_i x_{i,t}$$

for a set of constants δ_t, $i = 0, 1, \ldots, n$, then χ_t is I(d*), where $d^* = max(d_i)$.

3.3.3 Cointegration: static equilibrium with I(1) variables

Thus it appears that a linear combination of two I(1) variables is also I(1), not I(0). But, if static equilibrium is to exist between I(1) variables, a linear combination has to exist that is not I(1), but I(0). For some processes, this cannot happen. But it can happen that there exists a special association between the processes such that a special linear combination does result in a stationary series. This is called a cointegrating combination.

Rather than look for the properties of a special pair of I(1) series such that cointegration can result, it is easier to construct such a pair directly.[12] To keep things simple, suppose z_t is a pure random walk given by

$$\Delta z_t = \epsilon_{z,t}$$

where $\epsilon_{z,t}$ is white noise. Let $\epsilon_{y,t}$ be another white noise process, uncorrelated with $\epsilon_{z,t}$, and define

$$y_t = z_t + \epsilon_{y,t}. \tag{3.44}$$

If y_t is I(1), then z_t and y_t are cointegrated since

$$y_t - z_t = \epsilon_{y,t}$$

is a linear combination of I(1) processes which, being white noise, is I(0). To show that y_t is indeed I(1), note that the right-hand side of (3.44) is the sum of an I(1) and an I(0) process. It has been shown that an I(1) plus an I(0) process is I(1). Thus, y_t is I(1) and so z_t and y_t are cointegrated.

Given that earlier it was shown that any linear combination of an ARIMA(p_z, d_z, q_z) and an ARIMA(p_y, d_y, q_y) is ARIMA(p^*, d^*, q^*), with p^*, d^*, q^* as defined by equations (3.43), some further explanation is needed here. In particular, this states that the order of integration is $d^* = max(d_z, d_y)$. In fact, all the orders reported in equations (3.43) are upper bounds on the orders of the model, because it is possible that simpler representations may exist if the moving average polynomial has some roots in common with those of the autoregressive polynomial, or if it has any unit roots. Such common roots are known as common factors. For example, suppose that z_t and y_t have ARIMA(p_z, d_z, q_z) and an ARIMA(p_y, d_y, q_y) representations respectively, given by

$$\phi_z(L)\Delta^{d_z} z_t = \theta_z(L)\epsilon_{z,t},$$
$$\phi_y(L)\Delta^{d_y} y_t = \theta_y(L)\epsilon_{y,t},$$

and define the linear combination

$$\eta_t = \delta_z z_t + \delta_y y_t$$

Then η_t has the ARIMA representation

$$\phi^*(L)\Delta^{d^*}\eta_t = \theta^*(L)\epsilon_t \tag{3.45}$$

for some white noise process ϵ_t, where $\phi^*(L)$ and $\theta^*(L)$ are lag polynomials of orders p^* and q^* as defined by equations (3.43), and $d^* = max(d_z, d_y)$. But it is possible that $\phi^*(L)$ and $\theta^*(L)$ contain a common factor, say $(1 - \gamma L)$, so that

$$\begin{aligned}\phi^*(L) &= (1-\gamma L)\tilde{\phi}^*(L),\\ \theta^*(L) &= (1-\gamma L)\tilde{\theta}^*(L),\end{aligned} \tag{3.46}$$

the polynomials $\tilde{\phi}^*(L)$ and $\tilde{\theta}^*(L)$ being of orders $p^* - 1$ and $q^* - 1$ respectively. Substituting (3.1) into (3.45) gives

$$(1-\gamma L)\tilde{\phi}^*(L)\Delta^{d^*}\eta_t = (1-\gamma L)\tilde{\theta}^*(L)\epsilon_t \tag{3.47}$$

which on cancelling the common factor $(1 - \gamma L)$ becomes

$$\tilde{\phi}^*(L)\Delta^{d^*}\eta_t = \tilde{\theta}^*(L)\epsilon_t,$$

which is an ARIMA($p^* - 1$, d^*, $q^* - 1$) model. As a special case of this, the MA lag polynomial may have a unit root, and therefore be written

$$\theta^*(L) = (1-L)\tilde{\theta}^*(L).$$

Substituting this into (3.45) gives

$$\phi^*(L)\Delta^{d^*}\eta_t = (1-L)\tilde{\theta}^*(L)\epsilon_t$$

which on cancelling with one of the unit roots represented by Δ^{d^*} (and ignoring the generation of a possibly non-zero mean) gives

$$\phi^*(L)\Delta^{d^*-1}\eta_t = \tilde{\theta}^*(L)\epsilon_t.$$

In this case η_t is an ARIMA($p^* - 1$, $d^* - 1$, $q^* - 1$) process.

A special case of this last example is the case of cointegration where $\phi^*(L)$ has no unit roots, and $d_z = d_y = d^* = 1$. Then z_t and y_t are I(1) but η_t is I(0). Hence z_t and y_t are cointegrated.

The general result for the linear combination of two ARIMA processes is considered next.

3.3.3.1 Linear combinations of ARIMA processes

The linear combination of an ARIMA(p_z, d_z, q_z) and an ARIMA(p_y, d_y, q_y) process will be an ARIMA($\tilde{p}$, $\tilde{d}$, $\tilde{q}$) process

$$\tilde{\phi}(L)\Delta^{\tilde{d}}\eta_t = \tilde{\theta}(L)\epsilon_t$$

where $\tilde{\theta}(L)$ is invertible and $\tilde{\phi}(L)$ and $\tilde{\theta}(L)$ have no common factors, with the orders given by

$$\tilde{p} \leq p_z + p_y$$
$$\tilde{q} \leq max(p_y + q_z, p_z + q_y)$$
$$\tilde{d} \leq max(d_z, d_y)$$

with equality only if no common factors have been cancelled.

3.3.3.2 Example

Figure 3.1 shows a time series plot of two series generated artificially according to equations

$$\Delta z_t = \epsilon_{z,t}, \tag{3.48a}$$
$$y_t = 1 + z_t + \epsilon_{y,t}, \tag{3.48b}$$

where $\epsilon_{z,t}$ and $\epsilon_{y,t}$ are two independent NIID(0,1) series. Both z_t and y_t are I(1) and cointegrated by construction. Figure 3.2 shows the same data using a scatter plot.

The time series plots indicate the non-stationary nature of both series, and that, in this case, they track one another very closely. The latter property is not necessary for two series to be cointegrated. It is quite possible that an increasing gap may open up between them. This depends on exactly what the cointegrating combination is. The scatter plot strongly emphasizes the linear nature of the underlying long run, and in this case equilibrium relationship, which is, from (3.48b), $\bar{y} = 1 + \bar{z}$.

It is also important to realize that the cointegrating property depends on the selection of the correct linear combination. Using equation (3.48b), the linear combination generating cointegration is $\eta_{1,t} = y_t - z_t$. This is stationary by construction. But suppose instead, the combination $\eta_{2,t} = y_t - \frac{1}{2}z_t$ is considered. Subtracting $-\frac{1}{2}z_t$ from both sides of (3.48b) gives,

$$y_t - \frac{1}{2}z_t = 1 + \frac{1}{2}z_t + \epsilon_{y,t}. \tag{3.49}$$

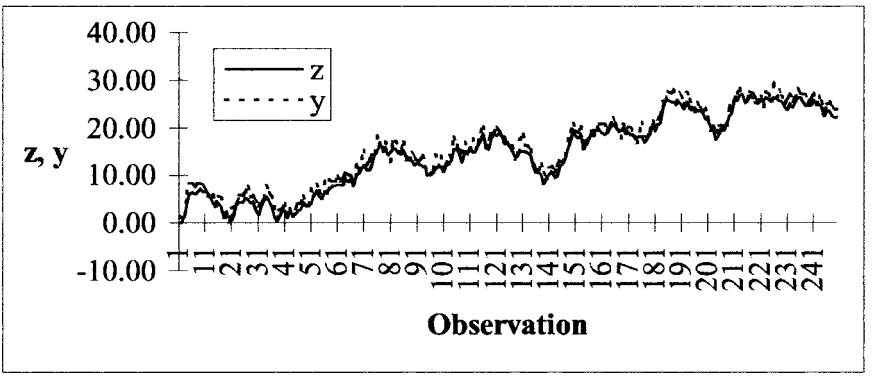


Figure 3.1 Time series plot of artificially generated cointegrated series

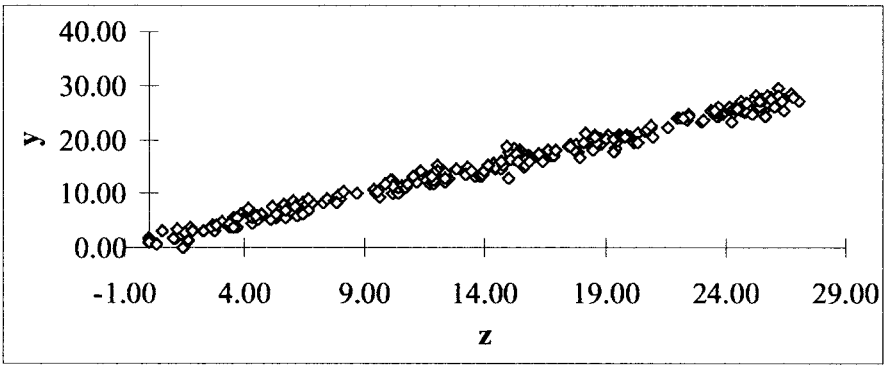


Figure 3.2 Scatter plot of artificially generated cointegrated series

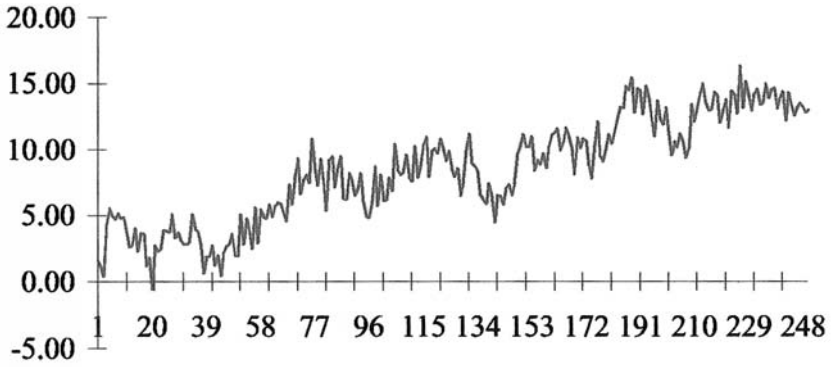


Figure 3.3 Plot of $\eta_{2,t}$, non-cointegrating combination of cointegrating variables

But $-\frac{1}{2}z_t$ is I(1) and $\epsilon_{y,t}$ is I(0), so from (3.49), $\eta_{2,t}$ is I(1), and so non-stationary. This illustrates a key point: to obtain cointegration where it exists, the correct linear combination must be used.

Figure 3.3 shows clearly that this combination is not stationary and so not a cointegrating combination. To illustrate the case where a cointegrating combination still results in a gap opening up between series represented on a time series plot, suppose instead of (3.48b), a series y_t^* is generated according to

$$y_i^* = 1 + \frac{1}{2}z_t + \epsilon_{y,t}. \tag{3.50}$$

Figure 3.4 is a time series plot of z_t and y_t^*. It would be wrong to conclude from this that just because the series are diverging that they are not cointegrated. It is simply that the difference between the two is not the cointegrating combination.

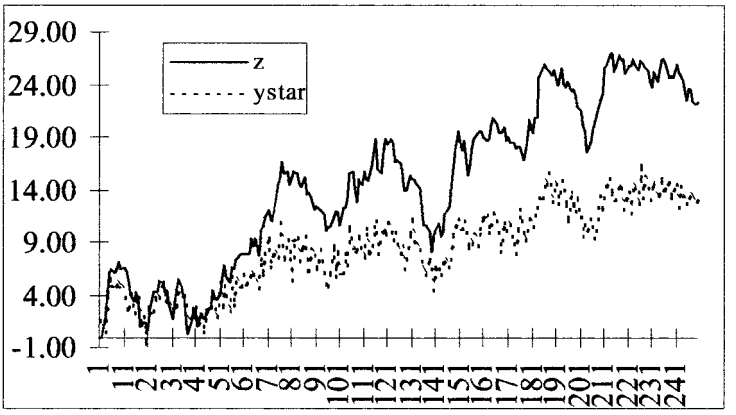


Figure 3.4 Time series plot of y_t^* and z_t generated according to (3.48a) and (3.50)

3.3.4 ADL models, cointegration, and equilibrium

There is a very strong link between cointegration and equilibrium when a set of I(1) variables are related according to an ADL model with stationary disturbances. To introduce the idea, consider a reparameterization of the ADL(1, 1) model (3.8) into an ECM:

$$\Delta y_t = \beta_0 \Delta z_t - (1-\alpha_1)\left(y_{t-1} - \frac{\mu}{1-\alpha_1} - \frac{\beta_0+\beta_1}{1-\alpha_1} z_{t-1} \right) + u_t. \tag{3.51}$$

Note that the static long run solution is

$$\bar{y} = \frac{\mu}{1-\alpha_1} + \frac{\beta_0+\beta_1}{1-\alpha_1}\bar{z} \tag{3.52a}$$

and for convenience, put

$$\eta_{t-1} = y_{t-1} - \frac{\mu}{1-\alpha_1} - \frac{\beta_0+\beta_1}{1-\alpha_1} z_{t-1} \tag{3.52b}$$

and substitute this into (3.51) as in section 2.4 to give

$$\Delta y_t = \beta_0 \Delta z_t - (1-\alpha_1)\eta_{t-1} + u_t.$$

Rearranging this last equation, η_t can be seen to be a linear combination of I(0) variables and is therefore itself I(0) as long as $1 - \alpha_1$ is materially different from zero, since

$$\eta_{t-1} = \frac{1}{1-\alpha_1}\Delta y_t - \frac{\beta_0}{1-\alpha_1} - \Delta z_t \frac{1}{1-\alpha_1} u_t. \tag{3.53}$$

All terms on the right-hand side of (3.53) are I(0): Δy_t and Δz_t because y_t and z_t are I(1), u_t by assumption, and, as long as $\frac{1}{1-a_1}$ is well defined, then

multiplying a variable by a constant does not change its order of integration. Hence, η_t is the sum of three I(0) processes, and so is I(0). From (3.52b), this means that $y_t - \frac{\mu}{1-\alpha_1} - \frac{\beta_0+\beta_1}{1-\alpha_1} z_t$ is I(0) and hence by the definition of static equilibrium in section 3.2.1.2, equation (3.52a) is a static equilibrium. It is also the cointegrating combination.

In this sense there is an intimate link between cointegration and equilibrium, and it is for this reason that the concept of cointegration is so appealing. It provides an empirically testable definition of equilibrium relationships amongst time series data. The general ADL result follows immediately along the same lines.

3.3.4.1 ADL models, cointegration and equilibrium

Let y_t and $x_{j,t}$, $j = 1, 2, \ldots, r$ be a set of I(1) variables related according to the ADL model

$$\alpha(L)y_t = \mu + \sum_{j=1}^{r} \beta_j(L) z_{j,t} + u_t,$$

where $\alpha(L)$ and $\alpha(L)$, $j = 1, 2, \ldots, r$ are lag polynomials of possibly different order and $\alpha(L)$, and u_t is I(0). The long-run static solution is both an equilibrium and a cointegrating combination of the variables. The deviations from the long-run values are therefore stationary and can be interpreted as deviations from equilibrium.

3.3.4.2 Example

Consider the ADL(1, 1) model

$$y_t = 1 + \frac{1}{2} y_{t-1} + \frac{3}{4} z_t - \frac{3}{8} z_{t-1} + u_t \tag{3.54a}$$

$$\Delta z_t = \epsilon_t \tag{3.54b}$$

where u_t and ϵ_t are uncorrelated white noise processes. First of all, z_t is I(1) since it is a random walk, and so is y_t since equations (3.54a) and (3.54b) imply

$$\Delta\left(1 - \frac{1}{2}L\right) y_t = \frac{3}{4}\left(1 - \frac{1}{2}L\right)\epsilon_t + \Delta u_t. \tag{3.55}$$

The right-hand side of (3.55) can be written as an MA(1) process, the left-hand side shows that the AR operator has a single unit root, so the process is I(1), and more fully is ARIMA(1,1,1). Equation (3.54a) has long-run solution

$$\bar{y} = 2 + \frac{3}{4}\bar{z}, \tag{3.56a}$$

and equilibrium error

$$\eta_t = y_t - 2 - \frac{3}{4} z_t, \tag{3.56b}$$

the ECM being

$$\Delta y_t = \frac{3}{4}\Delta z_{t-1} - \frac{1}{2}\left(y_{t-1} - 2 - \frac{3}{4} z_{t-1} \right) + u_t.$$

Figure 3.5 presents time series plots of z_t and y_t, and Figure 3.6 plots the equilibrium error η_t as given by equation (3.56b). The disturbance processes, u_t and ϵ_t are both NIID(0, 1).

The equilibrium errors in Figure 3.6 are stationary, but do not appear to be white noise. There are runs where the values remain continuously positive for a period of time, and others where negativity persists. This is consistent with autocorrelation. The time series properties of the process can be obtained from (3.56b) as follows.

First write the ADL of equation (3.54a) as

$$\alpha(L) y_t = 1 + \beta(L) z_t + u_t,$$

where

$$\alpha(L) = 1 - \frac{1}{2} L, \tag{3.57a}$$

$$\beta(L) = \frac{3}{4} - \frac{3}{8} L. \tag{3.57b}$$

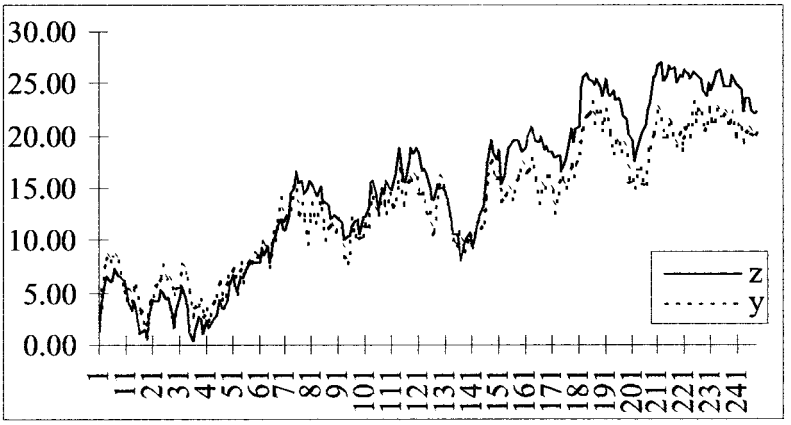


Figure 3.5 Cointegrated processes generated by (3.54a) and (3.54b)

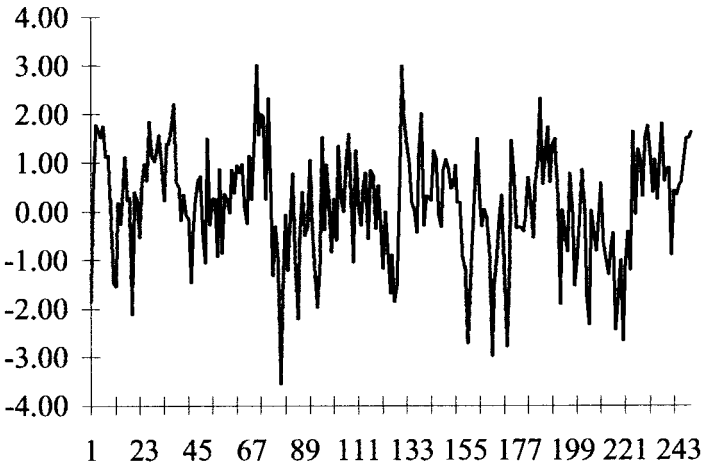


Figure 3.6 The equilibrium error sequence for (3.56b)

Then

$$\begin{aligned}\eta_t &= y_t - 2 - \frac{3}{4}z_t \\ &\Rightarrow \alpha(L)\eta_t = \alpha(L)y_t - \alpha(1)2 - \frac{3}{4}\alpha(L)z_t \\ &\Rightarrow \alpha(L)\eta_t = 1 + \beta(L)z_t - \alpha(1)2 - \frac{3}{4}\alpha(L)z_t. \end{aligned} \tag{3.58}$$

Note that α (1) = $\frac{1}{2}$ and hence α (1) 2 = 1, which, on substitution into (3.58) and rearrangement, gives

$$\alpha(L)\eta_t = \beta(L)z_t - \frac{3}{4}\alpha(L)z_t + u_t.$$

Using (3.57a) and (3.57b),

$$\beta(L) - \frac{3}{4}\alpha(L) = \frac{3}{4} - \frac{3}{8}L - \frac{3}{4}\left(1 - \frac{1}{2}L\right) = 0.$$

Hence α (*L*) $\eta_t = u_t$ or

$$\left(1 - \frac{1}{2}L\right)\eta_t = u_t.$$

Thus η_t is a stationary AR(1) process. Clearly the root of α (*L*) determines the persistence of the equilibrium errors. The closer it is to one, the more persistent they will be. This also determines the speed of adjustment, which is α (1). As the root tends to 1 this speed of adjustment will tend to 0. In the limit, the

long-run solution does not exist, and therefore neither does an equilibrium relationship or cointegration.[13]

It is straightforward to construct an alternative ADL(1, 1) model to (3.57a) and (3.57b) that has the same long-run solution, but much more persistent equilibrium errors and slower adjustment to equilibrium. To get the increased persistence and slower adjustment to equilibrium, replace (3.57a) by

$$\alpha(L) = 1 - 0.95L \tag{3.59a}$$

This has α (1) = 0.05 instead of 0.5. In order to obtain the same long-run solution, the intercept, μ, which was originally 1, and μ must be multiplied by 0.1. Although there are a number of ways to obtain the latter result, the easiest is to multiply the original operator given by (3.57b) by 0.1 to give

$$\beta(L) = 0.075 - 0.0375L, \tag{3.59b}$$

with the new value of μ being 0.1. Thus the DGP is now

$$\alpha(L)y_t = 0.1 + \beta(L)z_t + u_t,$$

with the polynomial lag operators defined by equations (3.59a) and (3.59b), while the DGP for z_t is still the random walk (3.54b). Figure 3.7 shows both the original errors (etaold) and the new much more persistent ones (etanew); note the scale of this and the earlier plot are different.[14] In fact, what seems to have happened is that the broad pattern of fluctuations has remained the same, but their amplitude has become much larger. For example, there are occasions where the low persistence series (etaold) is positive, but an individual shock (i.e. disturbance term) is sufficient to drive the sequence across the zero line so that the neighbouring value is negative. However, with increased persistence (etanew), the same shock is insufficient to drive the series into negativity because it is a lot further away from zero.

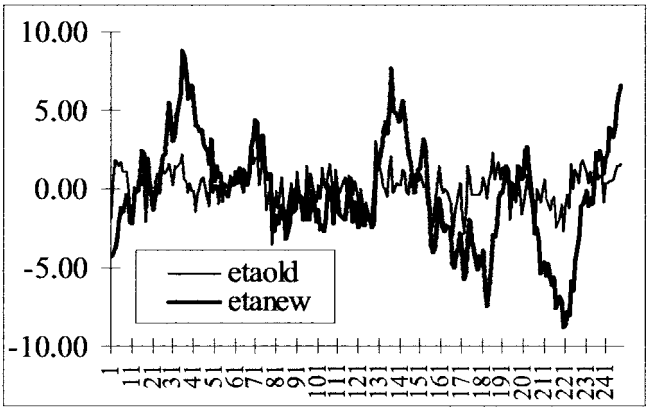


Figure 3.7 Time series plot comparing equilibrium errors

This result can be shown algebraically. For the models under consideration, the structure of the equilibrium errors is:

$$\eta_t = \alpha^t \eta_0 + \sum_{i=1}^{t-1} \alpha^{t-i} u_i + u_t.$$

Assuming the initial value, η_0 is zero gives

$$\eta_t = \sum_{i=1}^{t-1} \alpha^{t-i} u_i + u_t.$$

So the current equilibrium errors can be decomposed into two parts: that due to previous shocks, say

$$U_{t-1} = \sum_{i=1}^{t-1} \alpha^{t-i} u_i,$$

and that due to the current shock, u_t. In the case where u_t is white noise with variance σ_u^2, the variance of previous shocks is

$$Var(U_{t-1}) = \sum_{i=1}^{t-1} \alpha^{2(t-i)} \sigma_u^2 = \frac{\alpha^2(1-\alpha^{t-1})}{(1-\alpha^2)} \sigma_u^2$$

The ratio of the variance of the current shock to that of the component due to past shocks is therefore

$$r(\alpha) = \frac{(1-\alpha^2)}{\alpha^2(1-\alpha^{t-1})}$$

This is a decreasing function of α, and so as α increases, the variance of the current shock becomes smaller relative to that of the accumulated shocks. That is, increasing persistence will manifest itself by a decreasing importance of the current shock relative to the aggregate effect of all those that have preceded it in the evolution of the process.

3.3.5 Cointegration amongst I(d) variables

It is necessary to provide a definition of cointegration amongst variables that are integrated of the same order.

Definition: Cointegration Amongst Variables Integrated of Order d

Let $x_{i,t}$ $i = 1, 2, \ldots, n$ be a set of variables all integrated of order d. Let v_i, $i = 1, 2, \ldots, n$ be a set of constants. If there exists a linear combination of the variables given by

$$\eta_t = \sum_{i=1}^{n} v_i x_{i,t} \tag{3.60}$$

that is integrated of order $d - b$, where $0 < b = d$, then the variables $x_{i,t}$, $i = 1, 2, \ldots, n$, are said to be cointegrated of order (b, d) (or CI(b, d)). The coefficient

vector $v' = (v_1 \ldots v_n)$ is called the cointegrating vector, and $\sum_{i=1}^{n} v_i x_{i,t}$ is called the cointegrating combination of the variables.

The most important special case of this is where $d = b$ so that the linear combination is stationary. The ADL case discussed above is of this type with $d = 1$, so the variables in this case are CI(1, 1).

A useful thing to realize at this point is that a regression equation with disturbances can be written as a linear combination like (3.60). Suppose

$$x_{1,t} = \beta_2 x_{2,t} + \beta_3 x_{3,t} + \ldots + \beta_n x_{n,t} + u_t.$$

This can be written as

$$u_t = x_{1,t} - \beta_2 x_{2,t} - \beta_3 x_{3,t} - \ldots - \beta_n x_{n,t},$$

which is of the form (3.60) with $\eta_t = u_t$, $\theta_1 = 1$, $\theta_i = -\beta_i$ for $i = 2, 3, \ldots, n$. A regression where all the variables are I(1) but the disturbances are I(0) is called a cointegrating regression.[15]

3.4 Regression amongst cointegrated variables

3.4.1 Static regressions

The existence of cointegrating relationships between I(1) variables can be tested using a single static regression equation estimated by ordinary least squares (OLS). A static regression is one involving only contemporaneous values of the variables. There are strong reasons for preferring a multiple equation approach to this problem, but the single equation approach is described here due to its popularity. Its original application is due to Engle and Granger (1987).

Static regressions can be thought of as falling into three cases defined by the order of integration of the variables. These are: (i) variables I(0); (ii) variables I(1) but not cointegrated; (iii) variables I(1) and cointegrated. Case (i) is not discussed in detail, as it is the foundation case free of the complications due to non-stationarity. (For a clear discussion, see Patterson 2000).

Consider the bivariate case involving two variables z_t and y_t, both I(1), and consider the regression of y_t on z_t,

$$y_t = \beta z_t + \eta_t, t = 1,2,\ldots,T,$$

estimated by OLS. The OLS estimator of β, say $\hat{\beta}$, can be written

$$\hat{\beta} = \left(\sum_{t=1}^{T} z_t y_t\right)\left(\sum_{t=1}^{T} z_t^2\right)^{-1} = \beta + \left(\sum_{T=1}^{t} z_t \eta_t\right)\left(\sum_{t=1}^{T} z_t^2\right)^{-1}.$$

The large sample behaviour of the estimator $\hat{\beta}$ relative to β therefore depends on that of $\left(\sum_{T=1}^{t} z_t \eta_t\right)\left(\sum_{t=1}^{T} z_t^2\right)^{-1}$. Stock (1987) shows that, when z_t and y_t are

cointegrated with parameter β, so that $\eta_t \sim I(0)$, this term converges to zero at a rate of $Op\ (T^{-1})$.[16] This means that $\hat{\beta}$ tends to β at the same rate, and is therefore a consistent estimator. When the series are stationary, the convergence is only $Op\ (T^{-\frac{1}{2}})$. Stock's result is therefore known as a proof of super-consistency. It is unaffected by the disturbances, η_t, being autocorrelated (as long as they are I(0)), or being correlated with z_t. The consistency of $\hat{\beta}$ establishes that it can replace β in any model with no change to the asymptotic properties of other estimated parameters.

Standard inference on β is not available, however, as, appropriately normalized by multiplying by T, its asymptotic distribution is non-normal, being that of a random variable depending on Wiener processes (see Banerjee et al., 1993, p176, and Park and Phillips, 1988). Its t-ratio is also asymptotically non-normal in general.[17]

It is important to note that these results are asymptotic, and that in finite samples, significant biases may occur (Inder 1993; Banerjee et al. 1993). These arguments extend to multiple regressions of the form

$$x_{1,t} = \sum_{2=1}^{n} \beta_i x_{i,t} + u_t, t = 1,2,\ldots,T, \tag{3.61}$$

where $x_{i,t}$, $i = 1, 2, \ldots n$, are CI(1, 1).

When the dynamic model is relatively complex, containing a larger number of variables and higher-order lags, the structure of the disturbance term in (3.61) will be correspondingly complex. This greater complexity does not affect the super-consistency of OLS estimation, but it does increase the chances of considerable bias. This will impact on subsequent tests of cointegration, and upon any model, such as an ECM, estimated using the residuals from the static regression. In such cases, tests and estimations can be based upon the ADL model directly.

Analogously to the unit root testing problem, an alternative to, in effect, correcting the regression by adding lags, is to correct the estimates (and t-ratios) non-parametrically, as is done in the Phillips (1987) unit root test. This approach was developed by Phillips and Hansen (1990), and can be found described in detail in Hamilton (1994, p. 613). Super-consistency states that problems arising from autocorrelated residuals and endogeneity can be ignored in the limit. However, in finite samples these effects will still be present. Modification can be developed to reduce these problems, the resultant estimators being known as fully modified least squares (FMLS).

This approach is not discussed in detail here. Its application requires the calculation of variance matrices that can have rather poor properties with the result that simulation evidence varies according to the form of data generation process used. See Inder (1993), Patterson (2000), and Phillips and Hansen (1990) for more details.

3.4.2 Testing for cointegration in single equations

3.4.2.1 Tests based on static regressions

Equation (3.61) provides the basis for testing for cointegration of order CI(1, 1) Assume that $X_{i,t}$, $i = 1, 2, \ldots n$, are I(1). Then, in general, any linear combination of these variables will also be I(1). The exception is if they are cointegrated, in which case, estimating (3.61) by OLS (which minimizes the residual variance), should provide a good estimate of the cointegrating coefficients, β_i. It has already been argued that OLS provides a consistent estimate of β_i under cointegration. Thus one way to proceed is to estimate (3.61) by OLS, obtaining residuals $\hat{u}_t$, and testing the residuals for a unit root, since

$$\hat{u}_t = x_{1,t} - \sum_{i=1}^{n} \hat{\beta}_i x_{i,t},$$

is a minimum residual variance linear combination of observations on the variables. This means that any standard procedure for testing for unit roots or stationarity is available for testing the integratedness of the residuals, and hence whether or not the series are cointegrated. In the sense that the $\hat{\beta}_i$ are consistent for the β_i, the residuals $\hat{u}_t$ can be said to be consistent estimators of the disturbances u_t, so that the test is of the cointegration properties of the variables. Clearly $u_t \sim I\,(0)$ is equivalent to cointegration, while $u_t \sim I\,(1)$ is equivalent to non-cointegration. Unit root tests are tests of the null hypothesis of a unit root and are thus tests of the null of non-cointegration when the alternative is I(0), while stationarity tests have stationarity as the null, corresponding to cointegration.

Tests of the former type are the usual augmented Dickey–Fuller tests, or the Phillips Z tests. The asymptotic properties of these and other residual based tests of non-cointegration are discussed by Phillips and Ouliaris (1990). A leading example of the latter type of test, being one of the null of cointegration against an alternative of non-cointegration, is the test given in Kwiatowski et al. (1992).

The asymptotic distributions of these test statistics are altered as a result of the estimation of β_i, and finite sample distributions vary with both T and n. Critical values may be calculated using the response surfaces of MacKinnon (1991).[18]

3.4.2.2 Test based on dynamic model

It has been found that unit root tests lack power in finite samples. That is, they do not reject the null of a unit root sufficiently frequently as the autoregressive parameter approaches unity from below (e.g. Schwert, 1989). This problem transfers to their use as tests of non-cointegration, where the null of non-cointegration is not rejected with sufficient frequency when the residuals are close to being non-stationary, but are in fact stationary (Banerjee et al. 1993; and Inder 1993). The power of such tests can be improved by correcting

for neglected structure in the disturbances of the test regression (Kremers et al. 1992).

To illustrate, consider the bivariate ECM of equation (3.8), but without an intercept for simplicity,

$$y_t = \alpha_1 y_{t-1} + \beta_0 z_t + \beta_1 z_{t-1} + u_t,$$

which has the ECM form

$$\Delta y_t = \gamma_1 \Delta z_t + \gamma_2 (y_{t-1} - \gamma_3 z_{t-1}) + u_t,$$

with $\gamma_1 = \beta_0,\ \gamma_2 = -(1-\alpha_1),\ \gamma_3 = \frac{\beta_0 + \beta_1}{1-\alpha_1}$. If $\gamma_2 = 0$ then the series are not cointegrated. In practice, the cointegrating coefficient γ_3 is unknown, and so a test on an estimated coefficient requiring knowledge of its value appears impracticable. However, the ECM may be rewritten as

$$\begin{aligned} \Delta y_t &= \gamma_1 \Delta z_t + \gamma_2 (y_{t-1} - z_{t-1}) + \gamma_4 z_{t-1} + u_t, \\ \gamma_4 &= \gamma_2 (1 - \gamma_3). \end{aligned} \tag{3.62}$$

The test is of H_0: $\gamma_2 = 0$ in (3.62), the test statistic being the usual OLS t-ratio. If the actual cointegrating coefficient happens to be equal to 1 (i.e. $\gamma_3 = 1$), then the term in $\gamma_4 z_{t-1}$ will be superfluous. Banerjee et al. (1993, table 7.6) provide some critical values for this test statistic.

In the case of more than two variables the approach generalizes to specifying any potential equilibrium error as the variable of which γ_2 is the coefficient, and add correcting terms in the first lag of each of the variables other than y_{t-1}. More complex dynamics are allowed for by adding lagged differences of the variables, as in the original ECM.

An alternative approach is to reparameterize the ECM to a form an analogue of the equation in the disturbances used to perform the ADF test. Transforming the ECM:

$$\begin{aligned} \Delta y_t &= \gamma_1 \Delta z_t + \gamma_2 (y_{t-1} - \gamma_3 z_{t-1}) + u_t \\ &\Rightarrow \Delta y_t - \gamma_3 \Delta z_t = \gamma_2 (y_{t-1} - \gamma_3 z_{t-1}) \\ &\quad + (\gamma_1 - \gamma_3) \Delta z_t + u_t \end{aligned} \tag{3.63a}$$

But $\eta_t = y_t - \gamma_3 z_t$ is the cointegrating combination, the consistent estimates of which are the residuals from the static regression on which the ADF test is performed. However, rather than the usual ADF regression, (3.63a) is a further augmentation:

$$\begin{aligned} \Delta \eta_t &= \psi_1 \eta_{t-1} + \psi_2 \Delta z_t + u_t, \\ \psi_1 &= \gamma_2, \psi_2 = (\gamma_1 - \gamma_3). \end{aligned} \tag{3.63b}$$

This suggests that the ADF regression should be further augmented by the lagged difference of the right-hand side variable in the static regression. It also shows that the usual ADF regression assumes a restriction applies to the

original ECM, namely $\gamma_1 = \gamma_3$, so that there is no requirement for the extra difference term, or, more accurately, ignores the fact that there will be a correlation between the disturbances and the regressors of the standard ADF regression, since both will include a component of z_{t-1}.[19] The test statistic is the usual ADF t-ratio on η_{t-1} in (3.63b). The equilibrium error, η_t, should be calculated using a consistent estimator of the cointegrating coefficients. These could come from the static regression, or, from the long run solution to the dynamic model.

Equation (3.63b) is modified for more complex dynamics and additional variables by adding differences of all explanatory variables and lagged differences of all variables (including lags of Δy_t).

Finite sample critical values have to be simulated. Illustrative values may be found in Patterson (2000, table 8.11). Inder (1993) has found that such tests display more power than the usual ADF residual based tests, and have additional desirable properties. They are more robust, because when $\psi_2 \neq 0$, then the finite sample performance of the tests are distorted by the exclusion of extra dynamic terms such as Δz_t.

3.4.3 Problems with the single equation approach

The single equation approach is problematical for a number of reasons:

(i) If there is more than one cointegrating vector, which is possible when there are more than two integrated variables, then the single equation approach is only likely to result in a linear combination of these.

(ii) Even if there is only one cointegrating relationship, all variables may be responding to deviations from equilibrium. Estimating a single equation only, ignores this and leads to inefficiencies in the estimation. This amounts to assuming that the right-hand side variables are weakly exogenous, so their dynamic equations exclude the cointegrating relationship.

Given that an approach allowing estimation of a system of equations describing multiple cointegrating relations is available, it is likely that this will provide more robust estimation and inference since it does not rely on conditions that it is difficult to test in practice. This is the subject of the following chapter.

3.5 Conclusion

The notion of cointegration developed via the integration of conventional time series analysis with econometric methods. Econometrics dealt initially

with models that were mainly static in nature, while the dynamic nature of data was implicit in univariate time series analysis. In univariate time series analysis data is differenced to induce stationarity, but this was not common practice in economics until the 1980s. One of the first articles to amalgamate a time series model with an econometric formulation with levels was the wage equation article produced by Sargan (1964), the model unlike many wage equations of the time considered the question of the dynamic specification of a wage inflation equation in the context of a model that is estimated by instrumental variables. The article is the first example of an error correction model, which was both well ahead of its time and highly influential in terms of the institutional modelling of UK wage equations. The ARMAX representation is the first example of cointegration as the error correction term is significant assuming the type of asymptotic normality of the t-test on the coefficient of the error correction term is accepted (Kremers et al. 1992). Granger and Newbold (1974) provided the first simulation experiments to consider the impact of non-stationarity on the diagnostics associated with nonsense regressions. The problem of nonsense correlation was well known to time series statistics through the work of Yule (1926) and should have been known to the econometrics literature because of the intervention of Keynes (1939), who discussed the potential for misanalysis when regressions between variables with intermediate causes were considered. Granger and Newbold considered the special case for which the intermediate cause was an independent stochastic trend. While time series analysts started to consider tests for non-stationarity (Dickey and Fuller 1979), econometricians in the UK started to implement models which exhibited error correction behaviour. Davidson et al. (1978) introduced the notion that the error correction term explained the long-run behaviour of economic series and these dynamic models are again cointegrating relationships in the sense of Kremers et al. (1992) as they include combinations of stationary variables. In one case the lag series renders the variable stationary, for the error correction term, the contemporaneous observation in a different time series does the same thing.

Granger (1983) introduced the term cointegration to the literature, while Sargan and Bhargava (1983) provided the first recognized test of existence of long-run behaviour. It was Granger, via his decomposition of the Wold representation of what are quasi over-differenced series, whi explained how dependent moving average behaviour might yield long-run relations with variable that are stationary. Engle and Granger's (1987) article provided a means by which bivariate relationships might be given error correction representations, though more generally this proposition does not follow from the results developed in the article. The two-step method developed by Engle and Granger shows that the long-run parameters in the case where there are two variables

or a single cointegrating vector can be estimated from a cointegrating regressin. In general, this is not the case, which suggests that the requirement to develop an approach that might be applied to a system. It is the systems approach developed by Johansen (1988a, 1988b) for the autoregressive model that will be considered in detail in the next chapter.

4
Multivariate Time Series Approach to Cointegration

4.1 Introduction

This chapter considers the case where there are a number of non-stationary series driven by common processes. It was shown in the previous chapter that the underlying behaviour of time series may arise from a range of different time series processes. Time series models separate into autoregressive processes that have long-term dependence on past values and moving average processes that are dynamic but limited in terms of the way they project back in time. In the previous chapter the issue of non-stationarity was addressed in a way that was predominantly autoregressive. That is, stationarity testing via the comparison of a difference stationary process under the null with a stationary autoregressive process of higher order under the alternative. The technique is extended to consider the extent to which the behaviour of the discrepancy between two series is stationary or not. In the context of single equations, a Dickey–Fuller test can be used to determine whether such series are related; when they are this is called cointegration. When it comes to analyzing more than one series then the nature of the time series process driving the data becomes more complicated and the number of combinations of non-stationary series that are feasible increases.

Here we consider in detail a number of alternative mathematical models that have the property of cointegration. Initially we discuss representations that derive from the multivariate Wold form. This is the approach first considered by Granger (1983) and Granger and Weiss (1983), in which the Granger representation theorem is developed. From this theorem there are a number of mathematical decompositions, which transform moving average models into vector autoregressive models with multivariate error correction terms. From the perspective of the probability model from which the Wold form derives the VMA representation provides a more elegant explanation of non-stationary time series. First, the conditions associated with cointegration in

the VMA are more succinct and secondly implicit in the fundamental condition for cointegration in the VMA is the explicit conclusion that under cointegration the long-run levels relationships are stationary. An alternative mechanism of decomposing the VAR into an error correction form derives from Engle and Granger (1987), but beyond the single equation case inference about non-stationary processes, estimation of the long-run parameters and testing hypotheses about the long run all derive from the maximum likelihood theory developed by Johansen. When it comes to constructing dynamic models, then the approach developed by Johansen appears to provide a bridge between two main strands of econometric time series modelling: first, the VAR reduced form approach derived from the rational expectations literature by Sims (1980) and the error correction approach that has developed from the work of Sargan (1964), Davidson et al. (1978) and Ericsson, Hendry and Mizon (1998). The cointegration/error correction approach emphasizes the description of detectable economic phenomena in the long run. The cointegration approach assumes that short-run processes are not well defined by virtue of aggregation across agents, goods and time, differing forms of expectations, learning, habits and dynamic adjustment. Alternatively, the long run provides a useful summary of the non-detectable short-run dynamics, while the error correction approach in the confines of the VAR permits short-run policy analysis via the impulse response function and the ability to analyze both short-run and long-run causality and exogeneity. If the VAR defines a valid reduced form, then it allows the detection of the readily available structure.

More conventional econometric approaches (Pesaran et al. 2000) criticize the Johansen methodology for being ad hoc in the sense that it doesn't use as its starting point an econometric system of the type defined by the Cowles foundation, but, as is discussed in the context of RE models in chapter 6, it is still possible to introduce short-run restrictions in the confines of a VAR-style cointegration approach. The VMA approach appears to be less interested in the distinction between the long run and the short run, as to whether money causes inflation as compared with money leading to price rises, but still permits impulse response and short-run causality analysis. However, in the context of pure MA models, inference and detection of long-run behaviour is less well developed. Impulse response analysis emphasize the responsiveness of variables to and the effectiveness of policy. The use of the VAR and VMA for short-run analysis is discussed in detail by Lippi and Reichlin (1996).

Here, we define cointegration in terms of the Wold decomposition, then we consider the Johansen approach to testing and estimation, some empirical results are derived from the literature and discussed in the context of an increasing body of evidence based on Monte Carlo simulation. Alternative

representations are discussed along with the extension of the methods applied to multi-cointegration and polynomial cointegration.

4.2 The VMA, the VAR and VECM

The concept of cointegration is now well established following the seminal work of Engle and Granger (1987), and the development of practical estimation and inferential methods by Johansen (1988a,b). Although in the latter case, many different approaches are now available (for example, on estimation, see Gonzalo, 1994), it is the Johansen methodology that dominates empirical work. However, there is an uneasy relationship between the structure explored by Engle and Granger and those exploited by Johansen: the former is based on the Wold decomposition, that is, a potentially infinite order vector moving average (VMA) representation, while the latter employs a vector autoregressive (VAR) model.

It is not difficult to motivate consideration of moving average structure: there is much evidence in the literature for the poor performance of the Dickey–Fuller test (Said and Dickey 1984; Hall 1989) and some for the Johansen test (Boswijk and Franses 1992; Cheung and Lai 1993) under moving average errors. In the multivariate case, fundamentalness, for which invertibility is a necessary condition (see below and Lippi and Reichlin 1996), is required for impulse response analysis. Cointegrating VARs must be able to deliver reasonable approximations to an underlying VMA and prior testing must be reliable in identifying whether or not cointegration exists.

Interest in MA behaviour also has a generic basis in terms of univariate time series modelling, while what might be viewed as one of the earliest examples of error correction behaviour had an MA error that is, the wage–price model – developed by Sargan (1964). Further, moving average behaviour often defines measure or rather mismeasurement equations associated with rational expectations models. It is the principle reason for estimating dynamic Euler equations by generalized method of moments (GMM). More specifically, market efficiency in the confines of a model relating spot and futures contracts might be expected to display this form of over-differencing linked with a dependent relationship and the use of the Wold representation (Flôres and Szafarz 1995).

One problem to be addressed is how to obtain a VAR form in levels from a VMA form in differences. There are various ways of establishing the relationship. In general such theorems have become known as (Granger) representation theorems, after a working of the problem in Engle and Granger (1987). The application of the Smith–McMillan (SM) form to cointegrated systems is presented in Engle and Yoo (1991) and Yoo (1986). This approach is handled in detail in section 4.3.

4.2.1 The VAR and VECM models

Let z_t be an $n \times 1$ vector of time series, and ϵ_t a $n \times 1$ of vector white noise series having $E(\epsilon_t) = 0$ and

$$E(\epsilon_t \epsilon'_{t-j}) = \begin{cases} \Sigma & \text{for } j = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{4.1}$$

Σ being a positive definite matrix. Then x_t has a *vector autoregressive structure of order p (VAR(p))* if

$$A(L)x_t = \mu + \epsilon_t \tag{4.2}$$

where μ is a $n \times 1$ vector of constants, and $A(L)$ is an $n \times n$ matrix lag polynomial given by $A(L) = I_n - \sum_{i=1}^{p} A_i L^i$, I_n being the $n \times n$ identity matrix and A_i, $i = 1, 2, \ldots, p$, $n \times n$ coefficient matrices. Equation (4.2) can alternatively be written

$$x_t = \mu + \sum_{i=1}^{p} A_i x_{t-i} + \epsilon_t$$

A simple example for $n = 2$ and $p = 1$ is $x_t = [x_{1,t}\, x_{2,t}]'$ and $\epsilon_t = [\epsilon_{1,t}\, \epsilon_{2,t}]'$

$$x_t = \begin{bmatrix} .5 & .25 \\ 1 & .5 \end{bmatrix} x_{t-1} + \epsilon_t$$

Just as with scalar autoregressive models, the VAR(p) may be reparameterized into differences and a single lagged levels term. Any p^{th} order $n \times n$ matrix polynomial of the form $A(L) = I_n - \sum_{i=1}^{p} A_i L^i$ may be written in the form

$$A(L) = -\prod L + A^*(L)(1-L)$$

$$\text{where } A^*(L) = \begin{cases} I_n - \sum_{i=1}^{p-1} A_i^* L^i & \text{if } p > 1 \\ I_n & \text{if } p = 1 \end{cases},$$

$$A_i^* = -\sum_{j=i+1}^{p} A_j,\ i = 1,\ 2,\ \ldots,\ p-1 \text{ and } \prod = -A(1)$$

Applying this reparameterization to (4.2) yields

$$(-\prod L + A^*(L)(1-L))x_t = \mu + \epsilon_t$$

On rearrangement this gives

$$\Delta x_t = \mu + \prod x_{t-1} - \sum_{i=1}^{p-1} A_i^* \Delta x_{t-i} + \epsilon_t, \tag{4.3}$$

where the summation on the right-hand side does not appear if $p = 1$. Equation (4.3) is known as the *vector error correction representation* of the VAR. It exists irrespective of the orders of integration of the processes $x_{i,t}$, $i = 1, 2, \ldots, n$. This is commonly used in the analysis of cointegrated variables (Johansen 1995) and is then often known as a *vector error correction model (VECM)*.

4.2.2 The VMA model

Let ϵ_i be a vector white noise process as defined in equation (4.1). Let x_t be a $n \times 1$ vector of stationary variables. Then there exists a multivariate version of Wold's representation theorem that states that x_t may be represented as (a possibly infinite order) vector moving average (VMA) process. If x_t has a non-zero mean this may be introduced to the model. Thus x_t may be represented

$$x_t = \mu + \Theta(L)\epsilon_t \tag{4.4}$$

where μ is a $n \times 1$ vector of constants and $\Theta(L)$ is a $n \times n$ matrix lag polynomial given by $\Theta(L) = I_n + \sum_{i=1}^{q} \Theta_i L^i$ where q may be infinite. Equation (4.4) can be written less compactly:

$$x_t = \mu + \epsilon_t + \sum_{i=1}^{q} \Theta_i L^i \epsilon_t \tag{4.5}$$

and (4.5) define a VMA process of order q (***VMA(q)***).

4.2.3 The Granger representation theorem: systems representation of cointegrated variables

Chapter 2 established a link between cointegration of order (1,1) and the single equation error correction model. It was shown that cointegrated variables could be represented either as an ADL or as an ECM. Cointegration can be characterized in a systems context in a number of ways and manifests itself as a set of restrictions on a general model. Naturally there is more than one way in which this may be achieved. The first characterization of this nature is due to Granger (1983) and as such is known as the Granger representation theorem. Subsequent treatments develop alternative representations (Johansen 1988a, 1995) and generalize the cases considered.

4.2.3.1 Cointegration starting from a VMA and deriving VAR and VECM forms

This was the first approach to explaining how to characterize cointegration in the context of a multiple time series model. It is in many ways the most natural for two reasons. First, it builds on an established representation theorem, the multivariate version of the Wold representation. Secondly, it naturally restricts cases under examination to whatever orders of integration are the subject of investigation. Suppose x_t is a $n \times 1$ vector of time series each

element of which is *I(1)*. Then the first difference of the vector, Δx_t, is I(0). As such it has a Wold representation,

$$\Delta x_t = C(L)\epsilon_t.^1 \tag{4.6}$$

The task is to determine how this relationship can give rise to cointegration. This follows by application of the reparameterization to $C(L)$, then $C(L) = C(1)L + C^*(L)(1 - L)$, for some $C^*(L)$ of order one less than $C(L)$. Substituting this into (4.6) gives

$$\Delta x_t = (C(1)L + C^*(L)(1-L))\epsilon_t. \tag{4.7}$$

Cointegration requires that there exists a $n \times 1$ vector β, such that $\beta' x_t$ is $I(0)$. Equation (4.7) can be used to obtain an expression for $\beta' x_t$ by pre-multiplying by β', and remembering that Δ is a scalar linear operator. Thus

$$\beta'\Delta x_t = \beta'(C(1)L + C^*(L)(1-L))\epsilon_t \tag{4.8}$$

$$\Delta\beta' x_t = \beta' C(1)\epsilon_{t-1} + \beta' C^*(L)\Delta\epsilon_t \tag{4.9}$$

Equation (4.9) can be used to develop the following theorem.

Theorem 1 *x_t is cointegrated if and only if C(1) is singular.*

Proof. Singularity of $C(1)$ implies cointegration since $C(1)$ *is singular* $\Leftrightarrow$ *rank*$(C(1)) = r < n \Rightarrow$ there exists a vector α such that $\beta' C(1) = 0$. It follows from substitution of this term in (4.9), that

$$\Delta\beta' x_t = \beta' C^*(L)\Delta\epsilon_t$$

$$\Rightarrow$$

$$\beta' x_t = \mu_0 + \beta' C^*(L)\epsilon_t \tag{4.10}$$

The last implication can be thought of as cancellation of the differencing operator, but in the non-stationary context is better thought of as summation, effectively the discrete analogue of integration. However described, this process generates a constant of integration, μ_0, that is a function of the initial value of the processes involved. Equation (4.10) shows that $\beta' x_t$ has a moving average representation and is therefore stationary. However, x_t itself is I(1) and is therefore also CI(1, 1)

Cointegration implies singularity of C (1) since $x_t \sim$ CI(1, 1) $\Rightarrow$ there exists a vector β such that

$$\beta' x_t \sim I(0)$$

Therefore $\beta' x_t$ has an invertible MA representation,

$$\beta' x_t = M(L)\nu_t$$

$$\Rightarrow$$

$$\Delta\beta' x_t = \Delta M(L)\nu_t$$

and hence, from 4.9,

$$\Delta M(L)\nu_t = \beta' C(1)\epsilon_{t-1} + \beta' C^*(L)\Delta\epsilon_t.$$

Summing this last expression from 1 to t gives

$$M(L)\nu_t = \gamma_0 + \beta' C(1)\sum_{i=0}^{t-1}\epsilon_i + \beta' C^*(L)\epsilon_t, \tag{4.11}$$

where γ_0 is the constant of integration, and so

$$\beta' C(1) = 0 \tag{4.12}$$

since otherwise the right-hand side of (4.11) would be the sum of the I(1) process $\left(\sum_{i=0}^{t-1}\epsilon_i\right)$ and the I(0) process ($\gamma_0 + \beta' C^*(L)\epsilon_t$) and hence $x_t \sim I(1)$. But $M(L)\nu_t$ is I(0) process, which means, by contradiction, that an I(1) process must not enter (4.11) and hence $\beta' C(1)$ must be zero. However:

$$\beta' C(1) = 0 \Leftrightarrow rank\ (C(1)) < n \Leftrightarrow C(1)\ is\ singular. \qquad \blacksquare$$

A moving average process such as (4.7) with singular C (1) may be called a reduced rank moving average process. Next, the link between a reduced rank moving average and a VECM needs to be established. This is done by first establishing that x_t has a *vector autoregressive moving average* (VARMA) representation. A VARMA process is a VAR with VMA disturbances, so may be written

$$A(L)x_t = B(L)\epsilon_t \tag{4.13}$$

$$where\ A(L) = I - \sum_{i=1}^{p} A_i L^i \text{ and } B(L) = I - \sum_{j=1}^{q} B_j L^j$$

where x_t is $n \times 1$, A_i, B_j $n \times n$ coefficient matrices, and ϵ_t a $n \times 1$ vector white noise process as defined in equation (4.1). Equation (4.13) defines a VARMA process of order (p, q). In order to derive the VARMA structure from the Wold representation of equation (4.6), two results for polynomial matrices are developed in Appendix A (Engle and Granger, 1987).

4.2.4 VARMA representation of CI(1, 1) variables

In order to obtain a VARMA representation for x_t, it appears that C (L) in

$$\Delta x_t = C(L)\epsilon_t$$

must be inverted. However, since x_t is CI(1, 1), it follows that C (1) is singular. That is, C (L) has unit roots preventing its inversion (see Appendix A.3). In addition, a representation of x_t rather than Δx_t is required. The problem is overcome by factoring out the unit root components from C (L), although scalar factors are not available. Even so, the approach still very neatly allows

the cancellation of the differencing operator as required. Thus both objectives are achieved. The results in Appendix B may be used to obtain a VARMA form of the Wold representation, (4.6). Using notation consistent with equation (B.1), let *rank* $(C(1)) = n - r$, $1 \le r \le n$. Therefore the result developed to derive (B.5) may be applied to $C(L)$, which is a q^{th} order polynomial in L $(m = q)$. It follows that there exists a matrix lag polynomial $\tilde{H}_c(L)$ of order $b \le q^{n-1} - r + 1$, and a scalar lag polynomial $\tilde{g}_c(L)$ of order $a \le q^n - r$ such that

$$\tilde{H}_C(L)C(L) = \Delta\tilde{g}_C(L)I_n.$$

Pre-multiplying the Wold form above by $\tilde{H}_c(L)$ transforms the VMA into a VARMA:

$$\begin{aligned}\tilde{H}_C(L)\Delta x_t &= \tilde{H}_C(L)C(L)\epsilon_t \\ &= \Delta\tilde{g}_C(L)I_n\epsilon_t\end{aligned}$$

Dividing through by the difference operator

$$\tilde{H}_C(L)x_t = \gamma_0 + \tilde{g}_C(L)\epsilon_t$$

where γ_0 is a constant of integration, which for appropriately set initial conditions or data transformation:

$$\tilde{H}_C(L)x_t = \tilde{g}_C(L)\epsilon_t \tag{4.14}$$

This is a unique VARMA representation of x_t, for the case where the order of cointegration is (1, 1) and $\tilde{g}_c(L)$ is a scalar polynomial.

To further motivate this result consider the following example.
Let $q = 1$, $n = 3$ and

$$C(L) = \begin{bmatrix} 1 & 1/2L & 1/2L \\ -1/2L & 1-5/4L & -1/4L \\ 1/4L & 1/8L & 1-7/8L \end{bmatrix}.$$

Then

$$C(1) = \begin{bmatrix} 1 & 1/2 & 1/2 \\ -1/2 & -1/4 & -1/4 \\ 1/4 & 1/8 & 1/8 \end{bmatrix}.$$

It is easy to see that $C(1)$ is rank deficient, because the rows and columns of this matrix are scalar multiples of each. For example, using the notation $C(1)_{i.}$ to denote the i^{th} row of $C(1)$,

$$\begin{aligned}C(1)_{1.} &= \begin{bmatrix}1 & 1/2 & 1/2\end{bmatrix} = -2C(1)_{2.} = -2\begin{bmatrix}-1/2 & -1/4 & -1/4\end{bmatrix} \\ &= 4C(1)_{3.} = 4\begin{bmatrix}1/4 & 1/8 & 1/8\end{bmatrix}.\end{aligned}$$

By definition the rank of a matrix is the number of linearly independent rows or columns, which in this case is 1. The decomposition requires pre-

multiplication of $C(L)$ by the matrix the $\tilde{H}_c\ (L)$ where the adjoint of $C(L)$ is given by $C^a(L) = (1 - L)\tilde{H}_c(L)$. Calculation of the adjoint follows from the transpose of the usual matrix of minors (further detail see Dhrymes 1984). Therefore

$$C^a(L) = \begin{bmatrix} -\frac{17}{8}L + \frac{9}{8}L^2 + 1 & -\frac{1}{2}L + \frac{1}{2}L^2 & -\frac{1}{2}L + \frac{1}{2}L^2 \\ \frac{1}{2}L - \frac{1}{2}L^2 & -\frac{7}{8}L - \frac{1}{8}L^2 + 1 & \frac{1}{4}L - \frac{1}{4}L^2 \\ -\frac{1}{4}L + \frac{1}{4}L^2 & -\frac{1}{8}L + \frac{1}{8}L^2 & -\frac{5}{4}L + \frac{1}{4}L^2 + 1 \end{bmatrix}$$
$$= (1-L)\begin{bmatrix} 1-9/8L & -1/2L & -1/2L \\ 1/2L & 1+1/8L & 1/4L \\ -1/4L & -1/8L & 1-1/4L \end{bmatrix} = (1-\tilde{L})H_C(L)$$

This establishes the AR operator of the VARMA (4.14). To obtain the scalar MA operator note that, from the results on reduced rank polynomials, $|C(L)| = \Delta\tilde{g}_c\ (L)$. In this case

$$|C(L)| = 1 - \frac{17}{8}L + \frac{5}{4}L^2 - \frac{1}{8}L^3 = (1-L)^2(1-1/8L) = (1-L)[(1-L)(1-1/8L)], \quad (4.15)$$

and therefore $\tilde{g}_c(L) = (1 - L)(1 - 1/8L)$. Hence the VARMA representation is:

$$\begin{bmatrix} 1-9/8L & -1/2L & -1/2L \\ 1/2L & 1+1/8L & 1/4L \\ -1/4L & -1/8L & 1-1/4L \end{bmatrix} x_t = (1-L)(1-1/8L)\epsilon_t\,.$$

It should be noticed that the MA component is not invertible. In general the VMA does not directly transform into a VAR as only in special cases does $\tilde{g}_c(L)$ invert.

This completes the numerical example.

An important reason for wanting to re-express a cointegrating VMA in differences is that a VAR in levels follows from the widely employed techniques of Johansen (1995a). These assume a (finite order) VAR representation. The VMA in differences is a very natural starting point since it employs Wold's fundamental representation of a stationary process. It also conveniently allows the scalar processes to have a unit root (be I(1)) and be cointegrated. Such properties are more difficult to impose starting from a VAR (Johansen 1995a).

From the Johansen point of view, the Engle–Granger approach to transforming a VMA in first differences to a VARMA in levels is inconvenient in that some moving average structure remains. The right hand side of equation (4.14) is a VMA with a scalar diagonal matrix lag operator. It is not a pure VAR as defined in equation (4.2). The advantage is that it applies to any cointegrating (CI(1, 1)) VMA.

4.3 The Smith–McMillan-Yoo form

Engle and Yoo (1991) show that if the lag polynomial operator of the original cointegrating VMA is rational (each element of the VMA operator is rational and may have a different denominator polynomial), then there exists a VAR representation where the right-hand side is white noise and the autoregressive operator is rational. As with the Engle–Granger transformation, the unit root moves from being explicit in the VMA to being implicit in the VAR, but now there is no autocorrelation of the disturbances, and there is no restriction that the denominator polynomials of the final VAR operator need all be the same. The Engle–Yoo approach also has the advantage that it extends fairly readily to other forms of cointegration.

The problem to be addressed is how to obtain a VAR form in levels from a VMA form in differences. There are various ways of establishing the relationship. In general such theorems have become known as (Granger) representation theorems, after a working of the problem in Engle and Granger (1987).

As in the univariate case there are a number of alternative time series representations. Each representation has different characteristics. Here the alternative forms are used to move between models where differencing eliminates strong autoregressive behaviour, but due to dependence among economic series some over-differencing remains in the form of moving average behaviour with unit roots. If this type of behaviour inverts to a model with autoregressive behaviour then there may be cointegration amongst the levels of the non-differenced data. It is the movement from the MA to the AR which is important.

The application of the Smith–McMillan (SM) form to cointegrated systems is presented in Engle and Yoo (1991). A rational operator is not in general finite, which is a problem for the Johansen methodology, although special cases exist where the left-hand side reduces to a finite order VAR. (See section 4.7.2 for a discussion of a situation where a finite order pure VAR is available for the first differences.) However, as the denominator polynomials in the Engle–Yoo representation have all their roots outside the unit circle, the operator coefficients tend to zero as the lag length increases. This approach is described below.

Before describing the approach in detail, it is useful to make some preliminary points.

(i) The Smith–McMillan (SM) form is a decomposition of a matrix polynomial. It can be applied to convert a VMA in differences to a VAR in levels or vice versa, despite the presence of unit roots.

(ii) It is limited in its application to matrix lag polynomial operators the individual elements of which are rational (one scalar polynomial divided

by another). While rational operators are in general of infinite order in the lag operator, there exist infinite order polynomials that cannot be represented in rational form. Strictly speaking, therefore, this form of decomposition, and hence conversion from VMA to VAR and vice versa applies only to a sub-class of models: those of rational form. (This does not rule out the special case of finite order polynomials as these are a special case of rational polynomials.)

(iii) The SM form allows the diagonalization of rational polynomial matrices, making their manipulation much easier. This is done in two stages. First, it is noted that there exists a diagonal form for all finite order polynomial matrices, called a Smith form. Secondly, a rational operator can be expressed as a scalar factor dividing a finite order polynomial matrix. The finite order polynomial can then be put in Smith form after which the result can be divided again by the scalar factor. This gives the SM form. The Smith form relies on the application of elementary row or column operations (see Appendix A for details), and it is this approach that restricts application to finite order polynomials, and hence restricts the decomposition of infinite order cases to those that are rational.

(iv) The diagonalization process requires the pre- and post-multiplication of the original matrix by polynomial matrices that are ALWAYS invertible. This has two consequences: problems of simplification (this is not really inversion as will be seen) focus entirely on the diagonalized form (this is called the SM form); and secondly, multiplication by these matrices or their inverses do not alter rank.

The distinctive feature of the Smith–McMillan–Yoo form is the factorization of all the unit roots from the VMA operator in such a way that, by pre-multiplication by an appropriate matrix, a single differencing operator may be isolated on the MA side of the equation.

4.3.1 Using the Smith form to reparameterize a finite order VMA

Consider the VMA

$$x_t = C(L)\epsilon_t,$$

where $C(L)$ is a finite order operator. The Smith form of the operator $C(L)$ is

$$C(L) = G(L)^{-1}C_S(L)H(L)^{-1}$$

where $C_S(L)$ is a diagonal finite order polynomial matrix and $G(L)$ and $H(L)$ are invertible polynomial matrices having unit determinant (called unimodular matrices, see the Appendix A.2 for details), representing the elementary

operations necessary to obtain the diagonalization. Applying this decomposition to the VMA gives

$$x_t = G(L)^{-1} C_S(L) H(L)^{-1} \epsilon_t$$

and hence

$$G(L)x_t = C_S(L)H(L)^{-1}\epsilon_t. \tag{4.16}$$

For example, it is shown in Appendix A.1 that the operator

$$C(L) = \begin{pmatrix} 1-\frac{3}{4}L & -L \\ -\frac{1}{8}L & 1-\frac{1}{2}L \end{pmatrix}$$

can be written

$$C(L) = \begin{pmatrix} 1 & -6 \\ \frac{1}{8}L & 1-\frac{3}{4}L \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1-\frac{5}{4}L+\frac{1}{4}L^2 \end{pmatrix} \begin{pmatrix} 1 & -(2L-6) \\ 0 & 1 \end{pmatrix}^{-1}.$$

The roots of $C(L)$ and the Smith form, $C_S(L)$, are the same since $G(L)$ and $H(L)$ are unimodular. Further, the diagonality of $C_S(L)$ allows any individual roots to be factored out into another diagonal matrix. In particular, unit roots may be factored out. In this example,

$$C_S(L) = \tilde{C}_S(L)D(L),$$

where $\tilde{C}_S(L) = \begin{pmatrix} 1 & 0 \\ 0 & (1-\frac{1}{4}L) \end{pmatrix}$ and $D(L) = \begin{pmatrix} 1 & 0 \\ 0 & (1-L) \end{pmatrix}$. By construction, $\tilde{C}_S(L)$ has all roots outside the unit circle (see Appendix A.3), and so can be inverted. So, equation (4.16) can be pre-multiplied by $\tilde{C}_S(L)^{-1}$ to give

$$\tilde{C}_S(L)^{-1} G(L) x_t = D(L) H(L)^{-1} \epsilon_t.$$

Through $D(L)$, the presence of a unit root is now much more apparent than was the case in the original VMA expression.

4.3.1.1 Reparameterizing a VMA in differences

A further stage in the decomposition of $C(L)$, useful when the VMA describes the differences of a process, is to isolate the differencing factor as a scalar term. To do this in the context of the example, define

$$D^*(L) = \begin{pmatrix} (1-L) & 0 \\ 0 & 1 \end{pmatrix}$$

so that

$$\begin{aligned} D^*(L)D(L) &= \begin{pmatrix} (1-L) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (1-L) & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} (1-L) & 0 \\ 0 & (1-L) \end{pmatrix} = (1-L)I_2 \end{aligned}$$

where I_2 is the 2×2 identity matrix. Clearly, such a simplification will always be available when $D(L)$ has diagonal elements that are either *1* or $\Delta = (1 - L)$. Thus

$$D^*(L)\tilde{C}_S(L)^{-1}G(L) = \Delta H(L)^{-1}.$$

Since Δ is a scalar, this may be rewritten

$$H(L)D^*(L)\tilde{C}_S(L)^{-1}G(L) = \Delta I_2.$$

Thus the VMA can be written

$$H(L)D^*(L)\tilde{C}_S(L)^{-1}G(L)x_t = \Delta\epsilon_t. \tag{4.17}$$

Continuing the example,

$$\tilde{C}_S(L)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & (1-\frac{1}{4}L) \end{pmatrix}^{-1} = \left(1-\frac{1}{4}L\right)^{-1}\begin{pmatrix} (1-\frac{1}{4}L) & 0 \\ 0 & 1 \end{pmatrix}$$

giving

$$H(L)D^*(L)\tilde{C}_S(L)^{-1}G(L) = \left(1-\frac{1}{4}L\right)^{-1}\begin{pmatrix} (1-\frac{1}{2}L) & L \\ \frac{1}{8}L & 1-\frac{3}{4}L \end{pmatrix}$$

which is a rational lag polynomial matrix.

Clearly (4.17) is not a VAR because the right-hand side is the difference of a white noise process, not a white noise process, so it cannot be said that the VMA has been inverted to form a VAR. It has been inverted as far as possible. That is, all components involving roots outside the unit circle have been inverted. Those parts of the VMA operator $C(L)$ that cannot be inverted, the unit roots, have been isolated and alone remain on the right hand side. For convenience, let

$$K(L) = H(L)D^*(L)\tilde{C}_S(L)^{-1}G(L)$$

so that

$$K(L)x_t = \Delta\epsilon_t.$$

But now consider the case where the original model was a VAR for the differences of a process, that is $x_t = \Delta y_t$, so that, after rearrangement

$$K(L)\Delta y_t = \Delta\epsilon_t.$$

Then, apart from initial conditions, the differencing operator can be cancelled to give

$$K(L)y_t = \epsilon_t.$$

This is a VAR in levels corresponding to the VMA in differences

$$\Delta y_t = C(L)\epsilon_t.$$

The VAR is of infinite order, but is rational.[2]

4.3.2 The Smith–McMillan form in general applied to a rational VMA: the Smith–McMillan–Yoo form

The manipulation above starts with a finite order operator. It can be generalized by allowing every element of $C(L)$ to be a rational (and hence infinite order) operator. This requires the generalization of the Smith form to the SM form, the latter being a diagonalized form of a rational polynomial matrix.[3] The SM form has a strong structure. Let $C(L)$ be an $n \times n$ rational polynomial matrix.

Assumption A1: $C(L)$ is rational.
If assumption 1 holds, there exist a set of elementary row and column operations represented by unimodular matrices $U(L)$ and $V(L)$ respectively such that

$$C_{SM}(L) = U(L)C(L)V(L) \tag{4.18}$$

where $C_{SM}(L)$ is a diagonal rational matrix given by

$$C_{SM}(L) = diag\left(\frac{f_i(L)}{g_i(L)}\right) \tag{4.19}$$

where:

(i) $f_i(L)$ and $g_i(L)$ have no common factors;
(ii) $f_i(L)$ is a factor of $f_{i+1}(L)$, $i = 1,2, \ldots, n-1$;
(iii) $g_{i+1}(L)$ is a factor of $g_i(L)$, $i = 1,2, \ldots, n-1$.

There are a number of implications of this result, usefully summarized by Hatanaka (1996). Let z be a general complex argument.

(I1) For any specific value z_0 of z, the rank of $C(z_0)$ is equal to that of $C_{SM}(z_0)$.

(I2) If z_0 is not a root of $f_{n-r}(z)$ nor of $g_1(L)$, but is a root of $f_{n-r+1}(z)$, then z_0 is a root of $f_{n-r+i}(z)$, $i = 2, 3, \ldots, r$.

(I3) The roots of $f_i(L)$, $i = 1, 2, \ldots, n$ are the roots of $C(z)$.

If $C(L)$ has any unit roots, then it follows from implications (I2) and (I3) that they can be associated only with a set of consecutive $f_i(L)$, and that this sequence must extend to the $(n, n)th$ element. (One such case is associated with *rank* $(C(1)) = n - r$, this is defined as a necessary condition for cointegration.)

Assumption A2: $C(1)$ has rank $n - r$.
Under assumption 2 the rank of $C_{SM}(1)$ is $n - r$ by implication (I1). Therefore $n - r$ of the $f_i(1)$ must be non-zero, meaning that $n - r$ of the $f_i(L)$ cannot have a unit root. The remaining $f_i(1)$ must be zero, meaning that the corresponding $f_i(L)$ do have a unit root. Implication (I2) establishes that those having the

unit root must be $f_i(L)$, $i = n - r + 1, \ldots, n$ as otherwise there may be too many unit roots.

This observation is applied by Engle and Yoo (1991) to obtain the Smith–McMillan–Yoo (SMY) form. Define a set of scalar polynomial lag operators, $\tilde{f}_i(L)$, such that $\tilde{f}_i(L) \neq 0$, and

$$f_i(L) = (1-L)^{d_i}\tilde{f}_i(L),\ i = 1,\ 2,\ \ldots,\ n; \tag{4.20}$$

$$d_i = 0 \text{ for } i \leq n-r \text{ and } d_i > 0 \text{ for } i > n-r. \tag{4.21}$$

Since it is diagonal, $C_{SM}(L)$ can be factorized into the product of two diagonal matrices, one of the divisor polynomials, $g_i(L)$, and one of the $f_i(L)$. That is

$$C_{SM}(L)(L) = G(L)^{-1}F(L), \tag{4.22}$$

where

$$G(L) = diag\ (g_i(L)), \tag{4.23}$$

$$F(L) = diag\ (f_i(L)). \tag{4.24}$$

Using equations (4.20) and (4.21), $F(L)$ may be written:

$$F(L) = \tilde{F}(L)D(L)$$

where $\tilde{F}(L) = diag\ (\tilde{f}_i(L)$,

$$D(L) = \begin{pmatrix} I_{n-r} & & 0' & & \\ & (1-L)^{d_{n-r+1}} & 0 & \ldots & 0 \\ & 0 & (1-L)^{d_{n-r+2}} & \ldots & 0 \\ 0_r & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & \ldots & (1-L)^{d_n} \end{pmatrix}, \tag{4.25}$$

0_r an $r \times (n-r)$ matrix of zeros. It follows that

$$C_{SM}(L) = G(L)^{-1}\tilde{F}(L)D(L). \tag{4.26}$$

By construction, in this expression, the roots of $\tilde{F}(L)$ are the non-unit roots of $C(L)$. There is, however, no control over the roots of $G(L)$. By inverting the unimodular matrices in equation (4.18), $C(L)$ may be written

$$C(L) = U(L)^{-1}C_{SM}(L)V(L)^{-1}. \tag{4.27}$$

Substituting (4.25) into (4.27) gives

$$C(L) = U(L)^{-1}G(L)^{-1}\tilde{F}(L)D(L)V(L)^{-1}$$

and

$$G(L)U(L)C(L) = \tilde{F}(L)D(L)V(L)^{-1}.$$

As long as all the non-unit roots of $C(L)$ lie outside the unit circle, it follows that all the roots of $\tilde{F}(L)$ lie outside the unit circle.

Assumption A3: The roots of $C(L)$ are either equal to unity or lie outside the unit circle.
Then $\tilde{F}(L)^{-1}$ exists, implying

$$\tilde{F}(L)^{-1}G(L)U(L)C(L) = D(L)V(L)^{-1}$$

Since the roots of $D(L)$ are all unit roots it cannot be inverted. If $C(L)$ is the matrix lag polynomial of the VMA

$$x_t = C(L)\epsilon_t, \tag{4.28}$$

then pre-multiplying by $\tilde{F}(L)^{-1}\ G(L)\ U(L)$ gives

$$\tilde{F}(L)^{-1}G(L)U(L)x_t = D(L)V(L)^{-1}\epsilon_t \tag{4.29}$$

This makes the presence of the unit roots explicit but is not in VAR form. In order to take the problem further, specific cases must be considered.

4.3.2.1 The Smith–McMillan–Yoo form and cointegration of order (1, 1)

The starting point for the analysis is the VMA representation in differences,

$$\Delta x_t = C(L)\epsilon_t \tag{4.30}$$

Let assumptions A1, A2 and A3 hold, then from (4.28)

$$\tilde{F}(L)^{-1}G(L)U(L)\Delta x_t = D(L)V(L)^{-1}\epsilon_t. \tag{4.31}$$

Since $C(L)$ is rational, it may be possible to draw out a factor from the denominator that has a unit root. Mathematically, write

$$C(L) = \left\{\frac{\alpha_{i,j}(L)}{\beta_{i,j}(L)}\right\}.$$

If any of the $\beta_{i,j}(L)$ polynomials have a unit root, then this can be factored out of $C(L)$. Suppose $\beta_{m,n}(L) = (1-L)\ \beta^{\star}_{m,n}(L)$, so that it has a unit root, then:

$$C(L) = (1-L)^{-1}C^{\star}(L) \tag{4.32}$$

where

$$C^{\star}(L) = \left\{\frac{\tilde{\alpha}_{i,j}(L)}{\tilde{\beta}_{i,j}(L)}\right\},$$

$$\tilde{\alpha}_{i,j}(L) = \begin{cases} \alpha_{i,j}(L) \text{ if } i = m,\ j = n \\ (1-L)\alpha_{i,j}(L) \text{ otherwise} \end{cases}$$

and

$$\beta_{i,j}(L) = \begin{cases} \beta^{*}_{i,j}(L) & \text{if } i = m,\ j = n \\ \beta_{i,j}(L) & \text{otherwise} \end{cases}$$

Substituting (4.32) into (4.30) gives

$$\Delta x_t = (1-L)^{-1} C^*(L)\epsilon_t$$

or

$$\Delta^2 x_t = C^*(L)\epsilon$$

This gives rise to x_t being I(2), in direct contradiction of the assumption that this process is CI(1, 1). The following assumption is therefore required to exclude this possibility.

Assumption A4: All the roots denominator polynomials of the elements of $C(L)$, $\beta_{i,j}(L)$, $i, j = 1, 2, \ldots, n$, must lie outside the unit circle.

This assumption is worded so as to exclude not only unit roots, but also any roots on or inside the unit circle. Thus assumption A4 is that all the poles of $C(L)$ lie outside the unit circle.[4] A more fundamental way of justifying this assumption is to recognize that if $\beta_{i,j}(L)$ has any roots on or inside the unit circle, then the coefficients of $\dfrac{\alpha_{i,j}(L)}{\beta_{i,j}(L)}$ do not converge and so, strictly speaking, the operator is not even defined, just as it can be argued that the operator $(1-L)$ cannot be inverted. In other words, it is meaningless for $C(L)$ to have any poles on or inside the unit circle.[5]

Assumption A4 implies that all the roots of $G(L)$ in equation (4.31) lie outside the unit circle. The objective is to re-express (4.31) (and hence 4.30) as a VAR in the levels of the process x_t. In order to do this it is necessary to find a way of cancelling the differencing operator from the left-hand side of (4.31). Since $V(L)$ has no unit roots (because it is unimodular), it is necessary and sufficient to find a matrix $D^*(L)$ such that

$$D^*(L)D(L) = \Delta I \tag{4.33}$$

since then, pre-multiplying (4.31) by $D^*(L)$ gives (apart from initial values)

$$\begin{aligned} D^*(L)\tilde{F}(L)G(L)U(L)\Delta x_t &= D^*(L)D(L)V(L)^{-1}\epsilon_t \\ &= \Delta V(L)^{-1}\epsilon_t \qquad (4.34) \\ &\Rightarrow \\ V(L)D^*(L)\tilde{F}(L)^{-1}G(L)U(L)x_t &= \epsilon_t, \qquad (4.35) \end{aligned}$$

which is of the required VAR form. However, such a $D^*(L)$ will not be available for all $D(L)$ of the form given in equation (4.25). To see what is required, write

$$\overline{D}(L) = \begin{pmatrix} (1-L)^{d_{n-r+1}} & 0 & \ldots & 0 \\ 0 & (1-L)^{d_{n-r+2}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & (1-L)^{d_n} \end{pmatrix} \tag{4.36}$$

so that

$$D(L)=\begin{pmatrix} I_{n-r} & 0_r' \\ 0_r & \overline{D}(L) \end{pmatrix} \tag{4.37}$$

and recall that $d_{n-r+i} \geq 1$ for $i = 1,2, \ldots, r$. Partitioning $D^*(L)$ conformably as

$$D^*(L)=\begin{pmatrix} D^*_{1,1}(L) & D^*_{1,2}(L) \\ D^*_{2,1}(L) & D^*_{2,2}(L) \end{pmatrix}$$

gives

$$D^*(L)D(L)=\begin{pmatrix} D^*_{1,1}(L) & D^*_{1,2}(L)\overline{D}(L) \\ D^*_{2,1}(L) & D^*_{2,2}(L)\overline{D}(L) \end{pmatrix}=\Delta I, \tag{4.38}$$

where

$$D^*_{1,1}(L)=\Delta I_{n-r},\ D^*_{1,2}(L)=0,\ D^*_{2,1}(L)=0,\ D^*_{2,2}(L)=I_r. \tag{4.39}$$

The constraints on $D^*_{i,j}(L)$ in (4.39) follow from the matrix equivalence of (4.33) and (4.38). Given (4.37), $D^*_{1,1}(L)$, $D^*_{1,2}(L)$, and $D^*_{2,1}(L)$ impose no restrictions on $\overline{D}(L)$. However, $D^*_{2,2}(L)\,\overline{D}(L) = \Delta I_r$ implies $D^*_{2,2}(L) = I_r$,[6] $\overline{D}(L) = \Delta I_r$, and hence

$$D(L)=\begin{pmatrix} I_{n-r} & 0_r' \\ 0_r & \Delta I_r \end{pmatrix}. \tag{4.40}$$

That is, if x_t is CI(1, 1) then $D(L)$ of (4.31) must be given by (4.40).[7] Furthermore, from the conditions on $D^*(L)$ in (4.39)

$$D^*(L)=\begin{pmatrix} \Delta I_{n-r} & 0_r' \\ 0_r & I_r \end{pmatrix}. \tag{4.41}$$

As a result $D^*(L)\,D(L) = \Delta I$ and substituting into (4.35) gives the VAR in levels corresponding to the VMA in differences when the variables are CI(1, 1).

This illustrates that if x_t is CI(1, 1) with cointegrating rank r (Assumption A2), then the system may be represented either as a VMA in Δx_t or a VAR in x_t, providing the VMA is rational (Assumption A1).

The SMY form of the VMA operator is given by

$$\begin{aligned} C(L) &= U(L)^{-1}C_{SM}(L)V(L)^{-1}, \\ C_{SM}(L) &= G(L)^{-1}\tilde{F}(L)D(L), \\ D(L) &= \begin{pmatrix} I_{n-r} & 0_r' \\ 0_r & \Delta I_r \end{pmatrix}, \end{aligned} \tag{4.42}$$

where $U(L)$ and $V(L)$ are unimodular matrices corresponding to sets of elementary row and column operations respectively.

In summary, the SMY form consists of the factorization of all the unit roots from the VMA operator ($C(L)$) in such a way (as $D(L)$, that, by pre-multiplication by an appropriate matrix ($D^*(L)$), a single differencing operator (Δ) may be isolated on the MA side of the equation. This may then be cancelled with the differencing operator on the AR side where the original VMA is for a differenced process. This is the process represented in (4.34) leading to the final representation of (4.35).

4.3.3 Cointegrating vectors in the VMA and VAR representations of CI (1, 1)

When a CI(1, 1) system is represented in VMA form, the rank of $C(1)$ is $n - r$. The $n \times 1$ cointegrating vectors, β are those such that

$$\beta' C(1) = 0. \tag{4.43}$$

There are r such vectors that are linearly independent. The space of such vectors is the null space (of the columns) of $C(1)$. This can be compared with the corresponding VAR representation. For convenience, put

$$A(L) = V(L)D^*(L)\tilde{F}(L)^{-1}G(L)U(L) \tag{4.44}$$

so that the VAR form of (4.35) may be written

$$A(L)x_t = \epsilon_t.$$

From (4.27) and (4.42)

$$C(L) = U(L)^{-1}G(L)^{-1}\tilde{F}(L)D(L)V(L)^{-1}$$

and hence

$$C(1) = U(1)^{-1}G(1)^{-1}\tilde{F}(1)D(1)V(1)^{-1},$$

where $U(1)$, $G(1)$, $\tilde{F}(1)$ and $V(1)$ are all of full rank, while (4.44) implies

$$A(1) = V(1)D^*(1)\tilde{F}(1)^{-1}G(1)U(1). \tag{4.45}$$

It is also straightforward to see that

$$C(1)A(1) = U(1)^{-1}G(1)^{-1}\tilde{F}(1)D(1)D^*(1)\tilde{F}(1)^{-1}G(1)U(1) \tag{4.46}$$

and

$$A(1)C(1) = V(1)D^*(1)D(1)V(1)^{-1}. \tag{4.47}$$

Now replacing L by *1* in (4.40) and (4.41),

$$D(1) = \begin{pmatrix} I_{n-r} & 0 \\ 0 & 0 \end{pmatrix} \text{ and } D^*(1) = \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix}$$

so clearly *rank* $(D(1)) = n - r$ and $rank(D^*(1)) = r$. Thus

$$rank(A(1)) = rank(D^*(1)) = r.$$

In addition,

$$D(1)D^*(1) = D^*(1)D(1) = 0$$

and substituting into (4.46 and 4.47) gives

$$C(1)A(1) = A(1)C(1) = 0.$$

It follows that the rank of $A(1)$ is r, the cointegrating rank, and its rows are cointegrating vectors and span the space of cointegrating vectors (meaning all cointegrating vectors can be constructed from a linear combination of the rows of $A(1)$).

4.3.3.1 A(L) as partial inverse of C(L) in the CI(1, 1) case

From the definitions of $A(L)$ and $C(L)$,

$$A(L)C(L) = C(L)A(L) = \Delta I_n. \tag{4.48}$$

When the VMA in differences is considered then

$$\Delta x_t = C(L)\epsilon_t$$

Pre-multiplying this by $A(L)$,

$$A(L)\Delta x_t = A(L)C(L)\epsilon_t = \Delta\epsilon_t,$$

which, on cancelling the differencing operator, gives the VAR form. Pre-multiplying again by $C(L)$ reverses the transformation:

$$C(L)A(L)x_t = C(L)\epsilon_t$$
$$\Rightarrow$$
$$\Delta x_t = C(L)\epsilon_t,$$

regenerating the VMA. Broadly speaking, then, the problem that has been solved to show that the VMA in differences can be expressed as a VAR in levels is to find a matrix $A(L)$ such that equation (4.48) holds. The solution is (4.44).[8]

4.3.4 Equivalence of VAR and VMA representations in the CI(1, 1) case

It has been shown that VMA in differences representation of a CI(1, 1) system implies a VAR in levels as long as assumptions A1–A4 hold. It is also straight forward to move back to the VMA representation again, since the VAR operator $A(L)$ (4.44) also satisfies the assumptions, with r replacing $n - r$ throughout. This follows since $A(L) = V(L)D^*(L)\tilde{F}(L)^{-1}G(L)U(L)$ is rational (due to presence of $\tilde{F}(L)^{-1}$), with rank r. Its roots are those of $G(L)$ (all outside the unit

circle because these are the poles of $C(L)$) and those of $D^*(L)$ (unit roots). Its poles are the roots of $\tilde{F}(L)$, and so are all outside the unit circle.

Now consider any other VAR in levels representation of a CI(1, 1) system, say

$$\tilde{A}(L)x_t = \epsilon_t.$$

Then as long as $\tilde{A}(L)$ satisfies assumptions A1–A4, then there exists a matrix $\tilde{C}(L)$ such that $\tilde{A}(L)\,\tilde{C}(L) = \tilde{C}(L)\,\tilde{A}(L) = \Delta I$ and, hence, pre-multiplying by $\tilde{C}(L)$, the VAR becomes

$$\tilde{C}(L)\tilde{A}(L)x_t = \tilde{C}(L)\epsilon_t \Rightarrow \Delta x_t = \tilde{C}(L)\epsilon_t,$$

which is a VMA representation. By arguments similar to those above, $\tilde{C}(L)$ will also satisfy the assumptions. It is therefore the case that, among the class of models having operators obeying assumptions A1–A4, the VMA in differences and VAR in levels are equivalent representations of a CI(1, 1) system, and that this sub-class of models is closed.

4.4 Johansen's VAR representation of cointegration

The Engle–Granger–Yoo approach begins by assuming, that is imposing, that the univariate processes are I(1) and that the vector moving average process has reduced rank. Johansen's (1995) approach reflects the assumption of Sims (1980) that the VAR, though not necessarily the correct underlying process, may in practice be the only type of model that can be reliably identified and estimated. This approach also eliminates a dichotomy that existed prior to our knowledge of cointegration, between dynamic time-series models that derive from the LSE approach to econometrics via Hendry and Sargan as compared with the approach based on expectations that views the VAR as a fundamental reduced form. The former approach emphasize the role of the underlying Data Generation Process (DGP) to model complex agent interaction at an aggregate level with the error correction revealing the long-run theoretical model. The VAR is a natural extension of the univariate time series approach to analyze the properties of a vector of time series. Johansen has amalgamated the time series methodology of the VAR to incorporate long-run relationships associated with cointegration and provided an approach to estimation and testing which determines the conditions necessary on the VAR for the processes to be I(1) and cointegrated. The required conditions are more complex than those for the VMA in differences, but the benefit lies in facility of estimation and an inferential procedure that derives from the conventional maximum likelihood approach both in confirming cointegration and in testing theoretical propositions on parameters.

The starting point is a VAR where the intercept has been set to zero for simplicity. That is

$$A(L)x_t = \epsilon_t, \tag{4.49}$$

where

$$A(L) = I + \sum_{i=1}^{p} A_i L^i.$$

It is also assumed that all the roots of $A(L)$ are either outside the unit circle or equal to unity. Thus while non-stationarity is allowed, this can only be due to standard unit roots.[9] This VAR may be written

$$x_t + \sum_{i=1}^{p} A_i x_{t-i} = \epsilon_t \tag{4.50}$$

and reparameterized as the VECM

$$\Delta x_t + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} = \prod x_{t-1} + \epsilon_t, \tag{4.51}$$

where $\prod = I + \sum_{i=1}^{p} \Gamma_i = -A(1)$ and $\Gamma_i = \sum_{j=i+1}^{p} A_i$.

4.4.1 Cointegration assuming integration of order 1

For simplicity, assume that $x_t \sim I(1)$. Then Π must be of reduced rank, and unless $\Pi = 0$, x_t must be cointegrated. This can be shown by contradiction by assuming that Π is of full rank. First note that the VECM can be written

$$\prod x_{t-1} = \Delta x_t + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} - \epsilon_t$$

which is I(0) since all terms on the right-hand side are I(0) when $x_t \sim I(1)$. Then Π must be of reduced rank, since if this were not the case then its inverse would exist and

$$x_{t-1} = \prod^{-1} \left[\Delta x_t - \sum_{i=1}^{p-2} \Phi_i^* \Delta x_{t-i} - \epsilon_t \right] \sim I(0)$$

which contradicts $x_t \sim I(1)$. The fact that $\Pi x_{t-1} \sim I(0)$ then establishes cointegration as long as $\Pi \neq 0$, the rows of Π being cointegrating vectors. If $\Pi = 0$ then it is immediate from the VECM that the process is not cointegrated. Note that Π is an $n \times n$ matrix, and let *rank* $(\Pi) = r$ where for cointegration $r<n$, so that Π is of reduced rank. Then there exist $n \times r$ matrices α and β both of maximum rank, r, such that

$$\prod = \alpha\beta' \tag{4.52}$$

Furthermore, since each row of Π is a linear combination of the rows of β', the rows of β' are cointegrating vectors. The rank of Π is known as the *cointegrating rank* of the system. This establishes the following result.

4.4.1.1 Cointegrated VARs with I(1) processes

Let $x_t \sim I(1)$ and obey the VECM (4.51) with *rank* (Π) = *r*. Then:

(i) $0 < r < n$;
(ii) the rows of Π are cointegrating vectors;
(iii) the rows of β' in the representation of equation (4.52) constitute a set of linearly independent cointegrating vectors.

4.4.2 Conditions for the VAR process to be I(1) and cointegrated

The difficulty with assuming $x_t \sim I(1)$ is that the order of integration in the VAR can be greater than 1. It is necessary to establish conditions for the processes being I(1), to check that these can be satisfied, and to begin to consider how to handle higher order integrated processes. Some further preliminaries are necessary.

(i) Defining

$$\Gamma(L) = I + \sum_{i=1}^{p-1} \Gamma_i L^i,$$

the VECM may be written,

$$\Gamma(L)\Delta x_t = \prod x_{t-1} + \epsilon_t,$$

where

$$\Upsilon = \Gamma(1). \tag{4.53}$$

Then $A(L)$ may be written

$$A(L) = -\prod + (\Upsilon + \prod)(1-L) + \Gamma^*(L)(1-L)^2 \tag{4.54}$$

where $\Gamma^*(L)$ is a polynomial of order $p - 2$. Thus, substituting (4.54) into (4.49), the VAR may be written

$$-\prod x_t + (\Upsilon + \prod)\Delta x_t + \Gamma^*(L)\Delta^2 x_t = \epsilon_t. \tag{4.55}$$

(ii) For any full rank $n \times r$ ($r \leq n$) matrix φ, define its orthogonal compliment, $\varphi_\perp$ dimensioned $n \times (n - r)$ with rank $n - r$ such that

$$\varphi'\varphi_\perp = 0,$$
$$\varphi_\perp = \begin{cases} 0 \text{ if } r = n \\ I \text{ if } r = 0 \end{cases}.$$

There are explicit formulations of $\varphi_\perp$, though sub-blocks of this matrix are arbitrary. Also define

$$\overline{\varphi} = \varphi(\varphi'\varphi)^{-1} \tag{4.57}$$

with the projection matrix

$$P_\varphi = \varphi(\varphi'\varphi)^{-1}\varphi' = \overline{\varphi}\varphi' = \varphi\overline{\varphi}', \tag{4.58}$$

and note that

$$\overline{\varphi}'\varphi = I_r = \varphi'\overline{\varphi}.$$

Johansen's key (necessary and sufficient) condition on the VAR such that the processes are integrated of order 1 and cointegrated, is expressed in terms of Υ, $\alpha_\perp$ and $\beta_\perp$. An outline of the derivation of this condition is provided below.[10] The result applies only to VARs the roots of which are either equal to one or lie outside the unit circle.

The approach used is to split the differenced process, Δx_t, into components relating to the directions of (potential) cointegration, η_t (which occur in differenced form) and non-cointegration, u_t (in levels). The difference process is then cumulated (summed from the first to the t^{th} values) to give an equation for the levels, x_t. The cumulation results in: the sum of the u_t, giving rise to a stochastic trend (a unit root process if u_t is stationary); the transformation of the differences of η_t to its levels; and the appearance of an initial value vector (analogous to a constant of integration). To keep the treatment simple, the initial values are ignored (set to zero).[11] Since in detail, η_t is a set of linear combinations of the components of x_t, if both the u_t and the η_t are I(0) then x_t is both I(1) (as a result of the stochastic trend involving u_t) and cointegrated (because then η_t is a linear combination of I(1) variables that is I(0)). So the proof revolves around showing that u_t and the η_t are I(0). The condition results from the need for the stationarity of these processes. Having shown this, it is fairly straight forward to show that cointegration of order (1,1) implies the condition, and hence it is established that the condition is both sufficient and necessary.

An outline of the statement and proof is provided here. The result is that a necessary and sufficient condition for x_t to be both I(1) and cointegrated (i.e. CI(1, 1)) is that

$$rank(\alpha'_\perp \Upsilon \beta_\perp) = n - r, \tag{4.59}$$

i.e. $\alpha'_\perp \gamma\beta_\perp$ is of full rank.

To show why this matrix is important, first decompose the difference process as

$$\Delta x_t = (P_{\beta_\perp} + P_\beta)\Delta x_t = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}\beta'\Delta x_t + \beta(\beta'\beta)^{-1}\beta'\Delta x_t \tag{4.60}$$

The second term on the right-hand side of (4.60) can be rearranged in terms of potentially cointegrating combinations of x_t. Define

$$\eta_t = \beta' x_t \tag{4.61}$$

these being the potentially cointegrating combinations. Also, arising from the first term on the right hand side of equation (4.61), define

$$u_t = \beta' \Delta x_t \tag{4.62}$$

Then, from (4.60)

$$\Delta x_t = \beta_\perp (\beta_\perp' \beta_\perp)^{-1} u_t + \beta (\beta' \beta)^{-1} \Delta \eta_t. \tag{4.63}$$

The process of interest is not Δx_t but x_t itself, obtained by summing the difference process up to the current period. When this is done, an initial value is also generated. In addition, in order to reuse t as the index for the current period, a different index has to be used on the process being summed. Thus,

$$\sum_{i=1}^{t} \Delta x_i = x_t - x_0.$$

Applying the same operation to the right-hand side of (4.63) yields

$$x_t - x_0 = \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \sum_{i=1}^{t} u_i + \beta (\beta' \beta)^{-1} (\eta_t - \eta_0) \tag{4.64}$$

Ignoring the initial values x_0 and η_0, this becomes

$$x_t = \beta_\perp (\beta_\perp' \beta_\perp)^{-1} \sum_{i=1}^{t} u_i + \beta (\beta' \beta)^{-1} \eta_t. \tag{4.65}$$

From this last equation, since

$$u_t \sim I(0) \Rightarrow \sum_{i=1}^{t} u_i \sim I(1)$$

it can be seen that $x_t \sim I(d)$, $d \geq 1$, depending on the order of integration of η_t. In particular, if $\eta_t \sim I(0)$ then, from (4.65), x_t is the sum of an I(1) and an I(0) process and so is itself I(1). But if $x_t \sim I(1)$ and $\eta_t \sim I(0)$, then by the definition of η_t (a set of linear combinations of x_t), x_t is also cointegrated. In brief,

$$u_t \sim I(0), \eta_t \sim I(0) \Rightarrow x_t \sim CI(1,1).$$

Thus it is sufficient to show that both u_t and η_t are I(0). It is in the process of obtaining this result that condition (4.59) arises.
Define

$$\tilde{x}_t = (\eta_t' \; u_t')'.$$

If a VAR representation for can be found for $\tilde{x}_t$, all the roots of which lie outside the unit circle, then $\tilde{x}_t$ is stationary.[12] The required VAR is obtained by:

(i) pre-multiplying equation (4.55) by $\overline{\alpha}'$ to give one new equation, and by $\overline{\alpha}'_\perp$ to obtain another;
(ii) substituting using 4.61 and 4.62 to give equations in u_t and η_t, though a term in $\Delta^2 x_t$ remains;
(iii) noting that the term in $\Delta^2 x_t$ can be expressed in terms of the differences of $\tilde{x}_t$;
(iv) expressing the resultant equation in terms of $\tilde{x}_t$ only.

The result is the VAR

$$\tilde{A}(L)\tilde{x}_t = (\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t, \tag{4.66}$$

where the operator $\tilde{A}$ (L) can be written

$$\tilde{A}(L) = \tilde{A}^{(1)}(L) + \tilde{A}^{(2)}(L)\Delta \tag{4.67}$$

and $\tilde{A}^{(1)}$ (L) is partitioned conformably with $\tilde{x}_t = (\eta'_t \, u'_t)'$ as

$$\tilde{A}^{(1)}(L) = \begin{bmatrix} -I & \overline{\alpha}'\Upsilon\overline{\beta}_\perp \\ 0 & \overline{\alpha}'_\perp\Upsilon\beta_\perp \end{bmatrix} \tag{4.68}$$

It remains to establish that $\tilde{A}$ (L) has all its roots outside the unit circle. This is done in two stages. Firstly it is established that any non-stationarity is due to unit roots (by showing that the roots of $\tilde{A}$ (L) and A (L) are the same, except that the number of unit roots may differ), and then showing that $\tilde{A}$ (L) has no unit roots. To show the relationship between the roots of $\tilde{A}$ (L) and A (L), note that

$$\left|\tilde{A}(z)\right| = (1-z)^{-(n-r)}\left|A(z)\right\|Q| \tag{4.69}$$

where Q is a $(n - r) \times (n - r)$ matrix depending on α, $\alpha'_\perp$, β and $\beta_\perp$.[13] Thus

$$\left|A(z)\right| = 0 \Rightarrow \left|\tilde{A}(z)\right| = 0 \; \textit{for} \; z \neq 1,$$

so all non-unit roots of A (L) are also roots of $\tilde{A}$ (L), but due to the presence of the factor of $(1 - z)^{-(n-r)}$, $z = 1$ may or not be a root of $\tilde{A}$ (L). Thus if the roots of A (L) are all either outside the unit circle or equal to 1, so are those of $\tilde{A}$ (L). To show that $\tilde{A}$ (L) has no unit roots, consider $\tilde{A}$ (1). The required condition is $|\tilde{A}$ $(1)| \neq 0$, or equivalently that $\tilde{A}$ (1) should have full rank. From (4.67), $\tilde{A}$ $(1) = \tilde{A}^{(1)}$ (1), and directly from (4.68):

$$\tilde{A}^{(1)}(1) = \begin{bmatrix} -I & \overline{\alpha}'\Upsilon\overline{\beta}_\perp \\ 0 & \overline{\alpha}'_\perp\Upsilon\beta_\perp \end{bmatrix}.$$

Thus $\left|\tilde{A}(1)\right| = \left|-\overline{\alpha}'_\perp\Upsilon\overline{\beta}_\perp\right|$ and since

$$\overline{\alpha}'_\perp\Upsilon\overline{\beta}_\perp \; \textit{full rank} \; \Rightarrow \left|\overline{\alpha}'_\perp\Upsilon\overline{\beta}_\perp\right| \neq 0 \Leftrightarrow \left|\tilde{A}(1)\right| \neq 0,$$

it follows that $rank(\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}) = n - r$ is a sufficient condition for $x_t \sim CI\ (1, 1)$.

Necessity is easily established. Start from the assumption that $x_t \sim I(1)$. This means that in the original VAR, $|A\ (1)| = 0$, that is $A\ (1)$ is not of full rank. The rank of $A\ (1)$ is r. Therefore, there exist full rank $n \times r$ matrices α and β such that $\Pi = \alpha\beta'$ and the variables $\eta_t = \beta' x_t$ and $u_t = \beta'_{\perp}\Delta x_t$ can be constructed, where $u_t \sim I(0)$ since $x_t \sim I(1)$. It also follows from the VECM

$$\Delta x_t = \prod x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t,$$

that Πx_t must be I(0), and hence that $\eta_t = \beta' x_t$ is I(0). Thus the VAR for $\tilde{x}_t = (\eta'_t u'_t)'$, still given by (4.66) must be stationary, so $|\tilde{A}\ (1)| \neq 0$. But as before,

$$\tilde{A}(1) = \begin{bmatrix} -I & \overline{\alpha}'\Upsilon\overline{\beta}_{\perp} \\ 0 & \overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp} \end{bmatrix}$$

and so $|\tilde{A}\ (1)| = |-\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}|$. Hence $|-\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}| \neq 0$, that is $\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}$ is of full rank.

Finally note that

$$\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp} = (\alpha'_{\perp}\alpha_{\perp})^{-1}\alpha'_{\perp}\Upsilon\beta_{\perp}(\beta'_{\perp}\beta_{\perp})^{-1},$$

where $(\alpha'_{\perp}\alpha_{\perp})^{-1}$ and $(\beta'_{\perp}\beta_{\perp})^{-1}$ are full rank $(n - r) \times (n - r)$ matrices. Thus

$$rank(\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}) = rank(\alpha'_{\perp}\Upsilon\beta_{\perp}),$$

so that the rank condition applies equivalently to the simpler matrix $\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}$, as required.

In summary, when the vector process follows a VAR given by (4.49), and where the only non-stationary roots are unity, $\overline{\alpha}'_{\perp}\Upsilon\overline{\beta}_{\perp}$ being of full rank is necessary and sufficient for cointegration of order (1,1). In this matrix, $\alpha_{\perp}$ and $\beta_{\perp}$ are the orthogonal compliments of α and β defined by (4.52), and Υ is given by (4.53).[14]

4.4.2.1 Discussion

This key condition is undoubtedly difficult to understand from an intuitive point of view. However, practically speaking, its function is to guard against the component processes being I(2). If it is assumed from the outset that the processes are I(1), then the required condition on the VAR is simply that Π is of reduced rank. The condition can be used to extend the analysis of cointegrated systems to cases where the processes can be I(2). Having established the condition for I(1) and cointegration, since this is necessary and sufficient, clearly $\alpha'_{\perp}\ \Gamma\ \beta_{\perp}$ must be of reduced rank in order for the processes to be of a higher order of integration.

4.4.3 The moving average representation

To obtain a VMA representation, note that equation (4.65) provides an expression for x_t in terms of η_t and u_t. Equation (4.66) is a stationary VAR for $\tilde{x}_t$ (i.e. η_t and u_t) and can therefore be inverted. Thus a solution is available for η_t and u_t in terms of the stationary disturbances ϵ_t. That is

$$\tilde{x}_t = \tilde{C}(L)(\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t. \tag{4.70}$$

Thus from (4.66), x_t can be expressed as a function of ϵ_t through expressing η_t and u_t in this way. In equation (4.65) u_t appears as the increment in the stochastic trend, and a little further investigation of this term is potentially useful. Note that

$$u_t = (0 \quad I)\tilde{x}_t. \tag{4.71}$$

and applying the usual reparameterization $\tilde{C}(z) = \tilde{C}(1) + (1 - z)\tilde{C}^*(z)$ where, since $\tilde{C}(z) = \tilde{A}^{-1}(z)$

$$\tilde{C}(1) = \tilde{A}^{-1}(1) = \begin{bmatrix} -I & \overline{\alpha}'\Gamma\overline{\beta}_\perp \\ 0 & \overline{\alpha}_\perp'\Gamma\overline{\beta}_\perp \end{bmatrix}^{-1} = \begin{bmatrix} -I & \overline{\alpha}'\Gamma\overline{\beta}_\perp(\overline{\alpha}_\perp'\Gamma\overline{\beta}_\perp)^{-1} \\ 0 & (\overline{\alpha}_\perp'\Gamma\overline{\beta}_\perp)^{-1} \end{bmatrix}. \tag{4.72}$$

Thus

$$\begin{aligned} u_t &= (0 \;\; I)\tilde{x}_t = (0 \;\; I)\tilde{C}(L)(\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t \\ &= (0 \;\; I)\left[\tilde{C}(1) + (1-L)\tilde{C}^*(L)\right](\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t \\ &= (0 \;\; I)\left[\tilde{C}(1) + (1-L)\tilde{C}^*(L)\right](\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t \\ &= (0 \;\; I)\tilde{C}(1)(\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t + (0 \;\; I)(1-L)\tilde{C}^*(L)(\overline{\alpha} \;\; \overline{\alpha}_\perp)'\epsilon_t. \end{aligned}$$

Using equation (4.72), pre-multiplying by $\overline{\beta}_\perp$ and letting $\tilde{C}^+(L) = \overline{\beta}_\perp (0 \; I)\, \tilde{C}^*(L)\, (\overline{\alpha} \;\; \overline{\alpha}_\perp)'$, gives:

$$\overline{\beta}_\perp u_t = \overline{\beta}_\perp(\overline{\alpha}_\perp'\Gamma\overline{\beta}_\perp)^{-1}\overline{\alpha}_\perp'\epsilon_t + \tilde{C}^+(L)\Delta\epsilon_t. \tag{4.73}$$

Summing terms in (4.73) and setting initial values to zero for simplicity:

$$\overline{\beta}_\perp \sum_{i=1}^{t} u_i = \overline{\beta}_\perp(\overline{\alpha}_\perp'\Gamma\overline{\beta}_\perp)^{-1}\overline{\alpha}_\perp' \sum_{i=1}^{t}\epsilon_i + \tilde{C}^+(L)\epsilon_t.$$

This can be substituted for the first term on the right-hand side of (4.65). The remaining term, $\overline{\beta}\eta_t = \beta(\beta'\beta)^{-1}\eta_t$ requires the expression of η_t in terms of ϵ_t. It follows from (4.70) and the fact that $\eta_t = (I \;\; 0)\tilde{x}_t$, this term may be written

$$\eta_t = D(L)\epsilon_{t-1} \tag{4.74}$$

$$\overline{\beta}\eta_t = D^+(L)\epsilon_t. \tag{4.75}$$

Expressions (4.74) and (4.75) can be substituted into (4.65) to give

$$x_t = \overline{\beta}_\perp(\overline{\alpha}_\perp'\Gamma\overline{\beta}_\perp)^{-1}\overline{\alpha}_\perp' \sum_{i=1}^{t}\epsilon_i + \tilde{C}^+(L)\epsilon_t(L)\epsilon_t + D^+(L)\epsilon_t. \tag{4.76}$$

This is further simplified by setting $C = \bar{\beta}_\perp (\bar{\alpha}'_\perp \Gamma \bar{\beta}_\perp)^{-1} \bar{\alpha}'_\perp$ and $C(L) = \tilde{C}^+(L) + D^+(L)$. Hence

$$x_t = C\sum_{i=1}^{t} \epsilon_i + C(L)\epsilon_t$$

This is the VMA representation corresponding to the VAR (4.49).[15]

4.5 Johansen's approach to testing for cointegration in systems

The Johansen methodology is based on the VAR representation and referred to as a maximum likelihood approach. This is because the underlying estimation method which provides the Johansen test statistics is in fact maximum likelihood.[16]

4.5.1 Testing for reduced rank and estimating cointegrating vectors

4.5.1.1 Review of source of reduced rank in cointegrated systems

Consider the VECM

$$\Delta x_t = \prod x_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t.$$

The rank conditions on Π are discussed above: the condition for x_t to be CI(1, 1), given $x_t \sim I(1)$ is *rank* $(\Pi) = r$, where $0 < r < n$. The cointegrating rank r is the number of linearly independent cointegrating vectors of the system.[17]

The Johansen approach to testing for cointegration (that is, testing to obtain an estimate of the cointegrating rank) exploits these properties. In the sections below, the two commonly used tests of Johansen are derived.

4.5.1.2 Using eigenvalues and eigenvectors in cointegration analysis

Eigenvalues can be regarded as a set of summary statistics of a matrix from which a number of key properties can be determined. Each eigenvalue is associated with an eigenvector. Any matrix can be expressed in terms of its eigenvalues and eigenvectors.

In Johansen's cointegration analysis, the key statistics are a set of non-negative eigenvalues. In testing for cointegration, or more accurately determining the cointegrating rank, interest focuses on those that are significant, that is, significantly greater than zero. Because of the association of an eigenvector with an eigenvalue, an eigenvector is insignificant if its eigenvalue is insignificantly different from zero. Thus the significance of an eigenvector can be tested through the significance of its eigenvalue. If the eigenvalue is significant it is meaningful go on to calculate, and work with the corresponding eigenvector. In the problem that arises in cointegration analysis, the eigenvectors are the cointegrating vectors.

4.5.2 The removal of nuisance parameters

The matrix that characterizes the cointegration properties of the system is Π. All other parameters of the model and associated variables are irrelevant. These nuisance terms can be removed by regressing both Δx_t and x_{t-1} on Δx_{t-i}, $i = 1, 2, \ldots, p-1$ using ordinary least squares. The residuals from these regressions will be purged of their correlation with the lagged differences. Let $R_{0,t}$ and $R_{1,t}$ be the $n \times 1$ residual vectors from the regressions with Δx_t and x_{t-1} respectively as dependent variables. Then the least squares estimate of Π in

$$R_{1,t} = \prod R_{0,t} + error$$

is the same as that from (4.51).[18] From the point of view of maximum likelihood estimation, this is equivalent to concentrating the likelihood function. As long as a Gaussian likelihood is used, the maximum likelihood estimator of Π is also unaffected, even under the restriction that the matrix is of reduced rank, r, $r < n$. That is, the estimates of α and β in

$$\prod = \alpha\beta'$$

are unaffected. This is explained in Appendix D. The requirement for Gaussianity means that the disturbances, ϵ_t, must be jointly normally distributed, an important assumption.

4.5.3 Estimating potentially cointegrating relations

The residual vectors inherit the integration properties of the dependent variables since all explanatory variables in the auxiliary regressions generating them are stationary. Thus $R_{0,t} \sim$ I(0), $R_{1,t} \sim I(1)$. One way of motivating the problem of the determination of the cointegrating vectors is to observe that the correlation between an I(0) variable and a linear combination of I(1) variables will be low unless the particular linear combination of the I(1) variables is itself I(0). In such a case, the coefficients of the linear combination constitute a cointegrating vector. So, by choosing the linear combination to maximize the correlation, if a cointegrating combination is available, the procedure should select this combination. Of course, it may not exist, in which case the correlation between $R_{0,t}$ and all linear combinations of $R_{1,t}$ will be low. Or, there may be more than one, in which case some will be more correlated with the stationary residuals than others. This problem is closely related to that of canonical correlation applied to $R_{0,t}$ and $R_{1,t}$. Canonical correlation involves the transformation of each vector using linear combinations of their elements such that the transformed vectors have the identity matrix as variance – covariance matrix and the elements of the transformed vectors have a diagonal correlation matrix. This is explained in detail in Appendix D. The resultant correlations are known as canonical correlations. At first glance, the canonical correlation problem seems to apply more structure than is required,

and does not seem to address the issue of correlation maximization. However, a close examination of the relationship between this and the complete maximum likelihood problem reveals that in fact the problems yield the same solutions (see Appendix D). Both the maximum likelihood (ML) and canonical correlation problem deal with the sample covariance matrix of the residual vectors. Define the sample covariance matrices

$$S_{i,j} = T^{-1}\sum_{t=1}^{T} R_{i,t}R'_{j,t} \; i,j = 0,1.$$

where T is the sample size. In each case the problem reduces to an examination of the eigenvalues and eigenvectors of the matrix $S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}$. The following points are relevant:

(i) The eigenvalues of this problem are the squares of the canonical correlations.
(ii) The corresponding eigenvectors are the potential cointegrating vectors, β.
(iii) The maximized value of the log-likelihood function depends only on the r largest eigenvalues and $S_{0,0}$, where the term in $S_{0,0}$ is additive and so does not appear in expressions for the difference between maximized log-likelihood functions for different r.
(iv) Estimates of α, called the adjustment coefficients, are available as a function of the estimates of β and $S_{0,1}$.

The eigenvalue problem in the ML context is often expressed in generalized form as

$$\left|\lambda I - S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}\right| = 0$$

where the eigenvalues lie in [0,1] and are denoted in ordered form as $0 \leq \lambda_n \leq \lambda_{n-1} \ldots \leq \lambda_2 \leq \lambda_1 \leq 1$. This eigenvalue problem is equivalent to the more usual problem

$$\left|\lambda I - S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}\right| = 0 \tag{4.77}$$

when $S_{1,1}$ is non-singular (Dhrymes 1984), and as such is the same as that for the canonical correlation problem.

For each eigenvalue that satisfies (4.77), there is an equivalent eigenvector, v_i, that is a solution to the following homogenous system of linear equations:[19]

$$(\lambda I - S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1})v_i = 0$$

or

$$\lambda_i v_i = S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}v_i \tag{4.78}$$

For $i = 1, \ldots, r$, the eigenvectors define r cointegrating relationships, that is $\beta_i = \nu_i$, and so

$$\lambda_i \beta_i = S_{1,1}^{-1} S_{1,0} S_{0,0}^{-1} S_{0,1} \beta_i$$

It follows from the algebra of the problem that $S_{0,1} S_{1,1}^{-1}$ is an estimator of Π so an estimate of α can be obtained from that of β since

$$\alpha = S_{0,1} \beta$$

In addition:

$$\lambda_i = \alpha_i' S_{0,0}^{-1} \alpha_i$$

where α_i is the i^{th} column of α. This result follows only where the normalization $\beta' S_{1,1} \beta = I$ is used. Equation (4.79) shows that a test of $\lambda_i = 0$ is equivalent to a test of $\alpha_i = 0$ that is, that the i^{th} column of α is zero. The restriction $\alpha_i = 0$ means that the i^{th} potentially cointegrating combination does not appear in the VECM, the reason being either that it is not a stationary combination, or that it is not significantly linearly independent of the combinations associated with the larger eigenvalues, λ_j, $j < i$.

The maximized log likelihood conditional on r, ignoring certain constants, is given by

$$\log \tilde{L}_{MAX} = -\frac{T}{2}\left[\log|S_{0,0}| + \sum_{i=1}^{r} \log(1 - \lambda_i) \right]$$

The estimation of the β_i is not dependent on r, although their interpretation is. If λ_j is insignificantly different from zero, then the corresponding canonical correlation, using β_j as the coefficients of the I(1) processes, is insignificantly different from zero. That is, β_j does not result in a stationary combination of the I(1) processes.

4.5.4 Testing cointegrating rank

Since it is a function of r, denote the maximized likelihood above as

$$\log \tilde{L}_{MAX}(r) = -\frac{T}{2}\left[\log \left|S_{0,0}\right| + \sum_{i=1}^{r} \log \left(1 - \lambda_i\right) \right], \; r = 0, \; 1, \; \ldots, \; n,$$

where the summation term does not appear if $r = 0$. The log likelihood $\log \tilde{L}_{MAX}(r_1)$ is a restricted version of $\log \tilde{L}_{MAX}(r_0)$ if $r_0 < r_1$. Thus, the likelihood ratio statistic for comparing H_0: $r \leq r_0$ with the alternative H_1: $r \leq r_1$ is

$$LR(r_0, r_1) = -2\left[\log \tilde{L}_{MAX}(r_1) - \log \tilde{L}_{MAX}(r_0) \right]$$

since $\log \tilde{L}_{MAX}$ (r^*) is the log-likelihood for a model where H_0: $r \leq r^*$. Substituting from the expression for the maximized log likelihood in terms of the eigenvalues, this can be written

$$LR(r_0, r_1) = -T\left[\sum_{i=r_0+1}^{r_1} \log\ (1-\lambda_i)\right]$$

If used in a conventional way, the null hypothesis would be rejected for large values of the test statistic, such a rejection being a statement that the eigenvalues λ_i, $i = r_0 + 1, \ldots, r_1$ were jointly significantly different from zero. The normal choices of r_0 and r_1 are:

(a) $r_0 = j - 1$, $r_1 = n$, $j = 1, 2, \ldots, n$;
(b) $r_0 = j - 1$, $r_1 = j$, $j = 1, 2, \ldots, n$.

In case (a), the test is of whether the eigenvalues λ_i, $i = j, \ldots, n$ are jointly zero. These are the $n - j$ smallest eigenvalues. In case(b), the test is of whether the eigenvalue λ_j, alone is zero.[20] In performing the two tests, the information exploited is different, and so the inferences may not always agree.

The test associated with (a) is known as the *trace* statistic, denoted λ_{trace} $(j - 1)$. The null (H_0) and alternative (H_1) hypotheses are, for $j = 1,2, \ldots, n$:

$$H_0 : r \leq j-1$$
$$H_1 : r \leq n.$$

The test statistic is

$$LR(j-1, n) = -T\left[\sum_{i=j}^{n} \log\ (1-\lambda_i)\right] = \lambda_{trace}(j-1)$$

The test related to (b) is known as the *maximal eigenvalue* statistic, denoted $\lambda_{\max}$ $(j - 1)$, and has the hypotheses

$$H_0 : r \leq j-1$$
$$H_1 : r \leq j$$

for which the test statistic is

$$LR(j-1, j) = -T\ \log\ (1-\lambda_j) = \lambda_{\max}(j-1).^{21}$$

Each test rejects the null hypothesis for large values of the test statistic, which must be positive. Thus, using cv to stand for the critical value of the test, and $\lambda\ (j - 1)$ to represent the test statistic, the form of the test is:

$$reject\ H_0\ if\ \lambda(j-1) > cv$$

The critical values for the two tests are different in general (except when $j = n$), come from non-standard null distributions and are dependent on the sample size and the number of cointegrating vectors being tested for. The distribution theory leading to the critical values of the test is described in

Appendix D.[22] Most computer packages that compute the test statistics also compute critical values for the tests.

The interpretation of these tests should be considered carefully.

The trace statistic always has as its unrestricted case, that the cointegrating rank is at most n. The restricted, or null, case is that the cointegrating rank is at most j-1. This is consistent with the statement of the hypotheses in terms of the eigenvalues as

$$H_0 : \lambda_i = 0,\ i = j, \ldots, n$$
$$H_1 : \lambda_i > 0 \text{ for at least one of } i = j, \ldots, n:$$

since in the alternative case at least one of the set of eigenvalues being tested must be non-zero. So, it might be that only the largest remaining, the j^{th}, is non-zero, hence that the cointegrating rank is j, or at the other extreme, it could be that all are, in which case the rank is n. Given that the cointegrating rank cannot exceed n, the simplest way to represent the case under the alternative is $r \geq j$.

The maximal eigenvalue test has the same restricted model, but the unrestricted model only considers a cointegrating rank one higher. Thus, the only case explicitly considered under the alternative is a cointegrating rank is greater by one. In terms of the eigenvalues the hypotheses become

$$H_0 : \lambda_j = 0$$
$$H_1 : \lambda_j > 0.$$

From the hypotheses expressed in terms of eigenvalues it can be seen that the trace test is a joint test of all eigenvalues smaller than λ_{j-1}, that is λ_j, λ_{j+1}, …, λ_n while the maximal eigenvalue test is of λ_j only. The hypotheses of the two tests are summarized in Table 4.1.

In the case of neither test is the cointegrating rank established uniquely. To determine the cointegrating rank it is necessary to focus down onto a particular value for r. This can be achieved by testing in sequence, moving in the direction of increasing cointegrating rank. Notice that when using the trace test, rejection of the null $r \leq s - 1$ leads to the conclusion that $r \geq s$. The next

Table 4.1 Hypotheses of the maximal eigenvalue and trace statistics, $j = 1, 2, \ldots, n$

Test	*Hypotheses*				*Test statistic*
	In terms of cointegrating rank		*In terms of eigenvalues*		
	Null	Alternative	Null	Alternative	
$\lambda_{\max}(j-1)$	$r \leq j-1$	$r = j$	$\lambda_j = 0$	$\lambda_j \neq 0$	$-T\log(1-\lambda_j)$
$\lambda_{trace}(j-1)$	$r \leq j-1$	$r \geq j$	$\lambda_j = 0$, $i = j$, $j+1, \ldots, n$	$\lambda_j > 0$ for at least one i, $i = j, j+1, \ldots, n$	$-T\left[\sum_{i=j}^{n} \log(1-\lambda_j)\right]$

Table 4.2 Sequential testing using the trace test

Null hypotheses		*Alternative hypotheses*	*Interpretation of inference*	
Actual	Sequential*		Rejection of null	Non-rejection of null
r = 0	r = 0	r ≥ 1	r ≥ 1, continue to next stage, test null r ≤ 1	Conclude r = 0. No further testing.
r ≤ 1	r = 1	r ≥ 2	r ≥ 2, continue to next stage, test null r ≤ 2	Conclude r = 1. No further testing.
⋮	⋮	⋮	⋮	⋮
r ≤ n – 1	r = n – 1	r = n	Conclude r = n	Conclude r = n – 1

Note:
*Sequential interpretation assumes rejection of previous null hypothesis.

null in the sequence is $r \leq s$, but since $r \leq s - 1$ has already been rejected, this reduces to $r = s$.[23] The alternative is $r \geq s + 1$. Rejection of the null again would lead to a test of the null $r \leq s + 1$ (in effect $r = s + 1$) against the alternative $r \geq s + 2$, and so on until the null is not rejected. This sequence and the interpretation of rejection or non-rejection at each stage is described in Table 4.2.

The maximal eigenvalue test may be used in an analogous way, as described in Table 4.3. Rejection or non-rejection of the null hypothesis should be treated cautiously. Rejection of the null hypothesis does not imply that the

Table 4.3 Sequential testing using the maximal eigenvalue test

Null hypotheses		*Alternative hypotheses*	*Interpretation of inference*	
Actual	Sequential*		Rejection of null	Non-rejection of null
r = 0	r = 0	r = 1	Apparently r = 1, but r > 1 not considered so continue to next stage, test null r ≤ 1	Conclude r = 0. No further testing.
r ≤ 1	r = 1	r = 2	Apparently r = 2, but r > 2 not considered so continue to next stage, test null r ≤ 2	Conclude r = 1. No further testing.
⋮	⋮	⋮	⋮	⋮
r ≤ n – 1	r = n – 1	r = n	Conclude r = n	Conclude r = n – 1

Note:
*Sequential interpretation assumes rejection of previous null hypothesis.

alternative should be accepted. Similarly with non-rejection of the null. For example, rejection may occur because untested assumptions about the data are being contravened – that is, the hypotheses are in effect more complex than is being stated. The point is stronger still when a test is being performed on a parameter, but the union of the sets of values under null and alternative hypotheses is not exhaustive. Under such circumstances, if the true or best approximating value of the parameter is not accounted for under either hypothesis, it is difficult to predict which of the two competing hypotheses will be preferred. However, if both hypotheses constitute similarly poor approximations, then the null will be favoured by the test, as in tests of this sort, the null hypothesis is reverted to in the absence of discriminating evidence.

At each stage, the trace test covers all possible values of the parameter r, so the non-exhaustive problem does not arise. However, the maximal eigenvalue test only covers all possible cointegrating ranks when testing the final null in the sequence, $r \leq n - 1$. The procedure based on this test may fail to reject the null because neither the null nor alternative includes the true cointegrating rank. The procedure tests for lower orders of cointegration first and so may well underestimate the cointegrating rank. Thus, intuitively, the method based on the trace test is to be preferred. Johansen (1995, chapter 12) shows formally that, asymptotically, the sequential procedure based on the trace test does not underestimate the cointegrating rank, and overestimates it with probability equal to the size of test at an individual step. Thus, a procedure using tests of size 5% at each step, would, asymptotically, select the correct order of cointegration with probability 95%. This is not to say that the maximal eigenvalue test is inferior at all stages. If the alternative of the maximal eigenvalue test is the true rank then it can be expected to have more power than the trace test since the latter will be considering it jointly with a set of irrelevant alternatives. While it is common practice to use both tests as the basis of sequential testing procedures, the trace test should be relied upon more heavily. This begs the question of how the two tests might be combined in a more useful way in a finite sample. Two suggestions are as follows:

(i) Use the $\lambda_{\max}$ test only to check the cointegrating rank determined by the trace procedure. Thus a confirmatory inference is achieved if the first non-rejection of the trace sequence occurs at $H_0 : r \leq j$ (interpreted sequentially as $r = j$) versus $H_1 : r \geq j + 1$ and the non-sequential $\lambda_{\max}$ test does not reject at $H_0 : r \leq j$ versus $H_1 : r = j + 1$. In this way, the test with the better power is used up to a point where a test of greater power is used to confirm the inference.

(ii) Rather than compute the statistics in sequence, it is possible to compute p-values for all cases. The preferred alternative would be that of the test

with the highest p-value. The interpretation from the maximal eigenvalue test is clear as this has a point alternative. That from the trace test is less obvious since the alternative is of a compound form. However, the natural interpretation is to select the lower bound since cases involving only higher orders of cointegration are not preferred.

Hendry (1995) and others have argued that a general to specific approach is to be preferred in model selection. The sequential testing procedure, however, begins by testing the most restricted case: that all eigenvalues are zero. The restrictions are then relaxed one eigenvalue at a time. This is a specific to general approach. It is also specific to general in that the lower the rank, the fewer coefficients are needed to parameterize the VECM.[24] Nonetheless, Johansen's result establishes that the specific to general approach is a valid method for determining the cointegrating rank.

4.6 Tests of cointegration in VAR models

In this section, we consider the application of the Johansen procedure. Although many alternative methods are available in the literature for testing cointegration and detecting the long-run parameters, the Johansen procedure tests the proposition that series are cointegrated, estimates the parameters, permits theoretical propositions and exogeneity to be tested. Here the test is considered within the confines of a simple case. Then some results from the literature are discussed and their economic interpretation. Firstly, we consider the simplest case where the underlying models are essentially random walks, then models of the UK exchange rate considered by Hunter (1992a), and Johansen and Juselius (1992) and the results based on an extended data set first presented in Hunter and Simpson (1995). In the next chapter, identification and exogeneity are discussed.

The Engle–Granger two-step procedure first considered the estimation of the long run from a single equation regression and then the residual from this model defines a cointegrating vector. As was explained in chapter 3 the lagged residual can be entered into a dynamic model and this is then described as an error correction or, more precisely, an equilibrium correction term. The equilibrium correction term has estimated parameters, while the parameters of an error correction term sets the coefficients in absolute terms to unity, (i.e. [1–1]). It must be re-emphasized that the Engle–Granger method will only generally be valid when there is $r = 1$ cointegrating vectors or there are only two equations in the system. Excepting very particular cases the method will be incorrect when there is more than one cointegrating vector and more than two equations in the system. In spite of this, there have been many attempts to improve the performance of the long-run estimator. Saikkonen (1991)

suggested the inclusion of further dynamics to improve the estimates of the long-run parameter, while Phillips and Hansen (1990), and Phillips (1991) provide non-parametric corrections robust to different types of error structure and the correction proposed by Marinucci and Robinson (2001) seems to perform well when the system includes weakly exogenous variables. However, the performance of these types of estimator has generally been found to perform well in Monte-Carlo studies applied to bivariate models.

In this section, results associated with the multivariate approach due to Johansen are considered. The results associated with the Johansen estimator are well defined when the conditions described in section 4.3 are satisfied:

(1) The error process is normally distributed.
(2) The underlying VAR is well defined.
(3) There are no structural breaks.
(4) All the series are of the same order of integration (usually I(1)).

The Johansen test can be significantly altered by non-normality. Non-normality can be observed, because the series follow a non-normal distribution, due to intercept shifts, structural changes or the type of error variance behaviour linked to volatility. Non-normality can often be rectified by the introduction of dummy variables, when it has a simple institutional or structural cause. However, the impact of a dummy on the Johansen test statistic is not always innocuous. We will discuss the question dynamic specification on the Johansen trace test in the next section. It will be assumed here that the VAR has been correctly specified. Structural breaks whose point of occurrence is unknown are more difficult to handle. Here, it is assumed that any breaks that do occur are associated with well-documented events. As the order of integration is not known by definition, a number of issues arise. Firstly, when the Johansen procedure is used I(0) and I(1) processes can be mixed when there are at least two I(1) variables in the system. Secondly, non-stationary series that are fractionally integrated require a different type of estimator (Robinson and Marinucci 1998). Thirdly, balanced I(2) behaviour can be incorporated. Fourth, more general I(2) processes require a different estimator as do higher order processes. Flôres and Szafarz (1996) consider an extended definition of cointegration where there is a mixture of I(1) and I(0) processes, this is much more readily dealt with by the Johansen approach (Juselius 1995). Robinson and Yajima (2002) consider processes with long-memory that are stationary but require fractional differencing for stationarity, this approach is not handled by the Johansen procedure. When series are integrated of an order in excess of (1), but the integer order of integration is the same, then both the Engle–Granger and Johansen approaches are still valid. The dynamic model is estimated by rendering the data stationary through differencing an

appropriate number of times, while the long run is estimated in the usual way from the residuals from equations in the lag of the original data. Otherwise, a more general estimator is required. Currently there is an appropriate estimator for the I(2) case, which will be covered in more detail in chapter 6. Here, we consider an example which satisfies the property of balanced I(2) behaviour. Either that data are all I(2) and the dynamic models are specified in their second differences or when the data is logarithmic, accelerations of all the series analyzed are specified as being I(1), and then the usual Johansen method is applied to I(1) series of which some may also be differenced. To confirm the appropriateness of the balanced I(2) case, the test for I(2) by Johansen (1992) is applied.

Finally, the current evidence on the performance of tests of cointegration is discussed.

4.6.1 Special cases of the Johansen test

If we assume the simplest procedure drives the underlying series, then the following special case provides a more intuitive explanation of the Johansen procedure. Let all series be generated by random walks, then the likelihood statistic due to Johansen (1991) is (see Appendix C):

$$LogL(.) = T\sum_{i=1}^{r}\log(1-\lambda_i)$$

where λ_i is solved from the determinantal equation $|\lambda_i S_{1,1} - S_{1,0}\,(S_{0,0})^{-1}\,S_{0,1}| = 0$, $S_{ij} = \sum_{t=1}^{n} R_{it}R_{jt}$, $i, j = 0, 1$. In the VAR(1) case:

$$R_{0,t} = \Delta x_t$$
$$R_{1,t} = x_{t-1}.$$

Essentially, the Johansen procedure generalizes these equations to transform a more complex dynamic model (VAR(i)) into two sets of equations that reduce to a multivariate first order autoregression: based on the above description of $R_{0,t}$ and $R_{1,t}$ the equations that are being implicitly estimated by the Johansen procedure are

$$R_{0,t} = \alpha\beta' R_{1,t} \qquad (4.80)$$
$$or$$
$$\Delta x_t = \prod x_{t-1}.$$

The latter equation is a VAR(1). Not only is this a VAR(1), but this equation can be readily viewed as a multivariate generalization of the model estimated by Dickey and Fuller, to test stationarity of a single series; the estimation of this type of model is briefly considered in Engle and Granger (1987). For a single equation, based on one or more regressors, Engle and Granger test cointegration using regression residuals, while the Johansen estimator requires a

variance decomposition. Consequently the two methods may produce different results (Haug 1993, 1996).

4.6.2 Empirical examples of the Johansen test

Now let us consider some empirical examples. The model of the exchange rate derived using the data presented in Johansen and Juselius (1992) and Fisher et al. (1990). The model estimated by Hunter (1992a). Estimates based on an extended data set for the purchasing power parity model considered by Johansen and Juselius (1992). And estimates of a UK exchange rate model with balanced I(2) based on Juselius (1995).

Firstly, for simplicity the six variable VAR(2) model,[25] estimated by Hunter (1992a), is considered. The model is an extension of the five-variable VAR model estimated in Johansen and Juselius (1992). The system contains the following variables all in logarithms: oil prices (p_{ot}), UK prices (p_{1t}), world prices (p_{2t}), the UK effective exchange rate (e_{12t}), UK treasury bill rate (i_{1t}) and the Eurodollar rate (i_{2t}). Johansen and Juselius (1992) wanted to confirm that the UK effective exchange rate satisfied the conditions for Purchasing Power Parity (PPP). The six variables are stacked into the following VAR(2) model with normal errors and unrestricted intercept:

$$(I + \Gamma_1 L)\Delta x_t = \prod x_{t-1} + \mu + \epsilon_t. \tag{4.81}$$

The hypothesis to be tested relates to the cointegrating rank,

$$H_1(r): \ \prod = \alpha\beta'.$$

This test determines how many cointegrating vectors or long-run relationships (r) exist in the system. In this case there are at the most $r = 6$ and at the minimum $r = 0$, none.[26] A number of hypotheses exist in relation to trends, unrestricted intercepts in the model operate as drift parameters in the same way as occurs when all series in the system are purely difference stationary, which is equivalent to saying $r = 0$. Otherwise the VAR can have a time trend. The model considered by Johansen and Juselius has unrestricted intercepts, which implies that there is drift. Let us consider the results for the Johansen test outlined above in the case of the six-variable VAR, which allows for drift and includes centred seasonals.

The max test is calculated as $\lambda_{\max}(i) = -T\log(1 - \lambda_i)$ for $i = 1, \ldots n.$ and the trace test is $\lambda_{trace}(i) = -T\sum_{j=1}^{i} \log(1 - \lambda_i)$ for $i = 1, \ldots n.$[27] If it were known a priori that all the series were stationary, then both the Johansen test statistics, that are essentially likelihood ratio tests, would follow a Chi-squared distribution. However, as was discussed above, when the series are I(1), then the distribution is non-standard. It has been common practice to compare the test statistics with their asymptotic critical values, which come from simulating a null

Table 4.4 Eigenvalues, Johansen test statistics for VAR due to Hunter (1992)

Eigenvalue	*Alternative hypothesis*	λ_{max}	*95% critical value*	λ_{trace}	*95% critical value*
0.571	r = 1	50.82*	39.43	119.69*	95.18
0..335	r = 2	24.48	33.26	68.86	69.38
0..289	r = 3	20.44	27.34	44.37	48.41
0.161	r = 4	10.52	21.28	23.93	31.25
0..128	r = 5	8.22	14.6	13.41	17.84
0.083	r = 6	5.18	8.08	5.18	8.08

Note:
* Indicates significant at the 5% level for critical values. For tables of the Johansen trace test with un-restricted intercept and T = 50 observations see Francis (1994).

distribution for the test that the series are multi-variate random walks. The tests are significant when the null hypothesis $r = i$ is rejected against the alternative for both tests that $r > i$. From the results presented in Table 4.1, both tests (λ max(1)=50.82>39.43 and $\lambda_{trace}(1) = 119.69>95.18$) yield the same conclusion that there are $r = 1$ cointegrating vectors. The test is only significant in the case where $r = 1$, otherwise none of the tests are significant. The test statistics are asymptotic and much of the research that has looked at the impact of testing would conclude that the performance of both tests in small samples is poor. Based on the suggestion that the trace test is more reliable than the max test and the fact that rejection of the proposition that there are two cointegrating vectors is very marginal ($\lambda_{trace}(2) = 68.86$), Johansen and Juselius whose results are for a restricted version of this model, suggested $r = 2$. Some theoretical and empirical evidence is presented in the next two sections as to why there may be over-rejection.

Johansen and Juselius (1992) used the same data, but they assumed that the oil price was strictly exogenous to the system, which means that it has no influence on the long-run. They estimate the following five-variable VAR conditional on changes in the oil price (this proposition is tested in the next chapter):

$$(I + \Gamma_1 L)\Delta x_t = \prod x_{t-1} + (\Xi_0 + \Xi_1 L)\Delta p_{0t} + \mu + \epsilon_t.$$

The results presented in Table 4.5 are based on the same model, except that it is estimated on the data set extended to 1991q4. The results and conclusions are not materially different from those of Johansen and Juselius (1992). As was concluded before, Johansen and Juselius suggested that there were $r = 2$ cointegrating vectors, even though the test statistics did not quite bear this out. In what follows the analysis is based on the Johansen trace test. The extended data set implies that Johansen and Juselius (1992) were correct to suggest that

Table 4.5 Eigenvalues and trace test statistics for Johansen and Juselius model

Eigenvalue	*Alternative*	λ_{trace}	*95% critical*
0.31	r = 1	84.3*	70.6
0.27	r = 2	55.4*	48.3
0.19	r = 3	31.3	31.5
0.13	r = 4	14.9	17.9
0.05	r = 5	4.4	8.2

*Indicates significant at the 5% level for tabulated values of the test statistic with trend and one exogenous variables. For similar values, see Pesaran et al. (2000).

there are $r = 2$, cointegrating vectors, because the trace test is significant for the proposition that r exceeds zero and one. It will be discovered that the VAR(2) model is not well formulated, but any opportunity to re-specify the models associated with Hunter (1992a) and Johansen and Juselius (1992) is limited by the number of observations. For further comparison with the results in Johansen and Juselius (1992), eigenvectors are calculated for the case in which $r = 2$.

The two vectors are normalized with respect to the first element, but the normalization is arbitrary and no suggestion is made that these vectors have any meaning. However, when compared with the results presented in Johansen and Juselius (1992), the unrestricted eigenvectors suggest that the following restriction (1 – 1 – 1) might be applied to both aggregate price series and the exchange rate. The restriction implies that there is a long-run correspondence between the terms of trade and the exchange rate (a condition for Purchasing Power Parity or PPP). This conclusion is quite consistent with the results in Johansen and Juselius (1992). This type of restriction is analyzed in more depth in the next chapter where identification and exogeneity are discussed. It is of interest to note that neither Johansen and Juselius (1992) nor Hunter (1992a) could force the first vector to be restricted to satisfy pure PPP; that is to say the proposition that the real exchange rate is stationary was not sustained by the data. And, unlike Juselius (1995), who considers similar results for Denmark and Germany, the interest rates that appear in the model

Table 4.6 Normalized significant eigenvectors

Equation	$\beta_{.1}$	$\beta_{.2}$
p_1	1.00	1.00
$-p_2$	–1.07	–1.6
e_{12}	–1.03	2.9
i_1	–3.34	–8.4
i_2	–0.31	14.5

do not yield a PPP vector augmented by uncovered interest rate parity (UIRP). However, as will be observed in the next chapter the second vector does appear to suggest UIRP. Here, it is not possible to interpret the unrestricted cointegrating vectors as they have not been appropriately identified. Three matrices (α, β and Π) were calculated they all have the same rank, which implies that only part of Π can be used to identify both α and β. Without restriction not all of the matrix pair (α, β) is identified. Alternatively, without restriction both matrices can be transformed to a square $r \times r$ arbitrary non-singular matrix (ζ). Therefore:

$$\alpha\beta' = \alpha\zeta\zeta^{-1}\beta' = \alpha^{*}\beta^{*'}.$$

The system with cointegrating vectors β has the same likelihood function as the system with cointegrating vectors β^{*}. Consequently $L(\beta|.) = L(\beta^{*}|.)$ and both systems conditional on the existing information are observationally equivalent* (Rothenberg 1971). Observational equivalence is a key criterion for non-identification, when models cannot be distinguished then neither can their parameters. Further, the Johansen test statistics are well defined when the criterion for cointegration are satisfied and the DGP well approximated. What is required is a set of single equations in the VAR, which satisfy the conventional regression criterion (Spanos 1986; Davidson and MacKinnon 1993; or Hendry 1995). More appropriately the criterion for well defined dynamic systems of equations is discussed by Hendry and Richard (1982, 1983), that is they should have spherical disturbances, define stable processes subject to appropriate conditioning on impulse or periodic dummies. Should the disturbances be non-normal, then a quasi-likelihood result is required, that is the sample should be sufficient for the test statistics derived from the estimator to tend to their asymptotic distributions. The Johansen test is viewed to be sensitive to different types of deviation from normality, the lag length of the VAR and unmodelled dynamic behaviour in structure of the variance–covariance matrix of the disturbances. When the dynamic isn't well defined or the residuals are non-normal then the Johansen test may not be optimal. The tests may have low power to discriminate against the local alternative of cointegration or be inappropriately sized. The latter problem implies that when the test is calculated and it is defined at a particular critical value (5%), the true probability in the tail of the distribution may exceed 5% (over-sized) or be less than 5% (under-sized).

The VAR model in Johansen and Juselius (1992) and the VAR estimated on the extended data set isn't well defined, as can be observed from the diagnostic tests presented in Table 4.7. Here, emphasis has been placed only on the tests that reject the null of correct specification (no serial correlation, normality, no autoregressive conditional heteroscedasticity), which leads to the possibility that the statistics may be a reflection of the testing as compared with

Table 4.7 Key diagnostic tests for the VAR(2) model with strictly exogenous oil prices

Equation	*Normality statistic*	*Fourth order serial correlation*
p_1	25.74**	7.33*
p_2	8.53*	2.28
e_{12}	6.92*	0.84
i_1	3.42	2.57
i_2	32.44**	15.43**

true mis-specification. One procedure to counter the possibility of false rejection is to use a broader criterion than usual (i.e., 1%) on each test. This is the Bonferroni Principle (Stock and Watson, 2003, p. 191), which implies that i tests applied at the a/i% level yield an overall rejection rate of a% for the tests as a whole. Special attention should be paid to the test that fail at the 1% level as they are likely to be the ones that imply misspecification. Consequently, the UK price equation and eurodollar equations are the ones to be concerned about as they would reject the null hypothesis of normal errors. In the case of the eurodollar equation no serial correlation is rejected irrespective of the cut-off point selected for the test. There is almost zero probability that the errors from these models are drawn from a normal distribution or in the later case are not serially correlated.

In this light, and given the possibility that the price series may be I(2), Hunter and Simpson (1995) used the same extended data, but dropped the oil price. To partially correct for I(2) behaviour they followed the structure employed by Juselius (1995) to model the Danish kroner relative to the deutsche mark and they found that dummy variables were more effective in accounting for non-normality than oil prices. The model is a five-variable VAR(2) of the same form as (4.81), except $x'_t = [\Delta p_{1t}, p_{1t} - p_{2t}, e_{12}, i_1, i_2]$.

According to the trace test there are $r = 4$ cointegrating vectors. With a correction for the number of observations due to Reimers (1994) the conclusion of this test is less clear and given the number of dummies the test may also have the wrong size.[28] The fourth vector is marginal when inclusion of the intervention dummies is considered, but the model is based on a longer data

Table 4.8 Eigenvalues and trace test statistics for Juselius model applied to UK data

Eigenvalue	*Alternative*	*Statistic*	*95% critical*
0.471	r ≥ 1	117.5	68.5
0.316	r ≥ 2	70.4	47.2
0.294	r ≥ 3	42.2	29.7
0.138	r ≥ 4	16.4	15.4
0.071	r = 5	5.4	8.2

Table 4.9 Normalized significant eigenvectors

Equation	$\beta_{.1}$	$\beta_{.2}$	$\beta_{.3}$	$\beta_{.4}$
$p_1 - p_2$	1.00	0.16	0.32	0.53
Δp_1	9.15	0.22	1.83	1.00
e_{12}	–0.76	0.05	1.00	–0.83
i_1	–3.03	–1.26	1.05	0.31
i_2	0.02	1.00	1.45	–2.13

set than Johansen and Juselius (1992) and the decision to include this vector is based on a statistic that is significant in conventional terms.

A valid analysis and interpretation of the results is left to the next chapter, after identification is discussed. However, the first vector would appear to be PPP augmented by a UK interest rate and the inflation rate, the second vector suggests an interest parity condition, while the interpretation of the other vectors is not clear.

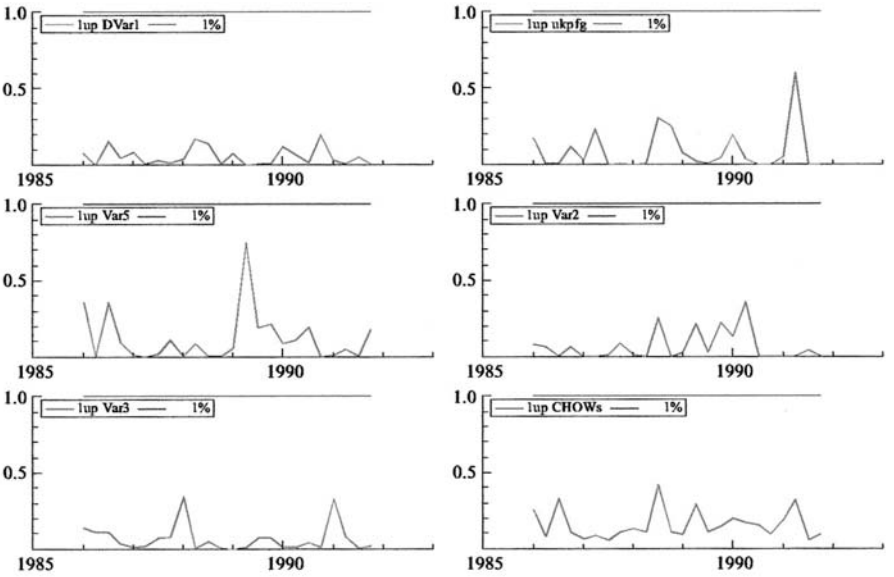
Thus far, three alternative VAR models have been devised to explain the UK effective exchange rate in association with a set of related variables. However, the conclusions drawn from this exercise depend on the performance of all the equations in the long-run and short-run models. Exclusion of variables in the long run depends on tests of exogeneity, long–run exclusion and restrictions associated with economic hypotheses that are likely to identify. Discussion of these issues is left for the next chapter, here the question of specification would appear to be a key mechanism to discriminate between models.

Comparison of the results in Table 4.7 with those in Table 4.10 suggest that the transformed model is better behaved (none of the diagnostics are significant at the 1% level). According to the system-wide diagnostic tests, the VAR(2) which includes the dummy variables is well specified, as can be observed from Table 4.10. If testing at the 1% level is considered acceptable, then none of the tests of dynamic specification are significant, which implies that each of the estimated equations is well specified. When tests are applied

Table 4.10 Single equation diagnostics for each equation in the VAR (2) model

Equation	*Normality statistic*	*Fourth order serial correlation*
$p_1 - p_2$	2.74	9.94*
Δp_1	7.19*	2.53
e_{12}	0.63	4.13
i_1	1.25	6.28
i_2	0.35	7.98

(**significant at the 1% level and *significant at the 5% level)



Figures 4.1–4.6 Recursive Chow tests for the 5 VAR equations and the VAR system

at the 5% level, then the serial correlation up to order four is marginally significant in the terms of trade equation and the normality test fails at the 5% level for the inflation equation. Again, 5% might be viewed as being overly harsh due to the risk of over-rejection, otherwise the model has performed better than any of the existing models analyzed with the extended data set. Beyond the tests described above, the models would also appear to have stable parameters based on the sequence of recursive Chow tests presented in Figures 4.1–4.6. The 1 step ahead Chow test is a one-period ahead forecast *F* test which is a variant of the Chow type 2 F test (Spanos 1986). This is an in-sample prediction test, which examines the model parameters over the data period for parameter constancy. It is scaled in such a way that critical values at each point in the sample are equal to unity. Hence the horizontal line at unity becomes the critical value to use for making inference about stability. If the parameters are found not to be consistent, then the model is spurious and the equation estimates meaningless. Details of the various recursive estimation tests can be found in PCFIML v10 (Doornik and Hendry 2001).[29] As can be observed from Figures 4.1–4.5 each of the short-run equations has stable parameters according to the sequence of Chow tests applied at the 1% level, which is reflected in the result in Figure 4.6 for the VAR as a system.

In section 4.4 the key condition for the existence of a cointegrated VAR according to Johansen (1995) is

$$rank(\alpha'_{\perp}\Gamma\beta_{\perp}) = n - r.$$

As is explained in Johansen (1995) a test of this rank condition can be undertaken in the usual way by transforming the data through pre-multiplying the data by $\alpha_\perp$ and then applying the Johansen test to the $n - r$ dimensioned system based on the difference of $x_t^+ = \alpha_\perp x_t$. Consider the model:

$$\Delta^2 x_t^+ + \sum_{i=1}^{p-1} \Gamma_i^* \Delta x_{t-1}^+ = \Gamma \Delta x_{t-i}^+ + \epsilon_t^+ \tag{4.82}$$

$$where\ \Gamma = I + \sum_{i=1}^{p} \Gamma_i^* = -\Gamma(1).$$

If $rank(\alpha'_\perp \Gamma \beta_\perp) < n - r$ then there are trends in the VAR that have not been accounted for. Taking the extreme case where $rank(\alpha'_\perp \Gamma \beta_\perp) = 0$, $\Gamma = 0$ and

$$\Delta^2 x_t^+ + \sum_{i=1}^{p-1} \Gamma_i^* \Delta x_{t-i}^+ = \epsilon_t^+,$$

$\Delta^2 x^+{}_t$ is a stationary process or there are linear combinations of x_t or more appropriately $n - r$ variables which are I(2). Failure of this condition breaks a fundamental criterion for cointegration in the VAR, which is that the system is predicated on variables that are at most I(1). If the test fails and $rank(\alpha'_\perp \Gamma \beta_\perp) < n - r$ then it is best to test for cointegration within the confines of a dynamic model, which accounts for the interdependencies associated with I(2) variables; this analysis is undertaken in chapter 6 for the data set used by Hunter (1992a) and Johansen and Juselius (1992).

For the case considered here $r = 4$ and $n - r = 1$, calculating $\alpha'_\perp$ from the singular value decomposition of α leaves a 1×5 vector, which is then used to multiply the variables in the VAR in second differences. Applying the Frisch–Waugh Theorem to (4.82), yields a Johansen test statistic, which under the null implies that rank(Γ) = 0. If the Johansen test statistic is significant at the 5% level then we accept the alternative that rank(Γ) = $n - r = 1$ and that the VAR(5) is a well-defined system in terms of I(1) variables that cointegrate. For the system developed here, the result in Table 4.11 implies that the alternative cannot be rejected and as a result there are no additional I(2) trends. This suggests, that we have, what is called by Johansen (1994), balanced I(2) behaviour. This seems to be the only coherent way, by which a single I(2) variable can cointegrate with other series that are I(1).

Table 4.11 Eigenvalues and trace test statistics for **I**(2) test

Eigenvalue	*Alternative*	*Statistic*	*95% Critical*
0.407	r = 1	36.36*	9.243

Note:
*indicates significant at the 5% level

Cointegration as a statistical construct exists, whether it is an effective procedure for approximating actual behaviour is open to question. The usefulness of any method depends on our capacity to detect this type of phenomena with the nature and quantity of data available. Soren Johansen did a great service to applied econometric and statistical research by providing a structure for estimation, inference and identification when series are close to random walks.

4.6.3 Evidence on the performance of the Johansen test

The usefulness of this approach depends on the properties and performance of the test statistics. There is a burgeoning literature on size, power and comparative performance of the Johansen test statistic based on simulation (Burke and Hunter 1998; Gonzalo and Pitarakis 1999; Hubrich et al. 2001; and Marinucci and Robinson 2001).

Performance of these tests generally relates to the quality and informativeness of the data, and the extent to which the underlying model satisfies the Gaussian properties that underlie the Johansen VAR. If the series to be considered have residuals that are approximately normal, the dynamic process can be well approximated by a finite order VAR, the residuals are not sensitive to dynamic behaviour in the variance and breaks in structure are limited to those that can be handled by dummy variables, then the Johansen methodology would appear to work reasonably well (Hubrich et al. 2001). Whether series are normal is not known a priori, but normality is testable and some of the abhorrent features associated with non-normality can often be removed by transformation, aggregation or dummy variables. Financial time series are prone to significant non-normality, but evidence exists that aggregate returns tend to normality as the period of aggregation increases. Hence, daily returns that are non-normal will when aggregated at a quarterly frequency appear normal (Barndorff-Nielsen and Shephard 2001). Conventional finance theory argues that share prices are log normal and that logarithms of share prices follow Wiener processes in the limit (Hull 2002). Hence, log transformations of income, share prices or wealth statistics are likely to be closer to normality. Shocks that induce large errors will cause series to fail normality tests, which has led to the use of dummy variables as a correction (Hendry 1995). However, in the latter case the distribution of the Johansen test is altered by certain types of intercept correction (Johansen 1995; Hubrich et al. 2001). When the distribution of the error satisfies appropriate regularity conditions (see Appendix D), then the Johansen test statistic will converge to the asymptotic distribution, but the rate of convergence depends on the information content or innate variability of the data. When series are highly informative convergence may be fast, statistics based on some underlying distributions converge more readily to normality (e.g. means calculated from data

generated from a uniform distribution converge to normality after thirty observations, while the t_1 or Cauchy Distribution never converges to the normal).[30]

In finance, there is significant interest in volatility and financial series are often viewed as being t_6 or a mixture of normals. This implies that such series may not behave in the way that is desirable. Spanos (1994) has suggested that the conditional t with small degrees of freedom provides an appropriate statistical model for financial time series that are volatile. An alternative model of volatility arises when there is a dynamic structure in the variance, this is often modelled by ARCH models (Engle 1982). With dynamic behaviour in the variance, then the Johansen test statistics may not perform well. Bauwens et al. (1997) have suggested correction for the two-step Engle–Granger approach, but no correction appears to exist for the Johansen procedure. One suggestion is to apply a GLS correction to the first step of the Frisch–Waugh procedure and then use the small sample tail correction to the test statistic suggested by Doornik and Hendry (2001); this approach yields recalibrated p-values for the calculated Johansen trace test, the same correction may apply when dummy variables are included in the model.

The order of the VAR is often difficult to determine, but is critical to the performance of the Johansen test. Too many lags will affect the small sample performance of the test, while too few lags will imply that the model is not well specified. Often information criterion are used to suggest lag length, this derives from univariate analysis of time series. However, such measures tend to perform less well, when a system is considered and the dynamic of the different variables in the system is not homogeneous. Firstly, any VAR model may be tested for the presence of serial correlation and should that be found then the dynamic model needs to be re-specified (Hendry 1995). Secondly, asymptotically the Johansen test is invariant to the number of lags in the VAR, which suggests a general to specific approach to derive the short-run dynamics:

(i) Specify a general model with $s \leq \frac{T}{3n}$ lags per equation.
(ii) Eliminating insignificant lags to order $p \leq s$.
(iii) Eliminating insignificant intermediate lags in each VAR equation.
(iv) Estimating the long-run relationships by applying the Frisch–Waugh Theorem to a restricted version of (4.80).

Another problem that is likely to arise in this case relates to the existence of what Caner and Kilian (2001) call hidden moving average behaviour. In section 4.2, the question of inversion of the Wold representation was discussed. It was stated that the VECM only derives from the Granger representation theorem when the system is bivariate. A more general transformation

exists when the matrix polynomial from the Wold representation ($C(L)$) is rational, but this proposition is still not testable from the VAR. An alternative inversion is considered in the next section, but this only yields a finite order VAR when $C(L)$ is first order. One solution is to apply the Johansen procedure to a Frisch–Waugh equation where the residuals are estimated using either a VARMA(1,q) or shorter order VARMA (Hunter and Dislis 1996; and Lütkepohl and Claessen 1993). Burke and Hunter (1998) have shown via simulation of models with quite simple moving average structure, that the size and size corrected power can be quite strongly affected by the existence of moving average errors and that this does not disappear as the sample size increases.

However, Marinucci and Robinson (2001) show that the Johansen trace test would appear to work quite well with samples of 100 observations, when compared with fully modified estimators, though there is some evidence for small systems that the Phillips modified estimator might perform better when the sample size is less than 100 (Hubrich et al. 2001). If the system is bivariate and one variable is weakly exogenous then the semi-parametric approach first applied by Robinson and Marinucci (1998) to fractionally integrated series appears to work well (Marinucci and Robinson 2001).

The number of observations likely to yield reasonable inference depends on the nature and complexity of the problem to be analyzed and the order of integration of the series. The advantage of the Johansen approach is that it still provides an inferential procedure, which permits the long run to be estimated and long-run systems to be identified, causal structure and endogeneity tested. None of the other approaches appear to do all of the above. The approach also generalizes to higher order cointegration.

In the next section we consider some further issues related to representations and in the next chapter issues of exogeneity and identification are discussed.

4.7 Alternative representations of cointegration VAR

It was observed that the switching between cointegrating forms in the Wold VMA and the Johansen VAR was not a straightforward exercise. One possible explanation is that VAR and the VMA are always approximations, the other is that the natural time series representation in the cointegration case is either a VAR or a VMA. However, the finite VMA that forms the basis of the Granger representation theorem and the Smith–McMillan–Yoo Form does not usually conform with a finite order VAR. In this section we develop an extension to the results previously considered, which derives from the literature on matrix polynomials (Gantmacher 1960; Gohberg et al. 1983). Based on some broad conditions for the extraction of divisor matrices from a matrix polynomial it follows that the VMA can be directly inverted. In this section the Generalized

Bézout Theorem and an extension that considers the unit root case are used to derive a VAR and VARMA representation for cointegration (Hunter 1989a, 1992). It is shown that under the conditions required for the extended Bézout Theorem, that the VMA(1) inverts exactly to a VAR(1), this result is demonstrated for a simple bivariate system, which is used by Burke and Hunter (1998) to develop their Monte Carlo study. The section concludes with a brief discussion of the articles by Haldrup and Salmon (1998) and Engsted and Johansen (1999).

4.7.1 The Sargan–Bézout factorization

From the Wold decomposition $C(L)$ is a finite matrix polynomial of degree q. Following the convention in the literature on matrix decompositions it is usual to look at the inverse of the spectral decomposition of $C(L)$:

$$Q(z) = (Q_0 z^q + Q1 z^{q-1} + \ldots Q_q) = z^q C(1/z).$$

If $Q(z)$ is a matrix polynomial, such that $Q_0 \neq 0$, then it follows from the Generalized Bézout Theorem:

Theorem 2 *If $Q_0 \neq 0$, then there exists a left-hand divisor $Q_0(z) = (Iz - F)$ such that $Q(z) = Q_0(z)Q_1(z)$, if and only if, $Q(F) = 0$.*

Proof: see Gantmacher (1960).

In the case where $Q(z)$ has a block of common roots, then the result devised by Sargan (1983a) to extract Matrix Common Factors in autoregressive error models can be applied to the case of common unit roots:

Theorem 3 *If $Q(z)$ has a block of common roots (for cointegration on the unit circle), then $Q(z)$ has a left-hand divisor $Q_o(z) = (Iz - F)$, if and only if $FQ(F) = 0$.*

Proof: Consider the quasi-monic matrix polynomial,

$$Q(z) = (Q_0 z^q + Q_1 z^{q-1} + \ldots Q_q)$$

where $Q_0 = I$ and $rank(Q(1)) = n - r$. If there is a left-hand divisor $Q_0(z) = (zI - F)$ of $Q(z)$, then

$$zQ(z) = Q_0(z)Q_1(z). \tag{4.83}$$

By comparison of the j^{th} polynomial powers of z on the left-hand and right-hand side of (4.83):

$$Q_{1(j)} = FQ_{1(j-1)} + Q_j \text{ for } j = 1 \ldots q. \tag{4.84}$$

Expanding the right-hand side of (4.83) into its component matrices when $Q_{1(1)} = I$, then:

$$(Iz - F)Q_1(z) = (Iz - F)(Iz^q + Q_{1(1)}z^{q-1} + \ldots Q_{1(q)}) = zQ(z) - FQ_{1(q)}. \tag{4.85}$$

It follows that $FQ_{1(q)} = 0$ is necessary and sufficient for (4.83) and (4.85) to be isomorphic. Replacing j by q and re-arranging (4.84):

$$FQ_{1(q)} = F^2Q_{1(q-1)} + FQ_q = 0$$

or equivalently

$$F^2Q_{1(q-1)} = -FQ_q.$$

By replacement of terms of the form $Q_{1(q-k)}$ it follows that:

$$F^{q+1} = -F\sum_{k=0}^{q-1} F^kQ_{q-k}. \tag{4.86}$$

Gathering together terms in powers of F on the left-hand side of (4.86) and extracting the common term in F gives rise to following polynomial in F:

$$FQ(F) = F\sum_{k=0}^{q} F^kQ_{q-k} = 0. \tag{a}$$

The existence of the left-hand divisor relies on $FQ(F) = 0$, which occurs either when the Generalized Bézout Theorem holds and $Q(F) = 0$ or when F is a left-hand annihilator of $Q(F)$. ■

This generalization implies that F lies in the null space of $Q(F)$ or when *rank* $(F) = r$, then *rank*$(Q(F)) = n - r$. Given *rank* $Q(F)) = n - r$ then there exists an $r \times n$ matrix K_1, which annihilates $Q(F)$. There is an arbitrary matrix K_2 of dimension $n \times r$ defined so that K_1K_2 is non-singular and without loss of generality $F = K_2(K_1K_2)^{-1}K_1$ is an idempotent matrix, which annihilates $Q(F)$. When F is idempotent, then by definition $F^k = F$ and:

$$FQ(F) = F\sum_{k=0}^{q} FQ_{q-k} = F\sum_{k=0}^{q} Q_{q-k} = FQ(1).$$

If F is idempotent, then the condition $FQ(F) = 0$ is equivalent to $FQ(1) = 0$ and F is the matrix, which annihilates $Q(1)$. Consequently, $Q(1)$ satisfies a necessary condition for cointegration that *rank*$(Q(1)) = n - r$ and F contains the cointegrating vectors K_1.

As $Q(z)$ is a simple inversion of the ordering of the spectral form of $C(z)$, then a similar result exist for $C(z)$:

$$z^qC(z^{-1}) = Q(z) = Q_0(z)Q_1(z).$$

Given that F is idempotent, $Q_0(z)$ has the following Smith rational form:

$$Q_0(z) = (Iz - F) = H^{-1}\begin{bmatrix}(z-1)I_r & 0\\ 0 & zI_{n-r}\end{bmatrix}H$$

and F the following canonical form

$$F = H^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} H.$$

Following Engle and Granger (1987), the necessary condition for cointegration is $K_1C(1) = 0$. For $K_1' = [K_{11}' : K_{12}']$, then any matrix K_2' which satisfies the condition that (K_1K_2) is non-singular can be used. It is convenient to select $K_2' = [K_{12}' : 0]$ as it is then straightforward to show that $H_1 = K_1$ and $H_1x_t = \epsilon_t$ defines an r vector of cointegrating variables as F then has the following form

$$F = \begin{bmatrix} I_r & K_{11}^{-1}K_{12} \\ 0 & 0 \end{bmatrix}.$$

The only condition required for Theorem 3 to go through is the existence of a block of common roots in $Q(z)$. In the cointegration case $Q(z)$ has a canonical form Λ with a block of unit roots and a sub-matrix Λ_1 with roots within the unit circle. Then Λ has a multiplicity of r common roots and the left-hand divisor has r unit roots and $n - r$ zero roots and moving from frequency domain to the time domain $C(L)$ can be decomposed in an equivalent manner to $Q(L)$ when there are sufficient zero roots[31]

$$\Delta x_t = C_0(L)C_1(L)\epsilon_t \tag{4.87}$$

where $Q(L) = L^{q-1}C(L^{-1})$, $Q_1(L) = L^qC_1(L^{-1})$, $Q_0(L) = LC_0(^{-1})$ and

$$Q_0(L) = H^{-1}\begin{bmatrix} (L-1)I_r & 0 \\ 0 & LI_{n-r} \end{bmatrix} H = LC_0(L^{-1})$$
$$= LH^{-1}\begin{bmatrix} (L^{-1}-1)I_r & 0 \\ 0 & L^{-1}I_{n-r} \end{bmatrix} H.$$

Therefore:

$$C_0(L) = H^{-1}\begin{bmatrix} (1-L)I_r & 0 \\ 0 & I_{n-r} \end{bmatrix} H.$$

Having defined a unique factorization, which extracts an appropriate number of unit roots the Yoo inversion procedure can be applied to (4.87). Therefore:

$$C_0(L)^{-1}\Delta x_t = C_1(L)\epsilon_t.$$

The non-invertible MA is eliminated by cancellation of the inverted difference operator in $C_0(L)^{-1}$:

$$C_0(L)^{-1} = H^{-1}\begin{bmatrix} \frac{1}{1-L}I_r & 0 \\ 0 & I_{n-r} \end{bmatrix} H.$$

By extracting a common factor $C(L)$ becomes quasi-invertible, because it removes a partial over-difference. Therefore:

$$A_0(L)x_t = C_1(L)\epsilon_t$$

where

$$A_0(L) = H^{-1}\begin{bmatrix} I_r & 0 \\ 0 & \Delta I_{n-r} \end{bmatrix} H = (\Delta I - H^{-1}\begin{bmatrix} LI_r & 0 \\ 0 & 0 \end{bmatrix} H) = (\Delta I - FL). \tag{4.88}$$

The above factorization is unique as long as (a) above holds and this prohibits the possibility of polynomial cointegration. The partial common factor $(1 - L)$ cancels to leave the following VARMA(1,q) in levels and differences:

$$(\Delta I - FL)x_t = C_1(L)\epsilon_t \tag{4.89}$$

or

$$\Delta x_t - Fx_{t-1} = C_1(L)\epsilon_t$$

where $F = K_2(K_1K_2)^{-1}\ K_1$ is an idempotent matrix, $FC(1) = 0$ and K_1x_t defines a block of r cointegrating vectors. Under cointegration when the K_1x_t processes are all I(0), then $C_1(L)$ is invertible and (4.89) has the following VAR representation

$$C_1(L)^{-1}\Delta x_t - C_1(L)^{-1}Fx_{t-1} = \epsilon_t$$

where $A(L) = C_1(L)^{-1}$. It is now straightforward to transform this into an error-correcting vector autoregressive (ECmVAR) representation: as the conventional reparameterization sets $A(L) = (A(0) + (1 - L)A^*(L))$. Therefore:

$$A(L)\Delta x_t - (A(0) + (1-L)A^*(L))C_1(L)^{-1})Fx_{t-1} = \epsilon_t$$

or

$$\Gamma(L)\Delta x_t = \prod x_{t-1} + \epsilon_t$$

where $\Gamma(L) = A(L) - A^*(L)\ FL$, $\Pi = A(0)F = \alpha\beta'$ and for the VAR to be equivalent to the VARMA(1,q) form it follows:

$$\begin{aligned} A(0)F &= A(0)H^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \\ &= A(0)H^{-1}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} H_1 \\ H_2 \end{bmatrix} \\ &= A(0)H_1^*H_1, \end{aligned}$$

$\alpha = A(0)\ H_1^*$, $\beta' = H_1$ and $H^{-1} = [H_1^* : H_2^*]$. As a consequence the cointegrating vectors are equivalent to those which result from the VARMA(1,q). Identification will be considered in more detail in the next chapter, but for the case considered identification stems from the existence of a number of

what will be called weakly exogenous variables.[32] If $A(0)$ has full rank and H^* has rank r this implies that α can be factorized so that there is an $n \times r$ block of well-defined elements. It is also of interest to notice that conditional on the knowledge of the number of cointegrating vectors, the VAR has the following structural representation:

$$\Gamma(L)\Delta x_t - A(0)Fx_{t-1} = \epsilon_t$$

or

$$\Gamma^+(L)\Delta x_t = Fx_{t-1} + \epsilon_t$$

where $\Gamma^+(L) = A(0)^{-1}\Gamma(L)$ which has the same cointegrating vectors as the VARMA(1,q) representation.

4.7.2 A VAR(1) representation of a VMA(1) model under cointegration

The following example is used to motivate the algebraic results presented above and assist the readers understanding. If the underlying process is a VMA(1), then the analytic result presented above yields a very simple alternative representation, which is a VAR(1) case:

$$\Delta x_t = C(L)\epsilon_t \tag{4.90}$$

$$C(L) = \begin{bmatrix} 1-\frac{1}{2}L & \frac{1}{2}L \\ \frac{1}{2}L & 1-\frac{1}{2}L \end{bmatrix}.$$

Using the notation of the section above

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ F = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \text{ and } H = h\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$$

where h is any non-zero scalar

As $C_1(L) = I$ in this case, then it follows from (4.88) that the VAR representation is

$$\left(I - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}L\right)x_t = \epsilon_t$$

the VECM being

$$\Delta x_t = -\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}x_{t-1} + \epsilon_t.$$

An alternative way of explaining these manipulations is to state that an operator, $A(L)$, is required such that

$$A(L)C(L) = \Delta I,$$

I is the 2×2 identity matrix. Pre-multiplying (4.90) by $A(L)$ yields

$$A(L)\Delta x_t = A(L)C(L)\epsilon_t = \Delta\epsilon_t.$$

The differencing operator then cancels, so that, apart from initial values,

$$A(L)x_t = \epsilon_t.$$

When $A(L)$ and $C(L)$ are first order, a sufficient condition on $C(L)$ is that the matrix lag coefficient must be idempotent. The required lag coefficient of $A(L)$ may then be solved for. In this case: $A(L) = \left(I - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} L \right)$, therefore:

$$\begin{aligned} A(L)C(L) &= \begin{bmatrix} 1-\frac{1}{2}L & -\frac{1}{2}L \\ -\frac{1}{2}L & 1-\frac{1}{2}L \end{bmatrix} \begin{bmatrix} 1-\frac{1}{2}L & \frac{1}{2}L \\ \frac{1}{2}L & 1-\frac{1}{2}L \end{bmatrix} \\ &= \begin{bmatrix} (-0.5L+1.0)^2 - 0.25L^2 & 0.0 \\ 0.0 & (-0.5L+1.0)^2 - 0.25L^2 \end{bmatrix} \\ &= \begin{bmatrix} 1-L & 0 \\ 0 & 1-L \end{bmatrix}. \end{aligned}$$

It is also of interest to note from the Granger reparameterization applied to the AR and the MA representation, that the above condition implies:

$$\begin{aligned} A(L)C(L) &= (A(1)L + \Delta A^*(L))(C(1) + \Delta C^*(L)) = \\ &= A(1)C(1)L + \Delta A^*(L)C(1) + \Delta A(1)C^*(L)L + \\ &\quad \Delta^2 A^*(L)C^*(L) \\ &= \Delta I. \end{aligned} \tag{4.91}$$

It is necessary and sufficient for the above result to hold that the following conditions apply

$$A(1)C(1) = \alpha\beta' C(1) = 0$$
$$\Delta A^*(L)C(1)L^2 + \Delta A(1)C^*(L)L + \Delta^2 A^*(L)C^*(L) = \Delta I.$$

For the example you will observe that:

$$A(1)C(1) = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0.$$

Which derives from the condition for cointegration $\beta' C(1) = 0$. If we now look at the second term, then this yields the difference operator that cancels and for this example $A^*(L) = I$ and $C^*(L) = I$. Therefore:

$$\Delta A^*(L)C(1)L + \Delta A(1)C^*(L)L + \Delta^2 A^*(L)C^*(L) = \Delta C(1)L + \Delta A(1)L + \Delta^2.$$

Applying this result to the matrices for the VAR(1) example above, demonstrates the result presented in (4.91):

$$\begin{aligned} &\Delta C(1)L + \Delta A(1)L + \Delta^2 \\ &= \frac{1}{2}\begin{bmatrix} 1-L & 1-L \\ 1-L & 1-L \end{bmatrix} L + \frac{1}{2}\begin{bmatrix} 1-L & -1+L \\ -1+L & 1-L \end{bmatrix} \\ &\quad + \begin{bmatrix} 1-2L+L^2 & 0 \\ 0 & 1-2L+L^2 \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} -2L + L^2 + 2L(-\frac{1}{2}L + \frac{1}{2}) + 1 & L(-\frac{1}{2}L + \frac{1}{2}) + L(\frac{1}{2}L - \frac{1}{2}) \\ L(-\frac{1}{2}L + \frac{1}{2}) + L(\frac{1}{2}L - \frac{1}{2}) & -2L + L^2 + 2L(-\frac{1}{2}L + \frac{1}{2}) + 1 \end{bmatrix}$$

$$= \begin{bmatrix} -L + 1 & 0 \\ 0 & -L + 1 \end{bmatrix}.$$

4.7.3 Further discussion of representation and estimation.

Beyond what is discussed here, a number of other attempts have been made to explain non-stationary series that cointegrate. A variant of what has been termed the Bewley transformation was developed by Wickens and Breusch (1988), Gregoir and Laroque (1994) develop a definition for time series processes under cointegration that embodies polynomial cointegration. While Haldrup and Salmon (1998) have developed a number of decompositions of $C(L)$ that separate out components with different orders of integration, based on forms that generalize the Smith-McMillan form developed by Yoo (1986). The theory of monic matrix polynomials (Gohberg et al. 1983) has been used to factor the Wold form and transform the VMA into a VECM (Engsted and Johansen 1999).

However, none of these alternative representations has yielded a structure for inference and estimation to match the Johansen methodology. The reason for seeking an alternative approach derives from an inability to properly invert either VMA or VARMA representations. The Johansen procedure may be severely compromised by significant MA behaviour and this is not alleviated by increasing the number of observations. More specifically, should there be a VMA process generating the data then the order of VAR that approximates the VMA will increase with the number of observations. In the limit, the VAR order that properly approximates MA or ARMA behaviour is infinite.

Should that be the case then one might consider approximating the error behaviour using a semi-parametric estimator. Phillips and Hansen (1990) and Marinucci and Robinson (2001) have developed this approach to estimation and inference for long-run behaviour. Toda and Phillips (1994) suggest a triangular representation to identify the long-run relationships, which suggests that the Phillips and Hansen approach can be applied sequentially to estimate the long-run equations of a system. However, this type of structure will only by chance embed the types of restriction that economic theory might suggest and such systems might be viewed as a long-run reduced form. As was stated above, much of the evidence in support of modified estimators relates to bivariate or trivariate systems and often still requires the use of a test of cointegration.

Otherwise, the VARMA is preferred to the VAR when it defines a parsimonious time series representation of the data. The existence of VMA errors is likely when the original data has been differenced, as often strongly autoregressive univariate time series exhibit signs of some over-differencing. The above is generally observed as a spike in the autocorrelation function at the frequency of the difference. An advantage of the factorization presented in the last section is that it maintains a minimum order for the lag-length of the

MA and AR components. When compared with the VAR derived using the Smith–McMillan–Yoo form, the VARMA defines a unique factorization, which can be made robust to the choice of r the number of cointegrating vectors and when r is known, the long-run parameters can be estimated in one step. It is also feasible that a Johansen type procedure can be applied in this case (Hunter and Dislis 1996). The VARMA approach associated with this decomposition selects unique linear combinations of variables which are stationary when $FC(1) = 0$. Where an exact VARMA procedure to be used, then it is possible to handle roots on or inside the unit circle (Phadke and Kedem 1978). A similar approach has been adopted by Lütkepohl and Claessen (1993), though they estimate the long run using the Johansen procedure and then estimate the short-run model using a VARMA model.

4.8 Conclusion

In this chapter, cointegration associated with series that are I(1) or may be transformed to being I(1), has been considered. Granger (1983) first specified cointegration in terms of VMA processes which have been over-differenced. If one considers such over-differencing, then it is mirrored in the error processes, which then exhibit moving average behaviour with unit roots. The theory was developed for a system of equations and from the reparameterization of the VMA polynomial follows the fundamental result for cointegration that $rank(C(1)) = n - r$. This implies that there are r over-differences or r unit roots in the moving average representation of the differenced data. The over-differences relate the series that cointegrate or form linear combinations that are stationary, while the remaining $n - r$ series, require differencing to be made stationary. In the Granger representation theorem it is shown that the linear combinations that are stationary are associated with error correction terms or cointegrating vectors that transform the non-stationary series to stationarity. The cointegrating vectors transform the series to stationarity under the Wold form, because they annihilate $C(1)$, which leads to the r cointegrating variables having a multivariate moving average representation with all roots outside the unit circle.

Unfortunately, it is not easy to show that the VMA in differences inverts to a VAR in levels. The result developed by Engle and Granger (1987) is only valid for bivariate systems. Yoo (1986) developed a factorization based on Smith–McMillan forms, but these are only correct when $C(L)$ is a rational polynomial. In this chapter an alternative approach is developed, which gives rise to an exact inversion of the VMA to an error correcting VAR, but this requires a matrix F that is idempotent and which annihilates $C(1)$. It follows that F contains the cointegrating vectors.

Johansen, in a sequence of papers that are best summarized in Johansen (1991, 1995), decided to adopt an approach based on the VAR to estimate and test for cointegration. The difficulty with the VAR is that it is difficult to prove that it exists and to show that the cointegrating vectors define stationary series excepting when all the series are by definition I(1). This is not the case for the Granger representation theorem as the Wold form always exists and when the cointegrating vectors annihilate $C(1)$, then they always yield new series that are I(0). However, in the context of the VAR inference and estimation are relatively straightforward. It is also easy to undertake inference on the cointegrating vectors once the appropriate order of the VAR has been selected.

The Johansen procedure has been an enormously useful tool for modelling non-stationary time series and, as may be observed from the plethora of articles based on this methodology, it has been much used in economics and finance. With sufficient data, the tests have relatively good size and power properties as long as the underlying disturbances are Gaussian and the order of the VAR is finite. Should there be non-normality or should the VAR length be difficult to determine then the approach might be jeopardized and the tests ill sized (Hubrich et al. 2001). It would appear possible to correct for some of these problems by altering the model specification to correct for ARCH behaviour in the variance and outliers to capture the non-normality. It is even possible to correct the Johansen method for moving average errors (Hunter and Dislis 1996). However, all of these corrections require alternatives tables for the Johansen test statistic.

In the next chapter the question of exogeneity and identification are considered in the context of the Johansen VAR.

5
Exogeneity and Identification

In this chapter, we consider the question of long-run exogeneity and the related issue of identification. In the authors' opinion, detection of the exogenous variables in either the long run or the short run is a precursor to any attempt to structurally identify economic or financial phenomena.

In the preceding chapters, such issues were not addressed because single equations are always identified to a normalization and VAR models are viewed as being multicausal. Economic theory often determines that certain variables are viewed as being exogenous to the system, but, given the inherent interrelatedness of economic systems, it may prove too arbitrary to purely permit theory to select what is exogenous as compared with endogenous.

It is difficult for the economist to concede that the theory might not be paramount in this context or that there may well be systems where theory may have no prior view as to the variables that are exogenous. The requirement to devise a theory of exogeneity, which is about the model within which variables are embedded, has led to the development of a range of alternative notions of exogeneity based on the principle that models are almost invariably incomplete. Engle et al. (1983) defined such notions in the short run, while a similar sort of discussion for the long run occurs in Ericsson (1994) and further consideration is given to these ideas here.

The notion of exogeneity combined with the existence of a set of exogenous variables is viewed as a preface to any process of identification. A distinction is drawn between the theoretical (or generic) concept of identification and empirical identification. Generic identification relates to the technical feasibility of being able to detect the parametric structure of the model. The process of generic identification may or may not reveal operational conditions either necessary, sufficient or both necessary and sufficient to identify. Empirical identification relates to an ability to detect by a range of measurable conditions the parameters of a model. Consequently, even though a model may be generically identified, empirically this might not be the case and vice versa.

In this chapter, the idea of exogeneity is first discussed in broad terms and it is then considered relative to the long-run parameters. When compared with the short run some of the long-run concepts are directly testable. Identification is then discussed in terms of a conventional system of equations and finally in terms of the long-run parameters of the model.

5.1 An introduction to exogeneity

In terms of cointegration Johansen (1992) first defined the conditions on the matrix of loadings (α) for weak exogeneity when the matrix of cointegrating vectors defines the parameters of interest. Hunter (1992a) extended the discussion in Johansen to deal with weak exogeneity for a sub-block (β_i) of parameters in the VAR and cointegrating exogeneity. Cointegrating exogeneity implies a separation between the cointegrating vectors or long-run non-causality between the exogenous variables (z) and endogenous variables (y).

As is reported in Hendry and Mizon (1993), a necessary condition for weak exogeneity is a block triangular α matrix. This would suggest that cointegrating exogeneity is an exact long-run analogue of strong exogeneity (see Engle et al. 1983) as it combines weak exogeneity for a sub-vector with long-run non-causality. The statement above is valid when further restrictions are applied to the α matrix. One such type of restriction leads to the quasi-diagonal form first discussed in Hunter (1992). Hall and Wickens (1994) point out an observational equivalence between the triangular and diagonal forms of α associated with cointegrating exogeneity. As is stated in Hunter and Simpson (1995), this algebraic result holds only under very special conditions and it is only consistent with the definition of weak exogeneity of z for a specific sub-block of parameters (β_i) when strong exogeneity is accepted.

It follows, from Engle et al. (1983), that exogeneity is model-dependent in the sense that variables are exogenous for a particular parameterization of a model. This is of interest as in the context of the long run the standard definitions of exogeneity can be directly tested (see Ericsson and Irons 1994). In particular weak exogeneity, cointegrating exogeneity, strong exogeneity for both a sub-block and for β depend on restrictions on α and β.

5.1.1 Conditional models and testing for cointegration and exogeneity

In this section we formulate a VAR system and relate it to an error correction model. The conditions for cointegration are specified in terms of the levels parameters in the error correction model. Cointegration imposes a restriction on the matrix of long-run parameters, implying that questions about the nature of exogeneity need to be discussed in this context. At the end of this section, we look at cointegrating exogeneity and the restrictions on the

long-run parameters associated with cointegrating and weak exogeneity. Consider the n variable, k^{th} order VAR in levels with Gaussian errors:

$$A(L)x_t = \mu + \epsilon_t \text{ and } \epsilon_t \; NIID(O, \Sigma) \text{ for } t = 1, \; \ldots T \tag{5.1}$$

where $A(L) = (I + A_1L + A_2L^2 + \ldots A_kL^k)$. In error correction form:

$$\Gamma(L)\Delta x_t = \prod x_{t-1} + \mu + \epsilon_t \tag{5.2}$$

where $\Gamma(L) = (I + \Gamma_1L + \Gamma_2L^2 + \ldots \Gamma_kL^k)$ and Πx_{t-1} is a set of non-zero stationary linear combinations of x_{t-1}. The hypothesis of r cointegrating vectors is:

$$H_1(r) : \; \prod = \alpha\beta'$$

where rank (Π) = rank(α) = rank(β) = r and $0 \leq r \leq n$. Conditional on the rank of Π and thus r, we can test further restrictions on Π to determine whether the variables are cointegrating exogenous and or weakly exogenous. Engle et al. (1983) distinguish between a number of concepts of exogeneity: strict, strong, weak and super. The cointegration literature has mainly dealt with the weak exogeneity of a variable z_t for β (Johansen 1992b). Weak exogeneity is defined in terms of specific parameters of interest and formulated in terms of the distribution of observable variables. The joint density of x_t in (5.2) can be partitioned into a conditional density of y_t given z_t and a marginal density of z_t (Engle et al. 1983):

$$D(x_t | X_{t-1}, \Theta) = D(y_t | z_t, X_{t-1}, \phi_1) D(z_t | X_{t-1}, \phi_2)$$

where $x_t = [y_t, z_t]$ and $X_t = (X_0, x_1, x_2, \ldots x_t)$. Weak exogeneity requires that the parameters of interest depend on only the parameters of the conditional density of y_t and that there is a sequential cut of the parameter spaces for ϕ_1 and ϕ_2 (Florens, Mouchart and Rolin 1990). If so, the marginal density for z_t can be ignored without loss of information when conducting statistical inference about the parameters of interest. Strong exogeneity combines weak exogeneity with Granger non-causality, so that the marginal density for z_t becomes $D(z_t|Z_{t-1}, \phi_2)$. Super exogeneity requires weak exogeneity and that the parameters of the conditional process for y_t are invariant to changes in the process for z_t. Weak exogeneity can either be defined in terms of the β matrix as a whole or in terms of a sub-block $\beta_{.1}$.

5.1.2 Cointegration and exogeneity

Here, we take as our point of departure the matrix of long-run parameters (Π) in the vector auto-regression (VAR) in error correction form. To define more precisely the different forms of exogeneity we partition Π into blocks of cointegrating vectors associated with y_t and those related to z_t:

$$\Pi = \begin{bmatrix} \Pi_{1,1} & \Pi_{1,2} \\ \Pi_{2,1} & \Pi_{2,2} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}' = \begin{bmatrix} \alpha_1\beta_1' & \alpha_1\beta_2' \\ \alpha_2\beta_1' & \alpha_2\beta_2' \end{bmatrix}$$

where α_1 is $n_1 \times r$, α_2 is $n_2 \times r$, β_1 is $n_1 \times r$ and β_2 is $n_2 \times r$. Hendry and Mizon (1993) emphasize the exogeneity conditions associated with the short-run parameters of the VAR. It also holds in the long run that a necessary condition for weak exogeneity of z_t for a sub-block $\beta_{.1} = [\beta_{1,1}\, \beta_{2,1}]$ is a block triangular α matrix

$$\alpha = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ 0 & \alpha_{2,2} \end{bmatrix}.$$

However, the triangular form is not sufficient for weak exogeneity which means that we require further restrictions for appropriate long-run inference in a conditional model. When triangularization is combined with (i) below (see Hendry and Mizon 1993) then these two conditions are necessary and sufficient for weak exogeneity of z_t for the sub-block $\beta_{.1}$:

$$\alpha_{1,2} = \Sigma_{1,2}(\Sigma_{2,2})^{-1}\alpha_{2,2} = 0 \quad or \quad \alpha_{1,2} = -\Sigma_{1,2}(\Sigma_{2,2})^{-1}\alpha_{2,2}. \tag{i}$$

Cointegrating exogeneity augments the triangular α with non-causality between y and z at the level of the system. Hence, the long-run relationships for z do not depend on the levels of y. It follows that z_t is cointegrating exogenous for the sub-vector $\beta_{.1}$, if and only if:

$$\Pi_{2,1} = 0. \tag{ii}$$

Following Hunter (1990), this form of separating cointegration is quite arbitrary in that any orthogonal combination of β_1 and α_2 satisfy (ii). However, when $\beta_{.1}$ defines the parameters of interest, then the partition $\alpha_2 = [0 : \alpha_{11}]$ and $\beta_1 = [\beta_{1,1} : 0]$ is the only one that is relevant. This gives rise to following matrix of long-run parameters (Hunter 1992):

$$\begin{aligned} \Pi &= \begin{bmatrix} \Pi_{1,1} & \Pi_{1,2} \\ \Pi_{2,1} & \Pi_{2,2} \end{bmatrix} = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ 0 & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} \beta'_{1,1} & \beta'_{2,1} \\ 0 & \beta'_{2,2} \end{bmatrix} \\ &= \begin{bmatrix} \alpha_{1,1}\beta'_{1,1} & \alpha_{1,1}\beta'_{2,1} + \alpha_{1,2}\beta'_{2,2} \\ 0 & \alpha_{2,2}\beta'_{2,2} \end{bmatrix} \end{aligned} \tag{iii}$$

where $\alpha_{i,j}$ is $(n_i \times r_j)$ and $\beta'_{i,j}$ is $(r_j \times n_i)$, and the following vectors: $\eta_{it} = \beta'_{.1}x_t$ $\eta_{2t} = \beta'_{2.2}z_t$ define r_1 and r_2 blocks of stationary variables.

If conditions (i) and (ii) above hold, then cointegrating exogeneity in this form is an exact analogue of strong exogeneity as in the usual setting of dynamic models weak exogeneity is combined with non-causality (see Engle et al. 1983). Unfortunately the restrictions implied by (i) are not easy to impose which leads to the alternative special case of diagonalization first discussed in Hunter (1992). Diagonalization or quasi-diagonalization of the system requires (ii) in combination with (iv) below.

$$\alpha_{1,2} = 0 \tag{iv}$$

However, (iv) is sufficient for $\alpha_{1,2} = \Sigma_{1,2}\,(\Sigma_{2,2})^{-1}\,\alpha_{2,2} = 0$ as this condition implies that $\Sigma_{1,2} = 0$.[1] Once the quasi–diagonal form is accepted, then weak exogeneity of z_t for $\beta_{.1}$ is equivalent to weak exogeneity of z_t for the first n_1 blocks of Π. As a result the first sub-block of cointegrating vectors can be estimated from the y sub-system. Hall and Wickens (1994) discuss a special case of the above result which occurs when $\alpha_{1,1}$ is non-singular. As a result, the quasi-diagonal form is observationally equivalent to the cointegrating exogenous case. This occurs when rank $(\alpha_{1,1}) = n_1 = r_1$, because it is then possible to reparameterize Π in the following way

$$\Pi = \begin{bmatrix} \alpha_{1,1} & 0 \\ 0 & \alpha_{2,2} \end{bmatrix} \begin{bmatrix} \beta'_{1,1} & b \\ 0 & \beta'_{2,2} \end{bmatrix} = \begin{bmatrix} \alpha_{1,1}\beta'_{1,1} & \alpha_{1,1}b \\ 0 & \alpha_{2,2}\beta'_{2,2} \end{bmatrix}.$$

This diagonal form is equivalent to (iii) above; when $b = \beta'_{2,1} + (\alpha_{1,1})^{-1}\,\alpha_{1,2}\,\beta'_{2,2}$ and:[2]

$$\Pi = \begin{bmatrix} \alpha_{1,1}\beta'_{1,1} & \alpha_{1,1}\beta'_{2,1} + \alpha_{1,2}\beta'_{2,2} \\ 0 & \alpha_{2,2}\beta'_{2,2} \end{bmatrix}.$$

However, when $\alpha_{1,1}$ is non-singular, then b is a linear combination of some minimal or more primitive set of cointegrating vectors of which $\beta_{2,1}$ and $\beta_{2,2}$ are sub-blocks. This difficulty in interpretation does not arise when z_t is weakly exogenous for $b_{.1} = [\beta_{1,1}: b]$, but weak exogeneity implies a sequential cut in the parameter space, which only occurs when $(\alpha_{1,1})^{-1}\,\alpha_{1,2}\,\beta'_{2,2} = 0$ as otherwise $b_{.1} = f(\beta_{.2})$, which violates the condition for a sequential cut. If $(\alpha_{1,1})^{-1}\,\alpha_{1,2}\,\beta'_{2,2} = 0$ then either z_t is weakly exogenous for $\beta_{.1}$ or $\beta_{2,2} = 0$. In the latter case $r_2 = 0$, $r = r_1$ and the system is decomposed into r_2 difference stationary variables and $r_1 = n_1$ stationary variables.

It is more usual to start from the proposition that z_t is weak or cointegrating exogenous for some parameters of interest $\beta_{.1}$, block triangularity implies and is implied by (ii) when $\beta_{.1}$ define the parameters of interest. However, z_t is only weakly exogenous for $\beta_{.1}$ when $\alpha_{12} = 0$ or z_t is strongly exogenous. The invariance of $\beta_{.1}$ when a block diagonality restriction is applied is an indicator that the diagonal form is valid.

5.1.3 Tests of long-run exogeneity

For such tests see Johansen (1991a) and Mosconi and Giannini (1992). Long-run exclusion (Juselius 1994) and weak exogeneity tests can be readily applied, while cointegrating exogeneity can be implemented using the procedure in PCGIVE (Doornik and Hendry 2001). A more detailed explanation of the tests of cointegrating exogeneity is given in Hunter (1992). Johansen and Juselius (1990) show that conditional on the choice of r, a likelihood ratio test, which is asymptotically distributed chi-squared can be used to test these hypotheses. A range of tests associated with α and β, which are related to the restrictions

discussed both above and subsequently, are discussed in more detail by Johansen and Juselius (1992) and Mosconi and Giannini (1992). Such tests were categorized as follows in Hunter (1992):

$$
\begin{aligned}
&H_{4\beta} : \beta = H_{4\beta}\varphi.\ H_{4\beta}(n\times s), \varphi(s\times r).\\
&H_{6\beta} : \beta = (H_{6\beta}\varphi_1\psi_2),\ H_{4\beta}(n\times s),\ \varphi_1(s\times r_1),\ \psi_2(n\times r_2).\\
&H_{7\beta} : \beta = (\psi_1, H_{7\beta}\varphi_2),\ H_{7\beta}(n\times s),\ \varphi_2(s\times r_2),\ \psi_1(n\times r_1).\\
&H_{4\alpha} : \alpha = H_{4\alpha}\vartheta.H_{4\alpha}(n\times s),\ \vartheta(s\times r).\\
&H_{6\alpha} : \alpha = (H_{6\alpha}\vartheta_1, \kappa_2)\ H_{4\alpha}(n\times s),\ \vartheta_1(s\times r_1),\ \kappa_2(n\times r_2).\\
&H_{7\alpha} : \alpha = (\kappa_1, H_{7\alpha}\vartheta_2).\ H_{7\alpha}(n\times s), \vartheta_2(s\times r_2),\ \kappa_1(n\times r_1).\\
&\textit{where } r \le s \le n \text{ and } r_1 + r_2 = r
\end{aligned}
$$

Tests of weak exogeneity, long-run exclusion, strict exogeneity, cointegrating exogeneity and diagonalization are presented in Table 5.1 and implemented in PCGIVE (Doornik and Hendry 2001).

Let us consider the results presented in Hunter (1992)[3] who tests for WE, CE and diagonalization in the context of a six-variable VAR(2) model, which is an extension of the VAR model presented in Johansen and Juselius (1992). As is discussed in Johansen (1992), the diagnostic tests are conditional on the cointegrating rank being assumed to be the same as that selected by Johansen and Juselius (1992) as $r = 2$. Variables in logarithms are: oil prices (p_{ot}), UK prices (p_{1t}), world prices (p_{2t}), the UK effective exchange rate (e_{12t}), UK treasury bill rate (i_{1t}) and the Eurodollar rate (i_{2t}).[4]

Before undertaking the usual tests Hunter (1992) checks whether the oil price can be excluded from the long-run behaviour of the model. The test applied is termed general exclusion by Juselius (1995) or strict exogeneity by Hunter and Simpson (1995). The test implies that the oil price is WE for β, which means that none of the cointegrating vectors appear in the short-run oil price equation and long-run exclusion (LE), which means that the oil price is excluded from all the cointegrating vectors. The restriction for weak exogeneity implies that the first row of α is set to zero, which based on the framework presented above, requires a 5×2 matrix of freely estimated parameters ϑ and a 6×5 selection matrix $H_{4\alpha}$ the exact form, which is:

$$
\alpha = H_{4\alpha}\vartheta = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ & & I_5 & & \end{bmatrix}\vartheta,\ \vartheta = \begin{bmatrix} 0 & 0 \\ & \vartheta_2 \end{bmatrix}.
$$

Similarly for long-run exclusion:

$$
\beta = H_{4\beta}\varphi = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ & & I_5 & & \end{bmatrix}\varphi,\ \varphi = \begin{bmatrix} 0 & 0 \\ & \varphi_2 \end{bmatrix}.^{5}
$$

Hence, $H_{4\beta}$ is a 6×5 selection matrix and φ is a 5×2 matrix of unrestricted parameters. Testing for strict exogeneity requires the application of

the restrictions associated with LE ($H_{4\beta}$) and WE ($H_{4\alpha}$). Using the results presented in Table 4 of Hunter (1992), the restriction does not hold as $\chi^2(4) = 23.83$ exceeds the critical value (9.49). As a result of the above finding, all subsequent tests were applied to a model, which included all six variables. By applying to each variable the same type of restriction as $H_{4\alpha}$ above, Hunter (1992) finds that three variables out of six might be viewed as being WE for β.[6] Subsequently, WE tests are applied to groups of variables. In particular, for the case where (e_{12}) and (i_1) are tested, then ϑ is a 4×2 matrix of parameters $H_{4\alpha}$ is a 6×4 selection matrix:

$$\alpha = H_{4\alpha}\vartheta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \vartheta = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ 0 & 0 \\ 0 & 0 \\ \alpha_{61} & \alpha_{62} \end{bmatrix}$$

and the WE variables are associated with the 4^{th} and 5^{th} rows of $H_{4\alpha}$ and α respectively. The test is not significant as $\chi^2(4) = 4.04$ does not exceed the critical value.

To test whether i_1 and i_2 are CE for the first cointegrating vector implies the following restrictions:

$$\begin{bmatrix} \alpha_{1,1} \\ \alpha_{1,2} \end{bmatrix} = \begin{bmatrix} & I_4 & \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \vartheta_1$$

$$\begin{bmatrix} \beta_{1,2} \\ \beta_{2,2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \varphi_2.^{7}$$

The restrictions are accepted as $\chi^2(6) = 7.82$ is less than the critical value at the 5% level.

Here emphasis is placed on long-run non-causality, the short-run concept relates to a combination of restrictions associated with CE, that $\Pi_{1,2} = 0$ and those on the short-run dynamics. Mosconi and Giannini (1992) apply the test of non-causality in a short-run sense, while here the emphasis is solely on the long run. Non-causality in the long-run relations associated with the variables in the cointegrating equations implies a recursive structure to β, whereas co-integrating exogeneity also implies that the equations associated with the CE

variables do not include the CE vectors associated with the non-CE variables. Cointegrating exogeneity implies that long-run forecasts can be made conditional on the CE variables.

In practice, all the restriction applied above can be undertaken using the general restrictions approach dealt with in Hendry and Doornik (2001):

$$H_{g\alpha\beta} \; : \; \alpha = \alpha(\theta) \cap \beta = \beta(\theta)$$

where α and β take a general form, which permits non-linear restrictions of the form $\theta_1\theta_4 + \theta_2\theta_3 = 0$. Furthermore, the restrictions can apply both within and across equations. Such restrictions can be imposed on the parameters of the following matrices:

$$\alpha(\theta) = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ 0 & 0 \\ 0 & 0 \\ \theta_1 & \theta_2 \end{bmatrix}, \; \beta(\theta) = \begin{bmatrix} \beta_{11} & 0 \\ 1 & 0 \\ -1 & 0 \\ -1 & -1 \\ \theta_3 & \theta_4 \\ -\theta_3 & -\theta_4 \end{bmatrix}.$$

The restriction editor in PCFIML permits the imposition of a wide range of non-linear cross equation restriction (i.e., $\theta_1\theta_2 = 0$). The exact basis of the likelihood comparison is considered in more detail in Appendix F. Based on the above procedure tests of WE, LE, SE and CE are applied to the exchange rate system estimated by Hunter and Simpson (1995) and presented in the previous chapter. The I(0) variables are defined as $p_1 - p_2$, Δp_1, e_{12}, i_1, i_2 based on the same data set considered before, though estimated over the sample 1973q3–1991q4. Hunter and Simpson (1995) consider the reordering of the system by the degree of exogeneity or causal nature of variables and the relationship of this ordering to identification and identification/identifiability. The test results and associated restrictions are presented in Table 5.1. Firstly, whether variables influence the long run will be considered, that is a sub-set of variables might be either SE or WE for β or LE[8] from β. It can be seen from the results in Table 5.1 that the test of long-run exclusion is rejected for all five variables in the VAR. While strict exogeneity for β is accepted at the 1% level for i_2 (the Eurodollar rate) and WE of i_2 for β at the 1 and the 5% level. If i_2 is weakly exogenous for β, then the short-run equation does not include any of the cointegrating relationships. The Eurodollar equation is a difference equation, which means that the behaviour of Eurodollar rate is predominantly a random walk. Hunter and Simpson (1995) suggest that uncovering exogenous variables through tests of WE, LE and SE for β do not identify, because the restrictions are either common across and/or to all the long-run equations. Hence, the restrictions do not identify, though it will be observed

Table 5.1 Tests of weak and strict exogeneity, long-run exclusion and cointegrating exogeneity

Hypothesis		*Null*	*Statistic (95% critical value)*†	
Cointegration	r = 4	r ≤ 3	$-n\Sigma\ln(1-\lambda_i)$ = 16.4*	(15.4)
(WE)|r = 4	Δp_1	α_{1i} = 0, for i = 1,…4	$\chi^2(4)$ = 32.54**	(9.49)
	e_{12}	α_{2i} = 0, for i = 1,…4	= 17.62**	
	$p_1 - p_2$	α_{3i} = 0, for i = 1,…4	= 16.71**	
	i_1	α_{4i} = 0, for i = 1,…4	= 10.43*	
	i_2	α_{5i} = 0, for i = 1,…4	= 8.18	
(LE)|r = 4	Δp_1	β_{j1} = 0, for j = 1,…4	$\chi^2(4)$ = 39.94**	(9.49)
	e_{12}	β_{j2} = 0, for j = 1,…4	= 20.31**	
	$p_1 - p_2$	β_{j3} = 0, for j = 1,…4	= 29.97**	
	i_1	β_{j4} = 0, for j = 1,…4	= 25.73**	
	i_2	β_{j5} = 0, for j = 1,…4	= 9.52*	
(SE)|r = 4	Δp_1	$\alpha_{1i} = \beta_{j1}$ = 0, i,j = 1,…4	$\chi^2(8)$ = 43.66**	(15.51)
	e_{12}	$\alpha_{2i} = \beta_{j2}$ = 0, i,j = 1,…4	= 31.34**	
	$p_1 - p_2$	$\alpha_{3i} = \beta_{j3}$ = 0, i,j = 1,…4	= 38.37**	
	i_1	$\alpha_{4i} = \beta_{j4}$ = 0, i,j = 1,…4	= 29.88*	
	i_2	$\alpha_{5i} = \beta_{j5}$ = 0, i,j = 1,…4	= 18.56*	
(CE) e_{12} and i_1 for $\beta_{.3}$|r = 4		$\alpha_{23} = \alpha_{43}$ = 0, $\beta_{31} = \beta_{32} = \beta_{34}$ = 0	$\chi^2(5)$ = 2.50	(11.07)

Note: †Cointegrating exogeneity (CE), strict exogeneity (SE), weak exogeneity (WE) and long-run exclusion (LE). (* significant at the 5% level and ** significant at the 1% level.)

subsequently that uncovering exogenous variables or excluding variables from all cointegrating vectors can aid the identification process.

As part of their approach to identification Hunter and Simpson also test for CE in their revised exchange rate system. The test for CE implies a sub-block of Π for which condition (ii) holds ($\Pi_{2,1}$ = 0). Evidence for CE may be drawn from the insignificance of parameters in Π. It was decided on the basis of t-tests by Hunter and Simpson (1995) that the treasury bill rate and the exchange rate are CE for a cointegrating vector which can be restricted to satisfy PPP. It can be observed from Table 5.1 that to test this proposition, six restrictions are applied to α and β, but of these only five are binding. The test is not insignificant as χ^2 (5) = 2.5, so the restrictions for CE can be accepted for both the exchange rate and the UK treasury bill rate (i_1), for the cointegrating vector that satisfies PPP. That the terms of trade are caused by the exchange rate might be viewed as counter-intuitive, but it is quite consistent with the type of sticky price monetary model of the exchange rate recently considered by Charles Engle (2001). Turning to the cointegrating vector for which i_1 is CE, because this variable is also set to zero in the long-run relationship, long-run non-causality takes a trivial form.

Bauwens and Hunter (2000) discussed identification in association with conditions for exogeneity and by applying tests of WE for a sub-block of variables they showed that the model estimated by Hunter (1992) can be identified from restrictions on α alone. At this point, it is viewed appropriate to consider the tests of weak exogeneity and strong exogeneity presented in the article by Bauwens and Hunter (2000).

Weak exogeneity for a sub-block of cointegrating vectors ($\beta_{.1}$) implies that $\alpha_{2,1} = 0$ and $\alpha_{1,2} = \Sigma_{1,2} (\Sigma_{2,2})^{-1} \alpha_{2,2}$. The exchange rate and interest rates are WE for a long-run augmented PPP equation when: firstly the loadings for that cointegrating vector are zero in both the exchange rate and interest rate equations ($\alpha_{2,1} = 0$) and secondly, the three coefficients in α associated with the price equations are in proportion to the coefficients in $\alpha_{2,2}$.[9] The test is insignificant as $\chi^2(4) = 2.5132$ does not exceed the criterion at the 5% level. Bauwens and Hunter proceed to test for long-run strong exogeneity, this combines WE with long-run non-causality. For the model in Hunter (1992), both interest rates (i_1, i_2) and the exchange rate (e_{12}) are strongly exogenous in the long run for the interest rate augmented PPP vector, because (i_1, i_2, e_{12}) satisfy the WE restrictions, and the interest rate vector is not long-run caused by the real oil price (p_0) and goods prices (p_1, p_2).[10] The restrictions that are being applied are a combination of those required for WE ($\alpha_{2,1} = 0$ and $\alpha_{1,2} = \Sigma_{1,2} (\Sigma_{2,2})^{-1} \alpha_{2,2}$) and CE ($\alpha_{2,1} = 0$ and $\beta_{1,2} = 0$).

As has been observed above, it is possible to undertake direct parametric tests of WE, LE and SE for β and WE, CE and strong exogeneity for a sub-block of β. Otherwise, the observation that the long-run parameters are invariant to a sub-set of variables combined with WE implies that such variables are super exogenous, but no direct restrictions apply to the parameters in the long run. For further discussion of super exogeneity, see Ericsson and Irons (1994).

5.2 Identification

Parametric econometric identification is the capacity to appropriately detect model parameters from empirical observations. One can further discriminate between a conceptual capacity to determine parameters algebraically and an observational capability of the data to permit such a distinction. The former is termed generic identification by Johansen and Juselius (1994). Generic identification is concerned with the specification of conditions that permit parameters to be solved, discriminated or detected from an unrestricted system or estimable reduced form, and consequently such conditions may be defined prior to any analysis based on the innate structure of the model. The latter form is empirical identification, although it might be possible, based on some restrictions, to identify some parameters, the restrictions selected might not be empirically acceptable and as a result the model will not be identified

in practice. A further issue which limits our ability to identify is the notion of observational equivalence. Appropriate restrictions might be found and generic identification satisfied, the restrictions applied might be accepted, but it may not be possible to discriminate between one class of model and another model drawn from a different set of theoretical principles.

For linear models identification is usually straightforward, depending on simple order conditions and a rank restriction (Goldberger 1964). When one considers further degrees of non-linearity, then it becomes more difficult to prove generic identification and the process becomes more empirical in nature. Although certain advances have been made, the notion of observational equivalence is often all that is available to discriminate between identified and non-identified models (Rothenberg 1971). Rothenberg (1971) makes a further distinction between local and global identification. Local identification is described as the ability to discriminate between models with observationally distinct parameterizations within a neighbourhood of the optimum. Consequently, identification, by its very nature, becomes more empirical and any conclusions drawn are reliant on the parameterization of the problem. Generic identification often stems from the rank of the information matrix, which is a necessary criterion for safe optimization, though in practice highly ill conditioned problems may yield locally well-defined parameter estimates. The empirical and generic notions become intimately related. The ability to estimate some 'structure' consistently yields the possibility of a sub-category of models, which may be observationally equivalent. Usually, the minimum parametric form is a reduced form and from this more specific structural models can be identified.

It is a combination of such necessary and sufficient conditions that will be the main concern of the following sections of the chapter, in combination with the question of observational equivalence. These results are then applied to the identification of long-run relationships. In the above sense, generic identification depends on sufficient conditions derived from Rothenberg (1971) combined with an order condition necessary for identification. Identification and identifiability are viewed as being non-linear in nature, which implies that this treatment is both different and more general than that of Johansen (1995a) and Boswijk (1996). The treatment also permits the ready combination of restrictions on all the parameters associated with the long-run behaviour of the model.

Some of the conditions considered here stem from the article by Hunter (1998) where the question of non-identification is addressed. Sargan (1983a) emphasized what he defines as conditions for higher order identification, the very existence of which may depend on higher-order moments. In this context consistency and non-identifiability are not equivalent when identification depends on distributional assumptions. This renders the usual

condition on the Hessian or information criterion (Doornik 1995; Doornik and Hendry 1996) as a necessary, but not sufficient condition. Questions of distribution automatically open the door to a Bayesian treatment of the problem (see Bauwens, Lubrano and Richard 2000, for a discussion of this issue).

Firstly, the preliminaries of identification, identifiability and observational equivalence are discussed, and then their relation to cointegration is considered. Next the results of Johansen, Boswijk and Hunter are placed in context and discussed in relation to some simple cases.

5.2.1 I(0) systems and some preliminaries

For a generic notion of identification or identifiability, it is important to consider the issue of observational equivalence. It is this idea which forms the basis for most conditions and definitions of identification, even though the final condition may be far removed from this. It is this definition, which is most general in nature, though often less easy to consider in practice. If one considers the simultaneous equation model (SEM), it is common in the identification literature to take as point of departure a matrix of reduced form parameters P. Consider, the following structural form for a linear SEM (see Goldberger 1964):

$$By_t + \Gamma z_t = u_t \text{ and } u_t \sim D(0, \Sigma) \tag{5.3}$$

where B is an $n_1 \times n_1$ matrix of endogenous variable parameters, y_t an n_1 vector of endogenous variables, Γ an $n_1 \times n_2$ matrix of exogenous variable parameters, z_t an n_2 vector of predetermined variables, u_t an n_1 vector of structural errors and Σ an $n_1 \times n_1$ variance–covariance matrix. It is well known that the relationship between the reduced form parameters (P) and the structural form parameters is:

$$P = -B^{-1}\Gamma. \tag{5.4}$$

It is common to redefine (5.3) above, thus:

$$Ax = u_t$$

where $A = [B : \Gamma]$ and $x_t = [y_t : z_t]$. Identification usually follows from the acceptance of a number of linear restrictions of the form:

$$R_i a_i = 0 \; for \; i = 1, \; \ldots, \; n_1$$

where a_i is a column vector composed of the i^{th} column from A and R_i is a selection matrix that determines the variables to be restricted in the i^{th} equation. This leads to the classical result on identification attributed to Koopmans (1953), which implies satisfaction of a rank condition. This gives rise to the following theorem:

Theorem 4 *A necessary and sufficient condition that parameters* [A : Σ] *are identified is rank* $(R_i) = n_1 - 1$ *for* $i = 1, \ldots, n_1$.

As it stands, no account is taken of the impact of cross-equation restrictions. It follows from simple algebraic manipulation of (5.4), that:

$$BP + \Gamma = 0$$

or

$$\begin{bmatrix} B & \Gamma \end{bmatrix} \begin{bmatrix} P \\ I \end{bmatrix} = A\Phi = 0. \tag{5.5}$$

When (5.5) is transposed and a single row from A is considered then the rank condition becomes:

$$rank(\begin{bmatrix} \Phi' & R_i \end{bmatrix}) = n_1 - 1.$$

However, applying the above condition to identify generically is somewhat complicated as it requires the matrix P associated with the restricted model for each case, otherwise a rank test similar to that associated with cointegration is needed. However, there is an order condition that can be used to select the appropriate number of restrictions

$$j_i + n_2 \geq n_1 + n_2 - 1$$

or

$$j_i \geq n_1 - 1.^{11}$$

Furthermore, some of the types of restriction that violate the rank condition are well known, identification is lost when two equations use the same restriction as they are observationally equivalent and the same restriction applied to all equations simply reduces the number of operational variables in the model and thus a restriction is lost. However, the type of restriction discussed above is linear in nature and often restrictions might well be non-linear (i.e., the case of CE discussed above requires non-linear estimation). Prior to any discussion of cointegration we consider non-linear identification, based on the results in Rothenberg (1971) and Sargan (1988). The following theorem follows from Sargan (1975):

Theorem 5 *If there exists a regular point* $\theta_0 \in \Theta$ *where* Θ *is some well-defined parameter space, then* θ_0 *is locally unidentifiable when there exists a* $\theta_1 \in \Theta$ *such that*:

$$L(\theta_0 | X_t) = L(\theta_1 | X_t)$$

where L is the likelihood function and X_t *is the observed sample*.

Proof: Sargan (1975).

Consider the following reduced form system

$$y_t = Pz_t + v_t$$

where y_t and z_t are defined above, v_t an n_1 vector of reduced form errors and P is an $n_1 \times n_2$ matrix of reduced form parameters. A consistent estimator of P is:

$$\hat{P} = Y'Z(Z'Z)^{-1}$$

where Y and Z are matrices composed of T stacked observations of the vectors y_t' and z_t'.

Now take a non-linear function which maps $vec(P) = p$ (an $n_1 n_2$ vector) onto ξ (a vector of q structural parameters).

$$p = g(\xi) \tag{5.6}$$

To preclude a trivial over-parameterization, it is assumed that $q \leq n_1 n_2$. A necessary condition for identification, given by Rothenberg (1971), is that the Hessian matrix is non-singular.[12] An alternative approach, relates to the Jacobian of the transformation and this is described as first-order identification by Sargan (1983):

$$rank\left(\frac{\partial g}{\partial \xi'}\right) < q. \tag{5.7}$$

Again, the above condition is necessary for local identification and failure leads to a model that satisfies the full rank condition that is generally viewed as being unidentified. However, for non-linear models, where $\frac{\partial g}{\partial \xi}$ has full rank, it may still be possible to obtain solutions to (5.6), because the conditions for singularity or near singularity are less burdensome than those required to solve (5.6). This gives rise to the following theorem that derives from Sargan (1983a):

Theorem 6 *Given p a vector of reduced form coefficients, then a sufficient condition for the identification of a vector of structural parameters* ξ *is the existence of a unique solution* $\xi = \xi^*$ *to the vector function* $p^* = g(\xi^*)$.

Proof: If $g(.)$ is continuously differentiable within a neighbourhood of ξ^* and $rank(\frac{\partial g}{\partial \xi}) < q$, then $\xi = \xi^*$ is a solution (5.6) and ξ^* is identified. However, when $rank(\frac{\partial g}{\partial \xi}) \approx q$ and $\xi = \xi^*$ is a solution to (5.6), then ξ^* is still identified. ■

By simulation Sargan (1983b) shows that there may be near singular models that cannot be distinguished from singular models, but satisfy (5.6) and are thus identified. The convergence in distribution of estimators derived from such near singular cases turns out to be much slower than usual. Because of a larger than usual asymptotic variance they tend to be classified empirically as unidentified.

An order condition which is necessary for identification obviates the problem of over-parameterization, while conditions for the existence of a solution to (5.6) are sufficient for global identification for a broad range of non-linear models. For example, Hunter (1992) considers such conditions for rational expectations models.

5.2.1.1 The cointegration case

Equation (5.3) might equally well define a structural cointegration model when $y_t = \Delta x_t$ and $z_t = x_{t-1}$. Taking a range of specifications of u_t, (5.3) can represent any order of VAR or VMA process. In the cointegration case the matrix P defines the matrix of long-run coefficients usually termed Π. The latter, assumed to have rank equal to r (the number of cointegrating relations), is decomposed thus:

$$\Pi = \alpha\beta'$$

where α and β are $n \times r$ matrices of rank r. Cointegration takes as a starting point the identification of Π. The error correction form under the usual assumptions produces r dependencies between n variables. In theory *rank* $(\Pi) = r$, but excepting exact dependencies, all n^2 elements are commonly estimated as compared to the $2nr$ elements in α and β, which need to be identified. After normalization, there are only $(n-1)r$ unrestricted elements in β. Under cointegration, a comparison between the number of parameters in Π, equal to $n^2 - (n-r)^2$, and those in α and β equal to $nr + (n-1)\,r$, gives rise to the order condition that at least $r^2 - r$ elements of α and β must be restricted (when $r = 1$, there is no need for restriction beyond the normalization). For identification of the long-run parameters, we require $j = r^2 - r$ restrictions to reduce the number of redundant parameters in α and β. This defines a multi-equation version of the usual systems order condition:[13]

$$r \leq r^2 - j \text{ or } j \geq r^2 - r \tag{5.8}$$

Without a true knowledge of structure, the reduced rank condition on Π reduces the set of alternative long-run models. However, the order condition is only necessary for identification. Two issues arise: (a) certain types of restriction do not identify; (b) alternative structural models may be identified. The former is well known within the conventional identification literature (Goldberger 1964; Sargan 1988), specifically, the same restriction applied to each equation in turn or the same restrictions applied to two or more equations. Observational equivalence implies the existence of two models with different structure, which are statistically indistinguishable. In the context of cointegration, Johansen (1995a) shows that there is a non-null set of models, which are observationally equivalent depending on the nature of the restrictions imposed.[14]

There is a presumption in the identification literature related to cointegration that restrictions on α are a priori non-identifying. This would seem to be the sentiment in Johansen's work, while Pesaran et al. (2000) is stronger in condemnation of those who impose such restrictions. In the Bayesian literature there is a suggestion that this overly complicates the estimation process (Bauwens, Lubrano and Richard 2000). It must be stated that (5.3), from which (5.8) is defined, has no such limitation on the imposition of restrictions. However, P, as distinct from Π, presumes the existence of a partition of x into endogenous and exogenous variables (Ericsson and Irons (1994) and Section 5.1).

Prior to any discussion of the merits of alternative approaches to identification there must be some discussion of these issues. Firstly, there are some cases where there may be no natural restrictions on β, whereas there may be strong views over causality (Parker 1998). Secondly, both cause and restriction might be relevant to a specific theory which implies that identification depends on joint tests on both α and β (for example, tests of the monetary approach to the balance of payments imply PPP and causality from price to the exchange rate). Thirdly, exogeneity might be viewed as a precursor to any analysis, because weak exogeneity of some zs for β may be a sufficient condition for identification (Hunter 1998). Fourthly, both non-linearity and non-normality might imply a significant role for prior information in the identification process (Bauwens, Lubrano and Richard 2000).

5.2.2 A simple indirect procedure for generic identification

In this section, identification is handled by a procedure that is widely used in the literature for I(0) econometrics.[15] First, a sufficient number of restrictions needs to be selected and this follows from the order condition derived above (i.e., $j_i = r - 1$ restrictions per long-run equation). Secondly, for generic identification a solution needs to be found for the parameters of the structural form from some reduced form. Thirdly, empirical identification is checked by testing any over-identifying restrictions.

Rothenberg (1971) suggests that global conditions for identification depend on the relationship between the structural form (SF) and the reduced form (RF) parameters. It is well known that the relationship between the reduced form parameters (P) and the structural form parameters for the linear case is:

$$P = -B^{-1}\Gamma. \tag{5.9}$$

If P is unrestricted, then identification of P follows from our ability to estimate the long-run parameters.

A multivariate generalization of the conventional condition for the identification of a regression equation is required, that is a first moment matrix composed of some regressors has to have full rank. If $P = \Pi$ is calculated using the

Johansen procedure, then the VAR(1) transformation yields well-defined parameters for all the long-run equations in the system when $rank(R_1'R_1) = n$ where R_1 is an $n \times n$ matrix of regression residuals. Under cointegration, Π is typically a reduced rank matrix and $rank\ (R_1'R_1) = r$. Hence, the Johansen rank test for cointegration determines the extent to which the rows and columns of Π may be uniquely defined. Identification of Π is only necessary for the existence of the long-run parameters.

Under cointegration Π is a reduced rank matrix, which implies that $n - r$ rows and columns are dependent. Now, from the traditional treatment of matrix algebra (Dhrymes 1984), a matrix of rank r has r independent rows and columns. Partitioning Π such that

$$\Pi = \begin{bmatrix} \Pi_1 \\ \Pi_2 \end{bmatrix} = \begin{bmatrix} \alpha_1\beta' \\ \alpha_2\beta' \end{bmatrix}, \tag{5.10}$$

then Π_1 is $r \times n$ dimensioned and α_1 is an $r \times r$ dimensioned sub-matrix of α. Subject to Π, α and β all of rank r (cointegrating rank), then $rank(\alpha_1) = r \Rightarrow \beta$ is identified. Where α_1 with full rank is a sufficient condition for β to be identified from Π as:

$$\beta' = (\alpha_1)^{-1}\,\Pi_1 .$$

By a similar argument an analogous result exists for α, which may be identified when a square sub-matrix of β exists such that:

$$\alpha' = (\beta_1)^{-1}\,\Pi_{.1}$$

where $\Pi_{.1}$ is an $r \times n$ matrix composed of the first r columns of Π. A generic proof of identification for the model estimated by Hunter and Simpson (1995) follows from solving equations of a similar form to (5.9), except for the cointegration case $P = \alpha\beta'$. The solution to the restricted parameters is shown to exist for the model in Hunter and Simpson (1995) using the indirect result derived in Appendix G.

Identification stems from the imposition of a number of additional restrictions on α and β. In this case, identification was simplified by reordering the system using the tests of weak exogeneity presented in Table 5.1.

5.2.3 Johansen identification conditions

Johansen, in a series of papers with Juselius (1994) and (1995a), places emphasis on β as the parameters over which structural hypotheses are defined. However, cointegration doesn't differentiate between endogenous and exogenous variables, which negates the original Cowles foundation view that identification stems from information on exogenous or pre-determined variables. By concentrating on the long run with no exogenous variables then conventional cointegration has no predetermined variables.[16]

The Johansen approach to identification considers a series of rank conditions which yield a family of ordered tests of the form:

$$H_1 \subset H_2 \subset H_3 \ldots \subset H_i.$$

Identification follows from the acceptance of the sequence of tests. Let us look at some linear restrictions of the form:

$$R_i A_i = 0 \text{ for } i = 1, \ \ldots r.$$

As was observed in section 5.4, it is more common in the cointegration literature to formulate the restrictions as:

$$A_i = H_i \varphi.$$

And H_i is a selection matrix composed of zeros and ones, and φ an i dimensioned vector of unrestricted parameters. As a result:

$$A = [H_1 \varphi_1 H_2 \varphi_2 \ldots H_r \varphi_r].$$

Any linear statistical model with a set of restrictions may be defined thus:

$$L = A_{gxr}, \ \sum | H_i \varphi_i = 0, \ i = 1, \ \ldots r.$$

It is then of use to differentiate between models which are linear and restricted as compared with those which are identified:

$$M = A_{gxr}, \ \sum | H_i \varphi_i = 0, \ rank \ (R_i A_i) = r - 1, \ i = 1, \ \ldots r.$$

Notice that the restrictions associated with M are now non-linear by virtue of the rank restriction. When the set of all possible restrictions is considered, then the class of just identified models is likely to be large, though it will define a subset of the restricted models, so that $M \subset L$.

Theorem 7 *If L contains an identified parameter value, that is M is not empty, then M is an open dense subset of L.*

Proof: see Johansen (1995b).

The above result implies that there is a non-null set of models that cannot be distinguished on the basis of the likelihood and they define a family of observationally equivalent models, which satisfy the rank condition and thus correspond to a point in M with certainty. If these results are made particular to the cointegration case, then the parameter point given by the restrictions R_i for $i = 1, \ldots r$, with

$$A^{(i)} = [H_1 \varphi_1 H_2 \varphi_2 \ldots H_r \varphi_r],$$

defines a representation which is identified when:

$$rank(R_i H_1 \varphi_1 R_i H_2 \varphi_2 \ldots R_i H_r \varphi_r) = r - 1 \text{ for } i = 1, \ \ldots r.$$

Under cointegration a necessary condition for a specific set of linear restrictions to be identifying is given by the following theorem.

Theorem 8 *The linear statistical model L defined by the restrictions R_i for i = 1, … r is identifying if and only if for each i*:

$$rank(R_i H_{i_1} R_i H_{i_2} \ldots R_i H_{i_k}) \geq k$$

where k is set in turn to 1, 2 …, *r* – 1 *for all sequence of indices* $1 \leq i_1 \leq i_2 \ldots \leq i_k \leq r$ *such that* $i_j \neq i$.

Proof: Of necessity and sufficiency see Johansen (1995b).

Consider, for example, the model estimated by Hunter and Simpson (1995), which has $r = 4$ cointegrating vectors, $\beta = [H_1\varphi_1 \ H_2\varphi_2 \ H_3\varphi_3 \ H_4\varphi_4]$, then identification of the first cointegrating vector alone requires us to check:

$$rank(R_1' H_{i_1}) = 1, \text{ for } i_1 = 2,3,4$$
$$rank(R_1' H_2 R_1' H_{i_2}) = 2 \text{ for } i_2 = 3,4$$
$$rank(R_1' H_2 R_1' H_3 R_1' H_4) = 3.$$

In the case of the second cointegrating vector:

$$rank(R_2' H_{i_1}) = 1, \text{ for } i_1 = 1,3,4$$
$$rank(R_2' H_1 R_2' H_{i_2}) = 2 \text{ for } i_2 = 3,4$$
$$rank(R_2' H_1 R_2' H_3 R_2' H_4) = 3.$$

Similar types of rank condition need to be checked for each remaining cointegrating vector.

Consider the simpler case estimated by Hunter (1992) and used before in section 5.1.3. In this case $n = 6$, $r = 2$, $\beta = [H_1\varphi_1 \ H_2\varphi_2]$ and based on the order condition two restrictions are required to identify each cointegrating vector without normalization. For this section PPP is applied as a parametric restriction to the first vector [*, *a*, -*a*, -*a*, *, 0] in combination with a zero restriction on the eurodollar rate,[17] while the second vector is restricted to accept UIP, [0, 0, 0, 0, *b*, –*b*]. Hence, there are $j_1 = 2$ restrictions in the first vector, which without normalization is enough to just identify. And $j_2 = 5$ means that the second vector ought to be over–identified before normalization. Therefore:

$$H_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

$$R_1' = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & \mathbf{0} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } R_2' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

In this case both vectors are identified when $k = r - 1 = 1$ conditions are satisfied, for a block of homogeneous restriction of the form $R_k'\beta_k = 0$ or $R_k' H_k = 0$ for $k = 1, 2$. It follows that the Johansen approach to identification checks each combination of conditions $rank(R_i' H_j) = 1$ for $i \neq j$. In the case of the first vector, it follows that,

$$R_1'H_1 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

the two matrices are orthogonal, while for identification:

$$R_1'H_2 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

According to the Johansen conditions, the first equation is just identified, whereas identification of the second vector follows from $rank(R_2'H_1) \geq r - 1$:

$$R_2'H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence, the second vector is identified, because the matrix product above has 3 independent rows and columns.

As can be seen from the above derivations, the algebra becomes increasingly burdensome with r. The conditions also relate to the specific definition of generic identification described by Johansen (1995b) and the article does not address the issue of empirical identification or the more general notion of identification associated with observational equivalence.

5.2.4 Boswijk conditions and observational equivalence

Boswijk (1996) emphasizes the restriction on β rather than those on α, though he argues that similar results will also hold for α. To solve this problem, Boswijk (1996) provides two further conditions for what he terms identifiability. According to Boswijk, β is non-identifiable when the normalization fails or some of the remaining parameters are not significant. Therefore:

$$H_{02}: \beta \in B_3 \cup B_4 = \{\beta: rank(R_1^{*}\beta) \leq r-1\},$$

where $R_1^{*\prime}$ is the restriction matrix including the normalization and $B_3 \cup B_4$ defines the null associated with non-identifiability. Consider the following example, developed from Boswijk (1996), $n = 3$, $r = 2$ and $j = 2 = r^2 - r$ restrictions identify β:

$$\beta' = \begin{bmatrix} a & 0 & b \\ c & d & 0 \end{bmatrix} \text{ and } H_2 = \begin{bmatrix} a & o \\ c & d \end{bmatrix}.$$

Selecting the normalization, $a = 1$ and $d = 1$ it follows from Boswijk (1996), that the first vector in β' is identifiable when the matrix H_2 has full rank. To discriminate between failure of normalization and other types of failure, a further rank test is applied to an $r - 1$ dimensioned sub-matrix. Therefore:

$$H_{03}: \beta \in B_4 = \{\beta: rank(R_1'\beta) \leq r-2\}.$$

In the example rank failure occurs for H_{02} when $a = 0$ (normalization) and for the further restriction associated with H_{03}: $d = 0$. However, from the acceptance of the Johansen test for cointegration ($rank(\beta') = r$), β is identifiable as r linearly independent cointegrating vectors must exist and, given acceptance of the over-identifying restrictions, then the first vector is identified when $I(0)$ variables are precluded from the system. Using the approach of Boswijk, once the first vector is identifiable, then rank conditions need to be tested for each of the other vectors in turn.

Based on the results presented in Hunter and Simpson (1995) and those above, some of the problems associated with incorrect normalization may be

obviated by excluding normalizations associated with weakly exogenous and long-run excluded variables.

5.2.5 Hunter's conditions for identification

It is conventional in the identification literature to consider the relationship between reduced form and structural form parameters when the problem is non-linear (Rothenberg 1971). Given the existence of a set of row and column vectors of appropriate dimension selected from Π, then the question arises as to which such sub-blocks might be used to identify. In this light an orientation might be selected for the system which does not prejudge the possible normalization of the long-run parameters, but defines possible rows and columns from which the α and the β might be identified; it should be noted that there are $\frac{n!}{(n-r)!r!}$ possible alternative combinations of rows and columns from Π.

The following theorem provides sufficient conditions for the existence of a unique solution to a vector function relating the identifiable elements of Π, that is $\xi = vec(\Pi_r)$, where $\Pi_r = [\pi_{ij} \in \Pi_i \cup \Pi_{.j}]$,[18] to the unknown parameters in α and β, that may be stacked in θ.

Theorem 9 *Given (5.8) is satisfied and knowledge of r (the cointegrating rank) a sufficient condition for a solution to the vector function $\theta = g(\xi)$ is the existence of two $r \times r$ dimensioned non-singular sub-matrices A and B, in α and β respectively.*

Proof. *Rank(α) = r* is equivalent to the existence of a sub-matrix A such that *rank(A) = r*. There are $\frac{n!}{(n-r)!r!}$ possible alternative combinations of rows of α from which A might be formed. It follows that each A has a related sub-matrix Π_i of Π such that $rank(A) = r \Leftrightarrow rank(\Pi_i) = r$ and $\Pi_i = A\beta'$. Vectorizing Π_i implies that $vec(\Pi_i) = vec(A\beta') = (I_n \otimes A)vec(\beta')$ (see Dhrymes 1984, pp. 40–3 and chapter 4). Following the argument in Sargan (1983a, pp. 282–3), β is identifiable when A has full rank as firstly a unique solution results:

$$vec(\beta') = (I_n \otimes A)^{-1} vec(\Pi_i) \tag{5.11}$$

and secondly $rank\left(\frac{\partial vec(\beta')}{\partial vec(\Pi_i)'}\right) = nr$, if the normalization is ignored. By similar argument, α is identifiable when there exists two matrices $\Pi_{.j}$ and B for which $\Pi_{.j} = \alpha B'$ and B is non-singular. As a result, a unique solution for α exists of the form:

$$vec(\alpha) = (B \otimes I_n)^{-1} vec(\Pi_{.j}) \tag{5.12}$$

■

In the cointegration case, the existence of one or more solutions to (5.11) and (5.12) is sufficient for the existence of a solution to $\theta = g(\xi)$, which is what is required for identification given (5.8). Finding such solutions negates the need to undertake the test in Johansen (1995b).

Linearity, or the need to consider α and β, does not present a problem for the condition in Theorem 9 that may be applied sequentially to α and β to yield a sufficient set of solutions. Empirical verification of the generic result follows from a direct test of the over-identifying restrictions:

$$(I) \quad \begin{aligned} H_\beta &: \phi_\beta + R_\beta vec(\beta) = 0 \\ H_\alpha &: R_\alpha vec(\alpha) = 0. \end{aligned}$$

Now ϕ_β is a $j_\beta \times 1$ vector of known constants (normalizations), R_β and R_α are $j_\beta \times nr$ and $j_\alpha \times nr$ matrices, which select all the j_β and j_α restrictions on β and α respectively, and $j = j_\beta + j_\alpha$.[19] The degrees of freedom of the test are calculated from the number of solutions to (5.11) and (5.12). If (*I*) is rejected, then this is sufficient for non-identification, while identification requires a different set of restrictions. However, acceptance of (*I*) is only necessary for identification as there may be a sequence of models, that accept either the over-identifying restrictions or Johansen's test (Johansen, 1995b).

Here, an alternative approach follows from the sufficient conditions for a solution to (5.11) and (5.12) given in Theorem 9:

(*II*) Test identifiability: $rank(B) = r$ and $rank(A) = r$.

The existence of a solution to (5.11) and (5.12) implies the system is generically identified. As Boswijk suggests, on empirical grounds identification may fail due to insignificance of certain parameters. Here, identifiability follows from the existence of sufficient information in certain rows and columns of Π to identify α and β (Sargan, 1983). Clearly, many such orientations related to particular over-identifying restrictions may exist. However, it is sufficient to find one such orientation of the system to empirically accept the generic solution. Consider the example used above where for comparison with Boswijk we let $B = H_2$. When $rank(H_2) = r$,[20] then the condition in Boswijk (1996) is satisfied, but also the sufficient condition for the existence of a solution to (5.12) (a matrix B of full rank). From Theorem 9, the rank condition identifies α based on the restrictions in (*I*). Then conditional on (*I*), discovery of a matrix (B) with full rank is sufficient for identification of α.

If the variable chosen for normalization is invalid ($a = 0$ and $rank(H_2) < r$), then failure of the rank condition yields an additional restriction on the set of cointegrating vectors (β'). Therefore α can be identified from a new orientation:

$$\beta' = \begin{bmatrix} 0 & 0 & b \\ c & d & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} 0 & b \\ d & 0 \end{bmatrix}.$$

The system is now over-identified as $j = 3 > r^2 - r$. From acceptance of the Johansen rank test, $|B| = 0$ can only occur when $d = 0$, but this contradicts the proposition that $\text{rank}(\beta') = 2$. The structure of β' based on $d = 0$ gives x_1 and x_3 as the cointegrating vectors, so two series in x_t are I(0).[21]

Boswijk and Johansen emphasize a limited information approach associated with linear restrictions, that can only be applied to α and β in turn. In this section, restrictions can be applied to both α and β, they can be non-linear and they apply to the system as a whole.

In the next section, the results are extended further to take account of exogeneity.

5.3 Exogeneity and identification

Traditional econometric methodology assumes the existence of a set of exogenous variables, whereas the notion of cointegration and vector autoregressive (VAR) modelling negates this. Cointegration is multi–causal and the VAR treats all variables as endogenous but within such a system, it is feasible to test a number of notions of long-run exogeneity. The reader is directed to look at Ericsson and Irons (1994) and Ericsson et al. (1998). Now consider the impact of long-run exogeneity and identification on the system.

Let the system (5.2) be separated into two sub-models, corresponding to a partition of x_t into y_t and z_t of dimensions n_1 and n_2, respectively, and conformable partitioning of α and β:[22]

$$\Delta y_t = (\alpha_{1,1}\beta'_{1,1} + \alpha_{1,2}\beta'_{1,2})y_{t-1} + (\alpha_{1,1}\beta'_{2,1} + \alpha_{1,2}\beta'_{2,2})z_{t-1} + \epsilon_{1t} \tag{5.13}$$

$$\Delta z_t = (\alpha_{2,1}\beta'_{1,1} + \alpha_{2,2}\beta'_{1,2})y_{t-1} + (\alpha_{2,1}\beta'_{2,1} + \alpha_{2,2}\beta'_{2,2})z_{t-1} + \epsilon_{2t}, \tag{5.14}$$

where $(\epsilon'_{1t}\ \epsilon'_{2t})' \sim N(O, \Sigma)$ and independently over $t = 1, \ldots, T$. It is well known that when $[\alpha_{2,1}: \alpha_{2,2}] = [0 : 0]$, then z_t is weakly exogenous for β (Johansen 1992).

However, such restrictions do not directly assist in the identification of the long-run parameters as they apply to a part of α which is non-informative. In terms of the requirement to find a solution to (5.11) and (5.12), weak exogeneity is of direct use when there are $n - r$ weakly exogenous variables as the only basis for a choice of A is the matrix $[\alpha_{1,1}: \alpha_{1,2}]$, which is then by definition of rank r.

Otherwise, one might consider weak exogeneity associated with a sub-block of cointegrating vectors. To discuss issues of exogeneity it is useful to look at the conditional model for y_t given z_t (Johansen 1992):

$$\begin{aligned}\Delta y_t = {} & [(\alpha_{1,1}\beta'_{1,1} + \alpha_{1,2}\beta'_{1,2} - \omega(\alpha_{2,1}\beta'_{1,1} + \alpha_{2,2}\beta'_{1,2})]y_{t-1} + \omega\Delta z_t \\ & + [(\alpha_{1,1}\beta'_{2,1} + \alpha_{1,2}\beta'_{2,2} - \omega(\alpha_{2,1}\beta'_{2,1} + \alpha_{2,2}\beta'_{2,2})]z_{t-1} + \epsilon_{1t} - \omega\epsilon_{2t}\end{aligned} \tag{5.15}$$

where $\omega = \Sigma_{1,2}\Sigma_{2,2}^{-1}$. One set of sufficient conditions for weak exogeneity of z_t for $\beta'_{.1} = [\beta'_{1,1}: \beta'_{2,1}]$ is $\alpha_{1,2} - \omega\alpha_{2,2} = 0$ and $\alpha_{2,1} = 0$, see Lemma 2 in Ericsson et al. (1998). Combining (5.15) with (5.14) yields a system which, to a non-singular transformation matrix, is equivalent to the original VAR. If ($\alpha_{1,2} = 0$, $\alpha_{2,1} = 0$) is

applied to (5.13) and (5.14), then the VAR has a quasi-diagonal long-run structure (Hunter, 1992). For weak exogeneity additional restrictions may apply as $\alpha_{1,2} - \omega\alpha_{2,2} = 0$ is required. Should $\alpha_{1,2} = 0$, then $\omega\alpha_{2,2} = 0$ is sufficient for weak exogeneity. This result can be associated with three possible requirements: (i) $\omega = 0$; (ii) $\alpha_{2,2} = 0$; or (iii) ω is a left-hand side annihilation matrix of $\alpha_{2,2}$. Under cointegration, (ii) does not apply as *rank* $(\alpha_{2,2}) = r_2$. Case (i) is consistent with Lemma 2 in Ericsson et al. (1998). For case (iii), the quasi-diagonality restriction ($\alpha_{1,2} = 0$, $\alpha_{2,1} = 0$) combined with $\omega\alpha_{2,2} = 0$ is sufficient for weak exogeneity of z_t for $\beta_{.1}$.

Weak exogeneity for a sub-block implies that analysis may be undertaken at the level of the sub-system. More specifically, identification conditions now apply at the level of the sub-system, as previously at the level of the full system. Let Π_1 denote an $n_1 \times n$ sub-matrix of Π for which *rank* $(\Pi_1) = r_1$ and $n_1 > r_1 \geq 1$. If $\Pi_{1(r_1)}$ defines an $r_1 \times n$ sub-matrix of Π_1 for which the maximum rank is given by its smallest dimension, then an equivalent column matrix exists which is $n_1 \times r_1$ and has full column rank. Given the quasi-diagonality restriction, it follows that:

$$\Pi_1 = \alpha_{1,1}\beta'_{.1} \text{ and } \Pi_{1(r_1)} = A_1\beta'_{.1}, \tag{5.16}$$

where A_1 is a square matrix of full rank r_1 obtained from $\alpha_{1,1}$ (by selecting r_1 rows). To identify $\alpha_{1,1}$ and $\beta_{.1}$ subject to a standard normalization (i.e. r_1 restrictions) the following sub-system order condition now applies:

$$r_1 n + r_1 n_1 - r_1 \leq r_1 n + r_1 n_1 - r_1^2 \Leftrightarrow r_1^2 - r_1 \leq j_1,$$

where j_1 is the number of restrictions associated with the sub-system. Now, $r_1 - 1$ restrictions apply to each equation in the first sub-block as compared with $r - 1$ when the full system condition is used. Hence, r_2 variables are viewed as exogenous to the sub-system.

Theorem 10 *Given $r_1^2 - r_1 = j_1$ and knowledge of the sub-system cointegrating rank (r_1), a sufficient condition for the existence of a solution to the vector sub-system: $vec(\beta'_{.1}) = (I_n \otimes A_1)^{-1} vec(\Pi_1(r_1))$ is the existence of a matrix A_1 of full rank r_1 constructed by selection of r_1 rows of $\alpha_{1,1}$.*

Proof. By analogy with the proof of Theorem 9, $vec(\beta_{.1})$, which follows from vectorizing (5.16), is identifiable when A_1 has full rank. ■

A special case arises when $r_1 = 1$ and excepting the choice of normalization no further restrictions are required to identify $\beta_{.1}$.

Corollary 11 *If $r_1 = 1$, then subject to a normalization, weak exogeneity is sufficient for identification of the long-run parameters $\beta_{.1}$ associated with the first sub-block.*

If in addition, $r_2 = 1$, then for a specific normalization weak exogeneity is all that is required for the identification of β when $r_1 + r_2 = r$. It follows from weak exogeneity that identification is a natural consequence of the partition. In more general sub-systems, the type of conditions derived in the previous section are relevant.

It can readily be shown that a similar result to Theorem 10 applies to any subsequent sub-system. Hence, $vec(\beta_{.2})$ is identified when a sub-matrix A_2 of α_2 has full rank. There are now at least two sub-systems that can be separately estimated and identified based on the above conditions.[23] However, the quasi-diagonal form of weak exogeneity implies that while y is dependent in the long run on z in the first sub-block, then z is also dependent on y in the second block. The latter statement does not appear to be consistent with the idea that in the long run the notions of exogeneity and causality are coherent.

To address the above concern, attention is focused on cointegrating exogeneity, the restrictions $\beta_{1,2} = 0$ combined with $\alpha_{2,1} = 0$ imply that z is not long-run caused by y and as a result $\Pi_{2,1} = 0$. Restrictions associated with cointegrating exogeneity direct attention towards the identification of the long-run parameters in a sub-block. However, such restrictions only identify β to the sub-block as ($\beta_{1,2} = 0$) implies that the same restrictions are applied to all the rows of $\beta_{.2}$. However, the order condition per sub-block is now less onerous ($r_2 - 1$ restrictions). And when $r_2 = 1$, then $\beta_{2,2}$ is identified via a normalized coefficient. When compared with the impact of quasi-diagonalizing the system, cointegrating exogeneity applies only to the set of identified sub-system relationships. In terms of identifying that sub-block, the following relationship is of interest:

$$\Pi_{2,2} = \alpha_{2,2}\beta'_{2,2}.$$

If $rank(\Pi_{2,2}) = r_2$, then there is a sub-matrix $\Pi_{2,\,(r_2)}$ of dimension $r_2 \times n_2$, and a matrix of column vectors dimensioned $n_2 \times r_2$, both of rank r_2. Now the order condition for this sub-system is:

$$r_2 n + r_2 n_2 - r_2 \leq r_2 n + r_2 n_2 - r_2^2 \Leftrightarrow r_2^2 - r_2 \leq j_2.$$

Even with all of the zero restrictions in the second block of cointegrating vectors, the number of relevant restrictions in the order condition for the sub-block remains unchanged at the level of the sub-block. Subject to an appropriate number of identifying restrictions, then a sufficient condition for the existence of a solution to the system associated with $\beta_{2,2}$ is the existence of A_2, an $r_2 \times r_2$ sub-matrix of $\alpha_{2,2}$. By analogy with the result in Theorem 10, the following relationship exists for $\beta_{2,2}$:

$$vec(\beta'_{2,2}) = (I_{n_2} \otimes A_2)^{-1} vec(\Pi_{1,(r_2)}).$$

Further, when z_t is also cointegrating exogenous, then the long-run behaviour of the sub-system for z_t does not depend on the endogenous variables. If z_t is

both weakly exogenous for $\beta_{.1}$ and z_t is not long-run caused by y_t, then z_t is termed long-run strongly exogenous for $\beta_{.1}$ Therefore, strong exogeneity combines the restrictions associated with weak exogeneity and the restrictions appropriate for cointegrating exogeneity.

In the next section, the identification and identifiability of a model involving weak, cointegrating and strongly exogenous variables is addressed.

5.4 Empirical examples

To motivate the analytic solution and empirical results discussed in the last section, the approach is applied to the data set analyzed by Johansen and Juselius (1992) and Hunter (1992a).[24] The system of equations associated with Theorem 9 is observed to have a number of solutions, which directly relate to the correct degrees of freedom for the test of over-identifying restrictions. Emphasis is placed on a model, that is identified via restrictions on α discussed in section 5.3 and both weak exogeneity and cointegrating exogeneity are tested.

From the discussion in section 5.2, whether it is possible to identify the parameters in the long run follows from the ability to solve for α and β from well-defined rows and columns of Π. According to Theorem 6, this depends on the existence of what might be called a valid orientation of the system. If $\Pi_i = A\beta'$ and from the cointegrating rank test $rank(\beta') = r$, then it follows from the conditions on the rank of sub-matrices, that $rank(A) = r \Rightarrow rank(\Pi_i) = r$. Hence, determining an A matrix with full rank is equivalent to associating the solved system with well-defined parts of the matrix Π. The ability to identify the parameters empirically from the solution to the algebraic problem of the form (5.11) and (5.12) relies empirically on finding matrices A and B with full rank. Prior to undertaking such a test, a set of minimum restrictions will be defined and then tested.[25] For generic identification of a system with $r = 2$ cointegrating vectors $r^2 - r = 2$ restrictions are required with normalization and r^2 without. To test the over-identifying restrictions and identifiability, the likelihood ratio test discussed in Johansen and Juselius (1992) and implemented in Doornik and Hendry (1998, 2001) is used. Using the results in Section 5.3, α and β can be identified via a normalization and the restrictions associated with quasi-diagonal α also discussed in section 5.1.2:

$$\alpha' = \begin{bmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{42} & \alpha_{52} & \alpha_{62} \end{bmatrix}. \tag{5.17}$$

The only restrictions applied to β are those associated with the normalization ($\beta_{41} = -1$, $\beta_{52} = 1$).

$$\begin{matrix} & p_0 & p_1 & p_2 & e_{12} & i_1 & i_2 \end{matrix}$$
$$\beta' = \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} & -1 & \beta_{51} & \beta_{61} \\ \beta_{12} & \beta_{22} & \beta_{32} & \beta_{42} & 1 & \beta_{62} \end{bmatrix}.$$

Table 5.2 Tests of exogeneity and identification conditional on $r = 2$

Test	*Null*	*Statistic* [*p-value*]
(*I*) Quasi-diagonality	$\alpha_{i1} = 0$ for $i = 4, 5, 6$; $\beta_{41} = -1$ $\alpha_{i2} = 0$ for $i = 1, 2, 3$; $\beta_{52} = 1$.	χ^2 (4) = 3.9595 [0.4115]
(*IIa*) Non-identifiability	$\alpha_{31} = 0$, $\alpha_{62} = 0$	$\chi^2(2) = 30.0465$ [0.0000]
(*IIb*) Non-identifiability	$\alpha_{51} = 0$, $\alpha_{52} = 0$	$\chi^2(2) = 4.42$ [0.1097]
(*IIc*) Non-identifiability	$\beta_{41}\beta_{52} - \beta_{42}\beta_{52} = 0$	$\chi^2(1) = 3.9087$ [0.0481]
(*IIIa*) Weak exogeneity	$\alpha_{i1} = 0$ for $i = 4, 5, 6$; $\beta_{41} = -1$, $\alpha_{i2} = \omega_{i1}\alpha_{42} + \omega_{i2}\alpha_{52} + \omega_{i3}\alpha_{62}$ for $i = 1, 2, 3$; $\beta_{52} = 1$.	$\chi^2(4) = 2.5132$ [0.6423]
(*IIIb*) Strong exogeneity (Weak + Cointegrating Exogeneity)	$\alpha_{i1} = 0$ for $i = 4, 5, 6$ $\alpha_{i2} = 0$ for $i = 1, 2, 3$ $\beta_{i2} = 0$ for $i = 1, \ldots, 4$.	$\chi^2(8) = 12.708$ [0.1223]

It can be seen from the p-value associated with test (*I*) in Table 5.2 that the long run is identified: (i) six restrictions are imposed ($j = r^2 - r = 2$) and (ii) the test of over-identifying restrictions is accepted at the 5% level.[26]

Now consider the orientation of the system or the selection of the appropriate r-dimensioned square matrices A and B. A valid choice for A is based on the 3rd and 6th rows from α. For a solution, it is required that:

$$vec(\beta') = (I_6 \otimes A)^{-1} vec(\Pi_3). \tag{5.18}$$

Hence, any matrix A needs to be of full rank. Following the acceptance of the quasi-diagonality restriction then the identifiability of β depends on the rejection of the condition $|A| = 0$.[27] One possible orientation is:

$$A = \begin{bmatrix} \alpha_{31} & 0 \\ 0 & \alpha_{62} \end{bmatrix} \text{ and } \Pi_3 = \begin{bmatrix} \pi_{31} & \pi_{32} & \pi_{33} & \pi_{34} & \pi_{35} & \pi_{36} \\ \pi_{61} & \pi_{62} & \pi_{63} & \pi_{64} & \pi_{65} & \pi_{66} \end{bmatrix}.$$

This test is applied under a null of non-identifiability of β (Table 5.2, *IIα*), the test is $\chi^2(2)$ and the null is rejected at 5% and any other conventional level of significance. Should one consider the alternative orientation associated with the treasury bill rate (i_1) and the exchange rate equations, then both were jointly accepted to be weakly exogenous by Hunter (1992a). To compare this orientation with that used above it is of interest to note that when the restrictions $\alpha_{51} = 0$ and $\alpha_{52} = 0$ are used to augment $\alpha_{31} = 0$ and $\alpha_{32} = 0$ (Table 5.2, *IIb*) then when compared with a $\chi^2(2)$ statistic the null cannot be rejected at the 5% significance level. This implies that the fifth column does not yield an appropriate sub-matrix to orientate the system and by a similar argument the fourth column can also not be used.

A possible choice of B is based on the fourth and fifth columns of Π, so that

$$vec(\alpha) = (B \otimes I_6)^{-1} \begin{bmatrix} vec(\pi'_{.4}) \\ vec(\pi'_{.5}) \end{bmatrix} \text{ and } B' = \begin{bmatrix} -1 & \beta_{42} \\ \beta_{51} & 1 \end{bmatrix} \tag{5.19}$$

where $\pi'_{.j} = [\pi_{1j}\ \pi_{2j}\ \ldots\ \pi_{6j}]$ for $j = 4, 5$. Here the test of orientation for the identification of α is undertaken prior to the imposition of any restriction (see Table 5.2, *IIc*). Under the null the determinant of B is set to zero, the test is $\chi^2(1)$ and from the critical value non-identifiability can be rejected at the 5% level. It follows from Theorem 9 that $\Pi_{.j} = \alpha B'$ and from the cointegrating rank test $rank(\alpha) = r$, so $rank(B) = r \Rightarrow rank(\Pi_{.j}) = r$ also and the orientation with respect to α is valid.

It follows that a solution can now be derived from (5.18) and (5.19) based on the selected A and B matrices (see Appendix H):

$$\begin{aligned}
\theta &= [\alpha_{11}\ \ \alpha_{21}\ \ \alpha_{31}\ \ \alpha_{42}\ \ \alpha_{52}\ \ \alpha_{62}\ \ \beta_{11}\ \ \beta_{21}\ \ \beta_{31}\ \ \beta_{51}\ \ \beta_{61}\ \ \beta_{12} \\
&\quad \beta_{22}\ \ \beta_{32}\ \ \beta_{42}\ \ \beta_{62}] \\
&= [\tfrac{1}{\delta}\pi_{14} - \tfrac{\beta_{51}}{\delta}\pi_{15}\ \ \ \tfrac{1}{\delta}\pi_{24} - \tfrac{\beta_{51}}{\delta}\pi_{25}\ \ \ \tfrac{1}{\delta}\pi_{34} - \tfrac{\beta_{51}}{\delta}\pi_{35}\ \ \ -\tfrac{\beta_{51}}{\delta}\pi_{44} - \tfrac{1}{\delta}\pi_{45} \\
&\quad -\tfrac{\beta_{51}}{\delta}\pi_{54} - \tfrac{1}{\delta}\pi_{55}\ \ \ -\tfrac{\beta_{51}}{\delta}\pi_{64} - \tfrac{1}{\delta}\pi_{65}\ \ \ \alpha_{31}^{-1}\pi_{32}\ \ \ \alpha_{31}^{-1}\pi_{33} \\
&\quad \alpha_{31}^{-1}\pi_{35}\ \ \ \alpha_{31}^{-1}\pi_{36}\ \ \ \alpha_{62}^{-1}\pi_{61}\ \ \ \alpha_{62}^{-1}\pi_{62}\ \ \ \alpha_{62}^{-1}\pi_{63}\ \ \ \alpha_{62}^{-1}\pi_{64}\ \ \ \alpha_{62}^{-1}\pi_{66}] \\
&= g^{-1}(\xi), \text{ where } \delta = -1 - \beta_{42}\beta_{51}.
\end{aligned}$$

Theorem 9 implies that a sufficient condition for the existence of a solution to the vector system associated with the first v_1 cointegrating vectors is the existence of a matrix A_1 such that:

$$vec(\beta'_{.1}) = (I_6 \otimes A_1)^{-1} vec(\prod_{1,(r_1)}).$$

From Corollary 11, when $r_1 = 1$, then the existence of a block of weakly exogenous variables is a sufficient condition for identification of the cointegrating vectors in the first block. By analogy the second block is also identified, when $r_2 = 1$. The system is sequentially identifiable from the restrictions on α alone and the selection of the normalization. In this case, the long run is partitioned into two sub-systems for which $r_i = 1$ and consequently each vector is identified by the normalization alone.

5.5 Conclusion

In this chapter exogeneity and identification have been discussed. Exogeneity implies restrictions on the long-run parameters of the model. In the case of weak exogeneity for β, the requirement is that all cointegrating vectors are excluded from the equation for the weakly exogenous variable. This proposition is tested using a likelihood ratio test, which compares the model estimated using a VAR, which is only restricted by virtue of the rank restriction on Π, with models that, irrespective of the restriction, can be estimated using the generalized restriction estimator given in Appendix F (Doornik 1995). For weak exogeneity and long-run exclusion there are r restrictions on α and β respectively for each variable excluded, while for strict exogeneity, there are

$2r$ restrictions on α and β for each variable excluded. Such restrictions are binding and can be tested by a test, which is asymptotically distributed χ^2_{ir} for WE and LE, and χ^2_{2ir} for SE, where i denotes the number of variables excluded. Small sample corrections are available for these tests either via the bootstrap (Podivinsky 1993) or exact small sample correction (Johansen 2002).

Cointegrating exogeneity is comparable with Granger causality, in the sense that the non-linear restrictions, when applied, are associated with non-causation of the exogenous by the endogenous variables in the long run, but in the later case the restriction also applies to the short-run dynamics of the model. The variables not caused are termed cointegrating exogenous for β. Forecasts of the endogenous variables in the long-run can be made conditional on the forecasts of the cointegrating exogenous variables, because both β and Π have a block triangular structure. As pointed out by Toda and Phillips (1994), care must be taken in determining the degrees of freedom of this test, because there is an annihilation of parameters that implies that not all of these restrictions are binding. Doornik (1998) has implemented a procedure for checking the degrees of freedom, but in more general terms the problem is best viewed as one of identification. The restrictions for exogeneity only in very special cases identify the cointegrating vectors. Furthermore, such common restrictions applied to β only identify to a sub-block of equations.

The procedure for identification outlined can be applied using standard packages and identifiability is a product of the conditions required for generic identification. The procedure requires identification to be checked on an a priori basis. The test of the existence of the sufficient conditions associated with Sargan (1983a) stems from the application of restrictions to both α and β, and the whole approach can be made operational with a range of non-linear restrictions.

The method was applied to data well known in the cointegration literature. The discovery of a solution to the vector conditions associated with Theorem 9 verifies the restrictions as over-identifying and determines the degree of over-identification. Identifiability of α is accepted on the basis of a test similar to the H_{02} in Boswijk (1996). However, this test confirms that it is appropriate to solve the system using the selected rows and columns of Π. Hence, the orientation of the system and the solution uncovered are empirically identified. Identifiability of β follows from restrictions on α that relate to the exogeneity of the variables selected. The question of which variables are exogenous would appear to be of importance when the appropriateness of the normalization is at issue.

Based on the results in section 5.4, the system was identified by imposing a quasi-diagonality restriction on α and by normalizing with respect to r coefficients in β. It is shown that quasi-diagonality, subject to additional covariance

restrictions, implies weak exogeneity for a sub-block of β. Finally, the joint acceptance of weak exogeneity and cointegrating exogeneity tests for the interest rates implies that they are long-run strongly exogenous for the first cointegrating vector. Given the diagonalization of the system, this causal ordering further emphasizes that the interest rates are the exogenous variables in the system.

6
Further Topics in the Analysis of Non-Stationary Time Series

6.1 Introduction

In this chapter three further topics are considered in some detail: estimation of models with I(2) variables; forecasting; and structural models with short-run behaviour driven by expectations. Though mathematically the notions of order of integration and cointegration are exact, in practice they are valid to the best approximation or resolution that the data may permit. To define an order of integration as a specific integer quantity is to assume that the series is approximated by a single well-defined time series process across the sample. Time series data for developed economies have exhibited many features, from behaviour that might be viewed as purely stationary through to series that require first or second differencing to render them stationary. Some nominal series in first differences may require further differencing, which suggests that the original nominal series are of order I(2) or higher when further differencing is required. In this chapter, discussion is limited to processes up until I(2).

The condition required for a series to be considered to be I(1), as compared with one exhibiting further features only consistent with I(2) behaviour, is necessary and sufficient for cointegration amongst I(1) series, but beyond testing this condition, there is a well defined procedure for inference and estimation of I(2) processes (Johansen 1992, 1995). It might often be difficult to distinguish between an I(1) and an I(2) series, which suggests that series, which appear to be I(2), are being approximated to some order of accuracy by second differences. Alternatively, these series may be better modelled using non-integer orders of differencing (Granger and Joyeux 1980; Hosking 1981). To this end, the question of fractional processes and long-memory will be discussed briefly after the section on I(2) behaviour. A further reason why it might be difficult to detect the order of integration of a series may be due to the existence of structural breaks. This opens up a plethora of potential difficulties for any form of structural modelling. Breaks in structure have a

number of forms when conventional (I(0)) linear econometrics is considered, but beyond slope and intercept shifts, there are other types of intercept correction used in macro modelling (see Clements and Hendry 1998, 2001). The break may also apply to the cointegrating relations (co-breaks) or in the order of integration and cointegration. Testing was limited in chapter 4 to recursive break tests and tests with a known break in structure that could be corrected by the use of dummy variables. In this chapter forecast performance is compared by considering the difference between forecasts made with and without the imposition of cointegration. Specifically, the simulation results of Hendry and Clements and Lin and Tsay are evaluated.

Once the notion of forecast failure is considered, then issues associated with our ability to detect short-run structure arise. In this context, there can be no difference between estimating a structural relationship as compared with a reduced form, except for the added efficiency that might derive from the imposition of further restrictions on the long-run and short-run parameters. There are a number of approaches to defining structural models under cointegration of which the best defined follows from the work of Pesaran et al. (2000). The elegance of the Johansen approach is lost once the long-and the short-run coefficients are interrelated, as testing for a unit root in multivariate processes cannot be readily disentangled from the estimation of the long-run and short-run parameters. In particular, when the long-run parameters are embedded within the short run, as occurs with models with future expectations, then testing for cointegration is less straightforward. Here, the impact of forward-looking behaviour is considered in terms of exogenous processes that are weakly and cointegrating exogenous and then processes that have unit roots in the exogenous variables. The simple method suggested by Dolado et al. (1991) is considered along with an extension of this method to the multivariate context by Engsted and Haldrup (1997). An alternative maximum likelihood approach is discussed here, though the inference is contaminated by both the unit root and generated regressor problem.

6.2 Inference and estimation when series are not I(1)

In this section the I(2) approach advanced in Johansen (1992a), is considered along with some discussion of multi-cointegrated and fractional processes. Whether a series is I(1), close to I(1) in levels or differences, is a matter of debate. To some extent cointegration operates beyond the framework of this debate, because long memory processes may also interact, as has been observed recently by Abadir and Talmain (2002). From the original definition of cointegration due to Engle and Granger (1987) series of order I(j) cointegrate and I(1) and I(0) series may also combine in the manner described by Flôres and Szafarz (1996). One estimator, which combines I(0), I(1) and I(2)

processes is that given in Johansen (1992a). This assumes that differenced series are of integer order, which rules out the possibility that series such as inflation rates are fractional processes. The distinction between long memory and non-stationarity might be viewed as semantic for the data sets readily available, but one cannot dismiss the possibility that series may move across orders of integration from non-stationarity through long-memory to stationarity. In this light the series might never be purely stationary or non-stationary. Where this would appear to accord with sound economic principle then one might have to look for the best approximation.[1]

6.2.1 Cointegration when series are I(2)

Consider the cointegration case developed by Engle and Granger (1987), where all the series are I(2). It follows from our discussion of cointegration in chapter 4 that second differences have the following Wold decomposition:

$$\Delta^2 x_t = C(L)\epsilon_t$$

and Δx_t cointegrate when $\beta'_{I(2)}\, C\,(1) = 0$ and $\beta'_{I(2)}\, \Delta x_t = \eta_{I(2)t} \sim I(0)$. If a left-hand factor can be extracted in the manner described in section 4.5, then:

$$\Delta^2 x_t = C_0(L)C_1(L)\epsilon_t. \tag{6.1}$$

It is possible to transform the Wold form into an error-correcting VARMA when $FC(1) = 0$, and F is an idempotent matrix. Therefore:

$$(\Delta I - FL)\Delta x_t = C_1(L)\epsilon_t. \tag{6.2}$$

When $C_1\,(L)$ has no more unit roots, then an I(2) cointegrating VAR exists in second differences:

$$\Gamma(L)\Delta^2 x_t = \prod_{I(2)} \Delta x_{t-1} + \epsilon_t$$

where $\Pi_{I(2)} = \alpha_{I(2)}\beta'_{I(2)} = F$. This has been called balanced I(2) behaviour by Juselius (1995). Now consider the case where $C(1)$ has further unit roots, then it might be possible to undertake a further factorization when a left-hand term $C_{01}(L) = (I - GL)$ can be extracted and $GC_1(1) = 0$. Therefore:

$$(\Delta I - FL)\Delta x_t = (I - GL)C_{11}(L)\epsilon_t \tag{6.3}$$

$$(\Delta I - GL)(\Delta I - FL)x_t = C_{11}(L)\epsilon_t \tag{6.4}$$

The following I(2) representation can be readily derived from multiplying through the two left-hand divisors above. Therefore:

$$\Delta^2 x_t - F\Delta x_{t-1} - G(\Delta x_{t-1} - Fx_{t-2}) = C_{11}(L)\epsilon_t,$$

transforming to the VAR by inverting $C_{11}(L)$ and applying the reparameterization $(A(1)L + (1 - L)A^{*}(L))$ to produce terms in first differences and $(A(0) + (1 - L)A^{+}(L) + (1 - L)^2 A^{++}(L))$ terms in levels,

$$\Psi(L)\Delta x_t - A(1)F\Delta x_{t-2} + A(0)GFx_{t-2} - A(1)G\Delta x_{t-2} + A^{+}(0)GF\Delta x_{t-2} = \epsilon_t$$

or

$$\Psi(L)\Delta x_t = A(1)F\Delta x_{t-2} - A(0)(GF)x_{t-2} + (A^{x}(1)G - A^{x}(0)GF)\Delta x_{t-2}) + \epsilon_t \quad (6.5)$$

where $\Psi(L) = (A(L) - A^{*}(L)(F + G)L + A^{++}(L)GFL^2)$, $A^{x}(1) = A(0)^{-1}A(1)$ and $A^{x}(0) = A(0)^{-1}A^{+}(0)$. Assuming a VAR(2) system with $A(1)F = \Sigma\alpha_{\perp}(\alpha'_{\perp}\Sigma\alpha_{\perp})^{-1}\kappa'\tau'$, $F = H^{-1}M_{I(2)}H$, $M_{I(2)} = diag(1 \ldots 1, 0 \ldots 0)$, $A(0)G = \alpha\rho'\tau'$ and $(A^{x}(1)G - A^{x}(0)GF) = \alpha\psi'$, then (6.5) is a restricted version of the I(2) representation in Hansen and Johansen (1998):

$$\Delta^2 x_t = \sum \alpha_{\perp}(\alpha'_{\perp} \sum \alpha_{\perp})^{-1}\kappa'\tau'\Delta x_{t-1} + \alpha(\rho'\tau' x_{t-1} - \psi'\Delta x_{t-1}) + \epsilon_t. \quad (6.6)$$

In the notation of Hansen and Johansen, α is $n \times r$, ρ is $(r + s) \times r$, τ is $n \times (r + s)$, ψ is $n \times r$, κ is $(r + s) \times (n - r)$ and Σ is $n \times n$.

Next the approach due to Johansen (1992) is considered for testing for cointegration in I(2) systems, then an example is discussed along with identification and estimation.

6.2.1.1 The Johansen procedure for testing cointegrating rank with I(2) variables

Prior to any discussion of the appropriate method of estimation the more conventional VECM for the I(2) case is presented (Johansen 1995a):

$$\Delta^2 x_t = \alpha\beta' x_{t-1} - \Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta^2 x_{t-i} + N_0 D_t + \epsilon_t. \quad (6.7)$$

Where $\Gamma = \Sigma\alpha_{\perp}(\alpha'_{\perp}\Sigma\alpha_{\perp})^{-1}\kappa'\tau' + \alpha\psi'$, α and $\beta' = \rho'\tau'$ are the conventional loadings and cointegrating vectors for the case in which series of any order may collapse to a stationary linear combination. If $\Gamma = 0$, then this is the cointegration case considered by Engle and Granger (1987) where all the series are I(2) and:

$$\Delta^2 x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta^2 x_{t-i} + N_0 D_t + \epsilon_t. \quad (6.8)$$

Alternatively, when $\alpha\beta' = 0$ and the differenced I(1) series have linear combinations that are stationary:

$$\Delta^2 x_t = -\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1} \Psi_i \Delta^2 x_{t-i} + N_0 D_t + \epsilon_t \quad (6.9)$$

where $-\Gamma = (\alpha'_{\perp})^{-1}\kappa'\tau' = \alpha_{I(2)}\beta'_{I(2)}$ as $\alpha'_{\perp}$ has full rank, because $\alpha\beta' = 0$ implies $\alpha = 0$ and $\rho = 0$. The full I(2) case allows for the possibility of cointegration amongst I(2) series that become I(0) in combination, and cointegration amongst I(1) series that become I(0).

Clearly, (6.8) can be estimated using the Johansen procedure, except the regression that is purged of short-run behaviour in, for example the VAR(1) case is:

$$R_{0,t} = \alpha\beta' R_{1,t}$$

or

$$\Delta^2 x_t = \prod x_{t-1},$$

and decomposition and testing follows in the usual way (see sections 4.3–4.4).

Alternatively, for the VAR(1) case associated with (6.9) the estimation procedure is in every respect the same as that derived by Johansen (1991), except the data are first and second differenced. For the VAR(1) case this involves estimating the following model:

$$R_{0,t} = \alpha_{I(2)}\beta'_{I(2)}R_{1,t} = (\alpha'_{\perp})^{-1}\kappa'\tau'R_{1,t}$$

or

$$\Delta^2 x_t = -\Gamma\Delta x_{t-1}$$

This becomes more complicated when the two types of cointegration are combined, then (6.7) needs to be estimated, but this requires two blocks of reduced rank tests to be undertaken. One procedure for undertaking this analysis would be to consider the unit roots associated with cointegration amongst I(2) series whose first differences cointegrate. However: when $\alpha\beta' \neq 0$, then the model to be estimated will either require very long lags as the moving average terms $\beta' x_{t-1} = J(L)\varepsilon_{t-1}$ have been omitted or the Johansen approach might be applied to a VARMA(1,q) model. To see this re-write (6.2) as:

$$\Delta^2 x_t - F\Delta x_{t-1} = C_1(L)\epsilon_t. \tag{6.10}$$

If (6.10) were to be estimated, then the method must account for roots on the unit circle as when the level terms cointegrate, $C_1(L)$ contains further unit roots. Otherwise, the conventional VAR associated with this problem is of infinite order and not conventionally invertible. There is no unique way of deriving the estimator and in general the existence of the time series representation cannot be proven.

In general, the case with both I(2) and I(1) interdependencies can be handled by considering the solution to two reduced rank problems:

$$\prod = \alpha\beta'$$
$$\alpha'_{\perp}\Gamma\beta_{\perp} = \xi\gamma'$$

where ξ and γ are $(n - r) \times s$ dimensioned matrices. To simplify the exposition quadratic trends are not considered here. Johansen (1995) suggests the problem is made tractable by correcting the short-run behaviour firstly for the usual cointegration case as the I(2) series collapse to linear combinations that are stationary. When the Frisch–Waugh theorem is applied to purge the short-run relationship of the nuisance terms, then $\Delta^2 x_t$ and x_{t-1} are both regressed on Δx_{t-1} and Δx^2_{t-i} $i = 1, 2, \ldots, n - 1$ by ordinary least squares. The residuals from these regressions will not be correlated with the lagged second differences and the influence of the first form of cointegration will be removed. Again $R_{0,t}$ and $R_{1,t}$ are, in essence, the $n \times 1$ residual vectors from regressions with Δx_t and x_{t-1} as the dependent variables. The following regressions, yield estimates of the first long-run parameter matrix:

$$R_{0,t} = \alpha\beta' R_{1,t} = \prod R_{1,t}. \tag{6.11}$$

Now β is calculated by solving the conventional eigenvalue problem for the I(1) case and the usual I(1) analysis is undertaken to determine cointegrating

rank (section 4.4). To confirm that the I(1) analysis is valid the test for I(2) components discussed previously in 4.4.5 needs to be undertaken, this relates to the solution to the second reduced rank problem, that is rank($\alpha'_\perp\Gamma\beta_\perp$) = $n - r$. Should this matrix not have full rank, then there are I(2) components not accounted for. Next an analysis of the I(2) components of the model is undertaken, controlling for the I(1) variables. Subject to knowledge of (α, β, r) the I(1) terms are eliminated by pre-multiplying (6.7) by $\alpha'_\perp$:

$$\begin{aligned}\alpha'_\perp\Delta^2 x_t &= \alpha'_\perp\alpha\beta' x_{t-1} - \alpha'_\perp\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha'_\perp\Psi_i\Delta^2 x_{t-i} + \alpha'_\perp N_0 D_t + \alpha'_\perp\epsilon_t \\ &= -\alpha'_\perp\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha'_\perp\Psi_i\Delta^2 x_{t-i} + \alpha'_\perp N_0 D_t + \alpha'_\perp\epsilon_t. \end{aligned} \tag{6.12}$$

This is an $n - r$ dimensioned system and in the pure I(1) case $rank(\alpha'_\perp\Gamma) = n - r$. The test for further I(2) trends is undertaken by regressing $\alpha'_\perp\Delta^2 x_t$ and $\alpha'_\perp\Delta x_{t-1}$ on $\alpha'_\perp\Delta x^2_{t-i}$ $i = 1, 2, \ldots, n-1$. The residuals from the regressions of $R_{0,t}$ and $R_{1,t}$ for this case yield an eigenvalue problem that can be solved in the usual way. The Johansen test for this case determines the *rank* ($\alpha'_\perp\Gamma\beta_\perp$) = s, where $0 \le s \le n - r$ and associated with s significant eigenvalues is the $s \times n - r$ matrix of eigenvectors γ' that define common trends. If all the variables are I(1), then the system separates into r stationary variables ($\beta' x_{t-1}$) and $n - r$ common trends $\gamma'\Delta x_{t-1}$. Otherwise there are s common trends and $n - r - s$, $I(2)$ trends.

To complete the I(2) analysis, (6.7) is now multiplied by the $r \times n$ matrix $\overline{\alpha}'$:

$$\begin{aligned}\overline{\alpha}'\Delta^2 x_t &= \overline{\alpha}'\alpha\beta' x_{t-1} - \overline{\alpha}'\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\overline{\alpha}'\Psi_i\Delta^2 x_{t-i} + \overline{\alpha}' N_0 D_t + \overline{\alpha}'\epsilon_t \\ &= \beta' x_{t-1} - \overline{\alpha}'\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\overline{\alpha}'\Psi_i\Delta^2 x_{t-i} + \overline{\alpha}' F D_t + \overline{\alpha}'\epsilon_t \end{aligned} \tag{6.13}$$

where $\overline{\alpha}'\alpha = I_r$. Subtracting (6.13) from $\omega \times$ (6.12):

$$\begin{aligned}\overline{\alpha}'\Delta^2 x_t - \omega\alpha'_\perp\Delta^2 x_t = \beta' x_{t-1} - \overline{\alpha}'\Gamma\Delta x_{t-1} + \sum_{i=1}^{p-1}\overline{\alpha}'\Psi_i\Delta^2 x_{t-i} + \overline{\alpha}' N_0 D_t + \overline{\alpha}'\epsilon_t + \\ \omega(\alpha'_\perp\Gamma\Delta x_{t-1} - \sum_{i=1}^{p-1}\alpha'_\perp\Psi_i\Delta^2 x_{t-i} - \alpha'_\perp N_0 D_t - \alpha'_\perp\epsilon_t) \\ \overline{\alpha}'\Delta^2 x_t - \beta' x_{t-1} = \omega\alpha'_\perp\Delta^2 x_t - (\overline{\alpha}' - \omega\alpha'_\perp)(\Gamma\Delta x_{t-1} - \\ \sum_{i=1}^{p-1}\Psi_i\Delta^2 x_{t-i} - N_0 D_t) + (\overline{\alpha}' - \omega\alpha'_\perp)\epsilon_t \end{aligned} \tag{6.14}$$

where $\omega = \Sigma_{\alpha\alpha_\perp}\Sigma^{-1}_{\alpha_\perp\alpha_\perp}$, $\Sigma_{\alpha\alpha_\perp} = \overline{\alpha}'\Sigma\alpha_\perp$ and $\Sigma_{\alpha_\perp\alpha_\perp} = \alpha'_\perp\Sigma\alpha_\perp$ The errors of (6.12) and (6.14) are independent by construction. While the parameters of (6.12), ($\alpha'_\perp\Gamma$, $\alpha'_\perp\Psi_i\Sigma_{\alpha_\perp\alpha_\perp}$) and (6.14), ($\omega$,($\overline{\alpha}' - \omega\alpha'_\perp$)$\Gamma$,($\overline{\alpha}' - \omega\alpha'_\perp$)$\Psi_i$, ($\overline{\alpha}' - \omega\alpha'_\perp$)$N_0$) are variation free. It follows that the parameters (Γ, Ψ_i, N_0, Σ) can be disentangled from the

above reparameterization. If there are no further cross-equation restrictions on the higher-order dynamics and cointegration, then (6.12) and (6.14) can be analyzed separately, while the dependence that operates on the common trends applies to (6.12) alone.

The second reduced rank hypothesis is:

$$H_{r,s}: rank(\alpha_\perp'\Gamma\beta_\perp) = rank(\xi\gamma') = s$$

where $0 \leq s \leq n - r$. Using the identity $I = \bar{\beta}\beta' + \beta_\perp\bar{\beta}'_\perp$, the variables $\beta'\Delta x_{t-1}$ and $\bar{\beta}'\Delta x_{t-1}$ may be introduced into (6.12):

$$\begin{aligned}\alpha_\perp'\Delta^2 x_t &= -\alpha_\perp'\Gamma(\bar{\beta}\beta' + \beta_\perp\bar{\beta}_\perp')\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha_\perp'\Psi_i\Delta^2 x_{t-i} + \alpha_\perp' N_0 D_t + \alpha_\perp'\epsilon_t \\ &= -\alpha_\perp'\Gamma\bar{\beta}\beta'\Delta x_{t-1} + \alpha_\perp'\Gamma\beta_\perp\bar{\beta}_\perp'\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha_\perp'\Psi_i\Delta^2 x_{t-i} + \alpha_\perp' N_0 D_t + \alpha_\perp'\epsilon_t \qquad (6.15) \\ &= -\alpha_\perp'\Gamma\bar{\beta}\beta'\Delta x_{t-1} + \xi\gamma'\bar{\beta}_\perp'\Delta x_{t-1} + \sum_{i=1}^{p-1}\alpha_\perp'\Psi_i\Delta^2 x_{t-i} + \alpha_\perp' N_0 D_t + \alpha_\perp'\epsilon_t. \qquad (6.16)\end{aligned}$$

The parameters (Γ, Ψ_i for $i = 1, \ldots n - 1$ N_0) can be estimated by regressing $\bar{\alpha}'\Delta^2 x_t - \beta' x_{t-1}$ on $\alpha_\perp'\Delta^2 x_t$, Δx_{t-1}, $\Delta^2 x_{t-i}$ and D_t. The dependence amongst the s common trends can be determined from the regression:

$$R_{0,t} = -\xi\gamma' R_{1,t} \qquad (6.17)$$

where $R_{0,t}$ and $R_{1,t}$ are residuals based on regressing $\alpha_\perp'\Delta^2 x_t$ and $\bar{\beta}_\perp'\Delta x_t$ respectively on $\beta'\Delta x_{t-1}$, $\Delta^2 x_{t-i}$ for $i = 1, \ldots p - 1$ and D_t. The likelihood ratio test statistic is based on the solution to the eigenvalue problem $|\lambda S_{1,1} - S_{1,0}\, S^{-1}{}_{0,0} S_{0,1}| = 0$, calculated from sample product moments derived for the I(2) case using:

$$S_{i,j} = T^{-1}\sum_{i=1}^{T} R_{i,t}R_{j,t}' \text{ for } i = 0,1$$

It follows that s is selected by calculating the maximal eigenvalue test:

$$LR(s_0, s_1) = -T\left[\sum_{i=s_0+1}^{s_1}\log(1-\lambda_i)\right]$$

and for an appropriate choice of s the matrix γ' is the matrix whose columns are the eigenvectors associated with the first s significant eigenvalues.

An alternative approach is derived in Johansen (1997) and Hansen and Johansen (1998) using (6.6) where the parameters to be estimated that are variation free are (α, ρ, τ, Σ, κ, ψ).

6.2.1.2 An example of I(2)

Identification and model selection in the I(2) case is more complicated than in the I(1) case and partial consideration of the null of cointegration conditioned

on the notion that the series are all I(1) may not be valid (Paruolo 1996). When the series are I(2) they become stationary by virtue of a combination of I(1) and I(2) processes and from (6.6) the cointegrating relations have the following form:

$$\begin{aligned}\alpha(\rho'\tau'x_{t-1} - \psi'\Delta x_{t-1}) &= \alpha\rho'\tau'x_{t-1} - \alpha\psi'\Delta x_{t-1} \\ &= \alpha\beta'x_{t-1} - \alpha\psi'\Delta x_{t-1} = \alpha(\beta'x_{t-1} - \psi'\Delta x_{t-1}).\end{aligned}$$

Engle and Yoo (1989) defined cointegrating relationships of the form $\beta'x_{t-1} - \psi'\Delta x_{t-1}$ as polynomial cointegration. To observe this re-write the cointegrating vectors as a lag polynomial $\beta(L)$ in x:

$$\beta(L)'x_t = \beta'x_{t-1} - \psi'x_{t-1} + \psi'x_{t-2} = ((\beta' - \psi')I + \psi'L)x_{t-1}.$$

The cointegrating vectors reduce to linear combinations ($\beta'x_{t'}$) of x_{t-1} (Engle and Granger 1987, when either $\psi' = 0$ or $\psi' = \alpha_\perp$ and $\Gamma = \kappa'\tau'$. In general, (6.7) has r linear combination of I(2) variables that are I(0), s independent linear combinations of I(1) variables that are I(0) and $n - r - s$ variables that follow I(2) trends. If, in addition, $\Gamma = \kappa'\tau' = 0$, then $s = 0$ and there are $n - r$, I(2) trends rendered stationary by the second difference operator; the case considered by Engle and Granger (1987).

It was suggested in Hunter (1992a) that some of the series analyzed by Johansen and Juselius (1992) were I(2). In response to this suggestion Hunter and Simpson (1995) analyzed a system in which the UK inflation series enters the model in first difference form, but they based their analysis on a longer data set. Here, the extended VAR(2) model estimated by Hunter (1992a) is tested for I(2) behaviour. For this example, $n = 6$, $x_t' = [p_{0t}p_{1t}p_{2t}e_{12t}r_{1t}r_{2t}]$, the variables are described in section 4.3.1.2 and the statistics are calculated for the period 1973Q2–1987Q3. When the first reduced rank regression (6.11) is undertaken to calculate $\alpha\beta'$, the intercept is unrestricted and a trend is introduced into the model. At the second stage the trend is restricted to exclude quadratic trends. The problem is addressed firstly using the approach adopted by Paruolo (1996) and this is then compared with that described in Johansen (1995).

Paruolo (1996) derives critical values for the test of the joint hypothesis:[2]

$$H_{r,s} : rank(\Gamma) + rank(\Pi) = s + r.$$

The test statistic ($_1Q_{r,s}$) is compared with associated points on the null distribution, the comparison is made either with [*p.value*] calculated by PCGIVE 10.1 (Doornik and Hendry 2001) or 5% critical values ($c_{r,n-r-s}$ (5%)) taken from Paruolo (1996). It is suggested in Doornik and Hendry (2001) that testing is applied from the top left of the table, while Paruolo (1996) suggests progressing from the top to the bottom of each column to a point at which the null can no longer be rejected. Paruolo (1996) advises that tests are applied to the specific case, moving to the general or from the most restricted to less

Table 6.1 I(2) Cointegration tests

r	$1Q_{r,s}(Q^{??}_{r,s})$ [5% c.v.] [p.value]						Q^{*}_{r} c_{n-r}	$c_{r,n-r-s}$
n-r-s	6	5	4	3	2	1		
	314.01	254.23	199.22	163.69	141.7	126.62		
0	[194.32]	[134.54]	[79.53]	[44.0	[22.01]	[6.93]		
	240.35	203.12	174.83	148.54	126.69	109.21	119.69	93.92
	[0.0000]	[0.0000]	[0.0031]	[0.0105]	[0.0073]	[0.0028]		
		203.82	148.4	114.58	90.026	74.347		
1		[134.96]	[79.539]	[45.719]	[21.165]	[5.486]		
		171.89	142.57	117.63	97.97	81.93	68.861	68.68
		[0.0009]	[0.0429]	[0.1335]	[0.2082]	[0.1840]		
			124.56	88.233	65.029	49.417		
2			[80.184]	[43.857]	[20.653]	[5.041]		
			116.31	91.41	72.99	57.95	44.376	47.21
			[0.0226]	[0.1234]	[0.2247]	[0.2537]		
3				83.798	56.535	35.023		
				[59.868]	[32.605	[11.093]		
				70.87	51.35	38.82	23.938	29.38
				[0.0039]	[0.0176]	[0.1215]		
					48.922	27.513		
4					[35.512]	[14.103]		
					36.12	22.6	13.413	15.34
					[0.0016]	[0.0084]		
5						13.576		
						[8.392]	5.184	3.84
						12.93		
						[0.0601]		
c^{*}_{n-r-s}		75.33	53.35	35.07	20.17	9.09		

restricted cases. Following this approach, the first diagonal element implies $r = 0$, $n - r - s = 6$ and the test statistic for the case with unrestricted constant ($\mu_1 \neq 0$) is ${}_1Q_{0,0} = 314.01 > c_{0,6}$ (5%) = 240.35. Based on the calculated statistic the null hypothesis ($rank(\Gamma) = s = rank(\Pi) = r = 0$) cannot be accepted. Progressing to the next column, where $r = 0$ and $n - r - s = 5$, ${}_1Q_{0,1} = 254.23 > c_{0,5}(5\%) = 203.12$, the null is rejected, that $rank(\Gamma) = s = 1$ and $rank(\Pi) = r = 0$. At this point using Paruolo's (1996) suggestion to move down the column, $r = 1$, $n - r - s = 4$, $s = 1$, the joint test statistic ${}_1Q_{0,1} = 203.82 > c_{0,5}(5\%) = 177.89$ and the [p-value]=.0009 confirms that the null hypothesis cannot be accepted at either the 5% or the 1% level. Now the next column is considered, $r = 0$, $n - r - s = 4$, $s = 2$ and the [p-value]=0.0031 implies the null ($rank(\Gamma) = s = 2$, $rank(\Pi) = r = 0$) cannot be accepted.

Following this approach, testing stops and the correct decomposition of the long-run is detected once a null in the above table is accepted. Looking at the

[*p.values*] in the column headed $n - r - s = 4$, there is no case where the null hypothesis can be accepted. The final rejection of the null implies that there are at least $r = 2$ cointegrating vectors and $6 - r - s \leq 3$, I(1) trends. Now progression is from the top of the next column ($n - r - s = 3$) and again to a point at which the null cannot be rejected. From the size of the [*p.value*] = 0.1335, this occurs when $r = 1$, $n - r - s = 3$ and $s = 2$. The Paruolo approach implies that there are $r = 1$ stationary linear combinations (cointegrating vectors), $n - r - s = 6 - 1 - s = 3$, I(1) trends and $s = 2$, I(2) trends. Were one to follow the direction in Doornik and Hendry (2001), to progress down and to the right, then this suggests shifting to the next column at the point at which $r = 2$ and then progressing down that column.[3] The direction of Doornik and Hendry is consistent with the proposition that the first step of the Johansen I(2) estimator correctly determines the number, but not necessarily the exact nature of the cointegrating vectors.

In comparison, Johansen (1995a) suggests that the cointegrating rank calculated from the first step estimation is still reliable, which suggests testing the hypothesis associated with I(2) trends conditional on selecting a particular value for *r*. The null hypothesis that Johansen (1995a) tests is:

$$H_{r,s} \big| H(r) : rank(\xi\gamma') = s.$$

Based on the first rank test it is suggested that $r = 2$ is selected and then s is determined by moving along that row to the point at which the null cannot be rejected. The Johansen test along each row considers the specific case and moves towards the more general, but this now occurs for different values of $n - r - s$, which for fixed r imply different values of s. Given $r = 2$, the test statistic $Q_{2,s}$ is considered for $s = 0, 1, 2, 3$. Starting from the left $n - r - s = 6 - 2 - 0 = 4$, the Johansen tests statistic is $Q_{2,0} = 80.184$, which exceeds the 5% critical value ($c^{*}_{6-2-0} = 53.35$) taken from Johansen (1995a), implying that the null ($r = 2$, $s = 0$) cannot be accepted. Continuing along the row where $r = 2$, the null eventually cannot be rejected when $n - r - s = 6 - 2 - 2$ and $s = 2$ ($Q_{2,2} = 20.653 < c^{*}_{6-2-2} = 20.17$). In line with Doornik and Hendry, the Johansen testing procedure implies that there are $r = 2$ stationary linear combinations (cointegrating vectors), $n - r - s = 6 - 2 - s = 2$, I(1) trends and $s = 2$, I(2) trends.

The two test procedures advanced by Johansen (1995a) and Paruolo (1996) imply that $s = 2$, but they disagree about the number of cointegrating vectors and I(1) trends. Johansen (1995a) shows that by progressing from $s = 0, 1, 2, 3$, the $Q_{2,2}$ test has the same optimal properties in the limit as the Johansen test statistic for cointegration. Furthermore, looking at the Johansen I(2) tests presented in the table above ($Q_{r,s}$), when $r = 0, 1, 2$ the tests are not materially different whatever value $n - r - s$ is selected. Partial confirmation of the optimality of the test may be observed by comparing values of $Q_{r,s}$. For the

column headed $n - r - s = 3$, $Q_{0,3} = 44 \approx Q_{1,2} = 45.719 \approx Q_{2,1} = 43.857$ and all these values exceed the critical value ($c^*_{6-2-2} = 35.07$) at the 5% level.

Inspection of the roots of the companion matrix of the VAR is often viewed as a useful tool in determining the number of unit roots and as a result some idea of the likely number of non-stationary processes driving x_t (Johansen 1995a). The VAR(2) written as a first order model in state space from is:

$$x_t^* = \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = A_c x_{t-1}^* + \varepsilon_t^* = \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} A_1 x_{t-1} + A_2 x_{t-2} + \varepsilon_t \\ x_{t-1} \end{bmatrix}$$

or

$$\begin{bmatrix} A(L)x_t \\ x_{t-1} - x_{t-1} \end{bmatrix} = \begin{bmatrix} x_t - A_1 x_{t-1} - A_2 x_{t-2} \\ x_{t-1} - x_{t-1} \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}.$$

Dhrymes (1984) shows that the characteristic roots of the dynamic process described by the polynomial $A(L)$ can be calculated from the eigenroots of the companion matrix A_c. The eigenvalues (roots) for the VAR(2) model estimated above and for comparison a similar VAR(1) are given in Table 6.2.

The Australian exchange rate example in Johansen (1991a), summarized in Johansen (1995a), yields the clear-cut conclusion that there are three unit roots when $n - r = 5 - 2 = 3$. By contrast, the VAR(2) case considered here appears to reveal three roots close to the unit circle, a real root (.9719) and a complex conjugate pair of roots with modulus (.9001), but, according to the I(2) test produced by Johansen, $n - r = 4$. This suggests that detecting the

Table 6.2 Eigenvalues of companion matrix

VAR(2)			*VAR(1)*		
real	*imag*	*modulus*	*real*	*imag*	*modulus*
–0.01897	0.3874	0.3879			
–0.01897	–0.3874	0.3879			
0.1327	0.0000	0.1327			
0.4550	0.3193	0.5559			
0.4550	–0.3193	0.5559			
0.9719	0.0000	0.9719	0.9574	0.0000	0.9574
0.8877	0.1486	0.9001	0.9222	0.1115	0.9289
0.8877	–0.1486	0.9001	0.9222	–0.1115	0.9289
0.6553	0.2302	0.6946	0.6587	0.2145	0.6927
0.6553	–0.2302	0.6946	0.6587	–0.2145	0.6927
0.4910	0.0000	0.4910	0.9252	0.0000	0.9252
0.7729	0.0000	0.7729			

number of unit roots from the companion matrix is not always straightforward. Firstly, a VAR(2) system can be decomposed into two stationary processes ($r = 2$), two non-stationary processes (either $n - 2 - s = 2$ or $s = 2$) and a pair of common I(2) or I(1) trends driven by a single unit root. Secondly, should the roots of the VAR(1) be considered for comparison, then the estimates are quite consistent with the proposition that there are $n - r = 4$ unit roots. Analysis associated with both sets of eigenvalues for the two companion matrices does not appear to support the approach due to Paruolo (1996), which suggests $r = 1$ and $n - r = 4$.

Having found that some of the series are I(2), the usual cointegrating vectors may not be valid as the stationary linear combinations may require combinations of I(2) processes that are I(1) to make them stationary or polynomial cointegration. Consider these following suggestions for the long-run relationships associated with the VAR(2) system developed above. Based on the findings in Hunter (1992a) and Johansen and Juselius (1992), there are two cointegrating vectors that accept PPP and UIRP restrictions. The conclusion of the I(2) analysis for PPP is that the series may only be rendered stationary when the cointegrating vector is augmented by differences in I(2) variables. For example, relative movements in the cross-country inflation rates may be what is required. With $s = 2$ common I(2) trends driving the price series ($p_0 p_1 p_2$) then the cointegrating vectors could take the following form:

$$\beta' x_{t-1} - \psi' \Delta x_{t-1} = \left(\begin{bmatrix} 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ e_{12} \\ r_1 \\ r_2 \end{bmatrix}_{t-1} - \begin{bmatrix} 0 & 0 & \psi_{31} & -\psi_{31} & 0 & 0 \\ \psi_{12} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Delta \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ e_{12} \\ r_1 \\ r_2 \end{bmatrix}_{t-1} \right).$$

A similar type of long run occurs with polynomial cointegration (Engle and Yoo 1991; Gregoir and Laroque 1993):

$$\beta' x_{t-1} - \psi' \Delta x_{t-1} = \begin{bmatrix} 0 & 1 & \psi_{31} - 1 - \psi_{31} L & -1 - \psi_{31} + \psi_{31} L & \beta_{51} & \beta_{61} \\ (\beta_{12} - \psi_{12} + \psi_{12} L) & 0 & 0 & 0 & 1 & -1 \end{bmatrix} x_{t-1}$$

where $x'_t = [p_{0t}\, p_{1t}\, p_{2t}\, e_{12t}\, r_{1t}\, r_{2t}]$. The two forms of I(2) cointegration are equivalent when $\beta_{51} = 0$, $\beta_{61} = 0$ and $\beta_{12} = 0$. Unfortunately, prior to any evaluation

of the long run, the system needs to be identified, but identification of the type discussed in chapter 4 is considerably more complicated in the I(2) case as three sets of matrices lack identification:[121]

$$\begin{aligned}\alpha\rho'\tau' &= \alpha\varsigma\varsigma^{-1}\rho'\theta'\theta'^{-1}\tau' = \alpha^{*}\rho^{+'}\tau^{*'}\\ \alpha\psi' &= \alpha\varsigma\varsigma^{-1}\psi' = \alpha^{*}\psi^{*'}\\ \alpha_{\perp}(\alpha_{\perp}'\Sigma\alpha_{\perp})^{-1}\kappa'\tau' &= \alpha_{\perp}\upsilon^{-1}(\alpha_{\perp}'\Sigma\alpha_{\perp})^{-1}\upsilon\kappa'\theta'\theta'^{-1}\tau'\\ &= \alpha_{\perp}^{*}(\alpha_{\perp}'\Sigma\alpha_{\perp})^{-1}\kappa^{+'}\tau^{*'}\end{aligned}$$

Hence, the same likelihood can be defined for (6.6) using parameters [α, ρ', τ', κ', ψ', $\alpha_{\perp}'$] and [α^{*}, $\rho^{+'}$, $\tau^{*'}$, $\kappa^{+'}$, $\psi^{*'}$, $\alpha_{\perp}^{*'}$]. The two sets of parameterizations are observationally equivalent and observational equivalence leads to a fundamental loss of identification.

Although inflation seemed to be I(1) in the late 1980s and early 1990s the argument appears less compelling in a world where inflation is predominantly under control, which suggests that economic and financial time series might be better described as long-memory.

6.2.2 Fractional cointegration

The notion of fractional differenced series was introduced in chapter 2. When such processes are considered then the possibility of fractional cointegration ought to be entertained. Robinson and Yajima (2002) explain that this notion of fractional cointegration is quite consistent with the original definition of cointegration due to Engle and Granger (1987). Consider a pair of series x_{1t} and x_{2t} that require fractional differencing for them to be rendered stationary, then:

$$\Delta^{d}x_{it} = (1-L)^{d}x_{it} \sim I(0) \text{ for } i = 1,\ 2$$

where $(1-L)^{d} = \sum_{j=0}^{\infty}\frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}$. For $a > 0$, $\Gamma(a) = \int_{0}^{\infty} z^{l-1}e^{-z}dz$ and $a = -l$, $l = 0$, 1, ..., $\Gamma(a)$ has simple poles with residues $\frac{-1^{l}}{l}$, otherwise $\Gamma(a) = \Gamma(a+1)/a$. It follows that x_t is cointegrated when:

$$\beta'x_t = J(L)\varepsilon_t \sim C(i,d)$$

$$\text{when } x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}.$$

Proofs exist for the analysis of stationary fractional series with $-.5 < d < .5$ (Robinson and Yajima 2002). The conventional question arises over the rank of the matrix of cointegrating vectors, $rank(\beta) = r$. Do there exist r linear combinations of variables x_t that require the fractional difference operator $(1-L)^{d}$ to be applied for the series to be I(0). Robinson (1994) explains how to use non-parametric estimates of the dynamic process to calculate the cointegrating relationships when series have the same order of integration. Robinson

and Marinucci (1998) apply this approach to stationary fractionally integrated series to estimate the long-run parameters from the equation:

$$\left[1-\beta_1\right]\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = J(L)\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

The estimator is similar to that used by Phillips and Hansen (1990) to estimate long-run parameters when the series are I(1). The unknown moving average parameters in $J(L)$ are captured by a frequency domain estimator, which also appears to compare well with Phillips and Hansen (1990) when the series are I(1) (Marinucci and Robinson 2001). Although there is evidence that this type of approach is able to estimate long-run parameters when r is known or not large, the method, though efficient in calculating well-known long-run relationships, does not provide a formal test of the proposition that either fractional or integer integrated series are cointegrated. The method can determine the extent to which the variables in the regression are related by determining whether β_1 is significant or not. Clearly, any such conclusion is conditional on the appropriateness of this normalization.

Robinson and Yajima have attempted to determine the order of integration and cointegration by two different methods. They consider three different crude oil prices (WTI, Dubai and Brent). Based on an Augmented Dickey–Fuller test with an intercept, the three series are found to be stationary at the 5% level of significance. But when the order of difference is assumed to be fractional, the estimates of d for the three series are [.5336, .4367, .4538].[5]

Robinson and Yajima (2002) suggest two approaches to the problem of selecting the cointegrating rank, but they use one of them in their example. Consider the Vector Auto-Regressive Fractionally Integrated Moving Average (VARFIMA) model:

$$E(L)x_t = C(L)\varepsilon_t$$

where $E(L) = diag[(1-L)^{d_1}, (1-L)^{d_2} \ldots (1-L)^{d_n}]$.[6] The series are ordered on the basis of the prior estimate of the difference order. The test is based as is usually the case on the rank of the matrix $C(1)$, which, under conventional cointegration, has rank $n - r$ associated with the extent to which there is any over-differencing. The test, as is the case with integer cointegration, progresses from the most restricted model, where $C(1)$ has full rank, $n - r = n$ and $r = 0$, there is no cointegration to the cointegration cases, $r = 1, 2, 3$. The test for fractional cointegration is:

$$H_i : rank(G) = rank(C(1)) = n - r$$
$$where\ G = \frac{1}{2\pi}C(1)C(1)'.$$

To make the test operational, Robinson and Yajima use the following non-parametric estimator of G:

$$\hat{G} = \frac{1}{m_1}\sum_{j=1}^{m_1} \mathrm{Rc}\left\{\hat{\Lambda}(\lambda_j)^{-1} I_j \hat{\Lambda}(\lambda_j)^{-1}\right\}.$$

Where $I_j = \omega(\lambda_j)\omega(\lambda_j)'$, $\omega(\lambda_j) = (\omega_1(\lambda_j)\omega_2(\lambda_j) \ldots \omega_n(\lambda_j))'$, $Re\{\cdot\}$ is the real component, $\hat{\Lambda}(\lambda_j) = diag(e^{\frac{i\pi d_\cdot}{2}}\lambda_j^{-\bar{d}_\cdot}, \ldots e^{\frac{i\pi d_\cdot}{2}}\lambda_j^{-\bar{d}_\cdot}), \lambda_j = \frac{2\pi j}{T}$, and $m < \frac{T}{2}$. It has been assumed that d_a is replaced by a pooled estimate $\bar{d}_\cdot = (\hat{d}_1 + \hat{d}_2 + \hat{d}_3)/3$ and $\omega_a(\lambda_j) = \frac{1}{\sqrt{2\pi T}}\Sigma_{t=1}^{T} x_{at}e^{it\lambda_j}$ is the discrete Fourier transform of the original data. The effective bandwidth m_1 is set to increase at a faster rate than m to counteract the effect of using an estimate of d_a Robinson and Yajima (2002) provide estimates of G evaluated with $m = 13$ and $m_1 = 15$:

$$\hat{G} = \begin{bmatrix} .00493 & .00542 & .00575 \\ .00542 & .00625 & .00653 \\ .00575 & .00653 & .0073 \end{bmatrix},$$

where $\hat{G}$ has the following eigenvalues [.01807, .000275, .000124]. The most important eigenvector is associated with the largest root, which given that the other two roots are small suggests that $n - r = 1$ or with $n = 3$ variables then there are $r = 2$ cointegrating relationships. Robinson and Yajima (2002) proceed to analyze the case where the three series have two distinct orders of differencing. This suggests that the WTI oil price series is handled differently than that for Brent and Dubai. Once Brent and Dubai crude prices are considered together with two types of difference, the reduced rank calculation is applied to a 2×2 sub-matrix, which from the obvious rank deficiency in $\hat{G}$ above implies $r = 1$.

6.3 Forecasting in cointegrated systems

6.3.1 VMA analysis

Cointegration describes how, in the long run, the levels of a set of variables should move together. A similar property should therefore be expected of forecasts from such a system. That is, the forecasts of a set of variables from a cointegrated system should be related to one another such that, although individually subject to the implications of non-stationarity, there remain linear combinations of the forecasts that are zero, or constant (depending on the deterministic terms in the model). If valid long-run relationships are imposed on an empirical model of the data, this ought to improve the quality of long-run forecasts, as additional information is being exploited. But is the value of the long-run restrictions, in terms of forecast improvement, greater than for other types of restriction, or restrictions on stationary systems? Engle and Yoo (1987) provide an analysis of this problem in the CI(1, 1) case.

Consider the usual VMA representation of an $n \times 1$, CI(1, 1) system considered in section 4.2:

$$\Delta x_t = C(L)\epsilon_t, \tag{6.18}$$

where $C(L) = \sum_{i=0}^{\infty} C_i L^i$, *rank* $(C(1)) = n - r$, and $C_0 = I_n$. In order to obtain an expression for x_t, which is to be the object of the forecast, sum both sides of (6.18) from $i = 1, \ldots, t$ to give

$$x_t - x_0 = \sum_{i=1}^{t} C(L)\epsilon_i.$$

In addition, assume initial values x_0 and ϵ_q, $q = 0$ are zero. Then,

$$C(L)\epsilon_i = \sum_{j=0}^{\infty} C_j \epsilon_{i-j} = \sum_{j=0}^{i-1} C_j \epsilon_{i-j}$$

and so

$$x_t = \sum_{i=1}^{t} \sum_{j=0}^{i-1} C_j \epsilon_{i-j}. \tag{6.19}$$

Equation (6.19) can be rewritten in terms of ϵ_s, $S = 1, \ldots, t$

$$x_t = \sum_{s=1}^{t} \sum_{r=0}^{t-s} C_r \epsilon_s.$$

Moving forward another h periods,

$$x_{t+h} = \sum_{s=1}^{t+h} \sum_{r=0}^{t+h-s} C_r \epsilon_s = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s + \sum_{s=t+1}^{t+h} \sum_{r=0}^{t+h-s} C_r \epsilon_s$$

and redefining the index on the last summation to emphasize that it contains terms in the disturbances beyond t only, gives

$$x_{t+h} = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s + \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q}. \tag{6.20}$$

Equation (6.20) expresses x_{t+h} as the sum of two terms that partition the disturbances between those occurring up to and including time t, and later values.

The forecast of x_{t+h} based on information available at time t is the expected value of x_{t+h} given the information, and is denoted $x_{t+h|t}$. In this context, h is known as the forecast horizon and t is called the forecast origin. Using the fact that the conditional expectation of a future disturbance term is zero, and the conditional expectation of any current or past value is the expectation of a realized value, from (6.20),

$$x_{t+h|t} = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s. \tag{6.21}$$

This does not yet establish that the forecasts are linearly related. The requirement for this is for there to exist a linear combination of the forecasts that is zero (in the absence of deterministic terms). That is, there must exist an $n \times 1$ vector γ such that $\gamma' x_{t+h|t} = 0$. From (6.21), a sufficient condition for this is that

$$\gamma' \sum_{r=0}^{t+h-s} C_r = 0.$$

But this does not follow from the properties of the VMA, as it requires each of $\sum_{r=0}^{t+h-s} C_r$, $s = 1, \ldots, t$ to be of reduced rank and to have the same null space. However, cointegration is a long-run property and its implications can only be expected to follow in the long run. In a forecasting context, this means that any special properties of the forecast arising from cointegration can only be expected to become apparent as the forecast horizon, h, becomes large. So consider the limit of $\sum_{r=0}^{t+h-s} C_r$, as $h \to \infty$:

$$Lim_{h\to\infty} \sum_{r=0}^{t+h-s} C_r = \sum_{r=0}^{\infty} C_r = C(1), \tag{6.22}$$

and define what can be called the long-run forecast, $x_{\infty|t}$, as:

$$x_{\infty|t} = Lim_{h\to\infty}\left[x_{t+h|t}\right]. \tag{6.23}$$

Then, from (6.21) and (6.22), $x_{\infty|t}$, is given by

$$\begin{aligned} x_{\infty|t} = Lim_{h\to\infty}\left[x_{t+h|t}\right] &= Lim_{h\to\infty}\left[\sum_{s=1}^{t}\sum_{r=0}^{t+h-s} C_r \epsilon_s\right] \\ &= \sum_{s=1}^{t}\left[Lim_{h\to\infty}\sum_{r=0}^{t+h-s} C_r\right]\epsilon_s = C(1)\sum_{s=1}^{t}\epsilon_s. \end{aligned}$$

The long-run forecast therefore follows a linear combination of the realized value of a vector stochastic trend. But *rank* $(C(1)) = n - r$, and so there exist r linearly independent vectors, that is the cointegrating vectors, β, such that $\beta' C(1) = 0$. Therefore:

$$\beta' x_{\infty|t} = \beta' C(1)\sum_{s=1}^{t}\epsilon_s = 0. \tag{6.24}$$

The extent to which fixed horizon forecasts approximate to (6.24) depends how quickly the matrix coefficients $C_i, i = 0, 1 \ldots$, decay. From (6.21)

$$x_{t+h|t} = \sum_{s=1}^{t}\left[C(1) - \sum_{r=t+h-s+1}^{\infty} C_r\right]\epsilon_s = x_{\infty|t} - \sum_{s=1}^{t}\sum_{r=t+h-s+1}^{\infty} C_r \epsilon_s$$

and so

$$\beta' x_{t+h|t} = -\beta' \sum_{s=1}^{t} \sum_{r=t+h-s+1}^{\infty} C_r \epsilon_s.$$

Thus the smallest index on the C_r is $r = h + 1$, indicating that, assuming the C_r do decay with r, the greater is the forecast horizon, the smaller will be the deviation of the forecasts from their long-run relationship. Thus, empirically, the evidence for cointegration restrictions improving forecasts should be weaker for short horizons, than longer ones. The more rapidly the coefficients decay, the fewer steps ahead the forecasts need to be before they display a functional relationship similar to the cointegrating relations.

Turning to the h-step ahead forecast error, denoted $e_{t+h|t}$, and its variance, from (6.20) and (6.21), this error is

$$e_{t+h|t} = \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s + \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q} - \sum_{s=1}^{t} \sum_{r=0}^{t+h-s} C_r \epsilon_s = \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q} \tag{6.25}$$

and, since the disturbances are not autocorrelated

$$\text{var } (e_{t+h|t}) = \sum_{q=1}^{h} \left[\left(\sum_{r=0}^{h-q} C_r \right) \Omega \left(\sum_{r=0}^{h-q} C_r' \right) \right],$$

where $\Omega = E(\epsilon_t \epsilon_t')$, for all t. That is, the forecast error variance grows with h. Interestingly, it is also the case that the forecast errors are cointegrated, with precisely the same time series structure as the original process, x_t, under the condition that all forecasts are made using the same information, that available at time t. To see this use (6.25) to construct the forecast error difference process

$$\begin{aligned}
\Delta e_{t+h|t} &= e_{t+h|t} - e_{t+h-1|t} \\
&= \sum_{q=1}^{h} \sum_{r=0}^{h-q} C_r \epsilon_{t+q} - \sum_{q=1}^{h-1} \sum_{r=0}^{h-1-q} C_r \epsilon_{t+q} \\
&= C_0 \epsilon_{t+h} + \sum_{q=1}^{h-1} \left(\sum_{r=0}^{h-q} C_r - \sum_{r=0}^{h-q-1} C_r \right) \epsilon_{t+q} \\
&= C_0 \epsilon_{t+h} + \sum_{q=1}^{h-1} C_{h-q} \epsilon_{t+q} = \sum_{q=1}^{h} C_{h-q} \epsilon_{t+q} = \sum_{q=1}^{h} C_{h-q} \epsilon_{t+h-(h-q)} \\
&= \sum_{k=0}^{h-1} C_k \epsilon_{t+h-k} = C(L) \epsilon_{t+h}, \quad \epsilon_q = 0, \ q \le t,
\end{aligned}$$

where the initial values are now relative to the forecast origin, and consistent with the original VMA, have been set to zero. Thus

$$\Delta e_{t+h|t} = C(L) \epsilon_{t+h}$$

and hence, from the original VMA, all h-step ahead forecast errors are cointegrated of order (1,1). That is, the difference between the h-step ahead and the $h-1$-step ahead forecast errors, both made conditional on information available at time t, is stationary, but the sequence of h-step ahead forecast errors, for $h = 1,2, \ldots$, is I(1).

An intuition for the non-stationarity of the forecast error can be provided by expressing a future value of the process as a sum of the forecast and the forecast error,

$$x_{t+h} = x_{t+h|t} + e_{t+h|t}. \tag{6.26}$$

Since, $x_{t+h|t}$ depends only on realized values (the disturbance values at time t and before), it is non-stochastic. Thus the stochastic non-stationarity properties of x_{t+h} and $e_{t+h|t}$ must be the same, so they must both be integrated of order 1. Applying the initial value condition $\epsilon_q = 0$, $q \leq t$, equation (6.26) gives $x_{t+h|t} = x_{\infty|t}$ and hence:

$$x_{t+h} = x_{\infty|t} + e_{t+h|t},$$

from which, pre-multiplication by the cointegrating vector gives

$$\beta' x_{t+h} = \beta' x_{\infty|t} + \beta' e_{t+h|t} = \beta' e_{t+h|t}. \tag{6.27}$$

The left-hand side of (6.27) is I(0) from the VMA, and therefore so is $\beta' e_{t+h|t}$, hence $e_{t+h|t}$ is CI(1,1).

6.3.2 Forecasting from the VAR

The property that the long-run forecasts should be linearly constrained can also be obtained from a VAR. Again, let x_t be an $n \times 1$ CI(1,1) vector, this time having the VAR(p) structure

$$x_t = \sum_{i=1}^{p} A_i x_{t-i} + \epsilon_t. \tag{6.28}$$

Reparameterize this in the usual way as the VECM

$$\Delta x_t = \Pi x_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t \tag{6.29}$$

where, again $\Pi = \alpha\beta'$ with α and β dimensioned $n \times r$. Following Lin and Tsay (1996), in order to understand how the forecasts from (6.28) have the same long-run properties as the series themselves, note that Δx_t is I(0), and that forecasts of a stationary series converge to the expected value of the process as the forecast horizon tends to infinity. That is

$$Lim_{h\to\infty} \Delta x_{t+h|t} = \mu_{\Delta x} \tag{6.30}$$

where $\mu_{\Delta x} = E(\Delta x_t)$. The properties of the forecasts of the difference process are used to obtain those of the levels via the VECM. Using (6.29), the h-step ahead forecast equation for the difference process is

$$\Delta x_{t+h|t} = \Pi x_{t+h-1|t} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t+h-i|t}. \tag{6.31}$$

In order to derive the properties of the long-run forecasts, take the limit of (6.31) as $h \to \infty$, and substitute from (6.30) to give

$$\mu_{\Delta x} = \Pi\left[Lim_{h\to\infty} x_{t+h-1|t}\right] - \sum_{i=1}^{p-1} \Gamma_i \mu_{\Delta x}.$$

Rearranging, and using the notation of (6.23) for the long-run forecast of the level,

$$\Pi x_{\infty|t} = \left[I_n + \sum_{i=1}^{p-1} \Gamma_i\right] \mu_{\Delta x}. \tag{6.32}$$

The right-hand side of (6.32) is a constant matrix, and so shows that the long-run forecasts, $x_{\infty|t}$, are tied together. The analysis can be taken further to complete the analogy with equation (6.24) for the VMA case. Pre-multiplying (6.32) by α' and replacing Π by $\alpha\beta'$ gives

$$(\alpha'\alpha)\beta' x_{\infty|t} = \alpha'\left[I_n + \sum_{i=1}^{p-1} \Gamma_i\right] \mu_{\Delta x}$$

where $(\alpha'\alpha)$ is non-singular, so that

$$\beta' x_{\infty|t} = (\alpha'\alpha)^{-1}\left[I_n + \sum_{i=1}^{p-1} \Gamma_i\right] \mu_{\Delta x}.$$

This is directly comparable with (6.24) (except that in 6.24 initial values have been set to zero), and shows that each cointegrating vector constitutes a constraint on the long run forecasts.

6.3.3 The mechanics of forecasting from a VECM

In order to benefit from any perceived advantages to forecasting from cointegrated models, it is necessary to impose the cointegrating relationships. In the VAR setting, this may be undertaken as follows.

For given β and by implication, known cointegrating rank, r, construct cointegrating combinations $\eta_t = \beta' x_t$, and estimate the VECM, conditional on r, as

$$\Delta x_t = \alpha \eta_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t.$$

Estimation may be performed by OLS, to give

$$\Delta x_t = \hat{\alpha}\eta_{t-1} + \sum_{i=1}^{p-1} \hat{\Gamma}_i \Delta x_{t-i} + e_t = \hat{\Pi} x_{t-1} + \sum_{i=1}^{p-1} \hat{\Gamma}_i \Delta x_{t-i} + e_t,$$

where $\hat{\Pi} = \hat{\alpha}\beta'$. Now, rearrange the VECM as the VAR

$$x_t = \sum_{i=1}^{p} \hat{A}_i x_{t-i} + e_t,$$

$$\hat{A}_1 = I_n + \hat{\Pi} - \hat{\Gamma}_1, \quad \hat{A}_i = \hat{\Gamma}_i - \hat{\Gamma}_{i-1} \text{ and } \hat{A}_p = \hat{\Gamma}_{p-1}.$$

The *h*-step ahead forecasts can then be produced recursively using

$$x_{t+h|t} = \sum_{i=1}^{p} \hat{A}_i \Delta x_{t+h-i|h}, \tag{6.33}$$

where $x_{t+h-i|t} = x_{t+h-i}$ for $h \leq i$. If r and β are unknown, they may be replaced by values $\hat{r}$ and $\hat{\beta}$ estimated using the Johansen procedure. This is the approach used by Lin and Tsay (1996).

The order of the forecasting VAR in (6.33), and that used for the Johansen pre-whitening, should be the same, determined, for example, using an information criterion, such as the Schwarz (SIC) (see Reimers 1992; Lütkepohl 1991). Otherwise, as was explained in section 4.3.3, programs such as PCGIVE provide systems and single equation diagnostic test for each equation in the VAR (Doornik and Hendry 2001).

The details of information criteria vary according to the weight put on additional parameters, but they are generally of the form

$$IC = \ln\left|\frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t' \hat{\epsilon}_t\right| + mf(T), \tag{6.34}$$

where $f(T)$ is an increasing function of T, $m = pn^2$, the number of estimated coefficients in an unrestricted VAR, and $\hat{\epsilon}_t$ the vector of VAR residuals. A criterion which often preferred is the SIC, for which $f(T) = \frac{\ln(T)}{T}$. Amongst the criteria most commonly used, this penalizes additional parameters (increasing VAR order) the most heavily, leading to relatively parsimonious models. The favoured model is that for which the information criterion value is minimized. When used in this way, the SIC provides consistent model selection in the sense that, as the sample size tends to infinity, it will select the correct model order with probability tending to one.

6.3.4 Forecast performance

The imposition of cointegrating restrictions on a model of I(1) series should lead to forecast improvements for two reasons. Firstly, valid long-run relationships should improve the accuracy of long-run forecasts by exploiting information about the interrelatedness of the series. Secondly, fewer parameters are estimated. In the unrestricted VECM, Π has n^2 elements, whereas when restricted, it has $2nr$. However, a number of practical issues arise:

(i) How useful is the long-run information in providing long but finite time horizon forecasts?

(ii) How are short-run forecasts affected?
(iii) What are the costs of mistakenly identifying series as I(1) when they are really I(0)?
(iv) What is the cost of incorrectly estimating r?
(v) What is the cost of imposing invalid long-run restrictions (getting the cointegrating vectors wrong)?

These issues are discussed by Clements and Hendry (1995, 1998), Lin and Tsay (1996) and Engle and Yoo (1987), among others. The three studies report Monte Carlo results; their findings are summarized below.

6.3.4.1 Engle and Yoo

These authors consider a bivariate model (representable as a first-order VAR) and discuss two types of forecast that can be made from it, one ignoring any long-run restrictions, and one imposing them. These forecasts are based on an unrestricted VAR (UVAR) and the Engle and Granger two-step methodology (EG) respectively. In the latter case, at each replication, a preliminary static regression is used to estimate the cointegrating relations and the lagged residuals from this model being included as the lagged levels term in a dynamic ECM.[7] The putative long relations are not subject to prior testing for cointegration.

The sample size is 100 and the forecast horizon from 1 to 20, so that in this case, a long-run forecast is being defined as one with a horizon 20 per cent beyond the sample, if not less. The finding is that, in terms of the mean square forecast error as measured by the trace of the sample covariance matrix of the forecast errors (see section 6.3.4.4 for more detail on forecast evaluation), the unrestricted VAR provides superior forecasts up to and including the 5-step ahead forecast (5 per cent of sample size), thereafter, the imposition of estimated long run restrictions improve the forecast monotonically, to an advantage of 40 per cent over the unrestricted forecast at 20 steps ahead. This is, of course, against a background of worsening forecast performance as forecast horizon increases.

6.3.4.2 Clements and Hendry

In their book and earlier paper, Clements and Hendry (1998, 1995) generalize the study of Engle and Yoo. They present the results of a bivariate VAR(1) system estimated on 100 observations, but for a wider range of parameter values and models. In addition to UVAR and EG, they consider the Johansen maximum likelihood estimator (ML) and a misspecified model in differences alone (DV), the lagged levels term being excluded. The DV model can be used to forecast the level of the process by adding successive forecasts of the differences to the known value of the level at the forecast origin. They also

introduce another issue, which is the form of the process used to compare forecasts: the levels, the differences, or the stationary combinations. The last of these representations is obtained by transforming the model to one in terms of the cointegrating combinations and the differenced common trends. Thus, the number of processes is unaltered, and their integration and cointegration properties preserved. Their notation for the I(0) variables is w_t where $w'_t = (x'_t\beta \quad \beta'_\perp \Delta x_t)$. Consider the partition $\alpha' = (\alpha'_a \quad \alpha'_b)$ with α_a dimensioned $r \times r$ and α_b dimensioned $r \times (n - r)$ and defining

$$J' = (0 \quad I_{n-r}) \text{and} Q = (\beta \quad J)'$$

the representation is

$$w_t = Gw_{t-1} + \nu_t \tag{6.35}$$

$$\text{where } G = \begin{pmatrix} (I_r + \beta'\alpha) & 0 \\ \alpha_b & 0 \end{pmatrix} \text{and } \nu_t = \begin{pmatrix} \beta' \\ J' \end{pmatrix} \epsilon_t.$$

Clements and Hendry produce forecasts of x_t and Δx_t using each of the four estimation methods, UVAR, ML, EG, and DV. These primary forecasts are transformed to produce forecasts of each of x_t, Δx_t and w_t. That is, each forecast is one of x_t or Δx_t, initially, but all are transformed (as necessary) into x_t, Δx_t and w_t. The purpose of the exercise is to emphasize that the superiority of one forecast method over another depends not only on what model is used to produce the forecast, but also on what properties of the forecast are being compared.

In particular, in comparing EG and UVAR to forecast x_t, the level of the process, the importance of the imposition of a valid long-run restriction is examined. But the question then arises as to whether it matters that the restriction is specifically a long-run restriction. In other words, are the advantages available from the imposition of correct restrictions markedly different in a non-stationary cointegrated environment compared to a stationary one? The way to get at this issue is to transform the forecasts to stationarity before comparing them, effectively filtering out long-run variation. The appropriate transformation is that of equation (6.35), applied to the forecasts. This procedure is only available in the context of simulations (using parameter values from the DGP), since the UVAR, by its very nature, brings with it no estimation of the cointegrating combinations. It is still the case that the forecasts differ in the method of their production, but are now being compared on a more appropriately matched basis – that is, in stationary terms. If relative forecasting performance is different in stationary space, then it suggests that the long-run nature of the restrictions is relevant in determining forecast behaviour.

If it is the long run nature of the restrictions that improve the long-run forecasts, then direct comparisons of the forecasts of the level of the process

where the restrictions are, and are not imposed, should favour the forecasts made subject to the restrictions. However, if the long-run components are removed prior to comparison, these transformed forecasts should not differ significantly. Equation (6.35) is a very useful device for decomposing the causes of relative forecast behaviour.

In their simplest case (among 13 parameterizations), Clements and Hendry generate data according to a bivariate VECM model with a single lag,

$$\Delta x_t = \Pi x_{t-1} + \epsilon_t.$$

Forecast comparisons are made in a number of ways, the simplest of which is based on the trace of the estimated variance–covariance matrix of the forecast errors (see section 6.3.4.4 for more detail on forecast evaluation). One parameterization is very similar to that used by Engle and Yoo, and therefore comparable with the earlier results. It is shown that, at longer forecast horizons, material improvement in the levels forecast are available by imposing cointegrating restrictions. That is, EG and ML are superior in levels forecasting to UVAR when the forecast horizon is relatively long. In addition, the superiority is more marked with smaller sample sizes due to the enhanced role of the degrees of freedom saved by imposing the restrictions.

When the forecasts are transformed to stationarity (using equation (6.35)) and compared again, UVAR is no longer inferior. This suggests that the gains in forecast performance from the imposition of the restrictions are due to their long-run characteristics, as no further restrictions have been imposed. In contrast to these findings, the misspecified DV model performs only slightly worse than EG and ML (and therefore better than UVAR) in levels forecasts at longer forecast horizons, but notably under-performs the other three when the forecasts are compared in stationary space.

These findings must be interpreted with care because, in practice, VAR order and cointegration rank are decided from the data. In addition, systems will normally consist of more than two variables. Clements and Hendry summarize the results of their more widely parameterized study using response surfaces, presenting their conclusions with a number of warnings about the additional complexities that enter in the practical forecasting setting. The results represent a benchmark case only.

6.3.4.3 Lin and Tsay

Lin and Tsay (1996) generalize the model for forecast performance comparisons to one involving four variables. Their Monte Carlo study is necessarily restricted in terms of the parameter values used, but the DGPs used are chosen to mimic observed data characteristics, so in this sense are calibrated so as to apply to a relevant parameter space. The structures used have the following characteristics.

(i) All systems are second order (VAR(2)).
(ii) Five DGPs are considered in all, being respectively, from model 1 to model 5, strongly stationary, but with two roots close to the unit circle, stationary with two roots very close to the unit circle, non-stationary system with cointegrating rank 2, non-stationary and non-cointegrating.[8] Of these, the stationary and unit root non-cointegrated cases are diagonal.
(iii) The in-sample period consists of 400 observations, with 100 additional out-of-sample data points generated for forecasting comparison. Forecast horizons of 1 to 60 are used. Each replication gives rise to a set of forecasts at each forecast horizon.
(iv) All models are estimated as ECMs with cointegrating rank $r = 0, 1, 2, 3, 4$ using Johansen's (1988, 1991, 1995) approach, and then recast as VECMs for the purpose of forecasting the levels.
(v) The forecasting metric, $E(L)$, where L is the forecast horizon (see equation 6.36), is based on the trace of the estimated variance–covariance matrix of the forecast errors. Each replication gives rise to an estimated variance–covariance matrix of forecast errors, and these are then averaged across replications. The larger is the statistic, the poorer the forecast.

The results of these exercises are presented in Figure 6.1.

Lin and Tsay gather their conclusions on these results into the following principal points:

(i) When the system is stationary the long-run forecasts approach a constant quickly as the forecast horizon increases. (The size of the forecast errors, in terms of their variance is also relatively small.)
(ii) If the system is stationary, then under-specifying the rank of the long-run matrix leads to under-performance. That is, imposing long-run restrictions that do not exist in practice (which are not valid) damages long-run forecast performance. The more of these there are, the worse the performance of the forecasts.
(iii) Unless the system is very close to non-stationarity (the near non-stationary DGP is model 3), correct specification of the cointegrating rank is best.
(iv) Under specification of the cointegrating rank is not serious if the processes concerned are non-stationary. This should be contrasted with the stationary case, where, although cointegration is not defined, the rank of the long-run matrix still is, and where this is under-specified, there is a deterioration in forecast performance.

Clearly, non-stationary and near non-stationary systems are harder to forecast than stationary ones. As a matter of design, it should be noted that while

Figure 6.1a

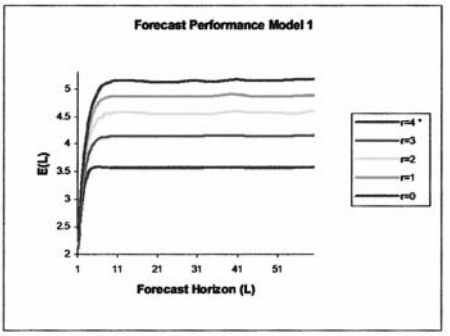


Figure 6.1b

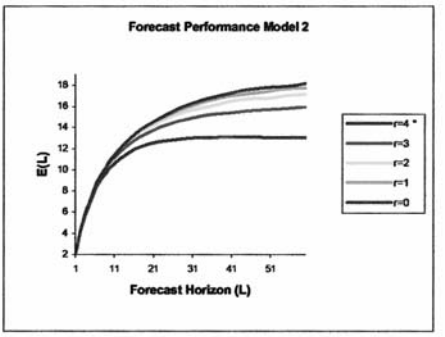


Figure 6.1c

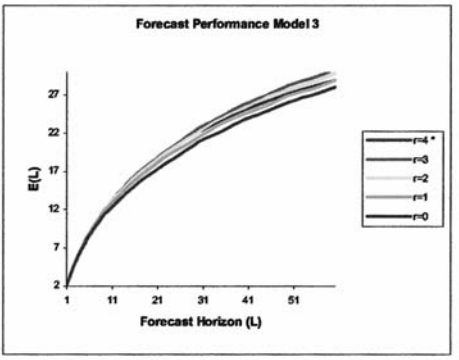


Figure 6.1d

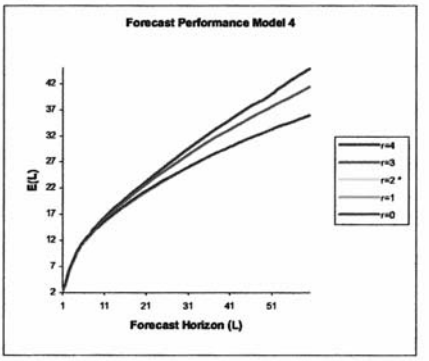


Figure 6.1e

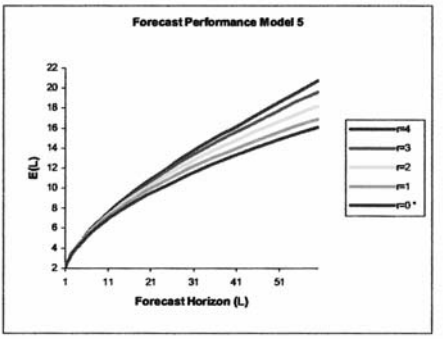


Figure 6.1 Forecasting performance from Lin and Tsay study, by model

Lin and Tsay control carefully for the roots of the processes involved, only their cointegrated structure displays common features, in this case of the unit root. All the other models are diagonal, meaning that, in the case of model 3 for example, although there are roots very close to being unit roots, they do

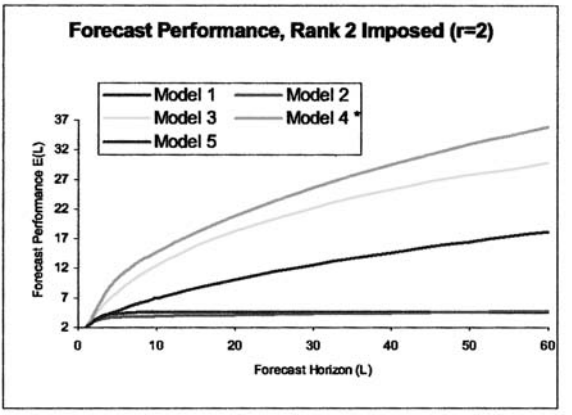


Figure 6.2 Lin and Tsay results, all models, rank 2 system

not constitute a common feature. For this to be so, the determinant of the VAR lag operator evaluated at that root would have to be less than full rank, but not zero. Diagonality results in its being zero (Burke 1996).[9] Model 3 also has the interesting property that the quality of forecasts is least affected by the choice of (cointegrating) rank.

By grouping these results differently, a further conclusion can be made. Instead of looking at the results by model and varying the cointegrating rank imposed, it is possible to fix the imposed cointegrating rank, and see which model is easiest or hardest to forecast for that restriction. Figure 6.2 demonstrates the case for the imposition of rank 2, which is correct for model 4. It is immediately obvious that, using the trace measure (see Forecast Evaluation below), the cointegrated system is the hardest to forecast at medium and long horizons. It is even harder to forecast than the non-stationary non-cointegrated case.[10] In fact, no matter what cointegrating rank is imposed (0 to 4), the cointegrated system is the most difficult to forecast, in the sense that it has the largest trace statistic. However, it remains the case that, if the system is cointegrated, it is best to impose the appropriate cointegrating rank (figure 6.1d).[11]

These forecast comparisons are more limited since they are compared in levels terms only. Clements and Hendry demonstrate that once transformed to stationarity, there is much less difference between forecasts based on different procedures. It is not clear from Lin and Tsay if the same transformation would result in less obvious distinctions between the forecasts based on the imposition of different cointegrating ranks at the estimation stage. Broadly speaking, the extension to the multivariate case is not found to undermine the findings on Clements and Hendry for the bivariate case. However, the four-variable setting makes it even more difficult to generalize the findings,

and the multiplicity of possible cases should lead to reticence when interpreting the results in a wider setting.

In order to reduce the impact of such criticisms, Lin and Tsay present two real data examples, one financial and one macroeconomic. They observe that the problem of roots close to the unit circle, but not actually being unit roots, is observable in data (that is, similarity to model 2, or, more extremely, model 3). In such circumstances, the under-specification of the rank (imposing unit roots that are not present) can be expected to result in poor long term forecasts.[12] Secondly, they observe that forecast error variances from a stationary system converge fairly rapidly as a function of forecast horizon. This is used to explore the stationarity of a system of bond yields. In this case, the unit root and cointegration tests performed suggest cointegration. This could be a case where the process is near non-stationary, and with a common feature, but the common feature is a root close to, but not on, the unit circle. It is clear from their investigations that, at a practical level, cointegrating restrictions cannot be assumed to improve long term forecasts, even where there is within-sample statistical evidence to support them.

6.3.4.4 Forecast evaluation

In both the Lin and Tsay and Clements and Hendry studies, the basic measure of forecast accuracy is the trace of the Monte Carlo estimate of the variance–covariance matrix of the forecasts. It has the following form. Let $e_{k,t}(j)$ be the j–step ahead vector forecast error made at time t arising from the k^{th} replication. Let the total number of replications be K. Then let

$$\hat{\Omega}_{\kappa,\tau}(j) = e_{k,t}(j)e_{k,t}(j)'.$$

One of the measures used by Clements and Hendry, and the one relevant to most of the results reported above, is

$$T(j) = trace\left[\frac{\sum_{k=1}^{K}\hat{\Omega}_{k,t}(j)}{K}\right],$$

which is referred to as the trace mean-square forecast error (TMSFE). Lin and Tsay use a modified version of this criterion since each replication gives rise to a set of j–step ahead forecasts, as a result of rolling forward the forecast origin within the same replication. They construct a within replication estimate of the forecast error variance–covariance matrix as

$$\hat{\Omega}_{k}(j) = \frac{\sum_{t=300}^{400-j}\hat{\Omega}_{k,t}(j)}{100-j+1}$$

This is then averaged across replications, the final measure being

$$E(j) = \sqrt{trace\left(\frac{\sum_{k=1}^{K} \hat{\Omega}_k(j)}{K} \right)} \tag{6.36}$$

Clements and Hendry (1998) discuss the choice of criterion, and use others in addition to TMSFE. An important aspect of these is their sensitivity to linear transformations of the data, although extensive use continues to be made of it.

6.3.4.5 Other issues relevant to forecasting performance in practice

In practice, forecasting will be subject to a number of other possible sources of error (Clements and Hendry, 1998, chapter 7, for a taxonomy). In the context of forecasting in cointegrated systems, these include the uncertainties associated with the selection of VAR order, the reliability of unit root and cointegration tests, and the estimation of the cointegrating vectors. This analysis has dealt exclusively with CI(1, 1) systems, elsewhere in this book, the case of cointegration in I(2) systems has been considered. This raises the question not just of how forecasting might be affected by choice of cointegration rank, but also types of (linear) cointegration, especially where there exists the possibility of variables being integrated of order up to 2.

All forecasting is predicated on at least two assumptions regarding model stability. That is, that the model structure has remained constant during the in-sample period, and that this same structure will remain into the forecasting period. Clements and Hendry (2001) have considered the implications for forecasting of some types of model instability in depth. Other procedures allow model switching (usually in a univariate setting, however), or non-linear adjustment to equilibrium. Any or all of these methods may be appropriate where a simple linear approximation fails to provide adequate forecasting performance.

Typically, the order of underlying VAR model is chosen by the optimization of some form of parsimonious information criterion, such as the SIC. These do not all have the same model selection properties, however (Reimers 1992). A potentially important variant of these criteria is to jointly select over VAR order and cointegrating rank. The criteria given by equation (6.34) are easily modified for this purpose. The VAR(p) can be estimated as a VECM as this does not alter the value of the $\left| \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}'_t \right|$, but cointegrating restrictions can be placed on the long-run matrix, via the Johansen procedure for example, such that

$$\underset{n\times n}{\Pi} = \underset{n\times r}{\alpha}\ \underset{r\times n}{\beta'}$$

such that there are only $2nr$ parameters of Π to be freely estimated. The information criterion is therefore of the form of (6.34) with $m = (p-1)n^2 + 2nr$, the selected model being that for which the criterion is minimized over a grid of values of p and $r = 0, 1 \ldots, n$ (the upper limit on the range of r allowing for stationarity). The evidence on the appropriate form of the penalty term, $f(T)$, is mixed (Reimers 1992), and while SIC can dominate, relative performance depends on simulation design. In practice, it is best to compute a range of criteria and search for corroborative evidence amongst them as to model order and cointegrating rank, and, if there is significant deviation in the findings, to check that subsequent inferences are not sensitive across the set of models suggested.[13]

Lin and Tsay (1996) point out that a model should be selected (and estimated) according to its purpose. In their paper they develop the idea that if the objective of the model is to forecast at a long-term forecast horizon, then it should be optimized to do this. Since standard methods of estimation and the form of information criteria are based on one step-ahead errors, it would not be surprising that such models were sub-optimal in terms of, say, 50-step ahead forecasts.

6.4 Models with short-run dynamics induced by expectations

A number of papers have considered the issue of estimating the linear quadratic cost of adjustment models under the type of dependence associated with cointegration (Dunne and Hunter 1998; Hunter 1989; and Engsted and Haldrup 1997). It should be understood that other forms of dependence might lead to similar types of problems. However, none of these are insurmountable. One issue which has been much discussed in the literature is the question of identification. As much of the analysis to date has concerned single equations, then the identification of the discount rate is of concern (Hendry et al. 1983; or Sargan 1982a). In general identification of parameters in structural or quasi-structural relationships is feasible (Arellano et al. 1999; Hunter 1989, 1992; and Pesaran 1981, 1987). A significant issue, as far as identification of forward-looking behaviour is concerned, is that both the IV and GMM estimators do not bind the solution based on the minimum of the optimization problem to the restrictions associated with the terminal condition (Nickel 1985; Hunter and Ioannidis 2000). Tests of over-identifying restrictions do not impose burdensome conditions on the estimator, and satisfaction of the necessary conditions follows without difficulty with the exception of highly non-persistent processes (Stock, Wright and Yogo 2002).

This section considers the impact of cointegration amongst endogenous and exogenous variables on rational expectations solutions and reveals a

computationally efficient estimation procedure that can readily be adapted to incorporate dependent I(1) processes either in the endogenous or the exogenous variables. The necessary and sufficient conditions for separation into two forms of long-run process is discussed in Hunter (1989, 1990), in terms of the types of condition discussed under cointegrating exogeneity in chapter 5. Otherwise efficient estimation of the long run requires the existence of a number of weakly exogenous variables either for the system or a sub-system for which behaviour is predominantly forward looking. This is intimately related to the notion of super exogeneity which may negate the practical use value of the Lucas critique (Lucas 1976; Hendry 1988; and Hendry and Favero 1992).

6.4.1 Linear quadratic adjustment cost models

Consider the following objective function based on Kollintzas (1985), though for ease of exposition the interaction between y_t and $(y_t - v_t)$ is excluded here:

$$E(\Im_t \Omega_t) = \sum_{t=o}^{T^*} E\{\delta^t (\Delta y_t' K \Delta y_t + (y_t - v_t)' H (y_t - v_t)) | \Omega_t)\}. \tag{6.37}$$

Let (6.37) define a control problem (Chow 1978), y_t is an n_1 vector of endogenous variables, v_t an n_1 vector of unobserved targets, that can be defined as a linear function of n_2 exogenous variables, z_t, where $v_t = Az_t + w_t$, A is a matrix of long-run multipliers, $w_t = z_t - E(z_t|\Omega_t)$ is a n_1 vector of white noise innovations and δ is the discount rate. With fixed initial conditions $y_0 = \bar{y}$, then from Kollintzas (1985) the Lagrange–Euler first-order condition after substituting out for v_t is:

$$E(\delta^t Q_0 y_t - \delta^{t+1} Q_1 y_{t+1} - \delta^t Q_1' y_{t-1} - \delta^t H A z_t | \Omega_t) = 0, \tag{6.38}$$

where $Q_0 = (1 + \delta) K + H$ and $Q_1 = K$.

Consider the process when it approaches its terminal value (at $T^* = T + N$):

$$E(\delta^{T^*} Q_0 y_{T^*} - \delta^{T^*+1} Q_1 y_{T^*+1} - \delta^{T^*} Q_1' y_{T^*-1} - \delta^{T^*} H A z_{T^*} | \Omega_t) = 0. \tag{6.39}$$

Stationarity is one precondition traditionally accepted for the transversality condition to be satisfied (Pesaran 1987), but when the structure includes a discount factor this assumption is too strong. In general all that is required is for (6.39) to be bounded as $T^* \to \infty$.

To reveal a standard symmetric solution to the forward-looking problem, (6.39) is scaled by $\delta^{-\frac{1}{2}(T^*+1)}$:

$$\begin{aligned} &E(\delta^{-\frac{1}{2}(T^*+1)} \delta^{T^*} Q_0 y_{T^*} - \delta^{-\frac{1}{2}(T^*+1)} Q_1 y_{T^*+1} - \\ &\delta^{-\frac{1}{2}(T^*+1)} \delta^{T^*} Q_1' y_{T^*-1} - \delta^{-\frac{1}{2}(T^*+1)} \delta^{T^*} H A z_{T^*} | \Omega_t = 0. \end{aligned} \tag{6.40}$$

Simplifying (6.40):

$$E(\delta^{-\frac{1}{2}}Q_0\delta^{-\frac{1}{2}(T^*)}y_{T^*} - \delta^{\frac{1}{2}(T^*+1)}Q_1 y_{T^*+1} - \delta^{\frac{1}{2}(T^*+1)}Q_1' y_{T^*-1} - \delta^{-\frac{1}{2}}HA\delta^{\frac{1}{2}T^*}z_{T^*}|\Omega_t) = 0. \tag{6.41}$$

Re-defining (6.41) in terms of $y^*_{T^*} = \delta^{-\frac{1}{2}(T^*)}\, y_{T^*}$ and $z^*_{T^*} = \delta^{-\frac{1}{2}(T^*)}\, z_{T^*}$ gives rise to the symmetric solution:

$$E(\delta^{-\frac{1}{2}}Q_0 y^*_{T^*} - Q_1 y^*_{T^*+1} - Q_1' y^*_{T^*-1} - \delta^{-\frac{1}{2}}HAz^*_{T^*}|\Omega_t) = 0. \tag{6.42}$$

In the limit (6.42) is bounded when the roots of the processes driving z_t and y_t are of mean order less than $\delta^{-\frac{1}{2}}$ as:

$$\underset{T^*\to\infty}{Lim}\, E(y^*_{T^*+1}|\Omega_t) \to 0 \;\; and \;\; \underset{T^*\to 0}{Lim}\, E(z^*_{T^*+1}|\Omega_t) \to 0.$$

Notice that (6.42) is bounded even when y and z have univariate time series representations that are non-stationary. Now consider the cointegration case. Dividing (6.38) by δ^t and transforming yields an error correction representation:

$$E(-\delta K\Delta y_{t+1} + K\Delta y_t + H(y_t - Az_t)|\Omega_t) = 0. \tag{6.43}$$

It follows that (6.43) is bounded in the limit when:

$$\underset{T^*\to\infty}{Lim}\{-\delta K\Delta y_{T^*+1} + K\Delta y_{T^*} + H(y_{T^*} - Az_{T^*})\} \to 0. \tag{6.44}$$

From the above discussion, a regular solution (see Pesaran 1987) to (6.42) exists, if and only if: (a) Q_o is symmetric; (b) K is non-singular; and (c) $\lambda < \delta^{-\frac{1}{2}}$. Dividing through (6.38) by δ^t yields the following difference equation:

$$E(Q_o y_t - \delta Q_1 y_{t+1} - Q_1 y_{t-1} - HAz_t|\Omega_t) = 0. \tag{6.45}$$

Redefining (6.45) using the forward (L^{-1}) and backward (L) lag operators:

$$Q(L)E(y_t|\Omega_t) = HAE(z_t|\Omega_t). \tag{6.46}$$

Now $Q(L) = (Q_o I - Q_1 L^{-1} - Q_1' L)$ has the following factorization:

$$Q_1 Q(L) = (I - G_1 L^{-1})(I - FL),$$

where $G_1 = \delta F$, $F = P\Lambda P^{-1}$ and Λ is a matrix whose diagonal elements are the stable eigenroots of the system. Therefore:

$$(I - G_1 L^{-1})(I - FL)E(y_t|\Omega_t) = K^{-1}HAE(z_t|\Omega_t). \tag{6.47}$$

It follows that the solution of the system can be written as:

$$y_t - Fy_{t-1} = \sum_{s=0}^{\infty}(G_1)^s F\, E(R_o Az_{t+s}|\Omega_t) + (G_1)^{-t}M_t + u_t \tag{6.48}$$

(Sargent (1978). Where $R_o = (\delta(F - I) + F^{-1} - I)$ and M_t satisfies the martingale property $E(M_{t+1}|\Omega_t) = (G_1)\, M_t$ (Pesaran 1987).

Reversing the transformation and applying it to (6.48):

$$\begin{aligned}
&(I - G_1L^{-1})(y_t - Fy_{t-1} - u_t)\\
&= (I - G_1L^{-1})(\sum_{s=0}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_t) + (G_1)^{-t}M_t)\\
&= \sum_{s=0}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_t) - G_1\sum_{s=0}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_{t+1})\\
&\quad + (I - G_1L^{-1})(G_1)^{-t}M_t.
\end{aligned}$$

The first two terms on the right-hand side simplify, while the Koyck operator annihilates the bubble behaviour. Therefore:

$$\begin{aligned}
&(I - G_1L^{-1})(y_t - Fy_{t-1} - u_t)\\
&= FR_oAz_t + \sum_{s=1}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_t) - \sum_{s=1}^{\infty}(G_1)^s F\,E(R_oAz_{t+s}|\Omega_{t+1})\\
&= FR_oAz_t + \sum_{s=1}^{\infty}(G_1)^s F(E(R_oAz_{t+s}|\Omega_t) - E(R_oAz_{t+s}|\Omega_{t+1})).
\end{aligned}$$

Assuming that there are no bubbles and a forcing process $z_t = B(L)w_t$ (w_t is white noise), then:

$$E(z_{t+s}|\Omega_t) - E(z_{t+s}|\Omega_{t+1}) = -B_{s-1}w_{t+1}$$

and

$$\begin{aligned}
(I - G_1L^{-1})(y_t - Fy_{t-1} - u_t) &= FR_oAz_t - \sum_{s=1}^{\infty}(G_1)^s F(B_{s-1}w_{t+1})\\
&= FR_oAz_t - FR_o(\sum_{s=1}^{\infty}(G_1)^s AB_{s-1})w_{t+1}.
\end{aligned}$$

Now reversing the Koyck lead and setting $(\sum_{s=1}^{\infty}(G_1)^s AB_{s-1}) = D$ gives rise to a forward-looking representation, which depends on future values of z_t. Therefore:

$$\begin{aligned}
(y_t - Fy_{t-1} - u_t) &= (I - G_1L^{-1})^{-1}(FR_oAz_t - FR_oDw_{t+1})\\
&= \sum_{s=1}^{\infty}(G_1)^s F(R_oAz_{t+s} - R_oDw_{t+s+1}).
\end{aligned}$$

It is possible to estimate the above model by FIML using the following recursion:

$$\begin{aligned}
(y_t - Fy_{t-1} - u_t) &= h_t \qquad (6.49)\\
h_t &= FR_oAz_t - FR_oDw_{t+1} + G_1h_{t+1}.
\end{aligned}$$

A fixed initial condition can be handled by recursively de-meaning the dependent variable (Taylor 1999), the problem of selecting an appropriate terminal

condition is solved by introducing a large enough future horizon or setting $(G_1)^s h_{t+s+1} = 0$.

Alternatively, the solution has the following backward representation, by substituting terms of the form $E(z_{t+s}|\Omega_t)$ using the Wiener–Kolmogorov prediction formula, which gives rise to the reduced form:

$$y_t - Fy_{t-1} = \Xi(L)z_t + u_t, \tag{6.50}$$

where $\Xi(L) = (\Xi_0 + \Xi_1 L + \ldots \Xi_{s-1}L^{s-1})$ is a function of δ, H, K, A and $\Theta(L) = (I + \Theta_1 L + \ldots \Theta_s L^s)$. However, this is a more complex set of non-linear relations to deal with (Hunter 1995; or Johansen and Swensen 1999).

It is also possible to give (6.49) a recursive structural form as long as K^{-1} exists. Notice that $R_o = K^{-1} H$ and:

$$\begin{aligned}(y_t - Fy_{t-1} - u_t) &= R_o FAz_t - R_o FDw_{t+1} + G_1 h_{t+1} \\ K(y_t - Fy_{t-1} - u_t) &= HFAz_t - HFDw_{t+1} + KG_1 h_{t+1}.\end{aligned} \tag{6.51}$$

As in a conventional system (Sargan 1988), to identify K, H and F, then $n_1 - 1$ additional restrictions are required (Hunter 1992a). Subject to knowledge of K and F, then H can be calculated from the following restriction $K\, R_o = K(\delta(F - I) + F^{-1} - I) = H$ as $R_o = K^{-1} H$ commutes. Essentially, identification of K follows from the additional restrictions, while identification of H follows from F, given knowledge of K and any additional restrictions to the system.

6.4.2 Models with forward behaviour and n_2 weakly exogenous I(1) variables.

If one considers the backward-looking form of the forward-looking model, then this is a VAR. The cointegrating VAR takes the from

$$\Delta y_t = [\alpha_{11} : \alpha_{12}] \begin{bmatrix}\beta_{1.} \\ \beta_{2.}\end{bmatrix} \begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \epsilon_{1t} \tag{6.52}$$

$$\Delta z_t = [\alpha_{21} : \alpha_{22}] \begin{bmatrix}\beta_{1.} \\ \beta_{2.}\end{bmatrix} \begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \epsilon_{2t}, \tag{6.53}$$

where any further dynamic can be incorporated in an appropriate time series representation of the error process. It follows for weak exogeneity relative to the long run, that $[\alpha_{21} : \alpha_{22}] = [0 : 0]$. As a result:

$$\Delta y_t = [\alpha_{11} : \alpha_{12}] \begin{bmatrix}\beta_{1.} \\ \beta_{2.}\end{bmatrix} \begin{bmatrix}y_{t-1} \\ z_{t-1}\end{bmatrix} + \epsilon_{1t} \tag{6.54}$$

$$\Delta z_t = \epsilon_{2t}, \tag{6.55}$$

where $\beta = [\beta'_{1.} \beta'_{2.}]$ and $\epsilon_{2t} = C(L)w_t$. Notice that inference on the short-run parameters is not appropriate as the coefficients of the ARMA error process forcing Δy_t depend on the MA process forcing ϵ_{1t}. It follows that the cointegrating relations are defined in the equations for Δy_t. Now consider the

solution to the forward-looking relationship given above, then the long-run behaviour that is important applies to the equation for y_t.

$$y_t - Fy_{t-1} = R_o F \sum_{s=0}^{\infty} (G_1)^s E(Az_{t+s}|\Omega_t) + u_t$$

where $R_oF = (\delta(F^2 - F) + I - F) = (\delta F(F - I) + I - F) = (I - F)(I - \delta F)$. It follows that:

$$\begin{aligned} & y_t - Fy_{t-1} - u_t \\ &= (I - F)(I - \delta F) \sum_{s=0}^{\infty} (G_1)^s E(Az_{t+s}|\Omega_t) \\ &= (I - F) \left(\sum_{s=0}^{\infty} (G_1)^s E(Az_{t+s}|\Omega_t) - \sum_{s=0}^{\infty} (G_1)^{s+1} E(Az_{t+s}|\Omega_t) \right) \\ &= (I - F)\{FAz_{t-1} + \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)\}. \end{aligned} \tag{6.56}$$

Now it follows from the results in Engsted and Haldrup (1997) that (6.56) has an error correction type representation in differences and levels. Furthermore:

$$\Delta y_t + (I - F)y_{t-1} - u_t = (I - F)\{Az_{t-1} + \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)\}$$

$$\Delta y_t + (I - F)\{y_{t-1} - Az_{t-1}\} - u_t = \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t).$$

In the error correction form $\alpha_1 = (I - F)$ and the cointegrating relations are normalized with respect of n_2 weakly exogenous variables as follows, $\beta = (I : A)$. The representations in Dolado et al. (1991) and Engsted and Haldrup (1997) rely on the existence of exactly n_2 weakly exogenous variables for the long-run to be estimated from the equations on y_t alone. It then follows that the above system can be estimated in two steps. Firstly the long run might be estimated using a regression or the Johansen Procedure, and then the short run relationship is estimated. There is no separate long-run relationship amongst the endogenous variables. Alternatively, consider a solved form similar to the one dealt with in sections 6.4.1:

$$y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t = \sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)\}$$

Reversing the Koyck transformation:

$$\begin{aligned} & (I - G_1 L^{-1})(y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t) \\ &= (I - G_1 L^{-1})(I - F)(\sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t)) \\ &= (I - F)\{\sum_{s=0}^{\infty} (G_1)^s E(A\Delta z_{t+s}|\Omega_t) - G_1 \sum_{s=0}^{\infty} (G_1)^s E(\Delta Az_{t+s}|\Omega_{t+1})\} \end{aligned}$$

or

$$(I - G_1L^{-1})(y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t)$$
$$= (I - F)A\Delta z_t + (I - F)\sum_{s=1}^{\infty}(G_1)^s(E(A\Delta z_{t+s}|\Omega_t) - E(A\Delta z_{t+s}|\Omega_{t+1})).$$

It follows from the Granger representation theorem that Δz_t has the following Wold form $\Delta z_t = C(L)w_t$ and

$$E(\Delta z_{t+s}|\Omega_t) - E(\Delta z_{t+s}|\Omega_{t+1}) = -C_{s-1}w_{t+1}.$$

Substituting back into the forward-looking model:

$$(I - G_1L^{-1})(y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t)$$
$$= (I - F)A\Delta z_t - (I - F)\sum_{s=1}^{\infty}(G_1)^s(AC_{s-1}w_{t+1})$$
$$= (I - F)A\Delta z_t - (I - F)(\sum_{s=1}^{\infty}(G_1)^s AC_{s-1})w_{t+1}.$$

Now reversing the Koyck lead and setting $(\sum_{s=1}^{\infty}(G_1)^s AC_{s-1}) = D^*$, gives rise to a forward-looking representation, which depends on future values of Δz_t:

$$y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t$$
$$= (I - G_1L^{-1})^{-1}((I - F)A\Delta z_t - (I - F)D^*w_{t+1})$$
$$= \sum_{s=0}^{\infty}(G_1)^s((I - F)A\Delta z_{t+s} - (I - F)D^*w_{t+s+1}).$$

Now decompose the last relationship as follows:

$$y_t - Fy_{t-1} - (I - F)Az_{t-1} - u_t$$
$$= \sum_{s=0}^{\infty}(G_1)^s((I - F)A\Delta z_{t+s} - \sum_{s=0}^{\infty}(G_1)^s(I - F)D^*w_{t+s+1})$$
$$= \sum_{s=0}^{\infty}(G_1)^s(I - F)Az_{t+s} - \sum_{s=0}^{\infty}(G_1)^s(I - F)Az_{t+s-1}$$
$$- \sum_{s=0}^{\infty}(G_1)^s(I - F)D^*w_{t+s+1}.$$

Therefore:

$$y_t - Fy_{t-1} - u_t = \sum_{s=0}^{\infty}(G_1)^s(I - F)Ax_{t+s} + (I - F)Ax_{t-1} - (I - F)Ax_{t-1}$$
$$- \sum_{s=1}^{\infty}(G_1)^s(I - F)Ax_{t+s-1} - \sum_{s=0}^{\infty}(G_1)^s(I - F)D^*w_{t+s+1}.$$

Re-writing the above into an equation purely in levels:

$$y_t - Fy_t - u_t = (I-F)\{\sum_{s=0}^{\infty}(G_1)^s Az_{t+s} - G_1\sum_{s=1}^{\infty}(G_1)^{s-1}(I-F)Az_{t+s-1}\}$$
$$-\sum_{s=0}^{\infty}(G_1)^s(I-F)D^*w_{t+s+1}.$$

Re-indexing the second sum and gathering terms, yields a levels relationship:

$$y_t - Fy_{t-1} - u_t = (I-F)(I-G_1)\sum_{s=0}^{\infty}(G_1)^s Az_{t+s} - \sum_{s=0}^{\infty}(G_1)^s(I-F)D^*w_{t+s+1}.$$

It is possible to estimate the above model by FIML using the following recursion:

$$y_t - Fy_{t-1} - u_t = h_t \qquad (6.57)$$
$$h_t = FR_oAz_t - (I-F)D^*w_{t+1} + G_1h_{t+1}.$$

In such circumstances the above relationship has the same forward recursion as was considered before, except the transversality condition relies on the existence of cointegration. Decompose (6.44) as follows:

$$\underset{T^*\to\infty}{Lim}\{-\delta K\Delta y_{T^*+1} + K\Delta y_{T^*} + H(y_{T^*} - Az_{T^*})\} =$$
$$\underset{T^*\to\infty}{Lim}\{-\delta K\Delta y_{T^*+1} + K\Delta y_{T^*}\} + \underset{T^*\to\infty}{Lim}\{H(y_{T^*} - Az_{T^*})\} \to 0.$$

The conditions for cointegration (Engle and Granger 1987) are sufficient for this to be satisfied. That is $y_t \sim I(1)$ and $(y_t - Az_t) \sim I(0)$, y_t and z_t cointegrate. Furthermore, (6.57) has an error correction form:

$$\Delta y_t - (I-F)(y_{t-1} - Az_{t-1}) - u_t = h_t$$
$$h_t = (I-F)A\Delta z_t - (I-F)D^*w_{t+1} + G_1h_{t+1}$$

In the next section the case with dependence amongst the endogenous variables is considered.

6.4.3 Models with forward behaviour and unit roots in the process driving y_t

There are a number of reasons for finding dependence amongst the endogenous processes, one of which would be cointegration, the other would be the type of dependence that exists amongst series that might satisfy an adding up type constraint. In the former case the cause of rank failure is the existence of a unit root and it can be shown that the original objective function can be solved in the usual way (Hunter 1989a).

Consider the loss function

$$E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{\delta^t(\Delta y_t'K\Delta y_t+(y_t-v_t)'H(y_t-v_t))|\Omega_t)\} \tag{6.58}$$

where the *rank*(*H*) = r_1 As a result, the following decomposition exists: $H = E'E$ and *rank*(*E*) = r_1. Now define *M* such that the matrix $[E' : M']$ has full rank. Now we can redefine the loss function in terms of new variables:

$$E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{\delta^t(\Delta y_t^{*'}K^*\Delta y_t^*+(y_t^*-v_t^*)'H(y_t^*-v_t^*))|\Omega_t)\} \tag{6.59}$$

where $y_t^{*'}=\left[y_{1t}^{*'}\quad y_{2t}^{*'}\right]=y_t'\left[E':M'\right], K^\star=\left[E':M'\right]^{-1}K\begin{bmatrix}E\\M\end{bmatrix}^{-1}$ and $v_t^\star$ conformable with $y_t^\star$. It follows that the loss function has the following form:

$$\begin{aligned}E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{&\delta^t\Delta y_{1t}^{*'}K_{11}^*\Delta y_{1t}^*+2\Delta y_{1t}^{*'}K_{12}^*\Delta y_{2t}^*+\\&\Delta y_{2t}^{*'}K_{22}^*\Delta y_{2t}^*+(y_{1t}^*-v_{1t}^*)'(y_{1t}^*-v_{1t}^*))|\Omega_t)\}.\end{aligned} \tag{6.60}$$

Re-writing the above relationship in terms of a new set of stationary variables, then $y_t^{+\prime}=[y_{1t}^{\star\prime}\quad \Delta y_{2t}^{\star\prime}]$ and here it is assumed that the long-run target for $v_{2t}^\star = 0$ and $\Delta\Delta y_{2t} = 0$. Therefore:

$$\begin{aligned}E(\Im_t|\Omega_t)=\sum_{t=o}^{T^\star}E\{&\delta^t(\Delta y_{1t}^{+'}K_{11}^*\Delta y_{1t}^++2\Delta y_{1t}^{+'}K_{12}^*\Delta y_{2t}^+ +\\&y_{2t}^{+'}K_{22}^*y_{2t}^++(y_{1t}^+-v_{1t}^+)'(y_{1t}^+-v_{1t}^+))|\Omega_t)\}.\end{aligned}$$

Now differentiating with respect to y_{1t}^+ gives rise to the following first-order condition:

$$E(\delta^tK_{11}^*\Delta y_{1t}^+-\delta^{t+1}K_{11}^*\Delta y_{1t+1}^+-\delta^t(y_{1t}^+-v_{1t}^+)-2\delta^tK_{12}^*(y_{2t}^+-\delta y_{2t+1}^+)|\Omega_t)=0, \tag{6.61}$$

and with respect of y_{2t}^+:

$$E(\delta^tK_{21}^*\Delta y_{1t}^++\delta^tK_{22}^*y_{2t}^+|\Omega_t)=0.$$

Subtracting the above equation from its forward value and re-writing:

$$E(\delta^tK_{21}^*(\Delta y_{1t}^*-\delta\Delta y_{1t+1}^*)+\delta^tK_{22}^*(\Delta y_{2t}^*-\delta\Delta y_{2t+1}^*)|\Omega_t)=0.$$

Now consider the system:

$$E(\delta^t(K^*(\Delta y_t^*-\delta\Delta y_{t+1}^*)+\begin{bmatrix}I_r&0\\0&0\end{bmatrix}(y_t^*-v_t^*|\Omega_t)=0.$$

Now divide through by δ^t and reverse the transformation:

$$E(K(\Delta y_t-\delta\Delta y_{t+1})+H(y_t-z_t)|\Omega_t)=0.$$

Hence, irrespective of the existence of cointegration, the same first-order condition exists as does the solution dealt with before, except that H is rank deficient. Therefore $R_0 = K^{-1}H$ is rank deficient, F has $n_1 - r_1$ unit roots and $R_0 = (I - F)(I - \delta F)F^{-1}$ is rank deficient as can be observed from the following decomposition:

$$\begin{aligned}(I - F)(I - \delta F) &= (I - P\Lambda P^{-1})(I - \delta P\Lambda P^{-1}) \\ &= P(I - \Lambda)(I - \delta\Lambda)P^{-1}.\end{aligned}$$

Where the $rank((I - F)(I - \delta F)) = r_1$, when there are $n_1 - r_1$ unit roots. Hence the rank of the matrix H determines the number of unit roots. Now it is probably better to consider the recursive representation (6.57):

$$\begin{aligned}\Delta y_t - (I - F)(y_{t-1} - Az_{t-1}) - u_t &= h_t \\ h_t = (I - F)A\Delta x_t - (I - F)D * w_{t+1} + G_1 h_{t+1}&.\end{aligned}$$

If $I - F$ is rank deficient, then there is also the possibility of cointegration amongst the endogenous and exogenous variables. Notice the dependence also feeds forward into the relations in differences.

6.4.4 Estimation and inference

The benefit of the above approach is that it reduces the dimension of the estimation problem when forward-looking behaviour needs to be considered. Especially in terms of the need to estimate and store future predictions. However, the downside is that inference is made more complicated.

As far as estimation is concerned, then the usual likelihood function applies, where:

$$LogL((\delta, H, K, A), \Sigma|.) = -Tn\log(2\pi) - \frac{1}{2}T\log|\Sigma| - \frac{1}{2}tr(\Sigma^{-1}\sum_{t=1}^{T} u_t u_t')$$

and $u_t = y_t - Fy_{t-1} - h_t$. Now concentrating out Σ yields the quasi-likelihood:

$$LogL_c((\delta, H, K, A)|.) = C - \log|\Sigma|$$

where $S = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ is a consistent estimate of Σ. The likelihood is maximized using a Quasi-Newton algorithm such as Gill, Murray and Pitfield (see Sargan 1988) or an equivalent method. The method due to Gill, Murray and Pitfield has the advantage of using the Cholesky factors from the inverse of the Hessian. They are then bounded to be positive definite subject to an appropriately conditioned Hessian matrix.

However, the conventional estimates of the parameter variance based on the information matrix are not valid, even when the model for the endogenous equations is estimated as a system. The correct estimate needs to take account of the generated regressors and their parameter estimates. The following algorithm is suggested to do this. Initial estimates of the exogenous

variables are estimated as a VAR, then the residuals are saved. The VMA representation is estimated by OLS using the method described by Spliid (1983). In state space form:

$$\Delta z = W\varsigma + w.$$

$$\text{where } \Delta z = \begin{bmatrix} \Delta z_1 \\ \Delta z_2 \\ \cdot \\ \cdot \\ \cdot \\ \Delta z_T \end{bmatrix}, W = \begin{bmatrix} w_{-1} & w_{-2} & \cdots & w_{-p} \end{bmatrix}, w = \begin{bmatrix} w_1 \\ w_2 \\ \cdot \\ \cdot \\ \cdot \\ w_T \end{bmatrix}$$

$$\text{and } \varsigma = \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ \cdot \\ C_p \end{bmatrix}$$

Hence, the OLS estimator of the parameters is given by:

$$\Psi^{(0)} = (W'_{(o)}W_{(o)})^{-1}W_{(o)}\Delta z,$$

where $W_{(o)}$ contains the initial estimates of the surprises, unobserved values of the residual are set to zero and $\Psi^{(0)}$ are the initial estimates of the parameters. Once the system has been estimated, then the likelihood is re-estimated based on $B = 200$, bootstrap re-samplings of the original residuals vector w, where each iteration reallocates a block of residuals w_i by the new residual set $w_{(b)}$ used to provide new estimates of the VMA parameters ($\Psi^{(b)}$ for $b = 1, \ldots, B$). Then given the maximum likelihood estimates of the parameters (δ, H, K, A) an empirical distribution for the estimated test statistics are generated from the bootstrap re-sampling regime. A sample of 400 is created by the use of antithetic variance technique, providing at each bootstrap replication a pair of residuals $w_{(b)}$ and $-w_{(b)}$ (see Hendry 1995). Then percentiles of the empirical distribution can be used to determine critical values for the estimated parameters.

6.5 Conclusion

In this chapter a number of more advanced issues have been addressed: cointegration amongst series with different orders of integration; forecasting with cointegrating relationships; and cointegration combined with short-run structure defined by rational expectations.

With orders of integration in excess of I(1), inference is similar to the I(1) case except that there are now three types of process that evolve to generate the data. Cointegration not only occurs in the usual way amongst the levels, but may also occur between levels and differenced series, there are I(1) common trends and also I(2) trends. However, identification is a fundamental problem for the estimation of long-run behaviour in the I(2) case as three sets of parameters are potentially ill-defined.

When the order of integration is less than 1, then series are not likely to have the same fractional order of differencing. One approach is to consider the average non-integer order of differencing for a group of series. Estimation of the cointegrating vectors can be undertaken in a similar way as that for I(1) series when a non-parametric approach is considered (Robinson and Marinucci 1998), but testing is more complex (Robinson and Yajima 2002). It is relatively straightforward to compare the order of difference between series and to calculate the cointegrating rank, but there is no conventional procedure for inference.

Forecasting in cointegrated systems occurs at two levels – the short run and the long run and cointegration influences both of these. Short-run forecasts are less influenced by cointegration, but long-run forecasts may be strongly influenced. The literature is unclear as to whether gains in forecast accuracy depend on the restrictions that cointegration imposes on the long-run process or the interrelationship that cointegration imposes on the long-run forecasts. It appears that there is little difference between long-run forecasts derived from models that imposed the long-run restrictions as the forecasr evolves when they are compared with forecasts that ex-post have the cointegrating restriction imposed on them. This might suggest that the benefits to long-run forecasting associated with cointegration follow from the imposition of the restriction rather than cointegration per se. This would appear to be an issue for further investigation, though the authors would conjecture that cointegration has a role in the accuracy of long-run forecasts.

Estimation of the structural parameters of optimizing models has become enormously popular. It has become common practice to suggest that the VAR is a solution to a forward-looking model, but then not to consider the relation between the long-run and the short-run behaviour of the model. However, both the Engle–Granger and the Johansen procedure have been applied to models with forward-looking behaviour. The final section of this chapter considered the impact of unit root processes in the endogenous and exogenous variables on the solution and estimation of forward-looking models with rational expectations. Inference is significantly more complicated in these cases and has thus far had to derive from the proposition that series are cointegrated.

7
Conclusion: Limitations, Developments and Alternatives

7.1 Approximation

Many economic theories, especially those in macroeconomics, are theories of the way in which economic processes interact with one another to provide a stable set of underlying equilibrium relationships. Failure to observe equilibrium is strong evidence against the theory predicting its existence.

In general, relatively little is said that would be useful to econometricians or policy makers about the detailed nature of these relationships. This is not necessarily a failure of the theories; they point the way towards what might be true in a general sense, in effect indicating to the applied researcher where to look.

Cointegration analysis is one tool in this search. It may or may not be useful, depending on the circumstances in which it is used. It is certainly not a definitive statement about the structure of an economic system. It is far more constructive to view the mathematical and statistical structures on which cointegration is based as being approximations to reality. In this case the question is how useful the approximation is given the ultimate aim of distinguishing between situations where equilibrium does and does not exist.

A reasonable requirement of a statistical tool is that it is internally consistent, in the sense that it works well in situations for which the approximation is exact. This requirement is satisfied by cointegration analysis, as demonstrated by Johansen (1995a), and many other statistical theorists. This is not really sufficient, however, and it is also necessary to examine the performance of the technique in situations where the underlying models are not an exact description of the actual processes generating the data. Such investigations indicate, as with all statistical techniques, that certain approximation failings, such as structural breaks, are more serious than others. The methods examined in this book use as the approximating process, one that is linear, has fixed coefficients and Gaussian disturbances. A further aspect of the approximation of the technique as a whole is that the distribution theory of the tests

is based on an arbitrarily large sample size – as is the analysis of the power of the tests.[1]

It is, of course, vacuous to state that cointegration does not exist in the real world. Cointegration is a model on which a set of methods are based, allowing inferences to be drawn from data about the existence and characteristics of equilibrium relationships. The reliability of the inferences varies, but it is essential to have a tool that is capable of making the inferences. And this methodology is capable of more. It can examine characteristics of equilibrium relationships, and the dynamic properties of the variations about equilibrium. As always, this is based on a set of approximations and unlikely findings should be viewed against the possibility of approximation failure. In this, cointegration does not differ substantially from any other statistical technique. Its benefit lies in its crucial ability to resolve the matter of the existence or otherwise of equilibrium, a key concept throughout the subject of economics. The huge and rapid adoption of cointegration methodology is evidence of its invaluable contribution.

7.2 Alternative models

Many developments of the basic model have taken place, such as the introduction of non-linear adjustment,[2] more detailed characterizations of non-stationarity (such as fractional cointegration considered briefly in the previous chapter). There have also been developments in other branches of times series econometrics, such as the modelling of higher order moments of the data, including variance, skewness and kurtosis. These models provide different means of analyzing data, not necessarily focussed on the concept of equilibrium. Even so, some of the features of their data generating processes have been used to investigate the robustness of the cointegration methodology. So for example, the issue arrises as to how reliably cointegration, or its absence, is identified in the presence of autoregressive conditionally heteroscedastic disturbances, or where the Gaussian disturbance structure is replaced by one with more frequently occurring extreme values (relative to a Gaussian distribution).[3] It is inevitable that eventually the methods will fail.

Cointegration analysis has also been extended to panel data models where the time series dimension is sufficiently large.

7.3 Structural breaks

Probably the main feature of economic time series that is capable of undermining cointegration analysis is that of structural breaks. Breaks in individual time series can lead to incorrect inference as to their order of integration. Thus data that are considered to be integrated of order 1, might in fact be stationary

in the sense that they consist of stationary stochastic deviations around a deterministic trend that displays jumps in value of changes in slope. Alternatively, it is conceivable that the nature of the cointegrating relationship between I(1) variables may change, or that the adjustment coefficients may change. That is, in the notation of the previous chapters, the structure of the intercept vector μ, Π and the Γ_i of

$$\Delta x_t = \mu + \Pi x_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \epsilon_t,$$

may change. If Π changes, then this may be represented as changes in α or β, and possibly the cointegrating rank (though this seems less acceptable since it suggests the appearance and disappearance of equilibrium relationships over time). Clearly this is not an exhaustive list.

The problem for cointegration analysis is that failure to allow for structural breaks, especially in α or β, is likely to result in an inference of non-cointegration, even where cointegration exists. Economically, this will result in the failure to infer the presence of an equilibrium relationship where in fact one exists.[4]

There is also a potential identification problem – where a structural break in Π occurs, is this parameterized as a change in α or β? Is it a change in adjustment to disequilibrium or the equilibrium that is being adjusted to?[5]

7.4 Last comments

There is no doubt about the impact methods for the empirical analysis of time series equilibrium have had on applied economics. The methods and models continue to develop, and the range of subjects to which it can be applied seems only to be limited by the availability of adequate data. Indeed, even relatively small samples have been analyzed via the use of bootstrapping techniques. Outside the realm of high frequency financial models, it is unlikely that a similar revolution in econometric time series analysis will occur in the near future.

Notes

1 Introduction

1 Muellbauer (1983) showed that a random walk model of consumption with innovations in income and interest rates can be nested in the ADL framework due to Davidson et al. (1978). However, the tests used do not take account of the underlying series being non-stationary.
2 As will be discovered in the last section of chapter 6, stationarity is overly strong. In addition, the types of model used by Sargent are excessively restrictive (Hunter 1989).
3 It should be noted that the impulse response function solved from the VAR is not unique (Lippi and Reichlin 1994) and any findings on causality depend on the variables in the VAR model estimated (Hendry and Ericsson 1990).
4 Keynes discusses the latent nature of expectations, the problems with dynamic specification, measurement error, the role of forecast performance and structural breaks.

2 Univariate and Single Equation Methods

1 The sudden drops in level of the series around 1973 and 1981 are typical of real economic time series and can cause a problem with their statistical analysis. They are called structural breaks, and their characterisation and impact on estimation and inference is a major concern. See, for example, Maddala and Kim (1999).
2 For a more precise definition of a stochastic process see Banerjee et al. (1993, p. 10).
3 The use of the word 'stationarity' can now be understood to refer to properties that are unchanging, hence 'stationary'. Changing the properties that are required to be fixed through time changes the definition of stationarity. The more that are required to be fixed, the stricter (and more impractical) the definition.
4 Covariance stationarity is also known as weak and second-order stationarity.
5 Although of no direct interest in this book, note that correlation measures only linear association. Concentration on linearity can be justified by a distributional assumption of normality.
6 A series that consisted of a linear time trend as the mean plus a stationary process would have this property. In its simplest form, this is known as a trend plus noise model.
7 In fact, the sequence of random variables underlying the time series observations is referred to as a stochastic process, and it is then the stochastic process that is labelled stationary.
8 In addition, the observed data will also not be a function of the DGP alone, but also of the observation process (Hendry 1995; and Patterson 2000) including errors and systematic distortion due to such procedures as seasonal adjustment (Wallis 1974).
9 This definition includes the requirement of zero mean. This is not really substantive, but keeps things simple. All white noise in this book is zero mean white noise.
10 The detailed theory draws a distinction between two components of a time series: that which is perfectly predictable from its own past, called a deterministic

component, and that which cannot be perfectly predicted from its own past. A purely non-deterministic process has no component that can be predicted from its own past, and it is this type of series to which this abbreviated version of the theorem refers.

11 See also Box and Jenkins (1976) and Granger and Newbold (1976).

12 In addition, ϵ_t is uncorrelated with future values of the process, x_{t+j}, $j>0$.

13 The initial values for this equation can be calculated from the process. See Hamilton (1994), chapter 3.

14 In fact, this derivation requires the autocorrelations to be non-time varying. In other words, equation (2.19) only applies in the stationary case. See section 2.3.7 below.

15 As described in any textbook dealing with difference equations, the other case that has to be considered, but which is less interesting, is that where the roots are repeated, in which case equation (2.19) has to be modified.

16 The AR(1) process $x_t = \phi x_{t-1} + \epsilon_t$ will have one root given by $\lambda_1 = \phi^{-1}$. Substituting this and p = 1 into (2.13) gives (2.11b).

17 This is, in fact, a linear trend, being a linear function of time. Higher-order polynomial functions of time, such as the quadratic, are also referred to as time trends. It is for the purposes of analogy that the linear case is used here.

18 This is not a very helpful piece of terminology as it seems to mix up the discrete and continuous time cases. Perhaps "summed" would have been a better, if more prosaic, choice.

19 There is another absurdity about this calculation. Although purporting to be a correlation, it is clear that this quantity is not restricted to [–1, +1], for if j is large but t small, then this quantity can fall below –1.

20 Similar arguments apply in the explosive case when the roots lie inside the unit circle.

21 Preserving the ordering so the inverse operator is the premultiplying factor on the right-hand side of (2.25) is not necessary in the univariate case, but is good practice since in the multivariate case discussed in section 4.2 it is important.

22 The zero lag coefficient does not have to be 1 but it simplifies things a little to consider this case, which is anyway appropriate for ARMA models.

23 As with all ACFs, $\rho_x(0) = 1$, and for all MA(1) ACFs, $\rho_x(i) = 0$ for j > 1, so only $\rho_x(1)$ is considered in this illustration.

24 Strictly this applies only to cases of distinct real roots. Complex roots will occur as complex conjugate pairs and both be replaced by their inverses in order that the process remain real. Repetition of roots will mean that fewer new parameterisations can be generated by inverting just one root.

25 Note also that the MA still has to be normalized so that $\theta(0) = 1$.

26 This definition deals with the case where the non-stationarity is due to a root of $z = 1$. As already stated, all that is required for non-stationarity is $|z| \leq 1$, so z = 1 is a special case.

27 Unless otherwise stated it is the case that ϵ_t is white noise and $\phi(0) = \theta(0) = 1$.

28 If the (first) differenced process does have a non-zero mean, then the undifferenced process will possess a linear deterministic trend. In other words, although a linear trend plus noise model (2.31) is not I(1), a random walk with drift is.

29 The key property is that, in the representation of the model using the initial values and the summed disturbance process, the order of integration of the purely stochastic component and the order of the polynomial time trend is the same. So, in a random walk with drift, $x_t = x_0 + bt + \sum_{j=1}^{t} \epsilon_j$. The time trend is first order (linear), and $\Delta \sum_{j=1}^{t} \epsilon_j = \epsilon_t$ and so is I(1).

30 A brief introduction may also be found in chapter 5 of Harvey (1993).
31 The tests referred to are tests of cointegration, the property that two or more time series share a common unit root driving process.
32 In practice, this would be exacerbated by the problem of approximating a MA with a near unit root by a finite order autoregressive process. To achieve a given level of AR approximation to the ACF of an MA process, more AR terms will be needed as the MA root approaches unity. See Burke (1994a) and Galbraith and Zinde-Walsh (1993).

3 Relationships Between Non-Stationary Time Series

1 See chapter 5 for a discussion of weak exogeneity.
2 In the static case there is no distinction between the disturbances to the relationship and the deviations from equilibrium because the relationship without the disturbance is the same as the long-run solution to the model.
3 Comparing (3.14) and (3.15) it can be seen that the intercept can be included either inside or outside the error correction term. When approached in this way, it is clear that the true equilibrium error must include the intercept, that is if the term in the lagged levels is supposed to represent the extent to which the system was out of equilibrium in the previous period then it must include the intercept. However, in general, the intercept could be divided between the constant in the equation and the lagged levels term with the only restriction being that they sum to the appropriate value. Thus the Error Correction Model (ECM) is sometimes written

$$\Delta y_t = \mu_0 + \beta_0 \Delta x_t - (1-\alpha_1)\left(y_{t-1} - \mu_1 - \frac{\beta_0 + \beta_1}{\alpha_1 - 1} x_{t-1} \right) + u_t \text{ where } \mu_0 + (1-\alpha_1)\mu_1 = \mu.$$

Being able to deduce μ_0 and μ_1 from μ, is an example of the identification problem.
4 David Hendry (1995) reserves the term error correction for the case where $\frac{(\beta_0 + \beta_1)}{(1-\alpha_1)} = 1.$
5 More generally, there is a steady-state growth rate for y and z to which the equilibrium adjusts, but except for the constant this does not affect the long-run relationship.
6 A previous footnote made reference to the identification problem in terms of the representation of the intercept in the ECM. Notice that there is no ambiguity in moving from an ECM with a constant outside the equilibrium correction term and within it. They will combine to form an intercept in the long-run solution.
7 Constant term outside the equilibrium correction term in this case, but inside in the other two equations listed here.
8 This does not imply that reversion to some mean value may never occur, but that the distribution of such reversions is so long tailed that the expected value does not exist.
9 In estimating equilibrium relationships it is important to include an intercept since failure to do so will bias the coefficient estimates of the relationship.
10 See Hamilton (1994, p. 106) for more details.
11 In general, such cancellation gives rise to an additive constant term, which depends on the initial values of the z and y processes. This contributes to the time series structure in the same way as summing a random walk does.
12 This approach has more obvious appeal when thinking in terms of a DGP that might give rise to the data, since this is almost certainly going to be causal in some way.

13 The case for general $\alpha(L)$ and $\beta(L)$ in the ADL demonstrates further the power of the lag polynomial notation. In this case the ADL can still be written as $\alpha(L)y_t = \mu + \beta(L)z_t + u_t$ and the equilibrium error will be η_t. Applying $\alpha(L)$ to both sides of the long-run relationship yields

$$\alpha(L)\eta_t = \alpha(L)y_t - \mu - \alpha(L)\frac{\beta(1)}{\alpha(1)}z_t.$$

Now substituting out for $\alpha(L)y_t$, implies

$$\alpha(L)\eta_t = \mu - \mu + \beta(L)z_t - \alpha(L)\frac{\beta(1)}{\alpha(1)}z_t + u_t,$$

$$= \{\beta(L) - \alpha(L)\frac{\beta(1)}{\alpha(1)}\}z_t + u_t.$$

Letting $\gamma(L) = \beta(L) - \frac{\beta(1)}{\alpha(1)}\alpha(L)$, it can be seen $\left\{\beta(L) - \frac{\beta(1)}{\alpha(1)}\alpha(L)\right\}$ has a unit root as $\gamma(1) = \beta(1) - \frac{\beta(1)}{\alpha(1)}\alpha(1) = 0$. As reparameterizing using $\gamma(L) = \gamma(1)L + \gamma^*(L)\Delta$, it follows that $\gamma(L) = \gamma^*(L)\Delta$. When substituted into the expression for $\alpha(L)\eta_t$ gives rise to $\alpha(L)\eta_t = \gamma^*(L)\Delta z_t + u_t$. If $\alpha(L)$ has all its roots outside the unit circle and z_t is at most I(1), then η_t is stationary. As a special case, if z_t is the random walk defined by (3.54b), then Δz_t is white noise and η_t is ARMA(p, q) where p is at most the order of $\alpha(L)$ and q is at most the larger of the orders of $\alpha(L)$ and $\beta(L)$ minus one (because the unit root has been factored out). This also shows that the closer are any of the roots of $\alpha(L)$ to unity, the more persistent will be the equilibrium errors. At the same time, $\alpha(1) \to 0$, so the speed of adjustment to equilibrium gets smaller. Note, however, that in addition to the roots of $\alpha(L)$ those of the autoregressive operator of the ARMA representation of Δz_t will also determine the behaviour of η_t. Thus if Δz_t displays persistence, so will η_t independent of the speed of convergence.

14 Exactly the same random number sequences are used in the two cases.

15 This terminology is also appropriate for any regression between I(d) variables with disturbances that are I(d–b). However, it is generally reserved for the case where the disturbances are stationary as in any case this is the case that is of most interest because of its equilibrium interpretation.

16 Speeds of convergence and $Op(.)$ are discussed in more detail in Spanos (1986, Chapter 10) and Patterson (2000, section 4.4.2) provides a brief introduction.

17 Asymptotic normality does apply if z_t is strongly exogenous for the estimation of b, that is, it is both weakly exogenous and z_t is not Granger caused by y_t.

18 If regressions involving not only I(1) but also I(2) variables are being considered, then the critical values of the tests must be further adjusted. The tests are still of the null that the disturbances are I(1) against the alternative that they are I(0), thus it is assumed that any I(2) processes are cointegrating to I(1). Haldrup (1994) discusses this problem and presents appropriate critical values.

19 The common factor restriction for autoregressive models in the error is discussed by Hendry and Mizon (1978). The ADF test applied to the cointegration case is a transformation of such autoregressive behaviour in the residual associated with the common factor restriction. The effect of such restrictions on ADF and ECM tests of cointegration is considered in Kremers et al. (1992).

4 Multivariate Time Series Approach to Cointegration

1 Without explaining or deriving the origin of the vector white noise process, this equality is best interpreted as meaning the autocorrelation structure of the processes on each side of the equation are the same.

2 This is an important point. Although the example has not explicitly applied the SM form, this is in fact being used. The rationality of the VAR operator means that the SM reparamaterization can be applied.

3 Accounts of the Smith–McMillan form only use the term "matrix polynomial" or "polynomial matrix" for matrices with finite order scalar polynomial elements.

4 The roots of the $\beta_{i,j}(L)$ are called the *poles* of the rational polynomial $\dfrac{\alpha_{i,j}(L)}{\beta_{i,j}(L)}$. Since these are elements of $C(L)$, their poles, for $i,j = 1,2, \ldots, n$ are called the poles of $C(L)$.

5 In addition, the Wold representation requires the coefficient matrices of the VMA to converge in the sense that

$$Lim_{g\to\infty}\left[\sum_{i=1}^{g} C_i C_i'\right]$$

must exist. This would not be the case if there were any poles on or inside the unit circle. That is, the only way a rational VMA form can be consistent with the Wold representation is for the operator to have all poles outside the unit circle.

6 If the j^{th} diagonal element of $D^*_{2,2}(L)\overline{D}(L)$ is

$d^*_{2,2,j}(L)(1-L)^{d_{n-r+j}} = (1-L)$

where $d^*_{2,2,j}(L)$ is the j^{th} diagonal element of $D^*_{2,2}(L)$, then it follows from (4.38) that

$d^*_{2,2,j}(L) = (1-L)^{1-d_{n-r+j}}$.

Since negative powers of $\Delta = (1-L)$ are not defined, $1 - d_{n-r+j} \geq 0$, but, as is stated above, $d_{n-r+j} \geq 1$. This implies $d_{n-r+j} = 1$ for $j = 1,2, \ldots r$, and $d^*_{2,2,j}(L) = 1$.

7 Clearly equation (4.40) will not hold for all $C(L)$. It implies conditions on $C(L)$. These are not discussed here.

8 The equation $A(L)C(L) = C(L)A(L) = \Delta I_n$ expresses very clearly the extent to which this process is not inversion. If A(L) were the inverse of C(L) then the relationship would be $A(L)C(L) = C(L)A(L) = I_n$. Instead, the inversion is only up to a scalar factor of Δ, so is a form of partial inversion, where all factors apart from Δ are cancelled.

9 Mathematically, this is written $|A(z)| = 0 \Leftrightarrow |z| > 1$ *or* $z = 1$.

10 The complete theorem is theorem 4.2, p. 49 of Johansen (1995a). The statement of this theorem is rigorous, and rather than simply refer to the I(0) property of the cointegrating combinations of the variables and of their difference, it refers to the condition required so that initial distributions may be given such that the processes are I(0). The reason for this is that the definition of stationarity used by Johansen (1995, p. 14) is such that despite the parametric condition, only specific manifestations of the initial values will deliver stationarity. However, without the parametric condition, none would suffice.

11 Johansen does not do this, but leaves in the initial values. They are reduced to zero later on in the proof anyway, by pre-multiplication.

12 See Johansen (1995a), theorem 2.2, p. 14.

13 The expression for this determinant is found on p. 51 of Johansen (1995). The additional step of factoring out the unit root term is achieved using the formula for the determinant of partitioned matrix provided by Dhrymes (1984, p. 37).
14 Johansen's theorem 2.2 establishes that the necessary and sufficient condition for a VAR to be stationary is that all the roots lie outside the unit circle.
15 The nature of the projection matrices is such that C may also be written $C = \beta_{\perp}(\alpha'_{\perp}\Gamma\beta_{\perp})^{-1}\alpha'_{\perp}$.
16 The method of maximum likelihood is not discussed here, although its relevance is described in Appendix C. For an introduction see Patterson (2000) or Sargan (1988).
17 Condition (4.59) is not usually considered in applied work. Instead, the series are individually tested to confirm that they are I(1).
18 This result is known as the Frisch–Waugh theorem. See, for example, Davidson and MacKinnon (1993, p. 19).
19 There are a number of standard programmes that can be used to solve eigenvalue problems, Doornik (1995) prefers the singular value decomposition which limits the problem to a solution in terms of positive/negative semi-definite matrices.
20 Note that since $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} \geq \lambda_n \geq 0$, then

$\lambda_j = 0 \Rightarrow \lambda_i = 0,\ i = j, \ldots, n.$

21 Note that the trace statistics can be written as the sum of a series of $\lambda_{\max}$ statistics:

$$\lambda_{trace}(j-1) = \left[\sum_{i=j}^{n} \lambda_{\max}(i-1)\right].$$

22 We will also see later that dummy variables and stationary variables may be included in the VECM and the number of these that are included also effects the critical values. This type of sensitivity is typical of tests and estimation procedures involving non-stationary processes.
23 This is not strictly correct since the rejection of the previous null was achieved using a different test. However, this is the way the non-rejection of the null would be interpreted in a sequential testing procedure, so it is stated as the null for convenience.
24 In addition, it is the last test of the sequence that examines whether the data is stationary or I(1), yet this is in practice a property that is pre-tested using unit root tests. That is, this is not the last but the first specification issue to be decided.
25 We would like to thank Paul Fisher and Ken Wallis for providing us with the data.
26 Hence, for r = 2, α and β are 6×2 dimensioned matrices.
27 For i = 1, with $\lambda_1 = .0827$, T = 60, the max test is $\lambda_{\max}(1) = -T\log(1 - \lambda_1) = -60\log(1 - .0827) \approx 5.18$, and for i = 2, $\lambda_{\max}(2) = -T\log(1 - \lambda_2) = 8.08$. The trace test is the sum of the max tests and for i = 2, $\lambda_{trace}(2) = 5.15 + 8.08 = 13.41$.
28 When the small sample adjustment due to Reimers (1994) is used to test whether there are r = 4 or more cointegrating vectors, then the revised test statistic is 14.2 and the test is marginally rejected at the 5% level. The test adjusts for the number of observations by correcting for degrees of freedom, but this corrected statistic is not necessarily any more reliable than the Johansen test statistic. More specifically, it is known that shift dummies will alter the distribution of the test statistic (Johansen (1995), while centred seasonal dummies do not. However, the critical values used here are based on $T = 50$ and are again taken from Frances (1994).
29 The one-step Chow test is based on recursive estimation starting with an initial sample of $M - 1$ observations and then re-estimated over samples M, $M + 1$ … T.

Here $M = 50$ and $T = 74$. To give a perspective on the choice of the initial sample, following the Sargan rule for model parameterization $k < \frac{T}{3}$. The minimum sample when $k = 18$ (a VAR has 8 constants/dummies and 2×5 lag coefficients) is $n = 3k$: = 54. For simplicity the recursive estimates were derived from $M = 50$ observations, but the first four calculations must in each case be viewed with caution.

30 The Cauchy is generated by a ratio of normals. Where nominal variables and prices are normally distributed then their ratio would not converge in distribution to normality.

31 In practice, $C(1)$ is not always of rank n – r, so that there may be insufficient zero roots. When C(L) has n – r zero roots, then $C(z) = C_0(z)C_1(z)$ and $C_1(z)$ is of degree q – 1. If there are insufficient zero roots, this can be rectified by extending the polynomial. Consider $zC^*(z)$, then $zC^*(z) = z^nC^*(z)$ and this extension introduces n additional null roots. For the extended model $C^*(z) = C_0(z)C_1(z)$ where $C_1(z)$ is of degree q and $C_0(z)$ is defined above.

32 Hunter and Simpson (1995) suggested that the system should be re-ordered on the basis of tests of weak exogeneity.

5 Exogeneity and Identification

1 I would like to thank Graham Mizon for his discussion of this issue.

2 For the cointegrating exogenous case in Hunter (1992a) $r_1 = 1$ and $n_1 = 3$.

3 Hunter (1992a) uses the same data set as Fisher et al. (1990).

4 Hence, α and β are 6×2 dimensioned matrices.

5 The restrictions for SE of the oil price are $\alpha_{ij} = 0$ and $\beta_{ij} = 0$ for $j = 1, 2$.

6 In fact the same test can be used by re-running the Johansen procedure with the variable to be tested for WE being placed first.

7 The restrictions for CE are $\alpha_{51} = 0, \alpha_{61} = 0$ and $\beta_{j2} = 0$ for $j = 1, \ldots, 4$.

8 Johansen (2002) provides a correction factor when α is known for linear restrictions of the form $\beta = H\phi$. This appears to work well when the correction is less than 2, which generally implies that there are about 100 observations in the sample available for estimation. The correction factor used to weight the Likelihood Ratio test is:

$$1 + \frac{1}{T}\left[(n_d + n_D + kn) + \frac{1}{2}(n + 1 + s - r)\right]$$
$$+ \frac{1}{Tr}\left[(n - 2r + s + 2n_D - 1)v + 2(c + c_\alpha)\right].$$

However, this is not trivial as c, c_α and v need to be evaluated. For the case where there are no higher-order dynamics, the non-stationary series are all random walks and r = 1, then:

$$c = -2\frac{\alpha'\beta(1 + \alpha'\beta)}{\beta'\Sigma\beta\alpha'\Sigma^{-1}\alpha}$$
$$c_d = -\frac{\alpha'\beta(2 + \alpha'\beta)}{\beta'\Sigma\beta\alpha'\Sigma^{-1}\alpha} = v.$$

Otherwise, $c_d \neq$ v and these terms are derived from the trace of products of the matrices in $\Gamma(L)$, β', α, Σ and Σ^{-1}. Calculations not readily available in existing software.

9 The WE restrictions for the model in Hunter (1992) are $\alpha_{i1} = 0$ for $i = 4, 5, 6$ and $\alpha_{j2} = \omega_{j1}\alpha_{42} + \omega_{j2}\alpha_{52} + \omega_{j3}\alpha_{62}$ for $j = 1, 2, 3$.

10 Strong exogeneity augments the sub-block WE restriction above by $\beta_{i2} = 0$ for $i = 1, 2, 3$.

11 If an equation has exactly $j_i = n_1 - 1$ restrictions, then it has enough restrictions to be exactly identified. When $j_i > n - 1$, enough restrictions to be over-identified, but without the appropriate number of restrictions it will be under or not identified.

12 The hessian is the second derivative of the likelihood, which provides an estimate of the variance–covariance matrix of the parameters. If some parameters are ill-defined, then the likelihood is flat and the hessian matrix singular. Then some parameters in the model are not identified. Perfect multi-collinearity is a special case of this and it occurs when two or more variables are related and their parameters cannot be independently estimated and as a result are not identifiable.

13 A number of authors have considered this issue, Hunter and Simpson (1995) and Juselius (1995).

14 Usually, two equations or blocks of equations with different parameterizations have the same value for their likelihoods. In general, there exists at least one model with r exactly identifying restrictions per equation with a likelihood value, the same as the unrestricted likelihood.

15 The approach described here was first outlined for the I(1) case in Hunter and Simpson (1995).

16 This does not account for multi-cointegration (Granger and Lee 1989) and polynomial cointegration (Yoo 1986), which does introduce lags into the long-run relationships.

17 The additional restriction is required to solve for all of the parameters and of the eight restrictions implied by $\beta_r' = \begin{bmatrix} \beta_{11} & \beta_{21} & -\beta_{21} & -\beta_{21} & 0 & \beta_{61} \\ 0 & 0 & 0 & 0 & \beta_{52} & -\beta_{52} \end{bmatrix}$ only six are binding. The test associated with this structure for β is $\chi^2_6 = 6.8291$, which is accepted at the 5% level based on a p-value = [0.3369].

18 For every row and column of Π selected, there is an equivalent r dimensioned sub-matrix of α and β. To determine an appropriate orientation of the system the sub-matrices selected need to be of full rank.

19 In the case where more complex restrictions apply, then the general restriction condition and procedure in Doornik and Hendry (1996) apply.

20 Here β is identifiable for the restrictions in (I) when the selected columns of Π yield a matrix A of rank r.

21 For $n = 4$, a more complex example, the approach discussed above can be shown to identify. Let:

$$\beta' = \begin{bmatrix} a & 0 & b & c \\ d & e & f & 0 \end{bmatrix} \text{ and } B = H_2 = \begin{bmatrix} a & 0 \\ d & e \end{bmatrix}.$$

Following Boswijk (1996), identifiability is lost when a normalization is invalid (i.e., $a = 0 \Rightarrow rank(H_2) < r$), but with this new restriction $[\alpha : \beta]$ is over-identified as $j = 3 > r^2 - r$. Selecting a new orientation, ensuring the generic result associated with Theorem 9 holds, then:

$$\beta'_{(1)} = \begin{bmatrix} 0 & 0 & b & c \\ d & e & f & 0 \end{bmatrix} \text{ and } B_{(1)} = \begin{bmatrix} b & c \\ f & 0 \end{bmatrix}.$$

This orientation is rejected when $x_t \sim I(1)$, $f = 0$ and α is not identifiable. But the following orientation for $x_t \sim I(1)$, implies:

$$\beta'_{(2)} = \begin{bmatrix} 0 & 0 & b & c \\ d & e & 0 & 0 \end{bmatrix}, \; B_{(2)} = \begin{bmatrix} 0 & b \\ e & 0 \end{bmatrix}$$

and $rank(B) = r$. Now, $[\alpha : \beta_{(2)}]$ is always empirically identified and identifiable.

22 The matrices α_{ij} and β_{ij} have the dimensions $n_i \times r_j$ for $i = 1, 2$ and $j = 1, 2$. For example, the matrix β is partitioned into two blocks of columns, $\beta_{.1}$ of dimensions $n \times r_1$, and $\beta_{.2}$ of dimensions $n \times r_2$, then each block is itself cut into two blocks of rows.

23 In the limit there are r such sub-blocks, which leads to the identification case considered by Boswijk (1992) where $\alpha = \begin{bmatrix} I \\ 0 \end{bmatrix}$.

24 The original source of the data is the National Institute of Economic Research, that has been kindly passed on to us by Paul Fisher and Ken Wallis.

25 The model in Hunter (1992a) is massively over-identified. It is possible to identify subject to restrictions on both α and β. Here we will concentrate on identification from α alone.

26 The discovery of four valid solutions implies that the model has four over-identifying restrictions.

27 If the determinant is tested for any sub-matrix of α then it is found that no such combination with non-zero determinant appears to exist.

6 Further Topics in the Analysis of Non-Stationary Time Series

1 Many series are often bounded to lie in the range [0, 1] as is the case for interest rates. The question of non-stationarity in this context is further complicated by the notion of what an extreme value might be. Maybe one should consider the performance of bond prices upon which the rate of return of the safe asset is derived. Then again the non-stationarity may be a function of the process of aggregation or the pricing formula. In practice all models are not identified, the models estimated are always approximations and the modellers task is to limit the degree of non-identification (Sargan 1983a).

2 Here our analysis is restricted to the case where trends are possible via unrestricted intercepts in the conventional cointegration analysis ($\mu_1 \neq 0$), but there are no quadratic trends. Otherwise, the second step of the I(2) estimator has a restricted intercept ($\mu_2 = 0$). This is the case considered by Johansen (1995) and, unlike Paruolo (1996), it restricts our discussion to a single table. In the emprical example considered by Paruolo (1996), he concludes that the selection of results associated with ${}_1Q_{r,s}$ is quite consistent when a pre-analysis of the data suggests that there are trends in the differences ($\mu_1 \neq 0$), but not the second differences ($\mu_2 = 0$).

3 For the example considered by Paruolo inference progressed in a straightforward manner, by sequentially moving past each test statistic a table at a time. For the case considered here, the progress is more complicated, even when one only considers the table of tests associated with $\mu_1 \neq 0$.

4 This is an extension of the I(1) case where $\alpha\zeta\zeta^{-1}\beta' = \alpha^*\beta^{*\prime}$ meaning that the estimated loadings and cointerating vectors are not distinguished from any non-singular matrix product. That is $[\alpha, \beta']$ and $[\alpha^*, \beta^{*\prime}]$ are observationally equivalent. Now this problem is further complicated in the I(2) case.

5 The values of d estimated are found to be sensitive to the bandwidth m. A common assumption made in the literature on evaluating standard errors in cointegrating regressions is to set the bandwidth to a third of the sample, $m = \frac{T}{3}$. Alternatively, Henry and Robinson (1996) provide some methods for the selection of m.

6 More generally $C(L) = \frac{\Theta(L)}{\Phi(L)}$ with the roots to the two finite polynomials all lying outside the unit circle.

7 In both cases, the lag specification of the models used is that known from the data generation process (DGP) and not determined empirically at each replication. This can be expected to improve the performance of long-run forecasts, but not

necessarily the short-run forecasts, as the sample dynamics may be better described by some other order than that used to generate the data.

8 Roots in the paper are the reciprocals of those normally reported, thus a root less than one in modulus is a stationary root. On this basis the roots of the process are, respectively: {0.5, 0.5, 0.5, 0.5}, {0.5, 0.5, 0.95, 0.95}, {0.5, 0.5, 0.99, 0.99}, {0.5, 0.5, 1.0, 1.0}, {1.0, 1.0. 1.0, 1.0}.

9 An alternative study would be one based on perturbations of the cointegrated model, model 4, that retained the common feature, but moved it from being at the unit root to being further outside the unit circle. This would mean that the processes became stationary, and more solidly so, but retained the reduced rank property key to cointegration. In this way, it is possible to isolate two aspects of the problem with potentially different impacts: stationarity and common features (reduced rank).

10 This conclusion can also be drawn by comparing the scaling on the vertical axes of Figures 6.1a–6.1e, whence it will be seen that much the smallest scale is employed in Figure 6.1c.

11 Figure 6.1d also shows clearly that, in this case, under-specification of the co-integrating rank is not harmful to forecast performance (including imposing unit roots), whereas over-specification leads to a deterioration in forecasting performance.

12 Though they do not establish whether it is the imposition of any false restriction that matters, or that of unit roots in particular. This is the point made by Clements and Hendry. They also do not consider if the near unit root is a common feature, or if restricting it to being so would be advantageous.

13 The information criterion can also be written in terms of the eigenvalues of the underlying problem, and hence in terms of the test statistics.

7 Conclusions

1 Johansen (2002b) provides a small sample correction to the rank test for cointegration $r = 0$ and $r = 1$. The correction factors are difficult to calculate, but based on the simulation results there can be considerable benefits to their use. Based on the study of a four-variable model of Danish Money, the critical values are adjusted by anything between 1.14 and 1.07 for $t = 50, 100$. For the empirical results in section 4.6.2 such adjustments would not affect the conclusions associated with the trace test for $r = 0$ and $r = 1$. Quite clearly such an adjustment might alter our conclusions when $r > 1$. Even so, the critical values used here were taken from Franses (1994), which assumed $T = 50$. Further, wrong rejection of the null might not be of paramount importance when over-rejection of the alternative of cointegration is what is critical to the applied researcher. Hence, were the true size of the test 10%, then over-rejection of the null might not be a problem, but cases where the size is considerably larger ought to be avoided. In particular, test properties are likely to be very poor when some series are I(2), because conventional tests of cointegration require all the series in the VAR to be no more than I(1). When there are I(2) series in the VAR, this violates the necessary and sufficient condition required for the cointegrating relationships to exist. Johansen shows that the correction increases in line with the true size of the test as the series tend to become I(2) and in the limit non-cointegration is always rejected. The reader is referred back to section 4.4.2 and 4.6.2.

2 For further discussion of such issues the reader is directed to Granger and Hallman (1991), Granger (1995) and Corradi et al. (2000).

3 As is mentioned in Haug (1996), non-gaussianity ought not to be crucial, as long as, sums of the residual vector of the Johansen VAR can be approximated by vector brownian motion. (Johansen 1991, Appendix E).
4 From the point of view of an analysis of an economic system, it is a moot point whether it is desirable to infer equilibrium where, although it exists, it does not have fixed long run coefficients or where the rate of adjustment towards equilibrium varies.
5 And further, the technique would not illuminate a situation where the adjustment coefficients and the cointegrating vectors change over time, but in such a way that the II matrix remains constant. This would correspond to a situation where the nature of the equilibrium relationships was developing, but being compensated for changing adjustment coefficients.

Appendix A

1 Notice that in this simple case the (2,2) element of $C_3(L)$ is $|C(L)|$.

Appendix C

1 The eigenvalue problem is solved with respect to both α and β under some of the restrictions considered in chapter 5, while the likelihood associated with general restrictions, applied to both α and β, is presented in Appendix F.

Appendix E

1 This statement implies that, for u > 0, $w(u) \sim N(0, u)$.
2 For $x > 0$, x can be written $x = X + \delta$, where X is a non-negative integer and $0 \leqslant \delta < 1$. Then $\langle x \rangle = X$.
3 To be more precise, let X be a random variable and x represent a value taken by X. Also, let X_T be a sequence of random variables. Let the distribution function of X be $F(.)$ and that of X_T be $F_T(.)$. Then $F_T(.)$ is said to converge weakly to $F(.)$ if

$$F_T(x) = \Pr(X_T \leq x) \rightarrow \Pr(X \leq x) = F(x) \quad as \quad T \rightarrow \infty.$$

4 For a proof of this result, see McCabe and Tremayne (1993, chapter 8).
5 See Johansen (1995, p. 151) for details.
6 Weak convergence, in contrast, indicates the convergence of one random variable to another.
7 Technically $\delta \xrightarrow{p} 0$, or, equivalently, δ is said to be $o_p(1)$.
8 For details, see Johansen (1995, p. 158).
9 These generalizations break down the residual product moment matrices in terms of components in the cointegrating space and orthogonal to it.
10 Davidson (1994) provides detailed discussion of different types of stochastic convergence.
11 Pesaran, Shin and Smith (2000) extended this set up by allowing exogenous I(1) variables, which distorts the distributions.
12 MacKinnon, Haug and Michelis (1999) find that using Monte Carlo simulations based on 400 observations leads to quite inaccurate results, especially when $n - r$ is large. They use a response surface estimated across a range of experiments using different sample sizes. This method calculates the relevant percentile, say the 95*th*,

appropriate for a test of 5%, for each set of Monte Carlo experiments using a particular DGP, and regresses this on the characteristics of the DGP. In the simplest form the dependent variables are an intercept and powers of the reciprocal of the sample size, such that the estimated intercept is the estimated asymptotic critical value of the test. Critical values for other sample sizes are obtained by using the estimated regression to predict substituting the relevant value for *T*. This approach is also used in MacKinnon (1991) for unit root and residual based cointegration tests.

13 Asymptotic tests are those based on finite samples of data but using asymptotic critical values.

Appendix G

1 The normalization adopted by Hunter and Simpson implies that the first vector is an inflation equation, the second an exchange rate equation, the third a terms of trade or real exchange rate equation and the fourth a real interest rate equation.

Appendix A: Matrix Preliminaries

A.1 Elementary row operations and elementary matrices

In what follows, the word row can be replaced by the word column to define an elementary column operation. There are three types of elementary row operation:

(i) The interchange of two rows.
(ii) Multiplication of one row by a constant.
(iii) Addition of one row to another row times a polynomial.

A left (right) elementary matrix is a matrix such that, when it multiplies from the left (right) it performs an elementary row (column) operation. The matrix formed from the product of such matrices therefore performs the same transformation as a sequence of such row (column) operations. For example, consider the use of row and column operations to diagonalize the 2×2 finite order polynomial matrix,

$$C(L) = \begin{pmatrix} 1 - \frac{3}{4}L & -L \\ -\frac{1}{8}L & 1 - \frac{1}{2}L \end{pmatrix}. \tag{A.1}$$

Row operation 1 (objective to alter the (1,1) element to unity): replace row 1 by row 1 minus 6 times row 2. This can be achieved by pre-multiplication by the matrix

$$\begin{pmatrix} 1 & -6 \\ 0 & 1 \end{pmatrix}.$$

The new matrix is

$$C_1(L) = \begin{pmatrix} 1 & -6 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 - \frac{3}{4}L & -L \\ -\frac{1}{8}L & 1 - \frac{1}{2}L \end{pmatrix} = \begin{pmatrix} 1 & -6 + 2L \\ -\frac{1}{8}L & 1 - \frac{1}{2}L \end{pmatrix}. \tag{A.2}$$

Row operation 2 on $C_1(L)$ (objective to alter the (2,1) element to zero): replace row 2 by row 2 plus $\frac{1}{8}L$ times row 1. This can be achieved by pre-multiplication by the matrix

$$\begin{pmatrix} 1 & 0 \\ \frac{1}{8}L & 1 \end{pmatrix}.$$

The new matrix is

$$C_2(L) = \begin{pmatrix} 1 & 0 \\ \frac{1}{8}L & 1 \end{pmatrix} \begin{pmatrix} 1 & -6 + 2L \\ -\frac{1}{8}L & 1 - \frac{1}{2}L \end{pmatrix} = \begin{pmatrix} 1 & -6 + 2L \\ 0 & 1 - \frac{5}{4}L + \frac{1}{4}L^2 \end{pmatrix}. \tag{A.3}$$

Column operation 1 on $C_2(L)$ (objective to alter the (1,2) element to zero): replace column 2 by column 2 minus $(2L - 6)$ times column 1. This can be achieved by post-multiplication by the matrix

$$\begin{pmatrix} 1 & -(2L - 6) \\ 0 & 1 \end{pmatrix}.$$

The new matrix is

$$C_3(L)=\begin{pmatrix}1 & -6+2L\\ 0 & 1-\frac{5}{4}L+\frac{1}{4}L^2\end{pmatrix}\begin{pmatrix}1 & -(2L-6)\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1-\frac{5}{4}L+\frac{1}{4}L^2\end{pmatrix}.$$

The elementary matrices of the row operations can be multiplied together (retaining the order of multiplication) as

$$G(L)=\begin{pmatrix}1 & 0\\ \frac{1}{8}L & 1\end{pmatrix}\begin{pmatrix}1 & -6\\ 0 & 1\end{pmatrix}=\begin{pmatrix}1 & -6\\ \frac{1}{8}L & 1-\frac{3}{4}L\end{pmatrix}$$

and writing the elementary matrix of the column operation as

$$H(L)=\begin{pmatrix}1 & -(2L-6)\\ 0 & 1\end{pmatrix}$$

the diagonalized matrix may be written

$$C_3(L)=G(L)C(L)H(L).$$

A.2 Unimodular matrices

Consider the matrices defining the elementary operations in the previous example. Note that both $G(L)$ and $H(L)$ are matrix polynomials. In general, the determinant of a polynomial matrix would be a polynomial in L. But in this case

$$|H(L)|=\left|\begin{pmatrix}1 & -(2L-6)\\ 0 & 1\end{pmatrix}\right|=1$$

and

$$|G(L)|=\begin{vmatrix}1 & -6\\ \frac{1}{8}L & 1-\frac{3}{4}L\end{vmatrix}=1$$

and so are not functions of L. Furthermore, because the determinant is non-zero, the matrices are invertible. Such matrices are known as *unimodular matrices* (having constant non-zero determinant). Usefully, all elementary matrices are unimodular and so therefore is the product of two or more elementary matrices. It is therefore possible to invert the transformation and express $C(L)$ in terms of $C_3(L)$ as

$$C(L)=G(L)^{-1}C_3(L)H(L)^{-1}.$$

A.3 Roots of a matrix polynomial

Let *A(L)* be an *n* by *n* matrix polynomial of order p, and let $|A(L)|$ be its determinant. Then z is a root of *A(L)* if $|A(z)| = 0$. The maximum number of roots possible is *np*. For example:

$$A(L)=\begin{bmatrix}1-\frac{3}{4}L & -L\\ -\frac{1}{8}L & 1-\frac{1}{2}L\end{bmatrix},$$

$$\text{so}\left|A(z)\right|=\left(1-\frac{3}{4}z\right)\left(1-\frac{1}{2}z\right)-\frac{1}{8}z^2=1-\frac{5}{4}z+\frac{1}{4}z=(1-z)\left(1-\frac{1}{4}z\right),$$

and the roots are therefore $z = 1$ and $z = 4$.

An important special case is that of a unit root. If $A(L)$ has a unit root then $|A(1)| = 0$, that is *A(1)* is singular. This the case in the example above by putting $z = 1$.

Appendix B: Matrix Algebra for Engle and Granger (1987) Representation

B.1 Determinant/adjoint representation of a polynomial matrix

Consider the matrix inverse, $A^{-1} = A^a/|A|$, where A^a is the adjoint matrix of A. Let z be a scalar complex number. The corresponding condition for the existence of the inverse of a square polynomial matrix $G(z)$ is that $|G(1)| \neq 0$. If this is satisfied, then the inverse polynomial may be written

$$G^{-1}(z) = G^a(z) / \left|G(z)\right|. \tag{B.1}$$

In particular, let $G(z)$ be an $n \times n$ polynomial matrix with $G(0) = I_n$. That is,

$$G(z) = I_n - \sum_{i=1}^{m} G_i z_i$$

where the G_i, $i = 1, 2, \ldots, m$ are $n \times n$ coefficient matrices. Then, denoting the determinant of $G(z)$ by $|G(z)|$ and its adjoint by $G^a(z)$:

$$G^a(z)G(z) = \left|G(z)\right| I_n. \tag{B.2}$$

Note that $G^a(z)$ is an $n \times n$ polynomial matrix or order at most $m \times (n - 1)$, and $|G(z)|$ is a scalar polynomial of order at most $m \times n$.

B.2 Expansions of the determinant and adjoint about $z \in [0, 1]$

The proof of this lemma may be found in Engle and Granger (1987). Now consider the case where $G(z)$ may be of reduced rank at $z = 0$, consider the expansion

$$G(z) = G(0) + zG^*(z).$$

Now, *rank* $(G(0)) = n - r$, $0 \leq r \leq n$ and $z \in [0,1]$. For the case considered here $G^*(0) \neq 0$ and the determinant of the polynomial in z is:

$$\left|G(z)\right| = z^r g(z),$$

where

$$g(z) = \sum_{i=0}^{a} g_i z^i, a \leq (m \times n) - r,$$

g_i being scalar coefficients, and the adjoint polynomial is

$$G^a(z) = z^{r-1} H(z),$$

where

$$H(z) = \sum_{i=0}^{b} H_i z^i.$$

It follows that the index on the sum is limited by $b = (m \times [n-1]) - r + 1$ with H_i being $n \times n$ coefficient matrices. If $G(z)$ is originally of infinite order, then a and b are also infinite.

B.3 Drawing out a factor of z from a reduced rank matrix polynomial

It is possible to extract a factor of z from a matrix polynomial $G(z)$ for the singular case where $G(0)$ is a reduced rank polynomial. If $G(0)$ is singular then $r \geq 1$, and substituting out for the adjoint and the determinant of $G(z)$ from (*B*.2) gives

$$G^a(z)G(z) = z^{r-1}H(z)G(z) = \\ = |G(z)|I_n = z^r g(z)I_n.$$

Dividing left and right by z^{r-1} and arranging the polynomials in z

$$H(z)G(z) = zg(z)I_n. \tag{B.3}$$

Pre-multiplying $G(z)$ by $H(z)$ extracts a factor of z and reduces the expression to a scalar diagonal form:

Application to lag polynomial to draw out a unit root factor

Let $A(L)$ be a $n \times n$ lag polynomial matrix or order m. This may be written instead as a polynomial of order m in $\Delta = (1 - L)$ using $1 - \Delta = L$ so that

$$A(L) = I_n - \sum_{i=1}^{m} A_i L^i = I_n - \sum_{i=1}^{m} A_i (1-\Delta)^i.$$

By application of the binomial expansion to the terms $(1 - \Delta)^i$ this can be shown to be a polynomial of order m in Δ. For easy application of equation (B.3), let $A(L) \equiv G(\Delta)$

$$A(L) = I_n - \sum_{i=1}^{m} A_i L^i = I_n - \sum_{i=1}^{m} A_i (1-\Delta)^i = G(\Delta).$$

Now consider the z transform by setting $z = \Delta$:

$$G(z) = I_n - \sum_{i=1}^{m} A_i (1-z)^i.$$

Now consider $G(z)$ evaluated at the zero frequency

$$G(0) = I_n - \sum_{i=1}^{m} A_i (1-0)^i = I - \sum_{i=1}^{m} A_i = A(1).$$

It is also important to recall that in the reduced rank case $G(0)$ must be singular as is $A(1)$. Assuming this condition to be satisfied, then replace z in equation (B.3) by Δ to give

$$H(\Delta)G(\Delta) = \Delta g(\Delta)I_n. \tag{B.4}$$

Both $H(\Delta)$ and $g(\Delta)$ may be written as polynomials of L (of unchanged order), say $\tilde{H}(L)$ and $\tilde{g}(L)$ respectively, and so (B.4) may be written

$$\tilde{H}(L)A(L) = \Delta\tilde{g}(L)I_n. \tag{B.5}$$

Equation (B.5) states that pre-multiplying $A(L)$ by $\tilde{H}(L)$ results in a scalar diagonal lag polynomial matrix with a scalar factor in the difference operator Δ.

Appendix C: Johansen's Procedure as a Maximum Likelihood Procedure

The starting point for obtaining the maximized log-likelihood function in terms of the relevant eigenvalues is a multivariate Gaussian distribution. From this assumption follow the maximum likelihood estimates of the cointegrating vectors as particular eigenvectors and the expression of the maximized likelihood in terms of the subset of the corresponding eigenvalues. This in turn leads to simple expressions for test statistics based on the comparison of maximized likelihoods, since these too will depend on the relevant eigenvalues. Not all distributional assumptions will lead to these results and, as such, the Johansen procedure can be said to depend on the Gaussian assumption. The distributional assumption is that the disturbances of the VAR follow a multivariate Gaussian distribution. That is:

$$\Delta x_t = \Pi x_{t-1} - \sum_{i=1}^{p-1} \Gamma_i^* \Delta x_{t-i} + \epsilon_t. \tag{C.1}$$
$$\Pi = \alpha\beta' \text{ and } \epsilon_t \sim NIID(0,\Sigma)$$

The individual disturbance vector ϵ_t has density

$$f(\epsilon_t|\Sigma) = (2\pi)^{-\frac{1}{2}n} \exp\left\{-\frac{1}{2}\epsilon_t'\Sigma^{-1}\epsilon_t\right\}$$

giving rise to the density for Δx_t from the VECM, conditional on past values, as

$$g(\Delta x_t|\alpha,\beta,\Gamma_i^*,\Sigma) = (2\pi)^{-\frac{1}{2}n}|\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\Delta x_t - \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\Gamma_i^*\Delta x_{t-i}\right)\right.$$
$$\left.\times\Sigma^{-1}\left(\Delta x_t - \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\Gamma_i^*\Delta x_{t-i}\right)\right\}.$$

The natural logarithm of the joint density of Δx_t $t = 1, 2, \ldots, T$, ignoring initial values for convenience, is

$$G(\Delta x_t, t = 1,2,\ldots,T|\alpha,\beta,\Gamma_i^*,\Sigma) = -\frac{1}{2}nT\log(2\pi) - \frac{1}{2}T\log(|\Sigma|)$$
$$-\frac{1}{2}\sum_{t=1}^{T}\left\{\left(\Delta x_t - \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\Gamma_i^*\Delta x_{t-i}\right)'\Sigma^{-1}\left(\Delta x_t - \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\Gamma_i^*\Delta x_{t-i}\right)\right\}.$$

Thus the log-likelihood of the VECM (conditional on the data), minus the constant term $-\frac{1}{2}nT\log(2\pi)$ is given by

$$\log L(\alpha,\beta,\Gamma_i^*,\Sigma) = -\frac{1}{2}T\log(|\Sigma|)$$
$$-\frac{1}{2}\sum_{t=1}^{T}\left\{\left(\Delta x_t - \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\Gamma_i^*\Delta x_{t-i}\right)'\Sigma^{-1}\left(\Delta x_t - \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1}\Gamma_i^*\Delta x_{t-i}\right)\right\}.$$

This expression and subsequent algebra is simplified by re-expressing the log likelihood in terms the following:

$$\bar{z}_{0,t} = \Delta x_t, \bar{z}_{1,t} = x_{t-1},$$
$$\bar{z}'_{2,t} = \left[\Delta x'_{t-1} \ \ldots \ \Delta x'_{t-(p-1)}\right]'$$

and $\Gamma = [\Gamma_1 \ldots \Gamma_{p-1}]$. Then the log likelihood can be written:

$$\log L(\alpha,\beta,\Gamma,\Sigma) = -\frac{1}{2}T\log\left(|\Sigma|\right) - \frac{1}{2}\sum_{t=1}^{T}\left\{\left(\bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} + \Gamma\bar{z}_{2,t}\right)'\Sigma^{-1}\left(\bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} + \Gamma\bar{z}_{2,t}\right)\right\}.$$

This function may be maximized with respect to Γ alone giving rise to an expression for the maximum likelihood estimator for Γ in terms of the data and the other parameters of the model.

Denote this estimator as $\bar{\Gamma}$. By differentiating the log likelihood with respect to $\bar{\Gamma}$ and solving the first-order conditions, $\bar{\Gamma}$ is given by

$$\bar{\Gamma} = M_{0,2}M_{2,2}^{-1} - \alpha\beta'M_{1,2}M_{2,2}^{-1}$$

where $M_{i,j} = \frac{1}{T}\sum_{t=1}^{T}\bar{z}_{i,t}\bar{z}'_{j,t}$. The values of α, β and Σ that maximize $\log L(\alpha, \beta, \Gamma, \Sigma)$ will also maximize this expression with Γ substituted for $\bar{\Gamma}$ – that is $\log L(\alpha, \beta, \bar{\Gamma}, \Sigma)$. The latter function is known as the concentrated likelihood function. Before writing it out in full, note that $\bar{\Gamma}$ appears in $\log L(\alpha, \beta, \bar{\Gamma}, \Sigma)$ only in the term $(\bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} - \bar{\Gamma}\bar{z}_{2,t})$ or its transpose, so $\bar{\Gamma}$ appears in the concentrated log-likelihood only in $(\bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} - \bar{\Gamma}\,\bar{z}_{2,t})$. But

$$\begin{aligned}\bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} - \bar{\Gamma}\bar{z}_{2,t} &= \bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} - (M_{0,2}M_{2,2}^{-1} - \alpha\beta'M_{1,2}M_{2,2}^{-1})\bar{z}_{2,t}\\ &= (\bar{z}_{0,t} - M_{0,2}M_{2,2}^{-1}\bar{z}_{2,t}) - \alpha\beta'(\bar{z}_{1,t} - M_{1,2}M_{2,2}^{-1}\bar{z}_{2,t}).\end{aligned}$$

Define

$$R_{0,t} = \bar{z}_{0,t} - M_{0,2}M_{2,2}^{-1}\bar{z}_{2,t} \quad \text{(C.2)}$$
$$R_{1,t} = \bar{z}_{1,t} - M_{1,2}M_{2,2}^{-1}\bar{z}_{2,t} \quad \text{(C.3)}$$

so that

$$\bar{z}_{0,t} - \alpha\beta'\bar{z}_{1,t} - \bar{\Gamma}\bar{z}_{2,t} = R_{0,t} - \alpha\beta'R_{1,t}$$

and note that $R_{0,t}$ and $R_{1,t}$ are the residuals from the least squares regression of $\bar{z}_{0,t}$ and $\bar{z}_{1,t}$ respectively on $\bar{z}_{2,t}$. Using this residual notation, the concentrated log-likelihood may be written

$$\log L(\alpha,\beta,\Sigma) = -\frac{1}{2}T\log\left(|\Sigma|\right) - \frac{1}{2}\sum_{t=1}^{T}\{(R_{0,t} - \alpha\beta'R_{1,t})'\Sigma^{-1}(R_{0,t} - \alpha\beta'R_{1,t})\}.$$

This likelihood function may be concentrated further to express it in terms of β only. Regarding β as fixed in $\log L(\alpha, \beta, \Sigma)$ above and solving the first-order conditions for α and Σ, their maximum likelihood estimators may be written

$$\hat{\alpha} = S_{0,1}\beta(\beta'S_{1,1}\beta)^{-1},$$
$$\hat{\Sigma} = S_{0,0} - S_{0,1}\beta(\beta'S_{1,1}\beta)^{-1}\beta'S_{1,0},$$
$$\text{where } S_{i,j} = \frac{1}{T}\sum_{i=1}^{T}R_{i,t}R'_{j,t}.$$

A property of a Gaussian log-likelihood such as log L (α, β, Σ), evaluated at the maximum, is that it may be expressed in terms of $\hat{\Sigma}$ alone, as

$$\log L_{MAX} = -\frac{T}{2}\left(\log\left(\left|\hat{\Sigma}\right|\right) + n\right).$$

Thus, finally, the only term of interest in the concentrated likelihood, i.e. in log L_{MAX}, is $\hat{\Sigma}$, which itself is a function only of β (and the data). It therefore remains only to maximize log L_{MAX} with respect to β. Clearly the value of β that maximizes log L_{MAX} also maximizes log $\tilde{L} = -\frac{T}{2}\log(|\hat{\Sigma}|)$, since the difference is a constant term (a multiplicative term in the likelihoods themselves). The problem is to obtain the value of β that maximizes log $\tilde{L}$. By definition, this will be the maximum likelihood estimator. Equivalently, the problem is to minimize

$$Q(\beta) = \left|\hat{\Sigma}\right| = \left|S_{0,0} - S_{0,1}\beta(\beta' S_{1,1}\beta)^{-1}\beta' S_{1,0}\right|.$$

The solution to this problem is obtained by first re-expressing $Q(\beta)$ using the formulae for the determinant of a partitioned matrix. In general, for any matrix

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$$

with invertible diagonal blocks

$$|A| = \left|A_{1,1}\right|\left|A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2}\right| = \left|A_{2,2}\right|\left|A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}\right|.$$

Equating these two expressions for $|A|$ and rearranging gives

$$\left|A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}\right| = \left|A_{1,1}\right|\left|A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2}\right| / \left|A_{2,2}\right|.$$

Setting $A_{1,1} = S_{0,0}$, $A_{1,2} = S_{0,1}\beta$, $A_{2,1} = A'_{1,2}$ and $A_{2,2} = \beta' S_{1,1}\beta$ gives rise to the following expression

$$\begin{aligned} Q(\beta) = \left|\hat{\Sigma}\right| &= \left|S_{0,0} - S_{o,1}\beta(\beta' S_{1,1}\beta)^{-1}\beta' S_{1,0}\right| \\ &= \left|S_{0,0}\right|\left|\beta' S_{1,1}\beta - \beta' S_{1,0}S_{0,0}^{-1}S_{0,1}\beta\right| / \left|\beta' S_{1,1}\beta\right|. \end{aligned}$$

The optimum is found by minimizing

$$\begin{aligned} \tilde{Q}(\beta) &= \left|\beta' S_{1,1}\beta - \beta' S_{1,0}S_{0,0}^{-1}S_{0,1}\beta\right| / \left|\beta' S_{1,1}\beta\right| \\ &= \left|\beta'(S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{0,1})\beta\right| / \left|\beta' S_{1,1}\beta\right|. \end{aligned} \tag{C.4}$$

Let $\hat{\beta}$ be the $n \times r$ matrix that minimizes $\tilde{Q}(\beta)$. Consider the solutions $\lambda = \lambda_i$ to the eigenvalue problem:

$$\left|\lambda I - S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}\right| = 0 \tag{C.5}$$

ordered so that $\lambda_1 > \lambda_2 > \ldots > \lambda_n$ Let $\hat{\beta}_i$ for $i = 1,2, \ldots, r$, be the eigenvectors corresponding to λ_i, $i = 1,2, \ldots, r$, the r largest eigenvalues. Then it is stated without proof that

$$\beta = \hat{\beta} = (\hat{\beta}_1 \ \ldots \ \hat{\beta}_r)$$

minimizes (C.4). Furthermore, the minimized function can be written:

$$Q(\hat{\beta}) = \left|S_{0,0}\right|\prod_{i=1}^{r}(1 - \lambda_i).$$

Thus, apart from constants, the maximized log-likelihood may be written

$$\log \tilde{L}_{MAX} = -\frac{T}{2}\left[\log\left|S_{0,0}\right| + \sum_{i=1}^{r}\log(1 - \lambda_i)\right]. \tag{C.6}$$

Since the $\hat{\beta}_i$ are eigenvectors, many normalizations are possible. A convenient choice for deriving the above expressions in terms of the eigen values follows from observing that the original eigen value problem is equivalent to solving

$$\left|\lambda S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{0,1}\right| = 0. \tag{C.7}$$

(This is known as solving for the eigenvalues of $S_{1,0}S_{0,0}^{-1}S_{0,1}$ in the metric of $S_{1,1}$.) Consequently the matrix of eigenvectors ($\hat{\beta}$) that diagonalizes $S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}$ also diagonalizes $S_{1,1}$ and $S_{1,0}S_{0,0}^{-1}S_{0,1}$ in the following manner:

$$\hat{\beta}'S_{1,1}\hat{\beta} = I,$$

$$\hat{\beta}'S_{1,0}S_{0,0}^{-1}S_{0,1}\hat{\beta} = \Lambda,$$

where $\Lambda = diag\ (\lambda_1, \lambda_2, \ldots, \lambda_r)$. It follows from the diagonalization that

$$\left|\beta'(S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{0,1})\beta\right| / \left|\beta'S_{1,1}\beta\right| = \left|I - \Lambda\right| / \left|I\right|$$

$$= \prod_{i=1}^{r}(1 - \lambda_i)$$

is the minimized value of (C.4), giving (C.6) as the maximized log-likelihood. Subject to conditioning the problem on the β, the values of $\hat{\alpha}$ and $\hat{\Sigma}$ can be calculated directly from the formulae above, but with this normalization reduce to:

$$\hat{\alpha} = S_{0,1}\hat{\beta}\left(\hat{\beta}'S_{1,1}\hat{\beta}\right)^{-1} = S_{0,1}\hat{\beta},$$

$$\hat{\Sigma} = S_{0,0} - \alpha\alpha'.$$

It follows that

$$\hat{\Pi} = \hat{\alpha}\hat{\beta}' = S_{0,1}\hat{\beta}\hat{\beta}'.$$

The determination of $\hat{\beta}$ and $\hat{\alpha}$ in this way ensures that $\hat{\Pi}$ is of rank $r \leq n$. Since the approach works regardless of whether $\hat{\Pi}$ is of full or reduced rank, the procedure is known as reduced rank regression. As will be indicated below, it is very closely related to the calculation of canonical correlations.

This analysis demonstrates that the Johansen approach rests on the Gaussian assumption in the following ways:

(i) Through concentration of the likelihood function it explains the generation of $R_{0,t}$ and $R_{1,t}$, and how this relates to the Gaussian likelihood.
(ii) The expression of the maximized likelihood in terms of the eigenvalues depends on the particular form of the concentrated likelihood function in terms of the ratio of the determinants of quadratic forms.[1]
(iii) The expressions for the likelihood ratio statistics in terms of the eigenvalues depends on the expression for the maximized log likelihood, and hence these too depend on the distributional assumption.

Appendix D: The Maximum Likelihood Procedure in Terms of Canonical Correlations

An interpretation of the maximum likelihood treatment described in Appendix B is available through the concept of canonical correlations. The problem can be viewed as one of finding the maximum extent of correlation between the two residuals series, $R_{0,t}$ and $R_{1,t}$ (C.2, C.3). However, rather than deal with the residuals as they are, arising as they do from individual equations in the system associated with a single process or its difference as the dependent variable, the correlations considered are between linear combinations of the $R_{0,t}$ and linear combinations of the $R_{1,t}$. It transpires that the correlations between the linear combinations of the residuals relate directly to the eigenvalues of the problem described in Appendix C (equation C.7), in such a way that maximizing the log-likelihood corresponds to choosing the r greatest correlations. The linear combinations of the $R_{1,t}$ that arise are the cointegrating combinations, where they exist.

The idea of canonical correlation is to transform two vectors of variables so that the elements of each vector have unit variance and are individually uncorrelated. In addition, only matching elements of the transformed vectors are correlated with each other. Since all transformed variates have unit variance, covariance matrices and correlation matrices are identical. This makes calculation easier.

Mapping the canonical correlation problem onto the maximum likelihood problem, the starting point is the residuals $R_{0,t}$ and $R_{1,t}$. These are transformed by pre-multiplication by $n \times n$ matrices A and B to give $R^{*}_{0,t} = AR_{0,t}$, and $R^{*}_{1,t} = B\,R_{1,t}$, where A and B must be chosen such that:

$$S^{*}_{i,j} = \frac{1}{T}\sum_{i=1}^{T} R^{*}_{i,t}R^{*'}_{j,t} = I \text{ for } i = j \text{ and } P \text{ otherwise}$$

where $P = diag\ (p_1, \ldots, p_n)$, $p_i > 0$, and, by appropriate ordering of the elements of $R_{0,t}$ and $R_{1,t}$, $p_1 \geq p_2 \geq \ldots \geq p_n$. As all the p_i are correlations and positive by construction, they lie on the [0,1] interval. They are called canonical correlations. The solutions to the problems of the selection of A and B are the solutions to two closely associated eigenvalue problems. Consider the matrices

$$H_0 = S^{-1}_{0,0}S_{0,1}S^{-1}_{1,1}S_{1,0}$$
$$H_1 = S^{-1}_{1,1}S_{1,0}S^{-1}_{0,0}S_{0,1}.$$

The eigenvalues of these two matrices are identical and given by the solution to equation (C.7) above. That is, they are the λ_i, $i = 1, 2, \ldots n$ of the maximum likelihood problem. The eigenvectors of H_0 are the solutions for the columns of A and are denoted a_i. They are chosen so that $a'_i\, S_{0,0}\, a_j = 1$ for $i = j$, *0* otherwise. The eigenvectors of H_1 have already been denoted β_i, and are normalized as before, so that $\beta'_i S_{1,1}\beta_j = 1$ for $i = j$, *0* otherwise. Thus B is an $n \times n$ matrix with i^{th}, column β_i. In addition, $R^2 = diag\ (\lambda_1, \ldots, \lambda_n)$, in other words, the eigenvalues are the squared canonical correlation. Thus, from the expression for the maximized log-likelihood of equation (C.6), the Johansen ML

procedure can be seen to be the calculation of the coefficients of the linear combinations of the non-stationary variables such that their correlation with the (canonically combined) stationary variables is maximized. For given r, the required linear combinations of the levels will be those using the eigenvectors, β_i, $i = 1, 2, \ldots, r$. In order to maximize correlation with stationary variables, the linear combinations of the I(1) variables will need to be as close to stationarity as possible. The problem is restricted by only considering the r most correlated combinations. The cointegrating rank r has to be determined by testing. The values of the model parameters are then obtained as outlined in Appendix C.

Appendix E: Distribution Theory

E.1 Some univariate theory

The fundamental building blocks of unit root asymptotic theory are convergence in distribution, or weak convergence, and the scalar Wiener process or Brownian motion, $b(u)$, defined as follows.

Let $b(u)$, $u \in [0, 1]$, be a continuous time stochastic process with $b(0) = 0$. The construct $b(u) - b(v)$ is called an *increment* of the process. Let $b(u_1) - b(u_2)$ and $b(u_3) - b(u_4)$ be two such non-overlapping increments, then

(i) $b(0) = 0$
(ii) $b(u) - b(v) \sim N(0, |u - v|) \quad \forall u \neq v$[1]
(iii) $E(b(u_1) - b(u_2))(b(u_3) - b(v_4)) = 0$.

The fundamental distributional result involving scalar Brownian motion is Donsker's theorem. In I(1) systems, partial sums of IID processes occur naturally and frequently. Donsker's theorem provides an approximate large sample distribution for such quantities in terms of Brownian motions. The term 'asymptotic' is normally used in place of 'approximate large sample'. For convenience, this simpler but not very informative abbreviation is used in what follows and in the main text.

Consider the sequence

$$\epsilon_t \sim IID(0, \sigma_\epsilon^2), \text{ for } t = 1, 2, \ldots, T$$

and partial sum

$$s_t = \sum_{i=1}^{t} \epsilon_i.$$

In order to employ Brownian motions to characterize the asymptotic distributions of such quantities, s_t has to be manipulated so as to relate to the unit interval. To do this, note that s_t can be thought of as the sum of the sequence up to some point a fraction of the way, say τ, into the complete sequence. To represent this notationally, let $\langle x \rangle$ represent the integer part of x.[2] Then for any $t = 1, 2, \ldots, T$ there is a $\tau \in [0, 1]$ such that $t = \langle T\tau \rangle$, allowing the partial sum to be written

$$s_{\langle T\tau \rangle} = \sum_{i=1}^{\langle T\tau \rangle} \epsilon_i.$$

As T increases, $s_{\langle T\tau \rangle}$ forms a sequence of random variables. In particular, in order to obtain the asymptotic distribution, interest is in the limit of this sequence as $T \to \infty$. Each of the random variables $s_{\langle T\tau \rangle}$ can be thought of as possessing a distribution function depending on T and τ, say $F_{T,\tau}(.)$. If there exists some distribution function $F_\tau(.)$ such that

$$F_{T,\tau}(.) \to F_\tau(.) \text{ as } T \to \infty,$$

then $F_\tau(.)$ is called the limiting distribution of the sequence $s_{\langle T\tau \rangle}$. It is said that $F_{T,\tau}(.)$ converges weakly to $F_\tau(.)$.[3] Notationally, $F_{T,\tau} \Rightarrow F_\tau$ means $F_{T,\tau}(.)$ converges weakly to

$F_\tau(.)$. Furthermore, if $S(\tau)$ is a random variable having distribution function $F_\tau(.)$, then $s_{\langle T\tau\rangle}$ is said to converge in distribution to $S(\tau)$. The notation for this ought strictly to be different since it involves random variables rather than their distribution functions, but the same symbol $\Rightarrow$ will be used. Otherwise, a commonly used notation for convergence in distribution is $\xrightarrow{D}$.

In fact, $s_{\langle T\tau\rangle}$ does not have a limiting distribution. In order to obtain convergence, it must be divided by $T^{\frac{1}{2}}$. Donsker's theorem defines the random variable to which $T^{-\frac{1}{2}} s_{\langle T\tau\rangle}$ tends in distribution in terms of Brownian motion. It states

$$T^{-\frac{1}{2}} \sum_{i=1}^{\langle T\tau\rangle} \epsilon_i \Rightarrow \sigma_\epsilon b(\tau) \quad for \quad \tau \in [0,1].^4$$

A further tool is needed since the asymptotic distributions required are those of functions of (normalized) partial sums. The continuous mapping theorem (CMT) states that, if a sequence of random variables tends in distribution to another random variable, then a given continuous function of the sequence tends to the same function of that limit. So, for any continuous function $g(.)$, the CMT states that

$$g\left(T^{-\frac{1}{2}} s_{\langle T\tau\rangle}\right) \Rightarrow g(\sigma_\epsilon b(\tau)).$$

An important example of a function to which the CMT applies is the integral. The CMT and Donsker's theorem can therefore be used to derive the Brownian motion characteristics of a wide range of random variables based on partial sums of IID random variables.

E.2 Vector processes and cointegration

The generalization from scalar to vector processes is necessary to deal with cointegration, but is straightforward. Standard vector Brownian motion with variance covariance matrix tI_n is defined as

$$B(\tau) = \begin{bmatrix} b_1(\tau) & \dots & b_n(\tau) \end{bmatrix}'$$

where the $b_i(\tau)$ are uncorrelated scalar Brownian motions. Thus, for $\tau \in [0, 1]$:

(i) $B(0) = 0$;
(ii) $B(\tau) \sim N(0, \tau I_n)$;

and the process has independent increments.

More generally, the process $W(\tau) = \Omega^{\frac{1}{2}} B(\tau)$ is encountered, having the same properties as $B(\tau)$ except that its variance-covariance matrix is $\tau\Omega$. Donsker's theorem now applies to partial sums of IID$(0,\Omega)$ vectors, ϵ_t, $t = 1, 2, \dots, T$, and is that

$$T^{-\frac{1}{2}} \sum_{i=1}^{\langle T\tau\rangle} \epsilon_i \Rightarrow W(\tau).$$

Johansen (1995, appendix B) discusses the key results needed to obtain the limiting distributions of the test statistics and estimators. Concerning the trace statistic, an important observation, allowing the application of the multivariate version of Donsker's theorem via the CMT, is that the eigenvalues on which the statistic depends, are continuous functions of product moments, the asymptotic distributions of which are available.

E.3 Testing the null hypothesis of non-cointegration

The main features in establishing the distribution of the test statistics are the following, where for simplicity, the case of the statistic for testing the null of zero cointegrating rank against the full rank alternative is considered, in the model with no deterministic terms.

(i) Establish the relationship between the eigenvalues that appear in the test statistic and the product moment matrices, $S_{i,j}$. This can be derived since the eigenvalues are the solutions to the problem $|\lambda I - S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}| = 0$. This is the standard eigenvalue problem, for which the solutions $\lambda = \lambda_i$, $i = 1, 2, \ldots, n$ are the eigenvalues of $S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}$ and so

$$\sum_{i=1}^{n} \lambda_i = tr\,(S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}).$$

(ii) Establish the asymptotic distributions of the $S_{i,j}$ under the null that the cointegrating rank is zero. These are.[5]

$$T^{-1}S_{1,1} \Rightarrow \int_0^1 WW'du \tag{E.1}$$

$$T^{-1}S_{1,0} \Rightarrow \int_0^1 W(dW)' \tag{E.2}$$

$$S_{0,0} \xrightarrow{\;p\;} \Omega$$

where $\xrightarrow{p}$ indicates 'convergence in probability', meaning that the random variable on the left-hand side tends to the deterministic quantity on the right.[6]

(iii) Replace λ_i by the more appropriate notation $\hat{\lambda}_i$ to emphasize that they are random variables, and apply the CMT to obtain the limiting behaviour of

$$\sum_{i=1}^{n} \hat{\lambda}_i = tr\,(S_{1,1}^{-1}S_{1,0}S_{0,0}^{-1}S_{0,1}).$$

This simply requires the substitution of the limit results at (ii) into the expression, and adjusting for the required normalization such that convergence is to a random variable. Thus,

$$\sum_{i=1}^{n} \lambda_i \Rightarrow tr\left(\left[T\int_0^1 WW'du\right]^{-1}\left[\int_0^1 W[dw]'\right]\Omega^{-1}\left[\int_0^1 dWW'\right]\right).$$

Clearly the right-hand side of this expression has a factor of T^{-1}, indicating that it tends to zero rather than a random variable. So, for weak convergence, both sides must be multiplied by T to give

$$T\sum_{i=1}^{n} \lambda_i \Rightarrow tr\left(\left[\int_0^1 WW'du\right]^{-1}\left[\int_0^1 W[dw]'\right]\Omega^{-1}\left[\int_0^1 dWW'\right]\right).$$

This expression can be written in terms of standard Brownian motion, $B(u) = \Omega^{-\frac{1}{2}}W(u)$ as

$$T\sum_{i=1}^{n} \lambda_i \Rightarrow tr\left(\left[\int_0^1 (dB)B'\right]\left[\int_0^1 BB'du\right]^{-1}\left[\int_0^1 B(dB)'\right]\right).$$

(iv) Next, establish how the trace statistic can be expressed in terms of $\sum_{i=1}^{n} \lambda_i$.

Note that, since $|\hat{\lambda}_i| < 1$, the usual expansion of the natural logarithm function applies, such that

$$-T\left[\sum_{i=1}^{n} \log\left(1-\hat{\lambda}_i\right)\right] = T\sum_{i=1}^{n} \hat{\lambda}_i + \delta$$

where δ is an asymptotically irrelevant term such that it can be ignored in the subsequent analysis, that is, $-T\left[\sum_{i=1}^{n} \log\left(1-\hat{\lambda}_i\right)\right]$ and $T\sum_{i=1}^{n} \hat{\lambda}_i$ converge weakly to the same random variable.[7] But $-T\left[\sum_{i=1}^{n} \log\left(1-\hat{\lambda}_i\right)\right]$ is the trace statistic for testing the null of non-cointegration against the full rank alternative.

(v) Thus the null distribution of the test statistic is given by

$$-T\sum_{i=1}^{n} \log\left(1-\hat{\lambda}_i\right) \Rightarrow tr\left(\left[\int_0^1 (dB)B'du\right]\left[\int_0^1 BB'du\right]^{-1}\left[\int_0^1 B(dB)'\right]\right). \tag{E.3}$$

E.4 Testing a null hypothesis of non-zero rank

This treatment has to be generalized to allow for null hypotheses of non-zero rank and for the various forms of trend that can be added to the basic VAR model. For testing the null of cointegrating rank *r*, the trace test statistic is

$$-T\left[\sum_{i=r+1}^{n} \log\left(1-\hat{\lambda}_i\right)\right].$$

The analysis proceeds by examining the behaviour of the $n - r$ smallest eigenvalues under the null. It is stated, without proof, that under the null hypothesis that the cointegrating rank is *r*, with appropriate normalization, the smallest $n - r$ eigenvalues converge to zero while the remaining *r* tend to positive constants.[8] It transpires that the problem is best addressed not in terms of the eigenvalues, λ_i but of $\rho_i = T\lambda_i$. For convenience, define

$$S(\lambda) = \lambda S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{0,1}.$$

The eigenvalues are the solutions to $|S(\lambda)| = 0$. Clearly the solutions are unchanged for the problem

$$\left|A'S(\lambda)A\right| = 0$$

for any non-singular matrix A. Now partition A such that $A = (A_1 \quad A_2)$, then:

$$\left|A'S(\lambda)A\right| = \left|H\right|\left|G\right|$$

where $H = (A_1'S(\lambda)A_1)$, $G = A_2'(S(\lambda) - S(\lambda)A_2[A_2'S(\lambda)A_2]^{-1}A_2'S(\lambda))A_2$.

The derivation of the distribution is obtained by choosing the partition of A such that, asymptotically, *H* is not a function of λ. Then, asymptotically, solutions for λ will arise only from

$$\left|G\right| = 0 \tag{E.4}$$

so that it is only necessary to consider G. But G can be broken down into a number of components whose asymptotic distributions can be derived, and hence, via the CMT, the distribution of the trace statistic is obtained.

Let $A_1 = \beta$, $A_2 = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}$ where β is $n \times r$ and $\beta_\perp$ is $n \times (n - r)$ and orthogonal to β. The derivation begins by showing that H is redundant. For this choice of A

$$\begin{aligned} H &= H_1 - H_2 \\ H_1 &= \lambda\beta' S_{1,1}\beta \\ H_2 &= \beta' S_{1,0} S_{0,0}^{-1} S_{0,1}\beta. \end{aligned}$$

H is seen to be a function of λ only through H_1. Now reparameterize the problem using $\rho = T\lambda$. Then, $H_1 = T^{-1}\rho\beta' S_{1,1}\beta$. The asymptotic limits taken from now on will be such that ρ remains fixed as $T \to \infty$, which means $\lambda \to 0$. Thus from this point on, the discussion is with respect to the eigenvalues normalized by T. Under this limit, $H_1 \to 0$, and so, asymptotically, H is not a function of λ, and the required solutions will follow from (E.4).

Now consider G, and for convenience put $D = \beta_\perp(\beta'_\perp\beta_\perp)^{-1}$ so that

$$\begin{aligned} G &= D'\Big(S(\lambda) - S(\lambda)\beta[\beta' S(\lambda)\beta]^{-1}\beta' S(\lambda)\Big)D \\ &= G_1 - G_2 - G_3 \end{aligned} \tag{E.5}$$

where $G_1 = \rho T^{-1} D' S_{1,1} D$, $G_2 = D' S_{1,0} S_{0,0}^{-1} S_{0,1} D$, $G_3 = \tilde{G}_3\ (\beta' S(\lambda)\beta)\tilde{G}'_3$, $\tilde{G}_3 = D' S(\lambda)\beta$. Further convergence results are now required (Johansen, 1995, lemma 10.3). These are, generalizing (E.1) and (E.2) respectively:

$$T^{-1} D' S_{1,1} D \Rightarrow \int_0^1 WW'du$$

$$D'(S_{1,0} - S_{1,1}\beta\alpha') \Rightarrow \int_0^1 W(dW)'$$

where W now has dimension $n - r$ dimension;[9] and

$$\beta' S_{1,1}\beta \xrightarrow{P} \Sigma_{\beta,\beta}, \tag{E.6}$$

$$\beta' S_{1,0} \xrightarrow{P} \Sigma_{\beta,0}, \tag{E.7}$$

$$S_{0,0} \xrightarrow{P} \Sigma_{0,0}, \tag{E.8}$$

$$D' S_{1,1}\beta = O_p(1). \tag{E.9}$$

The last equality means that the probability that $D' S_{1,1}\beta$ diverges from a constant value tends to zero, and hence that it can be regarded as a constant in the limit.[10] In the following the "=" sign represents either equality or weak convergence to the same random variable. Then, by (E.8)

$$G_2 = D' S_{1,0} \Sigma_{0,0}^{-1} S_{0,1} D \tag{E.10}$$

and

$$\begin{aligned} \tilde{G}_3 &= \rho T^{-1} D' S_{1,1}\beta - D' S_{1,0} S_{0,0}^{-1} S_{0,1}\beta = -D' S_{1,0} S_{0,0}^{-1} S_{0,1}\beta \\ &= -D' S_{1,0} \Sigma_{0,0}^{-1} \Sigma_{0,\beta} \end{aligned} \tag{E.11}$$

where the last equality follows from (E.8) and (E.7), and the previous one from (E.9). Then,

$$
\begin{aligned}
G_3 &= \tilde{G}_3\left(\beta' S(\lambda)\beta\right)^{-1}\tilde{G}_3' \\
&= \tilde{G}_3\left(\rho T^{-1}\beta' S_{1,1}\beta - S_{1,0}S_{0,0}^{-1}S_{0,1}\beta\right)^{-1}\tilde{G}_3' \\
&= -\tilde{G}_3\left(\beta' S_{1,0}S_{0,0}^{-1}S_{0,1}\beta\right)^{-1}\tilde{G}_3', \quad as \quad T\to\infty \\
&= -\tilde{G}_3\left(\beta' S_{1,0}S_{0,0}^{-1}S_{0,1}\beta\right)^{-1}\tilde{G}_3' \quad as \quad T\to\infty \\
&= -\tilde{G}_3\left(\beta'\Sigma_{\beta,0}\Sigma_{0,0}^{-1}\Sigma_{0,\beta}\beta\right)^{-1}\tilde{G}_3'
\end{aligned} \tag{E.12}
$$

where the last equality follows using (E.7, E.8). Substituting (E.10, E.11, and E.12) into (E.5) gives

$$
\begin{aligned}
G &= G_1 - G_2 - \tilde{G}_3\left(\beta' S(\lambda)\beta\right)^{-1}\tilde{G}_3' \\
&= \rho T^{-1}D' S_{1,1}D - D' S_{1,0}\Sigma_{0,0}^{-1}S_{0,1}D + \tilde{G}_3\left(\beta'\Sigma_{\beta,0}\Sigma_{0,0}^{-1}\Sigma_{0,\beta}\right)^{-1}\tilde{G}_3'
\end{aligned}
$$

or, using (E.11)

$$
\begin{aligned}
G &= \rho T^{-1}D' S_{1,1}D - D' S_{1,0}QS_{0,1}D = G_1 - G_4 \\
G_4 &= D' S_{1,0}QS_{0,1}D \\
G_4 &= D' S_{1,0}QS_{0,1}D \\
Q &= \Sigma_{0,0}^{-1} - \Sigma_{0,0}^{-1}\Sigma_{0,\beta}\left(\beta'\Sigma_{\beta,0}\Sigma_{0,0}^{-1}\Sigma_{0,\beta}\beta\right)^{-1}\Sigma_{\beta,0}\Sigma_{0,0}^{-1}.
\end{aligned} \tag{E.13}
$$

It can be shown that

$$
Q = \alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' = \alpha_\perp(Var(\alpha_\perp' W))^{-1}\alpha_\perp'
$$

and so,

$$
G_4 = D' S_{1,0}\alpha_\perp(Var(\alpha_\perp' W))^{-1}\alpha_\perp' S_{0,1}D.
$$

The asymptotic distribution of $D' S_{1,0}\alpha_\perp$ is given by

$$
D' S_{1,0}\alpha_\perp \Rightarrow \int_0^1 W(dW)'\alpha_\perp
$$

and so

$$
G_4 \Rightarrow \left(\int_0^1 W(dW)'\right)\alpha_\perp(Var(\alpha_\perp' W))^{-1}\alpha_\perp'\left(\int_0^1 W(dW)'\right)'. \tag{E.14}
$$

Similarly,

$$
T^{-1}D' S_{1,1}D \Rightarrow \int_0^1 WW'du \tag{E.15}
$$

and so

$$
G_1 \Rightarrow \rho\int_0^1 WW'du. \tag{E.16}
$$

Thus, substituting (E.14) and (E.16) into (E.13) gives

$$
G \Rightarrow \rho\int_0^1 WW'du - \left(\int_0^1 W(dW)'\right)\alpha_\perp(Var(\alpha_\perp' W))^{-1}\alpha_\perp'\left(\int_0^1 W(dW)'\right).
$$

It follows from the CLT that the solutions of the problem $|G| = 0$ converge in distribution to those of the problem

$$\left| \rho \int_0^1 WW'du - \left[\int_0^1 W(dW)' \right] \alpha_\perp (Var(\alpha'_\perp W))^{-1} \alpha'_\perp \left[\int_0^1 W(dW)' \right]' \right| = 0.$$

The solutions for ρ are unchanged if the matrix of which the determinant is being taken is pre- and post-multiplied by $\Omega^{-\frac{1}{2}}$, which leads to simplification since the outer occurrences of W become standardized as $B = \Omega^{-\frac{1}{2}}W$. Thus (E.17) may be replaced by

$$\left| \rho \int_0^1 BB'du - \left[\int_0^1 B(dW)' \right] \alpha_\perp (Var(\alpha'_\perp W))^{-1} \alpha'_\perp \left[\int_0^1 B(dW)' \right]' \right| = 0. \tag{E.18}$$

Finally, noting also that $(Var(\alpha'_\perp W))^{-\frac{1}{2}} (\alpha'_\perp W) = B$, equation (E.18) may be written

$$\left| \rho \int_0^1 BB'du - \left[\int_0^1 B(dB)' \right] \left[\int_0^1 B(dB)' \right]' \right| = 0 \tag{E.19}$$

where B is now $n - r$ standardized Brownian motion (dimension equal to the number of zero eigenvalues under the null). The trace statistic is

$$-T \left[\sum_{i=r+1}^{n} \log\left(1 - \hat{\lambda}_i\right) \right],$$

which is asymptotically equivalent to

$$T \sum_{i=r+1}^{n} \hat{\lambda}_i.$$

Equation (E.19) then gives

$$T \sum_{i=r+1}^{n} \hat{\lambda}_i \Rightarrow \sum_{i=r+1}^{n} \rho_i = tr \left\{ \left[\int_0^1 (dB) B' \right]' \left[\int_0^1 BB'du \right]^{-1} \left[\int_0^1 B(dB)' \right] \right\} \tag{E.20}$$

providing the required asymptotic distribution for the trace statistic for testing the null of cointegrating rank r against the alternative of rank n. This result specializes to that for testing cointegrating rank 0 against rank n by setting $r = 0$, as can be seen by comparing equations (E.20) and (E.3).

E.5 Distribution theory when there are deterministic trends in the data

The distribution has to be modified according to the deterministic components in the process. The processes with respect to which integration takes place are unchanged, but the integrands are modified.

The general model is

$$\Delta x_t = \Pi x_{t-1} - \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \mu_t + \epsilon_t$$

where $\mu_t = \mu_0 + \mu_1 t$ which, in its most general form, allows the process x_t to have a quadratic trend, and the cointegrating relations to have a linear trend (Johansen, 1991).[11] The deterministic components, in increasing order of complexity, are:

(i) no deterministic terms: $\mu_t = 0$
(ii) intercept only, in space of α: $\mu_t = \mu_0 = \alpha\kappa_0$
(iii) intercept only, not in space of α: $\mu_t = \mu_0$ $\alpha_\perp\mu_0 \neq 0$
(iv) time trend, slope in space of α: $\mu_t = \mu_0 + \alpha\kappa_0 t$
(v) time trend, slope not in space of α: $\mu_t = \mu_0 + \mu_1 t$, $\alpha_\perp\mu_1 \neq 0$.

These cases correspond to different solutions of the underlying process as follows.

(i) x_t has no deterministic terms and all stationary components have zero mean.
(ii) x_t has neither quadratic nor linear trend, but both x_t and $\beta' x_t$ have constant terms.
(iii) x_t has a linear trend, but this is eliminated in the cointegrating combinations.
(iv) x_t has no quadratic trend, but has a linear trend that is also present in the cointegrating relations.
(v) x_t has a quadratic trend, but the cointegrating relations have a linear trend only.

The asymptotic distribution of the trace statistic for testing the null of cointegrating rank r has the same generic form in each case, but the distributions have to be corrected differently. This form is

$$tr\left\{\left[\int_0^1 (dB)\,F'\right]'\left[\int_0^1 F\,F'du\right]^{-1}\left[\int_0^1 F\,(dB)'\right]\right\} \tag{E.21}$$

where B is an $n - r$ standard Brownian motion, and F is the same standard Brownian motion corrected for the deterministic components, with the final element (either the $n - r^{th}$ or $n - r + 1$st) consisting of the appropriate power of u corrected for the same components.

This is described in Table E.1. The coefficients a_i and b_i are fixed and required to correct for the included deterministic terms. All elements of the corrected Brownian motion, except the last, are, in effect, a residual having regressed the standard case on the deterministic terms. The last term, the q^{th} in the table below, corresponds to regressing the random variable u on the same terms. If the highest order deterministic term is orthogonal to a then the final term is $n - r + 1^{st}$, otherwise it is the $n - r^{th}$.

Tables of approximate asymptotic and finite sample distributions

Statistical tables are available for each of these cases reported in Table E.1, calculated for finite samples by simulation. Table E.2 indicates where each of the cases above may be found, with comments on their coverage. Johansen (1995) presents finite sample and approximate asymptotic critical values for the tests, employing the standard form of the test statistic. Osterwald-Lenum (1992) extends Johansen's tables to consider a wider range of dimensions for the process $n = 1, 2, \ldots, 11$, Doornik (1998, 2003) discusses an alternative method of obtaining approximate asymptotic critical values, the latter paper providing tables. It is now common practice for regression packages to compute critical values or p-values as required. MacKinnon, Haug and Michelis (1999) provide a response surface methodology for computing finite sample and approximate asymptotic critical values and p-values for all the standard cases.[12]

An alternative to providing tables for different sample sizes is to correct either the test statistic or the asymptotic critical values. Reinsel and Ahn (1992), Reimers (1992) and

Table E.1 Corrections to trace statistic distributions due to deterministic components $F = \{F_i(u)\}$

Case	*Deterministic Components*		*Corrected Standard Brownian Motion*		
	Intercept: μ_0	Slope: μ_1	q	$i = 1,2, \ldots, q$; $F_i(u) = B_i(u) - a_i - b_i u$	$i = q + 1$, $F_i(u)$
(i)	0	0	n–r	$a_i = 0$, $b_i = 0$	–
(ii)	$\alpha\kappa_0$	0	n–r	$a_i = 0$, $b_i = 0$	1
(iii)	$\alpha_\perp\mu_0 \neq 0$	0	n–r–1	$a_i = \int_0^1 B_i(u)du$, $b_i = 0^+$	$u - a_i$, $a_i = 1/2$*
(iv)	μ_0	$\alpha\kappa_0$	n–r	$a_i = \int_0^1 B_i(u)du$, $b_i = 0^+$	$u - a_i$, $a_i = 1/2$*
(v)	μ_0	$\alpha_\perp\mu_0 \neq 0$		$B_i(u) - a_i - b_i u$++	$u^2 - a_i - b_i u$**

a_i and b_i are fixed coefficients necessary to correct for the included deterministic terms. +Corrects $B_i(u)$ for a constant. *Corrects u for a constant. ++Corrects $B_i(u)$ for a linear time trend. **Corrects u^2 for a linear time trend.

Table E.2 Sources of tables for the trace test

Cases	*Source*		
	D (asymptotic)	J (finite sample)	OL
(i)	table 1	table 15.1	table 0
(ii)	table 2	table 15.2	table 1*
(iii)	table 3	table 15.3	table 1
(iv)	table 4	table 15.4	table 2*
(v)	table 5	table 15.5	table 2

Note:
D – Doornik (2003); J – Johansen (1995); OL – Osterwald-Lenum (1992).

Cheung and Lai (1993) suggest correcting for, in effect, the number of parameters estimated in the VAR. The correction is to replace T by $T - np$. Equivalently, the asymptotic critical values can be multiplied by $T/(T - np)$. The result is to correct a tendency of the asymptotic tests[13] to be over-sized. That is, when used naively, the tests reject the null hypothesis too frequently. When testing the null of non-cointegration, this results in findings of cointegration where it does not exist.

E.6 Other issues

The maximal eigenvalue statistic

This discussion has dealt with the results for the trace test. Hansen and Johansen (1998) discuss the results for the maximal eigenvalue test that can be derived using the same basic distributional results. Where the distribution of the trace statistic is given by the trace of

$$A(n,r) = \left[\int_0^1 (dB)\,F'\right]' \left[\int_0^1 F\,F'du\right]^{-1} \left[\int_0^1 F\,(dB)'\right]$$

as described in equation (E.21), the asymptotic distribution of the maximal eigenvalue statistic is, analogously, the maximal eigenvalue of $A(n,r)$. In practice, the maximal eigenvalue statistic would be used in the same sequential manner as the trace statistic, but it is important to note that there is no proof yet available of the consistency of this procedure for this statistic. It is therefore reasonable to place emphasis on the trace statistic.

Sequential testing and model selection

The distributions discussed above, whether asymptotic or finite sample, do not allow for distortionary effects of model selection or, in finite samples, that due to sequential tests. Each result assumes that the test takes place in isolation and is not subject to pre-testing. In practice, it is likely that tests will suffer both inflated size and reduced power if the critical values are not adjusted. That is, typically, the finite sample null distributions of the test statistics become more dispersed due to pre-testing or model selection.

Partial systems

The system discussed treats all variables as endogenous. There is no sense in which any of them plays a different causal role to any others. Johansen (1992) has discussed this, and, more recently, Harbo, Johansen, Nielsen, and Rahbek (1998), and Pesaran, Shin and Smith (2001) have considered the impact of exogenous I(1) variables on the asymptotic distribution of the test statistics. This generates a wider set of models for which the distributions must be calculated, depending not only on the total number of variables in the system (n), but also on the number of these that are endogenous (n_1, say). Thus $A(n,r)$ of equation (16), where B is of dimension $n - r$, and F depends on B as described in table A, is replaced by

$$\tilde{A}(n,k,r) = \left[\int_0^1 (d\tilde{B})\,\tilde{F}'\right]' \left[\int_0^1 \tilde{F}\,\tilde{F}'du\right]^{-1} \left[\int_0^1 \tilde{F}\,(d\tilde{B})'\right]$$

where B is now $k - r$ standard Brownian motion, and F is a modified $n - r$ standard Brownian motion, analogous to the modifications of table A for the purely endogenous case. The underlying models are conveniently explained in MacKinnon, Haug and Michelis (1999), where tables of critical values may also be found. Further tables may be found in Harbo et al. (1998), with modifications provided by Doornik (2003).

Appendix F: Estimation under General Restrictions

From the Frisch–Waugh form the system is written:

$$R_{0,t} = \alpha\beta' R_{1,t} + \varepsilon_t$$

or

$$\varepsilon_t = R_{0,t} - \alpha\beta' R_{1,t} = \left[I : \alpha\beta'\right]\begin{bmatrix} R_{0,t} \\ R_{1,t} \end{bmatrix}.$$

It follows from Doornik and Hendry (2001) that the Concentrated Likelihood for this multivariate least squares problem can be written:

$$\log L = K - \frac{T}{2}\log\left|\varepsilon\varepsilon'\right| = K - \frac{T}{2}\log\left|\left[I : -\alpha\beta'\right]\begin{bmatrix} R_0 \\ R_1 \end{bmatrix}\left[R_0' : R_1'\right]\begin{bmatrix} I \\ -\beta\alpha' \end{bmatrix}\right|$$

$$= K - \frac{T}{2}\log\left|\left[I : -\alpha\beta'\right]\begin{bmatrix} S_{0,0} & S_{0,1} \\ S_{1,0} & S_{1,1} \end{bmatrix}\begin{bmatrix} I \\ -\beta\alpha' \end{bmatrix}\right|$$

where $S_{i,j} = R_i R_j' = \Sigma_{t=1}^{n} R_{i,t} R_{j,t}'$. Now

$$\log L = K - \frac{T}{2}\log\left|S_{0,0} - \alpha\beta' S_{1,0} - S_{0,1}\beta\alpha' + \alpha\beta' S_{1,1}\beta\alpha'\right|.$$

Concentrating out the above likelihood for $\alpha = S_{0,1}\beta(\beta' S_{1,1}\beta)^{-1}$:

$$\log L = K - \frac{T}{2}\log\left|S_{0,0} - S_{0,1}\beta(\beta' S_{1,1}\beta)^{-1}\beta' S_{1,0}\right|$$

$$= K - \frac{T}{2}\log\left|S_{0,0}\right|\left|(\beta' S_{1,1}\beta)^{-1}\right|\left|\beta'(S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0})\beta\right|.$$

Subject to the normalization $\beta' S_{1,1}\beta = I$ and given that the solution to the likelihood problem with respect to β is invariant to $S_{0,0}$, then the likelihood problem is equivalent to solving the determinantal equation $|\beta'(S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0})\beta|$ which in the cointegration case is the reduced rank problem, $|S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0}| = 0$. What is required is a solution to the usual eigenvalue problem, $|\lambda S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0}| = 0$, where for each non-zero eigenvalue there is an eigenvector β_i such that:

$$(\lambda S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0})\beta_i = 0.$$

Stacking the eigenvectors associated with the non-zero eigenvalues into an $n \times r$ matrix β, then β is the matrix that diagonalizes $S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0}$. Therefore:

$$\left|\beta'(S_{1,1} - S_{1,0}S_{0,0}^{-1}S_{1,0})\beta\right| = \left|I - \Lambda_r\right| = \prod_{i=1}^{r}(1 - \lambda_i).$$

It is follows that the likelihood can be re-written thus:

$$\log L_{MAX} = -\frac{T}{2}\left[\log\left|S_{0,0}^{-1}\right| + \sum_{i=1}^{r}\log(1 - \lambda_i)\right].$$

As was stated in chapter 4, any test of parameters must compare the above likelihood, which imposes no restrictions on either α or β with one on which restrictions have been imposed. Therefore:

$$\begin{aligned}\log L(\theta|r, H_{g\alpha\beta} \quad &: \quad \alpha = f(\theta) \cap \beta = f(\theta)) \\ &= K - \frac{T}{2}\log\left|\left[I : -\alpha(\theta)\beta(\theta)'\right]\begin{bmatrix} S_{0,0} & S_{0,1} \\ S_{1,0} & S_{1,1}\end{bmatrix}\begin{bmatrix} I \\ -\beta(\theta)\alpha(\theta)'\end{bmatrix}\right|.\end{aligned}$$

The test is a likelihood ratio test:

$$LR(i) = 2\log L_{MAX} - 2\log L(\theta|r, H_{g\alpha\beta} : \alpha = f(\theta) \cap \beta = f(\theta)) \sim \chi^2_i.$$

Doornik and Hendry (2001) explain how to maximize the non-linear likelihood under a range of different restrictions.

Appendix G: Proof of Identification based on an Indirect Solution

Define α and β as consisting of α_{ij} and β_{ij} elements for $i = 1, \ldots 5$, and $j = 1, \ldots 4$, and Π as consisting of π_{ij} elements for $i = 1, \ldots 5$, and $j = 1, \ldots 5$. For (WE) of i_2,.$\alpha_5 = 0$ and $\Pi_5. = 0$, which excludes them from our deliberations. However, over-identification is sufficient for identification which implies that the conditions for over-identification are necessary for the preferred parameters to be identified. If we look at equation (5.10) and set $\alpha_2\beta' = 0$, then:

$$\begin{bmatrix} \Pi_1 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha_1\beta' \\ 0 \end{bmatrix}.$$

After imposing the same restrictions as Hunter and Simpson (1995) α_1 and β take the following form:

$$\alpha_1 = \begin{bmatrix} \alpha_{11} & 0 & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & 0 & \alpha_{24} \\ 0 & 0 & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & 0 & 0 & \alpha_{44} \end{bmatrix}, \ \beta' = \begin{bmatrix} 1 & \beta_{21} & 0 & 0 & 0 \\ 0 & 1 & 0 & -\beta_{52} & \beta_{52} \\ 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

With $r = 4$ cointegrating vectors, the requirement of the order condition is for $r^2 - r = 16 - 4 = 12$ restrictions with normalization. In α_1 and β above there are 20 restrictions without normalization.[1] Hence, there are enough a priori restrictions to identify α and β. However, based on the indirect least squares approach. we need to find whether there are enough solutions to the equation $\Pi_1 = \alpha_1\beta'$ to derive at least one estimate of α_{11}, α_{13}, α_{14}, α_{21}, α_{22}, α_{24}, α_{33}, α_{34}, α_{41}, α_{44}, β_{21} and β_{52}.

Multiplying β' through by α_1 yields the following matrix of restricted long-run parameters:

$$\alpha_1\beta' = \begin{bmatrix} \alpha_{11} - \alpha_{14} & \alpha_{11}\beta_{21} - \alpha_{13} & \alpha_{13} & \alpha_{14} & 0 \\ \alpha_{21} - \alpha_{24} & \alpha_{21}\beta_{21} + \alpha_{22} & 0 & -\alpha_{22}\beta_{52} + \alpha_{24} & \alpha_{22}\beta_{52} \\ -\alpha_{34} & -\alpha_{33} & \alpha_{33} & \alpha_{34} & 0 \\ \alpha_{41} - \alpha_{44} & \alpha_{41}\beta_{21} & 0 & \alpha_{44} & 0 \end{bmatrix}.$$

Comparing $\alpha_1\beta'$ with Π_1, where $\Pi_1 = [\pi_{ij}]_1$ for $i = 1, \ldots, 4$ and $j = 1, \ldots 5$, by matching parameters, it follows that:

$$\alpha_{13} = \pi_{13}, \ \alpha_{14} = \pi_{14}, \ \alpha_{33} = \pi_{33}, \ \alpha_{34} \textit{ and } \alpha_{44} = \pi_{44}.$$

Consequently:

$$\alpha_{11} = \pi_{11} + \pi_{14}, \ \beta_{21} = \frac{(\pi_{13} + \pi_{12})}{(\pi_{11} + \pi_{14})} \textit{ and } \alpha_{41} = \pi_{41} + \pi_{44}.$$

Furthermore:

$$\alpha_{21} = \pi_{21} - (\pi_{25} + \pi_{24}),\ \alpha_{24} = \pi_{24} + \pi_{25},\ \alpha_{22} - \pi_{22} - \alpha_{21}\beta_{21} \text{ and } \beta_{52} = \frac{\pi_{25}}{\alpha_{22}}.$$

The long-run restrictions imply that there are three over-identified parameters as there are three unused solutions associated with some of the parameters in the system:

$$\alpha_{34} = -\pi_{31},\ \alpha_{33} = -\pi_{32},\ \alpha_{41}\beta_{21} = \pi_{42}.$$

Hence, the parameters are slightly over-identified, which is surprising given the number of restrictions adopted, 20.

Appendix H: Generic Identification of Long-Run Parameters in Section 5.5

From (5.18), which can be written as:

$$\begin{bmatrix} vec(\beta_{11}\beta_{12}) \\ vec(\beta_{21}\beta_{22}) \\ vec(\beta_{31}\beta_{32}) \\ vec(\beta_{41}\beta_{42}) \\ vec(\beta_{51}\beta_{52}) \\ vec(\beta_{61}\beta_{62}) \end{bmatrix} = \begin{bmatrix} A^{-1}vec(|\pi_{31}\pi_{61}|) \\ A^{-1}vec(|\pi_{32}\pi_{62}|) \\ A^{-1}vec(|\pi_{33}\pi_{63}|) \\ A^{-1}vec(|\pi_{34}\pi_{64}|) \\ A^{-1}vec(|\pi_{35}\pi_{65}|) \\ A^{-1}vec(|\pi_{36}\pi_{66}|) \end{bmatrix}, A^{-1} = \begin{bmatrix} \frac{1}{\alpha_{31}} & 0 \\ 0 & \frac{1}{\alpha_{62}} \end{bmatrix},$$

and using the restrictions embodied in (5.17), we obtain:

$$\beta_{i1} = \alpha_{31}^{-1}\pi_{3i}, \text{ for } i = 1,2,3,5,6, \ \beta_{i2} = \alpha_{62}^{-1}\pi_{6i}, \text{ for } i = 1,2,3,4,6,$$
$$1 = \alpha_{31}^{-1}\pi_{34}, \ 1 = \alpha_{62}^{-1}\pi_{65}.$$

Similarly for (5.19):

$$\begin{bmatrix} vec(\alpha_{11}\alpha_{12}) \\ vec(\alpha_{21}\alpha_{22}) \\ vec(\alpha_{31}\alpha_{32}) \\ vec(\alpha_{41}\alpha_{42}) \\ vec(\alpha_{51}\alpha_{52}) \\ vec(\alpha_{61}\alpha_{62}) \end{bmatrix} = (B^{-1} \otimes I_6) \begin{bmatrix} vec(\pi_{14}\pi_{24}\ldots\pi_{64} \\ vec(\pi_{15}\pi_{25}\ldots\pi_{65} \end{bmatrix} \text{ and } B^{-1} = \begin{bmatrix} \frac{1}{\delta} & \frac{-\beta_{51}}{\delta} \\ -\frac{\beta_{42}}{\delta} & \frac{-1}{\delta} \end{bmatrix}$$

where $\delta = -1 - \beta_{12}\,\beta_{51}$. Solving the former equation, subject to the restrictions on α:

$$\alpha_{11} = \frac{1}{\delta}\pi_{14} - \frac{\beta_{51}}{\delta}\pi_{15}, \ \alpha_{21} = \frac{1}{\delta}\pi_{24} - \frac{\beta_{51}}{\delta}\pi_{25}, \ \alpha_{12} = -\frac{\beta_{42}}{\delta}\pi_{14} - \frac{1}{\delta}\pi_{15} = 0,$$
$$\alpha_{22} = -\frac{\beta_{42}}{\delta}\pi_{24} - \frac{1}{\delta}\pi_{25} = 0, \ \alpha_{31} = \frac{1}{\delta}\pi_{34} - \frac{\beta_{51}}{\delta}\pi_{35},$$
$$\alpha_{41} = \frac{1}{\delta}\pi_{44} - \frac{\beta_{42}}{\delta}\pi_{45} = 0, \ \alpha_{51} = \frac{1}{\delta}\pi_{54} - \frac{\beta_{42}}{\delta}\pi_{55} = 0,$$
$$\alpha_{61} = \frac{1}{\delta}\pi_{64} - \frac{\beta_{42}}{\delta}\pi_{65} = 0, \ \alpha_{32} = -\frac{\beta_{42}}{\delta}\pi_{34} - \frac{1}{\delta}\pi_{35} = 0,$$
$$\alpha_{42} = -\frac{\beta_{51}}{\delta}\pi_{44} - \frac{1}{\delta}\pi_{45}, \ \alpha_{52} = -\frac{\beta_{51}}{\delta}\pi_{54} - \frac{1}{\delta}\pi_{55}, \ \alpha_{62} = -\frac{\beta_{51}}{\delta}\pi_{64} - \frac{1}{\delta}\pi_{65}.$$

As the parameters are over-identified one only needs to consider the following results: α_{11}, α_{21}, α_{31}, α_{42}, α_{52}.

References

Abadir, K. and Talmain, G. (2002) Aggregation, persistence and volatility in a macro model. *The Review of Economic Studies*, 69 749–79.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59, 817–58.

Arellano, M., Hansen, L.P., and Sentana, E. (1999) Underidentification? Paper presented at the Econometrics Study Group Conference, Bristol, July.

Banerjee, A., Dolado, J.J., Galbraith, J.W., and Hendry, D.F. (1993) *Co-integration, Error-Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Barndorff-Nielsen, O.E., and Shephard, N. (2001) Modelling by Levy processes for financial econometrics, in *Levy Processes: Theory and Applications*. Barndorff-Nielsen, Ole E. Mikosch, Thomas Resnick, Sidney I., eds, Boston and Basel: Birkhauser, 283–318.

Bauwens, L., Deprins, D. and Vandeuren, J.-P. (1997) Bivariate modelling of interest rates with a cointegrated VAR-GARCH model. Discussion Paper CORE, The Catholic University, Louvain-La Nueve DP 9780.

Bauwens, L., and Hunter, J. (2000) Identifying long-run behaviour with non-stationary data. Discussion Paper CORE, The Catholic University, Louvain-La Nueve DP 2000/43.

Bauwens, L., Lubrano, M., and Richard J.-F. (2000) *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Barten, A.P. (1969) Maximum likelihood estimation of an almost complete set of demand equations. *European Economic Review*, 1, 7–73.

Blough, S.R. (1992) The relationship between power and level for generic unit root tests in finite samples. *Journal of Applied Econometrics*, 7, 295–308.

Boswijk, H.P. (1992) *Cointegration, Identification and Exogeneity: Inference in Structural Error Correction Models*. Amsterdam: Thesis Publishers.

Boswijk, H.P. (1996) Cointegration, identification and exogeneity: inference in structural error correction models. *Journal of Business and Economics and Statistics*, 14, 153–60.

Boswijk, H.P. and Frances, H. (1992) Dynamic specification and cointegration. *Oxford Bulletin of Economics and Statistics*, 54, 369–81.

Box, G.E.P. and Jenkins, G.E.M. (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Brockwell, P.J. and Davis, R.A. (1991) *Time Series: Theory and Methods* (second edition). New York: Springer Verlag.

Burke, S.P. (1994a) Confirmatory data analysis: the joint application of stationarity and unit root tests. *Discussion Papers in Quantitative Economics and Computing*, no. 20. Department of Economics, University of Reading.

Burke, S.P. (1994b) Unit root tests of the Phillips type with data dependent selection of the lag truncation parameter. *Discussion Papers in Quantitative Economics and Computing*, no. 11, University of Reading.

Burke, S.P. (1996) Some reparameterizations of lag polynomials for dynamic analysis. *Oxford Bulletin of Economics and Statistics*, 58, 373–89.

Burke, S.P. and Hunter, J. (1998) The impact of moving average behaviour on the Johansen trace test for cointegration. *Discussion Papers in Quantitative Economics and Computing*, no. 60, Department of Economics, University of Reading.

Caner, M., and Kilian, L. (2001) Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate. *Journal of International Money and Finance*, 20, 639–57.

Cheung, Y.-W. and Lai, K.S. (1993) Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. *Oxford Bulletin of Economics and Statistics*, 55, 313–28.

Chow, G.C. (1978) *Analysis and Control of Dynamic Economic Systems*. New York: John Wiley.

Clements, M.P. and Hendry, D.F. (1995) Forecasting in cointegrated systems. *Journal of Applied Econometrics*, 10, 127–46.

Clements, M.P. and Hendry, D.F. (1998) *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Clements, M.P. and Hendry, D.F. (2001) *Forecasting Non-Stationary Economic Time Series*. London: The MIT Press.

Corradi, V., Swanson, N.R., and White, H. (2000) Testing for stationarity-ergodicity and for comovements between nonlinear discrete time Markov processes. *Journal of Econometrics*, 96, 39–73.

Davidson, J.E.H. (1994) *Stochastic Limit Theory*. Oxford: Oxford University Press.

Davidson, J.E.H., Hendry, D.F., Srba, F., and Yeo, S. (1978) Econometric modelling of the aggregate time series relationships between consumers, expenditure and income in the United Kingdom. *Economic Journal*, 88, 661–92.

Davidson, R. and MacKinnon, J.G. (1993) *Estimation and Inference in Econometrics*. New York: Oxford University Press

Davidson, R. and MacKinnon, J.G. (1998) Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School*, 6, 1–26.

Deaton, A.S. and Muellbauer, J.N.J. (1980) An almost ideal demand system. *American Economic Review*, 70, 312–26.

Dickey, D.A. and Fuller, W.A. (1979) Distribution of the estimation for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427–31.

Dickey, D.A. and Fuller, W.A. (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, 1057–72.

Dickey, D.A, Hasza, D.P. and Fuller, W.A. (1984) Testing for unit roots in seasonal time series. *Journal of the American Statistical Association*, 79, 355–67.

Dickey, D.A. and Pantula, S.G. (1987) Determining the order of differencing in autoregressive processes. *Journal of Business and Economic Statistics*, 5, 455–61.

Dhrymes, P.J. (1984) *Mathematics for Econometrics*. New York: Springer-Verlag.

Dolado, J., Galbraith, J.W. and Banerjee, A. (1991) Estimating intertemporal quadratic adjustment costs models with dynamic data. *International Economic Review*, 32, 919–36.

Dornbusch, R. (1976) Expectations and exchange rate dynamics. *Journal of Political Economy*, 84, 1161–76.

Doornik, J.A. (1995) Testing general restrictions on the cointegration space. Mimeo, Nuffield College, Oxford.

Doornik, J.A. (1998), Approximations to the asymptotic distribution of cointegration tests. *Journal of Economic Surveys*, 12, 573–93.

Doornik, J.A. (2003) Asymptotic tables for cointegration tests based on the gamma-distribution approximation. Mimeo, Nuffield College, University of Oxford.

Doornik, J.A. and Hendry, D.F. (1996) *PCFIML 9*. London: Thompson International Publishers.

Doornik, J.A. and Hendry, D.F. (2001) *PCFIML 10*. London: Timberlake Consultants Press.

Dunne, J.P. and Hunter, J. (1998) The allocation of government expenditure in the UK: a forward looking dynamic model. Paper presented at the International Institute of Public Finance Conference, Cordoba, Argentina, August.

Elliott, G., Rothenberg, T.J and Stock, J.H. (1996) Efficient tests for an autoregressive unit root. *Econometrica*, 64, 813–36.

Engle, C. (2001) The responsiveness of consumer prices to exchange rates and the implications for exchange-rate policy: a survey of a few recent new open economy macro models. Mimeo University of Wisconsin.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.

Engle, R.F. and Granger, C.W.J. (1987) Co-integration and error-correction: representation, estimation and testing. *Econometrica*, 55, 251–76.

Engle, R.F. and Granger, C.W.J. (1991) *Long-Run Economic Relationships*. Oxford: Oxford University Press.

Engle, R.F. and Yoo, B.S. (1987) Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35, 143–59.

Engle, R.F. and Yoo, B.S. (1991) Cointegrated time series: an overview with new results. Chapter 12 in R.F. Engle and C.W.J. Granger (eds), *Long-run Economic Relationships*. Oxford: Oxford University Press.

Engle, R.F., Hendry, D.F. and Richard, R.F. (1983) Exogeneity. *Econometrica*, 51, 277–304.

Engsted, T. and Haldrup, N. (1997) Money demand, adjustment costs and forward looking behaviour. *Journal of Policy Modeling*, 19, 153–73.

Engsted, T. and Johansen, S. (1999) Granger's representation theorem and multicointegration, cointegration, causality, and forecasting. In *A Festschrift in Honour of Clive W. J. Granger*. Engle, Robert F. White, Halbert, eds., Oxford and New York: Oxford University Press, 200–11.

Ericsson, N.R. (1994) Testing exogeneity: An introduction, in *Testing Exogeneity*. Ericsson, N.R. and Irons, J.S., eds, Oxford University Press, 3–38.

Ericsson, N.R. and Irons, J.S. (1994) *Testing Exogeneity*. Oxford: Oxford University Press.

Ericsson, N.R., Hendry, D.F. and Mizon, G.E. (1998) Exogeneity, cointegration and economic policy analysis. *Journal of Business and Economics Statistics*, 16, 371–87.

Fama, E.F. (1970) Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, 25, 383–417.

Favero, C. and Hendry, D.F. (1992) Testing the Lucas critique: a review. *Econometric Reviews*, 11, 265–306.

Fisher, P.G., Tanna, S.K., Turner, D.S, Wallis, K.F., and Whitley, J.D. (1990) Econometric evaluation of the exchange rate in models of the UK economy. *Economic Journal*, 100, 1024–56.

Flôres, R. and Szafarz, A. (1995) Efficient markets do not cointegrate. Discussion Paper 9501, CEME, Université Libre de Bruxelles.

Flôres, R., and Szafarz, A. (1996) An extended definition of cointegration. *Economics Letters*, 50, 193–5.

Florens, I.P., Mouchart, M. and Rolin, J.-M. (1990) Sequential Experiments, Chapter 6 in *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Franses, P.H. (1994) A multivariate approach to modeling univariate seasonal time series. *Journal of Econometrics*, 63, 133–51.

Galbraith, J.W. and Zinde-Walsh, V. (1993) Autoregressive approximation of ARMA processes and choice of order in parametric unit root tests. Paper presented at EC^2, University of Oxford, December.

Galbraith, J.W. and Zinde-Walsh, V. (1999) On the distribution of the augmented Dickey–Fuller statistics in processes with moving average components. *Journal of Econometrics*, 93, 25–47.

Gantmacher, F.R. (1960) *Matrix Theory*, vol. I. New York: Chelsea Publishing Company.

Gohberg, I., Lancaster, P. and Rodman, L. (1983) *Matrix Polynomials*. New York: Academic Press.

Goldberger, A.S. (1964) *Econometric Theory*. New York: John Wiley and Sons.

Gonzalo, J. (1994) Comparison of five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics*, 16, 203–33.

Gonzalo, J. and Pitarakis, J.-Y. (1999) Dimensionality effect in cointegration analysis. In *Cointegration, Causality and Forecasting, A Festschrift in Honour of Clive Granger*, W.J. Granger, Engle, R.F., and White, H., eds, Oxford: Oxford University Press.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–38.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16, 121–30.

Granger, C.W.J. (1983) Cointegrated variables and error-correcting models. University of California San Diego Discussion Paper 83–13.

Granger, C.W.J. (1991) Developments in the study of cointegrated economic variables. Chapter 4 in R.F. Engle and C.W.J. Granger (eds), *Long-run Economic Relationships*. Oxford: Oxford University Press.

Granger, C.W.J. (1995) Modelling nonlinear relationships between extended-memory variables. *Econometrica*, 63, 265–79.

Granger, C.W.J. and Hallman, J.J. (1991) Long memory series with attractors. *Oxford Bulletin of Economics and Statistics*, 53, 11–26.

Granger, C.W.J. and Joyeux, R. (1980) An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15–39.

Granger, C.W.J. and Lee, T-H. (1989) Multicointegration. In *Co-integration, Spurious Regressions, and Unit Roots*. Fomby, T.B. and Rhodes, G.F., eds., Advances in Econometrics, vol. 8 Greenwich, Conn. and London: JAI Press. 71–84.

Granger, C.W.J. and Morris, M.J. (1976) Time series modelling and interpretation. *Journal of the Royal Statistical Society*, Series A, 139, 246–57.

Granger, C.W.J. and Newbold, P. (1974) Spurious regression in econometrics. *Journal of Econometrics*, 2, 111–20.

Granger, C.W.J. and Newbold, P. (1986) *Forecasting with Economic Time Series*. New York: Academic Press.

Granger, C.W.J. and Weiss, A.A. (1983) Time Series Analysis of Error-Correcting Models. in *Studies in Econometrics, Time Series, and Multivariate Statistics*. New York: Academic Press, 255–78.

Gregoir, S. and Laroque, G. (1994) Polynomial cointegration estimation and test. *Journal of Econometrics*, 63, 183–214.

Haldrup, N. (1994) The asymptotics of single equation cointegration regressions with I(1) and I(2) variables. *Journal of Econometrics*, 63, 153–81.

Haldrup, N. and Salmon, M. (1998) Representations of I(2) cointegrated systems using the Smith–McMillan form. *Journal of Econometrics*, 84, 303–25.

Hall, A. (1989) Testing for a unit root in the presence of moving average errors. *Biometrika*, 76, 49–56.

Hall, R.E. (1978) Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political Economy*, 86, 971–87.
Hall, S.J. and Wickens, M. (1994) Causality in integrated systems. Centre for Economic Forecasting, *Discussion paper, 27–93*, London Business School.
Hamilton, J.D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.
Hansen, B.E. (1995) Rethinking the univariate approach to unit root testing: using covariates to increase power. *Econometric Theory*, 11, 1148–72.
Hansen, L.P. and Sargent, T.J. (1982) Instrumental variables procedures for estimating linear rational expectations models. *Journal of Monetary Economics*, 9, 263–96.
Hansen, P. and Johansen, S. (1998) *Workbook for Cointegration*. Oxford: Oxford University Press.
Harbo, I., Johansen, S., Nielsen, B., and Rahbek, A. (1998) Asymptotic inference on cointegrating rank in partial systems. *Journal of Business and Economic Statstics*, 16, 388–399.
Harvey, A.C. (1989) *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge, Cambridge University Press.
Harvey, A.C. (1993) *Time Series Models* (second edition). London: Harvester Wheatsheaf.
Hatanaka, M. (1996) *Time-series-based Econometrics: Unit Roots and Cointegration*. Oxford: Oxford University Press.
Haug, A.A. (1993) A Monte Carlo study of size distortions. *Economics Letters*, 41, 345–51.
Haug, A.A. (1996) Tests for cointegration: A Monte Carlo comparison. *Journal of Econometrics*, 71, 89–115.
Hendry, D.F. (1988) The encompassing implications of feedback versus feed-forward mechanisms in econometrics. *Oxford Economic Papers*, 40, 132–49.
Hendry, D.F. (1995) *Dynamic Econometrics*. Oxford: Oxford University Press.
Hendry, D.F. and Ericsson, N.R. (1990) An econometric analysis of U.K. money demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna Schwartz. *American Economic Review*, 81, 8–38.
Hendry, D.F. and Favero, C. (1992) Testing the Lucas critique: a review. *Econometric Reviews*, 11, 265–306.
Hendry, D.F. and Mizon, G.E. (1978) Serial correlation as a convenient simplification not a nuisance: a comment on a study of the demand for money by the Bank of England. *Economic Journal*, 88, 549–63.
Hendry, D.F. and Mizon, G.E. (1993) Evaluating dynamic econometric models by encompassing the VAR. Chapter 18 in P.C.B. Phillips (ed.), *Models, Methods and Applications of Econometrics: Essays in Honour of A.R. Bergstrom*. Cambridge, MA: Blackwell Publishers, 272–300.
Hendry, D.F., Pagan, A. and Sargan, J.D. (1983) *Dynamic Specification: The Handbook of Econometrics*. Amsterdam: North Holland.
Hendry, D.F. and Richard, J.F. (1982) On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20, 3–33.
Hendry, D.F. and Richard, J.F. (1983) The econometric analysis of economic time series. *International Statistical Review*, 51 111–63.
Henry, M. and Robinson, P.M. (1996) Bandwidth choice in Gaussian Semi-parametric estimation of long-run dependence. In the *Papers and Proceedings of the Athens Conference on Applied Probability and Time Series Analysis*. Robinson, P.M. and Rosenblatt, M. eds. New York: Springer-Verlag, 220–32.
Hosking, J.R.M. (1981) Fractional differencing. *Biometrika*, 68, 165–76.
Hubrich, K., Lutkepohl, H., and Saikkonen, P. (2001) A review of systems cointegration tests. *Econometric Reviews*, 20, 247–318.

Hull, J. (2002) *Options, Futures and Other Derivatives*. London: Prentice Hall.
Hunter, J. (1989a) Dynamic modelling of expectations: with particular reference to the labour market. Unpublished PhD manuscript, London School of Economics.
Hunter, J. (1989b) The effect of cointegration on solutions to rational expectations models. Paper presented at European Econometrics Society Conference in Munich, September.
Hunter, J. (1990) Cointegrating exogeneity. *Economics Letters*, 34, 33–5.
Hunter, J. (1992a) Tests of cointegrating exogeneity for PPP and uncovered interest rate parity for the UK. *Journal of Policy Modelling*, Special Issue: Cointegration, Exogeneity and Policy Analysis 14, 4, 453–63.
Hunter, J. (1992b) Representation and global identification of linear rational expectations models. Paper presented at the European Econometrics Society Conference in Uppsala, CERF Discussion Paper, 92–03, Brunel University.
Hunter, J. (1994) A parsimonious cointegration representation of multi-cointegration. Paper presented at the European Econometrics Society Conference in Maastricht, CERF Discussion paper no 94–02, Brunel University.
Hunter, J. (1995) Representation and global identification of linear rational expectations. Mimeo, Brunel University.
Hunter, J. and Dislis, C.D. (1996) Cointegration representation, identification and estimation. Brunel University, Centre for Research in Empirical Finance, Discussion Paper.
Hunter, J. and Ioannidis, C. (2000) Identification and identifiability of non-linear IV/GMM Estimators. Paper presented at the LACEA conference in Uruguay and the ECSG conference in Bristol, Brunel University Discussion Paper, DP07–00.
Hunter, J and Simpson M. (1995) Exogeneity and identification in a model of the UK effective exchange rate. Paper presented at the EC^2 Conference in Aarhus Dec. 1995 and the Econometrics Society European Meeting in Istanbul 1996.
Inder, B. (1993) Estimating long-run relationships in economics: a comparison of different approaches. *Journal of Econometrics*, 57, 53–68.
Johansen, S. (1988a) The mathematical structure of error correction models. *Contemporary Mathematics*, 80, 359–86.
Johansen, S. (1988b) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12, 231–54.
Johansen, S. (1991a) Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica*, 59, 1551–80.
Johansen, S. (1991b) A statistical analysis of cointegration for I(2) variables. University of Helsinki, *Department of Statistics Report*, no. 77.
Johansen, S. (1992a) Testing weak exogeneity and the order of cointegration in UK money demand data. *Journal of Policy Modelling*, Special Issue: Cointegration, Exogeneity and Policy Analysis, 14, 313–34.
Johansen. S. (1992b) Cointegration in partial systems and the efficiency of single equation analysis. *Journal of Econometrics*, 52, 3, 389–402.
Johansen, S. (1995a) *Likelihood-Inference in Cointegrated Vector Auto-Regressive Models*. Oxford: Oxford University Press.
Johansen, S. (1995b) Identifying restrictions of cointegrating vectors. *Journal of Econometrics*, 69, 111–32.
Johansen, S. (1995c) A statistical analysis of cointegration for I(2) variables. *Econometric Theory*, 11, 25–59.
Johansen, S. (2002a) A small sample correction for the test of cointegrating rank in the vector autoregressive model. *Econometrica*, 70, 1929–61.

Johansen, S. (2002b) A small sample correction for tests of hypotheses on the co-integrating vectors. *Journal of Econometrics*, 111, 195–221.

Johansen, S. and Juselius, K. (1990) Maximum likelihood estimation and inference on cointegration – with applications to the demand for money. *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

Johansen, S. and Juselius, K. (1992) Some structural hypotheses in a multi-variate cointegration analysis of the purchasing power parity and the uncovered interest parity for UK. *Journal of Econometrics*, 53, 211–44.

Johansen, S. and Juselius, K. (1994) Identification of the long-run and the short-run structure: An application to the IS/LM model. *Journal of Econometrics*, 63, 7–36.

Johansen, S. and Swensen A.R. (1999) Testing exact rational expectations in co-integrated vector autoregressive models. *Journal of Econometrics* 93, 73–91.

Juselius K. (1994) Do PPP and UIRP hold in the long-run? – An example of likelihood inference in a multivariate time-series model. Paper presented at Econometric Society European Meeting, Maastricht.

Juselius, K. (1995) Do purchasing power parity and uncovered interest rate parity hold in the long-run? – An example of likelihood inference in a multivariate time-series model. *Journal of Econometrics*, 69, 178–210.

Keynes, J.M. (1939) Professor Tinbergen's method. Reprinted in the *Collected Writings of John Maynard Keynes*, vol. XIV, 306–18.

Kollintzas, T. (1985) The symmetric linear rational expectations model. *Econometrica*, 53, 963–76.

Koopmans, T.C. (1953) Identification problems in economic model construction. In *Studies in Econometric Method*, Cowles Commission Monograph 14, Koopmans, T.C and Hood, W.C., eds. New York: John Wiley and Sons.

Kremers, J.J.M., Ericsson, N.R. and Dolado, J. (1992) The power of cointegration tests. *Oxford Bulletin of Economics and Statistics*, 54, 325–48.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992) Testing the null of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54, 159–78.

Lee, D., and Schmidt, P. (1996) On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics*, 73, 285–302.

Leybourne, S.J. and McCabe, B.M.P. (1994) A consistent test for a unit root. *Journal of Business and Economic Statistics*, 12, 157–66.

Leybourne, S.J., McCabe, B.P.M. and Tremayne, A.R. (1996) Can economic time series be differenced to stationarity? *Journal of Business and Economic Statistics*, 14, 435–46.

Lin, J.-L. and Tsay, R.S. (1996) Co-integration constraint and forecasting: an empirical examination. *Journal of Applied Econometrics*, 11, 519–38.

Lippi, M. and Reichlin, L. (1994) VAR analysis, non-fundamental representations, Blaschke matrices. *Journal of Econometrics*, 63, 290–307.

Lucas, R.E. (1976) Econometric policy evaluation: a critique. In *The Philips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, vol. 1, Brunner K. and Meltzer A.H. (eds). Amsterdam: North-Holland.

Lütkepohl, H. (1991) *Introduction to Multiple Time-Series*. Berlin: Springer-Verlag.

Lütkepohl, H. and Claessen, H. (1993) Analysis of cointegrated VARMA processes. Paper presented at the EC^2 conference at the Institute for Economics and Statistics, Oxford, December.

MacKinnon, J.G. (1991) Critical values for cointegration tests. In *Long-Run Economic Relationships*, R.F. Engle and C.W.J. Granger (eds). Oxford: Oxford University Press.

MacKinnon, J.G., Haug, A.A. and Michelis, L. (1999) Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics*, 14, 563–77.

Maddala, G.S. and Kim, I.-M. (1998) *Unit Roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.

Marinucci, D. and Robinson, P.M. (2001) Finite-sample improvements in statistical inference with I(1) processes. *Journal of Applied Econometrics*, 16, 431–44.

McCabe, B. and Tremayne, A.R. (1993) *Elements of Modern Asymptotic Theory with Statistical Applications*. Manchester: Manchester University Press.

Mosconi, R. and Giannini, C. (1992) Non-causality in cointegrated systems: representation, estimation and testing. *Oxford Bulletin of Economics and Statistics*, 54, 399–417.

Muellbauer J. (1983) Surprises in the consumption function. *Economic Journal*, Supplement March, 34–50.

Nankervis, J.C., and Savin, N.E. (1985) Testing the autoregressive parameter with the t-Statistic. *Journal of Econometrics*, 27, 143–61.

Nankervis, J.C. and Savin, N.E. (1988) The student's t approximation in a stationary first order autoregressive model. *Econometrica*, 56, 119–45.

Nickell, S.J. (1985) Error-correction, partial adjustment and all that: an expository note. *Oxford Bulletin of Economics and Statistics*, 47, 119–29.

Newey, W. and West, K. (1987) A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–8.

Ng, S. and Perron, P. (1995) Unit root tests in ARMA models with data-dependent methods for selection of the truncation lag. *Journal of the American Statistical Association*, 90, 268–81.

Osterwald-Lenum, M. (1992) A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, 54, 461–71.

Park, J.Y. and Phillips, P.C.B (1988) Statistical in regressions with integrated processes: Part I. *Econometric Theory*, 4, 468–97.

Parker, S. (1998) Opening a can of worms: the pitfalls of time series regression analyses of income inequality. *Brunel University Discussion Paper*, 98–11.

Paruolo, P. (1996) On the determination of integration indices in I(2) systems. *Journal of Econometrics*, 72, 313–56.

Patterson, K. (2000) *An Introduction to Applied Econometrics: a Time Series Approach*. Basingstoke: Macmillan.

Patterson K. (2005) *Topics in Nonstationary Economic Time Series*. Basingstoke: Palgrave Macmillan.

Pesaran, M.H. (1981) Identification of rational expectations models. *Journal of Econometrics*, 16, 375–98.

Pesaran, M.H. (1987) *The Limits to Rational Expectations*. Oxford: Basil Blackwell.

Pesaran, M.H., Shin, Y. and Smith, R.J. (2000) Structural analysis of vector error correction models with exogenous I(1) variables. *Journal of Econometrics*, 97, 293–343.

Pesaran, B. and Pesaran, M.H. (1998) *Microfit 4*. Oxford: Oxford Electronic Publishing.

Perron, P. (1989) The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, 57, 1361–1401.

Perron, P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics*, 8, 153–62.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica*, 55, 277–302.

Phadke, M.S. and Kedem, G. (1978) Computation of the exact likelihood function of multivariate moving average models. *Biometrika*, 65, 511–19.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica*, 59, 283–306.

Phillips, P.C.B. (1994) Some exact distribution theory for maximum likelihood estimatiors of cointegrating coefficients in error correction models. *Econometrica*, 62, 73–93.

Phillips, P.C.B. and Hansen, B.E. (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies*, 57, 99–125.
Phillips, P.C.B. and Ouliaris, S. (1990) Asymptotic properties of residual based tests of cointegration. *Econometrica*, 58, 165–93.
Phillips, P.C.B. and Perron, P. (1988) Testing for a unit root in time series regression. *Biometrika*, 75, 335–436.
Podivinsky, J.M. (1993) Small sample properties of tests of linear restrictions on cointegrating vectors and their weights. *Economics Letters*, 39, 13–18.
Reinsel, G.C. and Ahn, S.K. (1992) Vector autoregressive models with unit roots and reduced rank structure: estimation likelihood ratio test, and forecasting. *Journal of Time Series Analysis*, 13, 353–75.
Reimers, H.-E. (1992) Comparisons of tests for multivariate cointegration. *Statistical Papers*, 33, 335–59.
Robinson, P.M. (1994) Semi-parametric analysis of long-memory time series. *Annals of Statistics*, 23, 1630–61.
Robinson, P.M. and Marinucci, D. (1998) Semiparametric frequency domain analysis of fractional cointegration. *STICERD discussion paper EM/98/348*, London School of Economics.
Robinson, P.M. and Yajima, Y. (2002) Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, 106, 217–41.
Rothenberg, T.J. (1971) Identification in parametric models. *Econometrica*, 39, 577–91.
Said, S.E. and Dickey, D.A. (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71, 599–607.
Saikonnen, P. (1991) Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory*, 7, 1–21.
Sargan, J.D. (1964) Wages and prices in the UK: a study in econometric methodology. In *Econometric Analysis for National Economic Planning*, P.E. Hart, G. Mills and J.K. Whitaker (eds). London: Butterworth.
Sargan, J.D. (1975) The identification and estimation of sets of simultaneous stochastic equations. *LSE discussion paper* no. A1.
Sargan, J.D. (1982) Alternatives to the Muellbauer method of specifying and estimating a rational expectations model. Florida University discussion paper 68.
Sargan, J.D. (1983a) Identification and lack of identification. *Econometrica*, 51, 1605–33.
Sargan, J.D. (1983b) Identification in models with autoregressive errors. In *Studies in Econometrics. Time Series and Multivariate Statistics*, S. Karlin, T. Amemiya and L.A. Goodman (eds). New York.: Academic Press, 169–205.
Sargan, J.D. (1988) *Lectures on Advanced Econometric Theory*. Oxford: Basil Blackwell.
Sargan, J.D. and Bhargava, A. (1983) Testing residuals from least squares regression for being generated by a Gaussian random walk. *Econometrica*, 51, 153–74.
Sargent, T.J. (1978) Estimation of dynamic labour demand schedules under rational expectations. *Journal of Political Economy*, 86, 1009–44.
Schwert, G.W. (1989) Tests for unit roots: a Monte Carlo investigation. *Journal of Business and Economic Statistics*, 7, 147–59.
Sims, C. (1980) Macroeconomics and reality. *Econometrica*, 48, 11–48.
Spanos, A. (1986) *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.
Spanos, A. (1994) On modeling heteroskedasticity: the Student's t and elliptical linear regression models. *Econometric Theory*, 10, 286–315.
Spliid, H. (1983) A fast estimation method for the vector auto-regressive moving average model with exogenous variables. *Journal of the American Statistical Association*, 78, 843–49.

Stock, J.H. (1987) Asymptotic properties of least squares estimates of cointegration vectors. *Econometrica*, 55, 1035–56.
Stock, J. and Watson, M. (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61, 783–820.
Stock, J.H. and Watson, M.W. (2003) *Introduction to Econometrics*. Boston: Addison Wesley.
Stock, J.H., Wright, J., and Yogo, M. (2002) A survey of weak instruments, weak identification in GMM. *Journal of Business and Economic Statistics*, 20, 518–29.
Taylor, A.M. (1999) Recursive mean adjustment to tests of the seasonal unit root hypothesis. *Birmingham University Discussion paper*, 99–11.
Theil, H. (1965) The information approach to demand analysis. *Econometrica*, 33, 67–87.
Tobin, J. (1950) A statistical demand function for food in the USA. *Journal of the Royal Statistical Society*, Series A, 113–41.
Toda, H.Y. and Phillips, P.C.B. (1994) Vector autoregression and causality: a theoretical overview and simulation study. *Econometric Reviews*, 13, 259–85.
Wallis, K.F. (1974) Seasonal adjustment and relations between variables. *Journal of the American Statistical Association*, 69, 18–32.
Wallis, K.F., Andrews, M.J., Bell, D.N.F., Fisher, P.G. and Whitley, J.D. (1984) *Models of the UK Economy*. Oxford: Oxford University Press.
White, H. (1980) A heteroskedasticity-consistent covariance estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–38.
Wickens, M.R. (1982) The efficient estimation of econometric models with rational expectations. *Review of Economic Studies*, 49, 55–67.
Wickens, M.R., and Breusch, T.S. (1988) Dynamic specification, the long run and the estimation of transformed regression models. *Economic Journal*, Conference Papers, 98, 189–205.
Wold, H. and Jureen, L. (1953) *Demand* Analysis. New York: Wiley.
Yoo, S. (1986) Multi-cointegrated time series and generalised error-correction models. University of San Diego working paper.
Yule G.U. (1926) Why do we sometimes get non-sense correlation between time-series? A study of sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89, 1–64.
Yule, G.U. (1927) On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philosophical Transactions* (A), 226, 267–98.
Zivot, E. and Andrews, D.W.K. (1992) Further evidence on the Great Crash, the Oil Price Shock and the unit root hypothesis. *Journal of Business and Economic Statistics*, 10, 251–70.

Index