An analysis of order submissions on the Xetra trading system using multivariate time series of counts*

Joachim Grammig† **Andréas Heinen**‡ **Erick Rengifo**§

March 2003

JEL Classification codes: C32, C35, G10.

Keywords: Market microstructure; Asymmetric Information, Liquidity; Multivariate count model.

Abstract

Using an innovative empirical methodology we analyze trading activity and liquidity supply in an open limit order book market and test a variety of hypotheses put forth by market microstructure theory. We study how the state of the limit order book, i.e. liquidity supply, as well as price volatility and limit order cancelations impact on future trading activity and identify those factors which explain liquidity supply.

*The authors would like to thank Luc Bauwens and Helena Beltran for helpful discussions and suggestions. We also wish to thank Helena Beltran for sharing her computer code, which was of great help for our computations. The usual disclaimers apply.

†Eberhard Karls University of Tübingen.

‡Corresponding author. University of California, San Diego and Center of Operations Research and Econometrics, Catholic University of Louvain, 34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium, e-mail: heinen@core.ucl.ac.be

§Center of Operations Research and Econometrics, Catholic University of Louvain.

1 Introduction

More and more trading venues are organized as open limit order markets in which assets are traded during continuous automated auctions. Unlike in the most prominent of stock markets, the New York Stock Exchange, no role is cast for a dedicated market maker responsible for managing liquidity supply. Whether traders asking for immediate transactions will be able to buy or sell their desired volume without having to bear large price impacts depends solely on the state of the electronic limit order book. The limit order book consists of previously submitted, non executed buy and sell orders with a given price limit which can be viewed as free options written by patient market participants (see Lehmann (2003)). In modern trading systems the state of the limit order book is disclosed (albeit sometimes not entirely) at each point in time. For large stocks traded in developed markets the trading process is rapid and highly dynamic with a limit order book permanently in flux. The arrival of new information induces changes in trading strategies which are implemented by cancelations, revisions and new submissions of limit and market orders. Microstructure theory has put forth a variety of propositions about how information processing affects trading activity, price formation and liquidity supply in limit order markets. Due to the increasing availability of detailed transaction data and the recent development of econometric techniques for the analysis of financial transactions data it is possible to test some of these hypotheses and draw conclusions for market design issues.

In this paper we employ a recently developed methodology for the econometric modelling of multivariate time series of counts to analyze the trading and liquidity supply and demand process in an open limit order book market. In an empirical analysis using data from the Xetra system, which operates at various exchanges in continental Europe, we analyse how the inside spread and the depth at the best quotes affect the submission of market or limit orders. We offer insights about the factors which explain the state of the limit order book (i.e. liquidity supply) and how liquidity supply, in turn, impacts on future trading activity. Furthermore, we study how price volatility affects the behaviour of limit and market order traders and analyze the role that cancelations of limit orders play in the process of information processing.

This is not the first paper that deals with those issues. Related work has focussed on whether a trader chooses a market or limit order and how market conditions affect this choice (see, e.g., Biais, Hillion & Spatt (1995), Griffiths, Smith, Turnbull & White (2000), and Ranaldo (2003)). Sandas (2001) uses Swedish order book data and estimates a version of Glosten (1994) celebrated limit order book model. More recently Pascual & Veredas (2004) analyse the limit order book information of the Spanish Stock Exchange and find that most of the explanatory power of the book concentrates on the best quotes. The interest in empirical microstructure and market design spurred the development of econometric techniques for the analysis of financial transactions data. Russell (1999) and Engle & Lunde (2003) have proposed multivariate econometric models to analyse financial markets. Hasbrouck (1999) discusses how to apply these models to analyse financial market microstructure processes. The present paper links and contributes to the literature in the following ways.

As in Biais et al. (1995) we study in detail the trading process in a developed limit order market. Following their approach we break up limit orders according to their aggressiveness and study the order submission, execution and cancelation processes. Additionally, we distinguish less aggressive limit orders in terms of their relative position in the limit order book with respect to the best quotes. We show that this constitutes an improvement over the categories proposed in Biais et al. (1995) as the analysis of the new order categories provides new insights into trader behaviour.

The empirical methodology is quite new in empirical finance. So far, most studies have used either ordered probit models to analyse the type of the next event (see Ranaldo (2003) for example), in which case the time series aspect is not adequately taken into account; or, alternatively, duration models for which it is difficult to account for multivariate aspects. The difficulty arises from the nature of financial transactions data, which are, by definition, not aligned in time. We propose employing an alternative methodology [1] which amounts to counting the number of relevant market events in a given time interval arrivals of orders of the different types in a given interval.

We work with the multivariate time series count model proposed by Heinen & Rengifo (2003) which enables us to take into account the relationship between the different components of the trading process. We assess empirically theoretical predictions of market microstructure models about trader behaviour, the incorporation of information into prices and liquidity dynamics. In particular, we determine whether agents time their trades in order to get a lower price impact (strategic placement of orders) and how much the various components of market activity react to changes in the state of the order book.

We have excellent data for these purposes. During the period of the sample the Xetra system displayed the whole order book to all market participants. This is in sharp contrast to other order-driven markets, where only the five or so best prices are shown and/or where hidden orders are allowed. For the purpose of this study we had access to a complete record of the submission/cancellation/execution events of different types of orders on stocks traded in the Xetra system operated by the Frankfurt Stock Exchange. Using the data and the Xetra trading rules we completely reconstruct the prevailing order book at any point in time. This provides the most detailed information about liquidity supply. The market events in which we are particularly interested in are market order entries, limit order submissions, submissions of marketable limit orders (a limit order submitted at a price which makes them immediately executable and in that respect indistinguishable from market orders) and cancelations of limit orders.

In the empirical analysis we study the effect of the state of the limit order book on the submission of the different types of orders. To measure the state of the book and liquidity supply we construct the following indicators. Most naturally we start with the inside spread and depth at the best quotes being the most widely used information about liquidity and asymmetric information. To condense the information present in the whole order book, we follow Beltran-Lopez, Giot & Grammig (2004) and employ a factor decomposition of the buy and sell side liquidity. This approach is similar to what is done in the term structure of interest rates literature. More precisely, the input for the factor decomposition are percentage average prices relative to the best quote that a market order of a given size would get if it were executed against the existing book. We then use Principal Components

[1] The CBIN model proposed by Davis, Rydberg & Shephard (2001) is another model that uses counts that its authors use to test for common features in the speed of trading, quote changes, limit and market order arrivals.

Analysis to summarise this information using only a small number of well interpretable factors, typically three, which enable us to explain most of the variation in the book. Furthermore, we study the impact of volatility on trading activity. Theory suggests that the level of volatility determines the choice between submitting a limit or a market order. We also analyse the impact of past order flow and test for the presence of diagonal effects, i.e. the hypothesis that events of the same type (e.g. submission of aggressive buy orders tend to follow each other. We also analyse the inter-dependence of market and limit orders and present an exhaustive analysis of the role that cancellations play in the trading process.

The empirical analysis delivers the following results. As predicted by theoretical models of financial market microstructure (Foucault (1999), Handa & Schwartz (1996)) we find that larger spreads reduce the relative importance of market order trading activity compared to limit order submissions (with a negative effect of the spread on general trading activity). Increasing depth at the best quotes stimulates submission of aggressive limit orders at the same side of the market as limit order traders strive for price priority. On the other hand, a larger depth on the opposite side of the market reduces aggressiveness of limit orders on the own side. This indicates that the presence of traders with a different asset valuation on the opposite side of the market makes it profitable to place less aggressive orders at the own side and wait for being hit. These insights, which are consistent with hypotheses that can be derived from the paper by Parlour (1998), can only be obtained when working with our finer classification of order types. Using the aggregated data these results are blurred. As in Beltran-Lopez et al. (2004), we find that not more than three factors explain a considerable fraction of the variation of market liquidity. We find, in line with the predictions derived from Foucault (1999) theoretical analysis, that one of the factors, which can be interpreted as the "informational factor", proves to be very useful for predicting the order submission process. For example, if the informational factor indicates a "bad news" regime, aggressive limit buy and market order trading increases whilst seller activity decreases. Consistent with theoretical predictions we also find that order aggressiveness is reduced and cancelation activity rises when price volatility increases. Analysing the dynamics of the multivariate system we find evidence for a diagonal effect, similar as in Biais et al. (1995). More precisely, an increase in the number of one minute counts of a certain type of event exerts a strong positive impact on the conditional mean of the same type of event in the next one minute interval. Furthermore, we report that buy (sell) market orders initiate submissions of sell (buy) limit orders indicating market resiliency. Finally, our results indicate that cancelations do matter in the sense that they carry information for predicting future market activity. Cancelations of aggressive limit orders (close to the best quotes) lead to reduced trading activity. However, cancelations also induce an increase in the submission of limit orders inside the first five quotes. This indicates that whilst liquidity is reduced, it does not vanish from the market entirely. These interesting facts about cancelations warrant future theoretical and empirical research.

The paper is organised as follows. Section 2 briefly describes the market structure and the Xetra system. Section 3 presents the data, and descriptive anaylsis as well as the the different categories of order types. Section 4 explains the econometric methodolgy. Section 5 presents the empirical results and a discussion. Section 6 concludes and provides an outlook for future research.

2 Market Structure

We use data from the automated auction system Xetra which was developed by the German Stock Exchange. After its introduction at the Frankfurt Stock Exchange (FSE) in 1997, Xetra has become the main trading venue for German blue chip stocks. The Xetra system is also the trading platform of the Dublin and Vienna stock exchanges as well as the European Energy exchange. The Xetra system represents the platform for a pure electronic order book market. The computerized trading protocol keeps track of the entries, cancelations, revisions, executions and expirations of market and limit orders. For blue chip stocks there are no dedicated market makers, like the specialists at the New York Stock Exchange (NYSE) or the Japanese saitori. For some small capitalized stocks listed in Xetra there may exist Designated Sponsors - typically large banks - who are obliged, but not forced, to provide a minimum liquidity level by simultaneously submitting competing buy and sell limit orders.

Xetra does face some local competition for order flow. The FSE maintains a parallel floor trading system, which bears some similarities with the NYSE. Furthermore, like in the US, some regional exchanges participate in the hunt for liquidity. However, due to the success of the Xetra system, the FSE floor, previously the main trading venue for German blue chip stocks, became less important. The same holds true for the regional exchanges. However, they retain market shares for smallest capitalized local firms. Initially, Xetra trading hours at the FSE extended from 8.30 a.m to 5.00 p.m. CET. From September 20, 1999 the trading hours were shifted to 9.00 a.m to 5.30 p.m. CET. The trading day begins and ends with call auctions. Another call auction is conducted at 12.00 p.m. CET. Outside the call auctions periods the trading process is organised as a continuous double auction mechanism with automatic matching of orders based on price and time priority. Bauwens & Giot (2001) provide a complete description of an order book market and Biais, Hillion & Spatt (1999) describe the opening auction mechanism employed in an order book market and corresponding trading strategies. Five other Xetra features should be noted.

- Assets are denominated in euros, and uses a decimal system, which implies a small minimum tick size (1 euro-cent).
- Unlike at Paris Bourse, market orders exceeding the volume at the best quote are allowed to "walk up the book". At Paris Bourse the volume of a market order in excess of the depth at the best quote is converted into a limit order at that price entering the opposite side order book, In Xetra, however, market orders are guaranteed immediate full execution, at the cost of getting a higher price impact in the trades.
- Dual capacity trading is allowed, i.e. traders can act on behalf of customers (agent) or as principal trader on behalf of the same institution (proprietary).
- Until March 2001 no block trading facility (like the upstairs market at the NYSE) was available.
- Before 2002, and during the time interval from which our data is taken, only round lot order sizes could be filled during continuous trading hours. A Xetra round lot was defined as a multiple of 100 shares. Execution of odd-lot parts of an order - this is an integer valued fraction of one hundred shares - was possible only during call auctions.

Besides these technical details, the trading design entails some features which render our sample of Xetra data (described in the next section) particularly appropriate for our empirical analysis. First, the Xetra system displays not only best quotes, but the contents of the whole limit order book. This is a considerable difference compared to other systems like the Paris Bourses CAC system, in which only the best five orders are displayed. Second, hidden limit orders (or iceberg orders) were not known until a recent change in the Xetra trading rules that permitted them. [2] As a result, the transparency of liquidity supply offered by the system was quite unprecedented. However, Xetra trading is completely anonymous. The Xetra order book does not reveal the identity of the traders submitting market or limit orders.[3]

3 Data

The dataset used for our study contains complete information about Xetra market events, that is all entries, cancelations, revisions, expirations, partial-fills and full-fills of market and limit orders that occurred between August 2, 1999 and October 29, 1999. Due to the considerable amount of data and processing time, we had to restrict the number of assets. Market events were extracted for three blue chip stocks, Daimler Chrysler (DCX), Deutsche Telekom (DTE) and SAP. The combined weight represented 30.4 percent in the DAX index at the end of the sample period. The three blue-chip stocks under study are also traded at several important exchanges. Daimler-Chrysler shares are traded at the NYSE, the London Stock Exchange (LSE), the Swiss Stock Exchange, Euronext, the Tokyo Stock Exchange (TSE) and at most of German regional exchanges. SAP is traded at the NYSE and at the Swiss Stock Exchange. Deutsche Telekom is traded at the NYSE and at the TSE. They are also traded on the FSE floor trading system, but this accounts for less than 5% of daily trading volume in those shares. Trading volume at the NYSE accounts for about 20% of daily trading volume in those stocks. As the prices for our three stocks remained above 30 euros during the sample period, the tick size of 0.01 Euros is less than 0.05% of the stock price. Hence, we should not observe any impact of the minimum tick size on prices or liquidity. Based on these market events we perform a real time reconstruction of the order book sequences. Starting from the initial state of the order book, we track each change in the order book implied by entry, partial or full fill, cancelation and expiration of market and limit orders. This is done by implementing the rules of the Xetra trading protocol outlined in Deutsche Börse AG (1999) in the reconstruction program. From the resulting real-time sequences of order books, snapshots at 1 minute interval during the continuous trading hours were taken. For each snapshot, the order book entries were sorted on the bid (ask) side in price descending (ascending) order.

The large number of marketable limit orders (MLO) compared to "true" market orders is remarkable. A MLO is a limit order which is submitted at a price which makes it immediately executable. In this respect it is indistinguishable from a "true" market order.

[2] Biais et al. (1995) show that the possibility of hiding part of the volume of a limit order leads to all sorts of specific trading behaviour, for example submitting orders to "test" the depth at the best quote for hidden volume.

[3] Further information about the organization of the Xetra trading process and a description of the trading rules that applied to our sample period is provided in Deutsche Börse AG (1999)

However, MLOs differ from market orders in that the submitter specifies a limit of how much the order can walk up the book. Hence, a MLO might be immediately, but not necessarily completely filled. The non-executed volume of the MLO then enters the book.[4] In our empirical analysis we therefore treat the either completely or partially filled parts of an MLO just like a market orders. When, for the sake of brevity, we will refer in the following refer to "market orders" what we precisely mean is "true market orders and completely/partially filled marketable limit orders".

For the purpose of this study we classify market and limit orders in terms of aggressiveness following Biais et al. (1995):

- Category 1: Large market orders, orders that walk up the book.
- Category 2: Market orders, orders that consume all the volume available at the best quote.
- Category 3: Small market orders, orders that consume part of the depth at the best quote.
- Category 4: Aggressive limit orders, orders submitted inside the best quotes.
- Category 5: Limit orders submitted at the best quote.
- Category 6: Limit orders outside the best quotes, orders that are below (above) the bid (ask).
- Category 7: Cancelations.

Moreover, we break up categories 6 and 7 according to their relative position with respect to the best quote, measured in either number of steps or in terms of a given percentage increment of the best price. The resulting series will be referred to as the "disaggregated" data whilst the data resulting from the Biais et al. (1995) classification will be referred to as the "aggregated" series. The disaggregated data will be useful to test hypotheses about the informativeness of the state of the order book. We then count the submission/cancellation events in the different categories during each one minute interval of the sample. The resulting multivariate sequence of counts provides the input for the econometric model described in the next section.

To avoid dealing with the change in trading times, and given the large number of observations, we restrict the whole sample to observations between August 20 to September 20, 1999. The data therefore contain information about 21 trading days with 510 one-minute intervals per day giving a total of 10730 one minute intervals. Due to space limitations we will only report the results for Daimler-Chrysler (DCX).[5] Sample statistics are presented in Table (1) where the main characteristics of the data can be appreciated. First, the number of buy (sell) limit orders is 3.35 (4.7) times larger than the number of market orders. The presence of MLOs is striking, especially in the two first categories on both

[4] MLOs therefore share some properties with Paris Bourse market orders.

[5] The results obtained with the other two assets confirm the findings we present here. These results are available upon request.

sides of the market. This gives some intuition about how the traders participate in the continuous trading period: they send MLOs to fix the maximum price impact they want to bear. The means of all the categories are very small giving us the baseline to decide the use of a discrete distribution rather than a continuous one such as the normal. Moreover, we appreciate that all series are overdispersed (the standard deviation is larger than the mean), which has implications for the appropriate statistical model to be used.

Figure (1) presents two days auto- and cross-correlograms of the aggregated series for Daimler-Chrysler (DCX). We consider buy and sell market orders, limit orders and total cancelations of both sides of the market. Observing the autocorrelations one can see that all series of counts show persistence in the occurrences. A visual inspection of the cross correlations between market buys and market sells shows that these are almost symmetric. This implies that that the tendency of market buys at time t to follow market sells of time $t-k$ is almost the same as the tendency of market sells to follow market buys. This indicates that the informational effects, found by Hasbrouck (1999) when analysing data from the TORQ dataset, are not detectable in our data.

Figure (2) presents an analysis of the daily seasonality of the aggregated variables. It should be noted that neither buy nor sell market order counts reflect the often reported U-shape of intra-day financial series. There is a small increase in the number of counts at about 2.30 p.m. CET which most likely corresponds to the NYSE opening time. The number of buy limit orders is large early in the morning, but decays quite fast. Then, limit orders at both sides of the book behave similarly in that we observe an increase in trading activity in the afternoon at the same time as the market order activity increases. We observe a similar pattern in the cancelation series.

4 The Model

In this paper we are interested in modeling the process of order submissions in minute detail. In order to do this, given the limitations of ordered probits in capturing the full dynamics of the order submissions and the difficulties associated with extending duration models to large multivariate systems, we choose to work with the number (counts) of all the different types of orders that are being submitted to the market in one minute intervals. As we are mainly interested in the dynamic interactions between the various components of the order flow, we want to work with a multivariate dynamic model. As can be seen from the descriptive statistics, the series we work with have very small means, which makes the use of a continuous and symmetric distribution like the Gaussian questionable. This is why we want to model discreteness explicitly. Finally, the series we consider usually have a variance which is larger than their mean. This property is referred to as overdispersion, and we want a model which is able to match this stylised fact.

Let us now describe in more detail the Multivariate Autoregressive Conditional Double Poisson (MDACP) model used in this paper. The MDACP was developed in Heinen & Rengifo (2003), and this section draws on that paper, but we refer the reader to the original paper for more technical details.

In order to model a $(K \times 1)$ vector of counts N_t, we build a VARMA-type system for the conditional mean. In a first step, we assume that conditionally on the past, the different series are uncorrelated. This means that there is no contemporaneous correlation and that

Table 1: **Descriptive statistics of the types of orders per 1-minute interval**

	Obs	**Mean**	**Std. Dev.**	**Disp.**	**Max.**	$Q(60)$
BUY ORDERS	**52712**	**4.91**	**4.37**	**3.89**	**68**	**37817**
Large MO of which	**3494**	**0.33**	**0.71**	**1.53**	**22**	**7888**
- Large MO	898	0.08	0.36	1.51	18	872
- Large MLO	2596	0.24	0.56	1.28	6	7396.8
MO of which	**3369**	**0.31**	**0.64**	**1.32**	**6**	**1629**
- MO	18	0.01	0.04	1.00	1	64.4
- MLO	3351	0.31	0.64	1.33	6	1627
Small MO of which	**5250**	**0.49**	**0.81**	**1.33**	**7**	**11106**
- Small MO	2564	0.24	0.54	1.22	6	8990
- Small MLO	2686	0.25	0.55	1.23	5	1344
Total MO	**12113**	**1.13**	**1.46**	**1.89**	**29**	**22759**
LO above the best bid (overbidding)	18312	1.71	1.85	2.00	17	21309
LO at the best bid	11411	1.06	1.33	1.68	18	14313
LO below the best bid	10876	1.01	1.28	1.62	11	8657
Total LO	**40599**	**3.78**	**3.35**	**2.96**	**39**	**33304**
Cancelations	**20534**	**1.91**	**2.03**	**2.15**	**18**	**13623**
SELL ORDERS	**43163**	**4.02**	**3.92**	**3.82**	**38**	**20498**
Large MO of which	**2263**	**0.21**	**0.53**	**1.36**	**6**	**1442**
- Large MO	524	0.05	0.23	1.12	3	472
- Large MLO	1739	0.16	0.45	1.25	5	1125
MO of which	**3077**	**0.29**	**0.63**	**1.38**	**8**	**2602**
- MO	94	0.01	0.11	1.33	5	305
- MLO	2983	0.28	0.62	1.36	7	2551
Small MO of which	**2241**	**0.21**	**0.52**	**1.32**	**10**	**833**
- Small MO	892	0.08	0.31	1.14	5	362.08
- Small MLO	1349	0.13	0.40	1.24	8	426
Total MO	**7581**	**0.71**	**1.15**	**1.86**	**15**	**5331**
LO below the best ask (undercutting)	15012	1.34	1.68	2.00	13	11184
LO at the best ask	10166	0.95	1.30	1.78	23	8660
LO above the best ask	10404	0.97	1.25	1.62	11	6738
Total LO	**35582**	**3.32**	**3.14**	**2.97**	**38**	**21272**
Cancelations	**20010**	**1.86**	**2.09**	**2.34**	**29**	**11379**

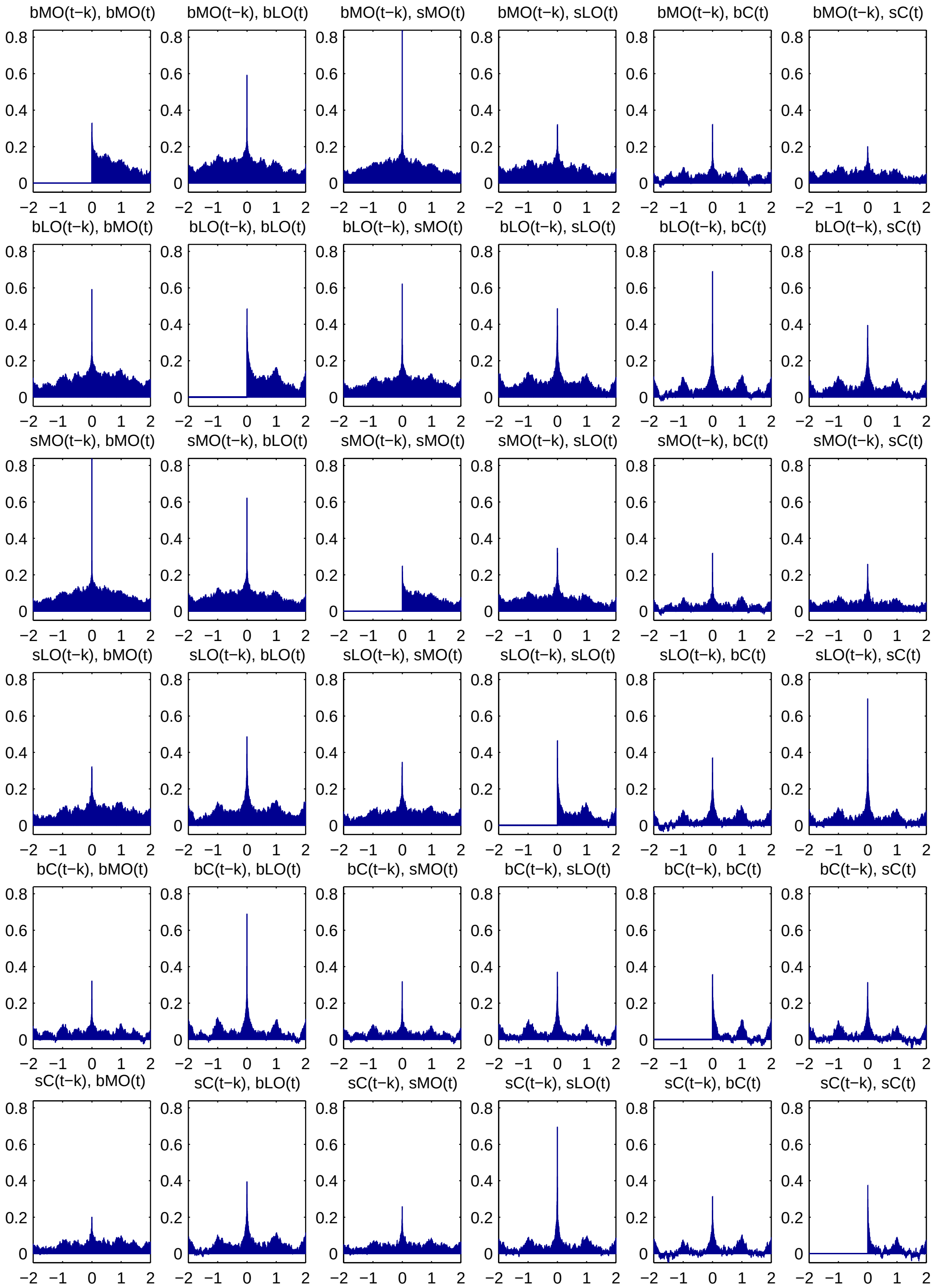

Figure 1: Cross-correlation of aggregated data of DCX.

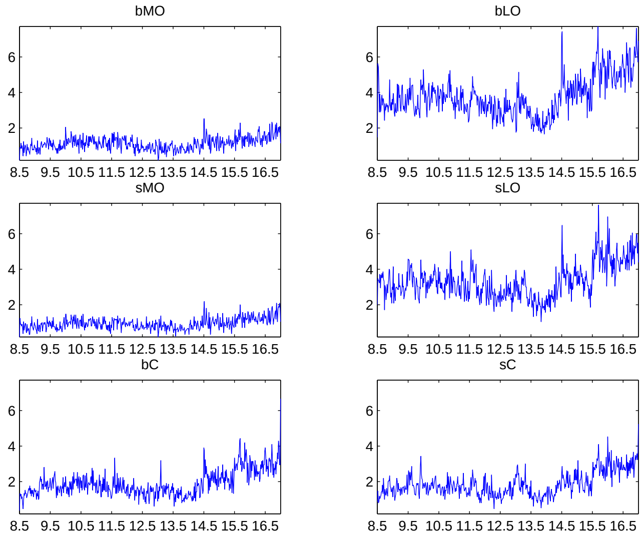

Figure 2: Seasonality of aggregated data of DCX.

all the dependence between the series is assumed to be captured by the conditional mean. Even though the Poisson distribution with autoregressive means is the natural starting point for counts, one of its characteristics is that the mean is equal to the variance, property referred to as equidispersion. However, by modelling the mean as an autoregressive process, we generate overdispersion in even the simple Poisson case. In some cases one might want to break the link between overdispersion and serial correlation. It is quite probable that the overdispersion in the data is not attributable solely to the autocorrelation, but also to other factors, for instance unobserved heterogeneity. It is also imaginable that the amount of overdispersion in the data is less than the overdispersion resulting from the autocorrelation, in which case an underdispersed marginal distribution might be appropriate. In order to account for these possibilities we consider the double Poisson distribution introduced by Efron (1986) in the regression context, which is a natural extension of the Poisson model and allows one to break the equality between conditional mean and variance. The advantages of using this distribution are that it can be both under- and overdispersed, depending on whether ϕ is larger or smaller than 1. We write the model as:

$$N_{i,t}|\mathcal{F}_{t-1} \sim DP(\mu_{i,t}\, exp(X_t\gamma_i), \phi_i)\ ,\ \forall i = 1, \ldots, K. \tag{4.1}$$

where $\mathcal{F}_{t-1}$ designates the past of all series in the system up to time $t-1$ and $X_t\gamma_i$ is the effect of some exogenous variables on series i[6]. With the double Poisson, the conditional variance is equal to:

$$V[N_{i,t}|\mathcal{F}_{t-1}] = \sigma^2_{i,t} = \frac{\mu_{i,t}}{\phi_i} \tag{4.2}$$

[6]It is shown in Efron (1986) (Fact 2) that the mean of the Double Poisson is μ and that the variance is approximately equal to $\frac{\mu}{\phi}$. Efron (1986) shows that this approximation is highly accurate, and we will use it in our more general specifications.

The coefficient ϕ_i of the conditional distribution will be a parameter of interest, as values different from 1 will represent departures from the Poisson distribution. The Double Poisson generalises the Poisson in the sense of allowing more flexible dispersion patterns. The conditional means μ_t are assumed to follow a VARMA-type process:

$$E[N_t|\mathcal{F}_{t-1}] = \mu_t = \omega + \sum_{j=1}^{p} A_j N_{t-j} + \sum_{j=1}^{q} B_j \mu_{t-j} \tag{4.3}$$

For reasons of simplicity, in most of the ensuing discussion, we will focus on the most common $(1,1)$ case and for notational simplicity, we will denote $A = \sum_{j=1}^{p} A_j$ and $B = \sum_{j=1}^{q} B_j$ and drop the index whenever there is no ambiguity.

We evaluate models on the basis of their log-likelihood, but also on the basis of their Pearson residuals, which are defined as: $\epsilon_t = \frac{N_t - \mu_t}{\sigma_t}$. If a model is well specified, the Pearson residuals will have variance one and no significant autocorrelation left.

5 Empirical Results

In this section we use the MDACP model and the count data from DCX. First of all, we determine the aspects that make our model a well specified one. Then, we analyse the state of the book, the volatility and the dynamics of order submissions.

Table 3 presents the results from the estimation of a system with six variables: Buy Market Orders (BMO), Buy Limit orders (BLO), Sell Market Orders (SMO), Sell Limit Orders (SMO), Buy Cancelations (BCANC) and Sell Cancelations (SCANC). This is the most aggregated system we analyse, but it already offers opportunities to test several predictions of theoretical models. Tables 4 and 5 present the results of a more detailed system, where orders are divided up according to aggressiveness into the six categories described earlier. Finally, tables 6, 7 and 8 present a system in which the behaviour and relations of cancelations is analysed.

5.1 Specification testing: Dispersion

For all variables the Poisson assumption (the null hypothesis that the dispersion coefficient is equal to one) is strongly rejected. All marginal distributions are significantly overdispersed, giving support for the use of the Double Poisson distribution, that is the underlying distribution of our model.

We evaluate the different systems on the basis of their log-likelihood, but also on the basis of their Pearson residuals. If a model is well specified, the Pearson residuals will have variance one and no significant autocorrelation left. In the last line of each table we present the variance of the Pearson residuals and figure 3 presents the autocorrelogram of the system presented in table 3 [7]. We can appreciate that the variances are very close to 1 and that there is no autocorrelation left, meaning that our models are well specified.

[7] Because of limitations of space we do not present the other Autocorrelograms, but they behaviour is similar to the one presented in the paper.

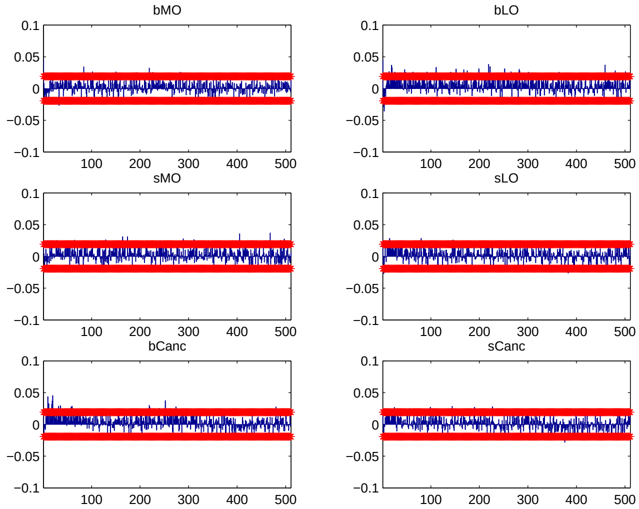

Figure 3: Autocorrelogram of the system presented in table 3.

5.2 The State of the Book

The limit order book gathers prices and volumes at which eager and patient traders are willing to engage in trades. As such it contains information related to liquidity, as well as potential price information. The most basic information given by the book and which is available in all markets, not only in highly transparent electronic limit order markets like Xetra is the inside spread. There exists a large theoretical and empirical literature, which documents that the inside spread is a measure of the amount of information-based trading that is going on in the market. The next piece of information, which is also available in most markets is the quoted depth at the best quotes. This represents the amount of stock that a market order can get without adversely affecting the price. Obviously the higher this is, the higher the liquidity is. Finally, as we have sufficiently precise data to reconstruct the full order book we use a third information set based on information in the book up to high volumes. We now proceed to analyse the impact on order flow composition of these three information sets and relate our empirical findings with theoretical results.

5.2.1 The Inside Spread

Theoretical models of Handa & Schwartz (1996) and Foucault (1999) predict that large spreads reduce the proportion of market orders relative to limit orders in the order flow. This result holds because higher spreads mean a higher price of immediacy, which therefore makes market orders less attractive relative to limit orders, which get a higher reward for providing liquidity to the market. This is confirmed in the empirical studies of Griffiths et al. (2000) and Ranaldo (2003).

In table 3 we observe that the spread has a negative effect on all six components of the order flow, therefore there seems to be a slowing down in overall order submissions and cancelations when the spread is high. However, the spread has a strong negative impact on market orders on both sides of the markets and a much smaller negative effect on limit orders. These results support the theoretical predictions in the sense that our findings show that even though the whole market slows down when the spread is high, the proportion of market orders in the order flow decreases while the proportion of limit orders increases. Analysing tables 4 and 5, in which the orders are divided according to their aggressiveness, we see that the spread has a significant and negative effect in the market orders and a negative but not significant effect on limit orders.

5.2.2 Volume at Best Quotes

The volume at the best quotes is linked with liquidity aspects of the market. Theoretically, Parlour (1998), Handa, Schwartz & Tiwari (2003) show that the size of the depth at the best quotes on one side of the market is related to the execution probability of limit orders on that same side. The larger the depth, the lower the execution probability. When the execution probability of a limit order is low, traders have more incentives to act more aggressively, submitting either limit orders inside the best quotes or market orders. Moreover, when the depth at the opposite side of the market is large, the aggressiveness on the own side decays. This is due to an informational effect, since traders interpret large depth on the opposite side as an indication that there is a high proportion of traders wanting to buy (sell) given their own valuation of the asset, based on common and private information. They also understand that this situation (large depth at the opposite side) will lead the opposite side traders to become more aggressive and either overbid (undercut) their best prices by submitting buy (sell) limit orders inside the best quotes or by submitting buy (sell) market orders to consume the shares available at the opposite side.

According to our findings from the aggregated system (table 3), the volume at the best quotes has a positive effect on all components of the order flow. The volume at the bid (ask) has a positive effect not only on the buy (sell) side but also on the opposite side. While the own side effect provides support for the theoretical models, the opposite side effect is contrary to what the models predict.

In order to get more insight into this, we analyse the results of the disaggregated system presented in tables 4 and 5.

Firstly, let us examine the same-side effect. The theoretical results are fully supported, i.e. as soon as the queue at the best quote is large, traders become more aggressive in order to get price-time priority. Depth at the bid (Bidvol) increases the number of buy market orders of categories 1 and 2, but decreases the number of small buy market orders (category 3). The volume at the ask (Askvol) has similar impacts on its own side with the only exception that it also positively influences the small sell market orders. On the other hand, in accordance with theoretical results, when the depth at the bid (sell) is large, traders prefer to send buy (sell) limit orders inside the best quote than limit orders at the best quotes. This is the case because the execution probability decreases when the queue at the best quote is large.

Secondly we analyze the traders' behaviour in response to changes in the opposite side depth, we observe from tables 4 and 5, that when the depth at the bid (ask) gets larger the two most aggressive sell (buy) market orders (categories 1 and 2 respectively) decrease and meanwhile all the other type of orders increases, i.e. order aggressiveness decreases as opposite side volume increases, giving support to theoretical models that make these predictions. This last result was not possible to observe in table 3 suggesting that analysing the relationships using only the aggregated data could give unprecise relations that only could be appreciated at a disaggregated level.

5.2.3 Information beyond the best quotes: Factors of the Limit Order Book

In order to analyse the impact of the state of the order book on traders' strategies, we use a factor decomposition of the limit order book, which is similar to what is done in the term structure of interest rates literature. We use as input the deseasonalised percentage average price (with respect to the best quote) that a market order for volume v would get if it were executed immediately against the existing book at time t. We compute this for all volumes on a grid of 1000 shares. We then use Principal Components Analysis (PCA) to summarise this information with a small number of factors. PCA is designed to reduce a group of variables into linear combinations that best represent the variation in the original data set. For example the first principal component is the normalized linear combination (the sum of squares of the coefficients being one) with maximum variance [8]. Note, that the linear combinations produced by the PCA are uncorrelated with each other. This will prove useful in the interpretation of our results when we will need to distinguish between informational and liquidity effects.

Table 2 presents the variance proportion of the first five components and figure 4 shows the graph of the factor weights of the first three components estimated using the percentage average price at the buy side (similar results are obtained for the sell side). Looking at the table we can appreciate that the first three components already explain 99% of the total variation of the data. Thus, these three factors enable us to explain virtually all the variation in the book. The first factor has nearly constant loadings for all volumes and we therefore interpret it as the mean effect. If the weight of this factor increases, this means that the percentage average price is increasing on average for all trades of any volume v. The second factor is typically negatively related to the mark up at small volumes and the factor loadings increase monotonically as the volume increases. This factor is therefore related to the slope of the price schedule. Finally the third factor has positive factor loadings for small and large volumes and negative loadings for the volumes in the middle of the range. This factor is thus related to the convexity of the price schedule. We do this analysis separately for the bid and the ask side.

Hall, Hautsch & Mcculloch (2003) propose to use the difference between the absolute bid and ask slope to study the imbalances between the buy and sell side. Thus, a positive difference implies higher liquidity on the ask side of the market. This corresponds to the imbalance between buyers and sellers of theoretical models like the ones of Foucault (1999) and Handa et al. (2003). Moreover, Hall et al. (2003) note that there is a trade off between an interpretation in terms of liquidity or information of the relation between liquidity and trading intensity. On one hand more liquidity implies a lower price impact, therefore inducing more trading. On the other hand, more liquidity on one side of the market could be associated with information about the future price of the asset affecting trade in an opposite way. Working with the orthogonal factors computed by PCA allows us to go beyond this possible confusion between informational and liquidity effects. Our

[8]For a detailed discussion of factorial analisis we refer to Anderson (1984)

Table 2: **Principal Components Analysis of the Limit Order Book.**
The table presents the eigenvalues, the percentage of the explained variance and the cumulative explained variance of the first five principal components estimated using the deseasonalised percentage average price (respect to the best quote).

	Principal Component Analysis				
	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5
Eigenvalue	32.83	3.90	0.80	0.24	0.09
VarianceProp.	0.864	0.103	0.021	0.006	0.002
CumulativeProp.	0.864	0.967	0.988	0.994	0.996

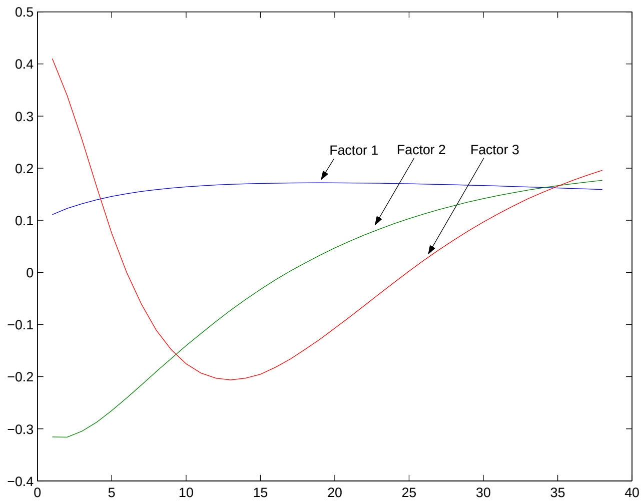

Figure 4: First three components of the percentage average price for volume v of DCX.

results show that the first factor, related linearly to the markup at all volumes is a liquidity factor, whereas the second factor, which represents the difference between the absolute values of the slopes of the book at the bid and at the ask is related to the informational aspects of the market.

Tables 4 and 5 present both results. Note that an increase in the first factor on the buy (sell) side implies more liquidity on that side, since an upward (downward) parallel movement of the percentage average price in the buy (sell) side implies that the price impact is decreasing. Thus, we expect that when the market is more liquid trading aggressiveness increases, since movements on one side give own side traders incentives to act more aggressively in order to get price-time priority affecting also positively to the other side traders in that they can obtain small price impacts on their trades. First, observing the own side effect of the first factor ($Bfact1$ and $Sfact1$), we see that an increase in the buy (sell) factor, increases the aggressiveness of the same side, meaning that traders submit relatively more market and aggressive limit orders to obtain price-time priority. The effect on the other side, due to the decrease of the price impact, is also positive, i.e. order aggressiveness of the opposite side increases. An increase of the first factor of the buy side affects sell market orders positively, while the effect of the first factor of the sell side increases not only buy market orders but also the most aggressive buy limit orders, even though the proportional effect on market orders is higher.

We do not use the second factors directly, instead we use the difference of the absolute values of the second factors at the bid and at the ask (diffslope). This variable will be positive when the book on the ask side is relatively flat and the book on the bid side is steep. Such a book is a signal that there is bad news about the stock and prices are expected to go down, as buyers are only willing to buy small amounts of shares and more volume can only be traded at much less favourable prices. Theoretical models predict that under these circumstances buyers (sellers) become less (more) aggressive. In the first case because traders do not want to pay higher prices for an asset whose price is expected to fall and in the second case because traders compete to obtain the best prices available in the market. This theoretical result is supported by the results presented in tables 4 and 5. We observe that if diffslope increases (possible bad news arriving to the market), buy orders become less aggressive, while sell orders become more aggressive, in accordance with the previous explanation.

5.3 Volatility

Foucault (1999) shows that when volatility increases, limit order traders ask for a higher compensation for the risk of being picked off, i.e. being executed when the market has moved against them. Bae, Jang & Park (2003) and Danielson & Payne (2001) find that the aggressiveness of the orders submitted is lower when the volatility is high. Griffiths et al. (2000) and Ranaldo (2003) report more aggressive trades when temporary volatility increases.

We measure volatility as the standard deviation of the midquote returns of the last 5 minutes. Following hypothesis 4 of Ranaldo (2003), the higher the volatility the less aggressive the orders. Table 3 shows that the influence of volatility in the aggregated buy orders support the theory, in the sense that the most aggressive orders (market orders) decreases and that the less aggressive orders (limit orders) increase. Moreover, volatility has a positive impact on cancelations on both sides of the book, possibly meaning that as volatility increases, traders cancel their positions to avoid being kicked off. However, with this analysis of the aggreagated counts, there are left many questions open, as the way the order aggressiveness react by changes in volatility. Tables 4 and 5 present this analysis. One can see that volatility affects negatively and significantly to the most aggressive market orders (category 1) and negatively but not significantly to categories 2 and 3. It has a higher positive impact on limit orders at or outside the best quotes (categories 5 and 6) and a negative effect on limit orders inside the best quotes (category 4). These results imply that the order aggressiveness decreases when volatility raises and that limit order traders ask for a higher compensation for the risk of being picked off submitting limit orders at and outside the best quotes and decreasing their limit order submissions inside the best quotes.

5.4 Dynamics of Order Submissions

5.4.1 Diagonal Effect

Biais et al. (1995) find that the probability of observing a certain event right after the same event just occurred is higher than its unconditional probability. They call this the diagonal effect. Using similar data but with a different econometric technique, Bisière & Kamionka (2000) do not find any evidence of this. In our setting we identify a similar type of effect, but it needs to be redefined somewhat. Let us rewrite equation (4.3) as:

$$\mu_t - \mu = A(N_{t-1} - \mu) + B(\mu_{t-1} - \mu)$$

where μ is the unconditional mean and remembering that as soon as μ_t follows a VARMA(1,1) process, $\mu = (I - A - B)^{-1}\omega$.
In this framework, the diagonal effect implies that the conditional mean of event i at time t is larger than its unconditional mean ($\mu_{i,t} > \mu_i$) if the number of events in the previous period was larger than the unconditional mean ($N_{i,t-1} > \mu_i$) and that the diagonal elements of A are positive and significant.

We can see from table 3 that we have a significant diagonal effect: in the vector autoregressive part of the equation (upper panel) we note that the coefficients on the diagonal (α's) are all positive and significant. For instance, assuming that in a given period, the number of buy market orders is larger than the unconditional mean (which can be interpreted as the mean in normal market conditions), we expect to have an increment in the conditional mean of this kind of market event above the unconditional mean, i.e. we expect to find that an increase in the number of a certain type of event in this one minute window has a strong positive impact on the conditional mean of the same type of event in the next interval. The same phenomenon is observed in the upper panels of tables 4 and 5.

5.4.2 Limit and Market Orders

In tables 3, 4 and 5 we see that buys and sells move together. We note that there are positive and significant coefficients for buy orders as a group as well as for sell orders. This suggests that all types of orders on one side of the market tend to arrive together. Traders provide and consume liquidity according to their private information but also by observing the state of the order book. During periods in which no information is arriving to the market, one observes a continuous consumption and provision of liquidity. As can be seen in table 3, buy (sell) market orders have a positive and significant effect on sell (buy) limit orders, which is a good sign, since it means that when liquidity gets consumed by market orders, there are new limit orders being submitted, i.e. the book is refilled. This guarantees that there is always liquidity in the book and that no liquidity crisis occur. The same comments apply when analyzing the results of the more disaggregated system in tables 4 and 5. In general market orders on both sides of the market positively affect all categories of limit orders in the opposite side of the book.

As mentioned by Bisière & Kamionka (2000), large buy orders tend to be followed by buy limit orders inside the best quotes, because the asset valuation is larger than the bid and the spread increases presenting a good opportunity for liquidity suppliers on the buy-side to overbid and compete for price-time priority and increasing their own asset valuation. It is also expected in normal times that liquidity providers from the sell side take advantage of this higher spread situation, sending orders that undercut the ask. Both scenarios assume that there is no new information in the market. Looking at table 4 we can appreciate that large buy market orders (category 1) have a positive and significant impact on the buy limit orders inside the best quotes (category 4). Moreover, observing table 5 we can appreciate that the effect that large buy market orders on sell limit orders is also positive and significant but only for categories 5 (at the ask) and 6 (above the ask). The influence of sell market orders to the sell limit orders (table 5) is the same as in the buy case. Finally, the most aggressive limit orders in both sides of the market have a positive impact on the market orders of their respective sides, showing that traders respond more aggressively to more aggressive limit orders in their own side.

5.4.3 Cancelations

There has not been much attention devoted to order cancelations in the theoretical literature. We nonetheless find that they do carry some information and have an impact on the order submission process. Furthermore, we hypothesize that their relative position respect to the best quotes matters in terms of the information they convey. We first have a look at the effect of cancelations in our most aggregated system in table 3. Next we run a system in which cancelations are classified according to their relative position in the limit order book, expressed as the number of steps away from the best quote. We consider three types of cancelations: type 1 occur in the first two steps. type 2, between the third and fifth step and, type 3 are all other cancelations.

In table 3 we have a first look at the behaviour of cancelations. Cancelations on both sides of the market have a negative and significant impact on market orders and a positive impact on limit orders (significant in the case of sell cancelations) on their own side. Thus cancelations reduce order aggressiveness. Cancelations also have a positive and significant effect on limit orders on the other side of the market. However, based on these results we can only draw a limited number of conclusions, as we do not know more precisely which cancelations and which limit orders are the ones that matter. Nonetheless it seems natural to hypothesize that buy cancelations mean bad news, while sell cancelations mean good news. When there is bad (good) news, traders do not want to submit buy (sell) market orders, but choose more passive buy (sell) limit orders.

In order to get more insight we analyse two new systems, one with market orders (table 6) and one in which we use the three types of cancelations described above, along with different categories of limit orders (table 7). In table 6 we see that the cancelations that have a significant negative effect on market orders are those at or within the next step of the limit order book (type 1). This confirms our information hypothesis in the sense that the most informative cancelations are the ones closest to the best quotes. Cancelations further away from the best quotes do not have a significant influence on market orders (even though on the bid they are significant at the 10% level).

In table 7, we observe an interesting pattern in the behaviour of cancelations with respect to the disaggregated limit orders on their own side. All the cancelations affect the most aggressive limit orders (categories 4 and 5) negatively and positively the least aggressive limit orders (category 6). This is consistent with the idea that cancelations carry information. When limit orders are canceled, traders get scared that the market is moving against them and that they might get picked off and as a consequence they stop undercutting (overbidding) and do not submit limit orders at the quotes. Instead, they prefer to submit limit orders away from the best quotes. An interesting question here is how far traders submit their orders respect to the best quote when cancellations in their own side increases. Table 8 presents a system in which the limit orders of category 6 have been divided into three categories, similarly to the one made on cancelations, i.e. type 1 occur in the first two steps. type 2, between the third and fifth step and, type 3 are all other limit orders of this category. Interestingly, cancelations of the first type have a positive and significant effect on the submissions of limit orders inside the best 5 quotes and a negative and significative impact on limit orders far away those quotes. This means that even though traders become less aggressive when cancelations of type 1 are arriving, they do not quit the market and instead they submit orders inside the best 5 quotes keeping with this the liquidity in the market.

As a conclusion, we can say that cancelations contain relevant information for traders. Moreover, this information depends on the relative position of the cancelation. The most aggressive type of cancelations (type 1) are the ones that exert a negative influence on market orders, while the other types do not have a significant effect. Cancelations of all types exert a negative influence on the most aggressive own side limit orders and a positive one on the least aggressive. However, traders do not submit orders far away in the book but within the best 5 quotes. All these results confirm that cancelations are worth analysing, and that their relative position in the limit order book matters a great deal.

6 Conclusions

In this paper we presented a detailed analysis of the trading process on the Frankfurt electronic stock exchange. We are interested in modeling the process of order submissions in minute detail. In order to do this, given the limitations of ordered probits in capturing the full dynamics of the order submissions and the difficulties associated with extending duration models to large multivariate systems, we choose to work with the number (counts) of all the different types of orders that are being submitted to the market in one minute intervals. One contribution to the existing empirical literature is that in our analysis of the state of the book we propose to use Principal Component Analysis (PCA) on the percentage average price respect to the best quote. Based on the orthogonality property of the linear combinations estimated by PCA techniques, we can study in a separate way the liquidity and information aspects present in the state of the book. We show that the first principal component has nearly constant loadings for all volumes and we therefore interpret it as the mean effect. If the weight of this factor increases, this means that the percentage average price is increasing on average for all trades of any volume v. The second factor is typically negatively related to the mark up at small volumes and the factor loadings increase monotonically as the volume increases. This factor is therefore related to the slope of the price schedule. Therefore, we link liquidity aspects to the first factor and information ideas to the second one. The results support this idea and they seem to be robust in that they are also present when analyzing the other two assets of our sample [9]. Another contribution of this paper is the analysis of the cancelations and their relation with order aggressiveness. We show that cancelations contain information. However, this information is related to their relative position in the order book. Cancellations among the nearest two steps are the ones that exert a significant and negative effect in the most aggressive orders (market orders in both sides of the market). Moreover, When limit orders are canceled, traders get scared that the market is moving against them and that they might get picked off and as a consequence they stop undercutting (overbidding) and do not submit limit orders at the quotes. Instead they move away from the best quotes.

This paper presents and application of the Multivariate autoregressive double Poisson model proposed by Heinen & Rengifo (2003). The results presented in this paper are only small examples of many other interesting questions that could be addressed and that because of space limitations we restrict our analysis to the discussed topics.

[9]The estimated results based on the data of DTE and SAP are available upon request.

References

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York.

Bae, K.-H., Jang, H. & Park, K. (2003), 'Trader's choice between limit and market orders: Evidence from nyse stocks', *Journal of Financial Markets* **6**, 517–538.

Bauwens, L. & Giot, P. (2001), *Econometric Modelling of Stock Market Intraday Activity*, Kluwer Academic Publishers.

Beltran-Lopez, H., Giot, P. & Grammig, J. (2004), How to summarize the information in the order book?

Work in progress. Center for Operations Research and Econometrics-CORE.

Biais, B., Hillion, P. & Spatt, C. (1995), 'An empirical analysis of the limit order book and the order flow in the paris bourse', *Journal of Finance* **52**, 655–681.

Biais, B., Hillion, P. & Spatt, C. (1999), 'Price discovery and learning during the preopening period in the paris bourse', *Journal of Political Economy* **107**, 1218–1248.

Bisière, C. & Kamionka, T. (2000), 'Timing of orders, orders aggressiveneess and the order book at the paris bourse', *Annales d'Economie et de Statistique* **56**, 43–72.

Danielson, J. & Payne, R. (2001), Measuring and explaining liquidity on an electronic limit order book: evidence from reuters d200-2.

LOndon School of Economics, Working paper.

Davis, R., Rydberg, T. & Shephard, N. (2001), The cbin model for counts: Testing for common features in the speed of treading, quote changes, limit and market order arrivals.

mimeo.

Denuit, M. & Lambert, P. (2002), Constraints on concordance measures in bivariate discrete data.

Discussion Paper 02-12, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Efron, B. (1986), 'Double exponential families and their use in generalized linear regression', *Journal of the American Statistical Association* **81**(395), 709–721.

Engle, R. & Lunde, A. (2003), 'Trades and quotes: a bivariate point process', *Journal of Financial Econometrics* **1**(2), 159–188.

Foucault, T. (1999), 'Order flow composition and trading costs in a dynamic limit order market', *Journal of Financial Markets* **2**, 99–134.

Glosten, L. R. (1994), 'Is the electronic open limit order book inevitable?', *The Journal of Finance* **49**(4), 1127–1161.

Griffiths, M., Smith, B., Turnbull, A. & White, R. (2000), 'The costs and determinants of aggressiveness', *Journal of Financial Economics* pp. 65–88.

Hall, A., Hautsch, N. & Mcculloch, J. (2003), Estimating the intensity of buy and sell arrivals in a limit order book market.

Mimeo, University of Technology, Sidney and University of Konstanz.

Handa, P. & Schwartz, R. A. (1996), 'Limit order trading', *The Journal of Finance* **51**(5), 1835–1861.

Handa, P., Schwartz, R. & Tiwari, A. (2003), 'Quote setting and price formation in an order driven market', *Journal of Financial Markets* **6**, 461–489.

Hasbrouck, J. (1999), Trading fast and slow: Security markets events in real time. mimeo.

Heinen, A. & Rengifo, E. (2003), Multivariate modeling of time series count data: An autoregressive conditional poisson model. CORE Discussion Paper 2003/25.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

Lehmann, B. (2003), Notes for a contingent claims theory of limit order markets. Second draft, Graduate School of International Relations and Pacific Studies, University of California at San Diego.

Nelsen, R. B. (1999), *An introduction to Copulas*, Springer, New York.

Parlour, C. (1998), 'Price dynamics in limit order markets', *Review of Financial Studies* **11**, 789–816.

Pascual, R. & Veredas, D. (2004), What pieces of limit order book information do are informative? an empirical analysis of a pure order-driven market. Mimeo. Dept. of Business. Universidad de las Islas Baleares.

Patton, A. (2002), Skewness, asymmetric dependence, and portfolios. mimeo.

Ranaldo, A. (2003), 'Order aggressiveness in limit order book markets', *Journal of Financial Markets* **7**, 53–74.

Russell, J. (1999), The econometric modeling of multivariate irregularly-spaced high-frequency data. Working Paper Tilburg University and CentER.

Sandas, P. (2001), 'Adverse selection and competitive market making: Empirical evidence from a limit order market', *Review of Financial Studies* **14**, 705–734.

Sklar, A. (1959), 'Fonctions de répartitions à n dimensions et leurs marges', *Public Institute of Statistics of the Univeristy of Paris* **8**, 229–231.

Table 3: **Estimation results of MDACP models with market and limit orders, as well as cancelation by side of the market and volatility as measured by the standard deviation of the last 5 minutes and spreads during the last period, depth at the best bid and at the best ask**

The table presents the Maximum Likelihood estimates of the Multivariate Autoregressive Conditional Double Poisson (MDACP) model, with the following mean:
$\mu^*_{t,i} = \mu_{t,i} \exp\left(X_{t-1}\eta_i + \sum_{p=1,2}(\psi_{c,p}\cos\frac{2\pi p\, Re[t,N]}{N} + \psi_{s,p}\sin\frac{2\pi p\, Re[t,N]}{N})\right)$, and
$\mu_{t,i} = \omega_i + \sum_{j=1}^{6}\alpha_{i,j}N_{t-1,j} + \beta\mu_{t-1,i}$, for $t = 1, \ldots, 10731$,
where $Re[t,N]$ is the remainder of the integer division of t by N, the number of periods in a trading session. X_{t-1} is the vector of explanatory variables. The seasonality parameters are not shown, but we show a Wald test $W(\psi's = 0)$ for joint significance of all the seasonality variables. $Var(\varepsilon_t)$ is the variance of the Pearson residual. Parameters that are significant at the 5% level appear in bold font for better readability.

Parameters	BMO	BLO	SMO	SLO	BCANC	SCANC
ω	**0.032**	**0.270**	**0.052**	**0.348**	**0.115**	**0.159**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BMO	**0.086**	**0.051**	0.002	**0.168**	-0.019	-0.017
	(0.00)	(0.00)	(0.64)	(0.00)	(0.08)	(0.12)
BLO	**0.023**	**0.178**	0.004	**0.041**	**0.102**	0.012
	(0.00)	(0.00)	(0.28)	(0.00)	(0.00)	(0.09)
SMO	0.011	**0.266**	**0.087**	**0.048**	-0.001	0.015
	(0.11)	(0.00)	(0.00)	(0.03)	(0.95)	(0.31)
SLO	-0.006	-0.008	**0.034**	**0.180**	**-0.015**	**0.121**
	(0.08)	(0.42)	(0.00)	(0.00)	(0.05)	(0.00)
BCANC	**-0.016**	0.023	-0.001	**0.027**	**0.104**	0.009
	(0.00)	(0.07)	(0.88)	(0.04)	(0.00)	(0.36)
SCANC	0.004	**0.064**	**-0.011**	**0.054**	**0.044**	**0.086**
	(0.41)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
β	**0.848**	**0.647**	**0.709**	**0.541**	**0.660**	**0.600**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Spread	**-1.019**	-0.134	**-1.736**	**-0.341**	-0.300	-0.391
	(0.00)	(0.43)	(0.00)	(0.05)	(0.17)	(0.06)
Bidvol	6.32E-6	3.96E-6	**1.015E-5**	**5.06E-6**	6.48E-7	6.57E-6
	(0.13)	(0.13)	(0.04)	(0.05)	(0.87)	(0.06)
Askvol	**7.26E-6**	2.95E-6	**1.65E-5**	**6.01E-6**	4.10E-6	4.12E-6
	(0.05)	(0.20)	(0.00)	(0.01)	(0.13)	(0.16)
Volat	**-0.643**	0.160	-0.492	**0.380**	**1.475**	**0.573**
	(0.01)	(0.48)	(0.15)	(0.02)	(0.00)	(0.01)
Disp	**0.713**	**0.525**	**0.752**	**0.511**	**0.614**	**0.604**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$W(\psi's = 0)$	**25.20**	**44.05**	**16.75**	**32.94**	**52.09**	**55.54**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Var(ϵ_t)	1.02	1.00	1.16	1.01	1.00	1.03
Log likelihood	-14661.9	-23535.2	-12049.8	-22717.2	-18410.6	-18254.2

Table 4: **Estimation results of MDACP models**

Parameters	**BMO-C1**	**BMO-C2**	**BMO-C3**	**BLO-C4**	**BLO-C5**	**BLO-C6**
ω	**0.008**	**0.050**	**0.007**	**0.150**	**0.045**	**0.060**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BMO-C1	**0.056**	0.009	**0.011**	**0.062**	0.019	-0.004
	(0.00)	(0.24)	(0.00)	(0.00)	(0.16)	(0.79)
BMO-C2	0.008	**0.047**	-3.13E-4	-0.011	0.001	-0.003
	(0.07)	(0.00)	(0.94)	(0.63)	(0.97)	(0.84)
BMO-C3	**0.022**	0.011	**0.035**	**0.066**	0.011	0.001
	(0.00)	(0.06)	(0.00)	(0.00)	(0.30)	(0.89)
BLO-C4	**0.009**	**0.038**	0.003	**0.110**	**0.018**	0.011
	(0.00)	(0.00)	(0.09)	(0.00)	(0.01)	(0.12)
BLO-C5	**0.011**	**0.011**	**0.006**	**0.052**	**0.091**	**0.031**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BLO-C6	-0.001	0.001	-0.001	**0.024**	**0.058**	**0.138**
	(0.38)	(0.79)	(0.68)	(0.00)	(0.00)	(0.00)
SMO-C1	0.001	-0.014	**-0.011**	**0.092**	**0.158**	**0.110**
	(0.84)	(0.12)	(0.02)	(0.00)	(0.00)	(0.00)
SMO-C2	**0.017**	0.002	-0.004	**0.104**	**0.070**	**0.051**
	(0.00)	(0.84)	(0.36)	(0.00)	(0.00)	(0.00)
SMO-C3	0.001	0.015	-0.003	0.020	**0.034**	0.010
	(0.86)	(0.09)	(0.46)	(0.35)	(0.04)	(0.56)
SLO-C4	-0.003	0.004	0.004	**0.027**	-0.001	0.011
	(0.11)	(0.35)	(0.07)	(0.01)	(0.86)	(0.16)
SLO-C5	-0.001	-0.003	0.001	-0.015	**0.017**	0.012
	(0.51)	(0.37)	(0.63)	(0.07)	(0.00)	(0.07)
SLO-C6	-0.003	0.007	-0.002	-0.003	-0.001	**0.015**
	(0.09)	(0.06)	(0.28)	(0.72)	(0.92)	(0.03)
β	**0.807**	**0.579**	**0.921**	**0.656**	**0.721**	**0.689**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Bfact1	2.24E-4	**0.010**	0.004	**0.005**	**0.010**	**-0.009**
	(0.94)	(0.00)	(0.15)	(0.01)	(0.00)	(0.00)
Diffslope	0.009	**-0.023**	0.002	**-0.016**	-0.008	**0.014**
	(0.19)	(0.00)	(0.78)	(0.00)	(0.20)	(0.01)
Bfact3	**0.039**	0.011	0.001	0.018	0.013	-0.014
	(0.01)	(0.53)	(0.93)	(0.10)	(0.34)	(0.22)
Sfact1	**0.018**	0.004	**0.010**	**0.005**	**0.005**	-0.003
	(0.00)	(0.11)	(0.00)	(0.03)	(0.02)	(0.16)
Sfact3	**-0.008**	-0.004	-0.006	-0.012	0.003	0.005
	(0.61)	(0.80)	(0.72)	(0.26)	(0.83)	(0.64)
Spread	0.150	**-2.549**	**-0.809**	0.086	-0.500	-0.404
	(0.62)	(0.00)	(0.02)	(0.71)	(0.06)	(0.11)
Bidvol	9.80E-6	**2.20E-5**	**-1.70E-5**	**1.46E-5**	**-1.82E-5**	2.70E-6
	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)	(0.57)
Askvol	-3.33E-6	**-3.21E-5**	**2.87E-5**	2.65E-6	2.66E-6	-2.38E-7
	(0.48)	(0.00)	(0.00)	(0.36)	(0.49)	(0.95)
Volat	**-0.622**	-0.591	-0.263	**-0.153**	**0.174**	0.257
	(0.03)	(0.65)	(0.42)	(0.03)	(0.04)	(0.07)
Disp	**1.177**	**1.146**	**1.004**	**0.652**	**0.790**	**0.762**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$W(\psi's = 0)$	**45.07**	**18.18**	**29.89**	**26.13**	**25.09**	**42.39**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Var(ϵ_t)	1.44	1.38	1.11	0.98	1.00	1.01
Log likelihood	-7318.7	-7425.4	-9486.6	-17499.5	-14001.5	-13891.0

Table 5: **Estimation results of MDACP models**

Parameters	SMO-C1	SMO-C2	SMO-C3	SLO-C4	SLO-C5	SLO-C6
ω	**0.017**	**0.027**	**0.024**	**0.151**	**0.057**	**0.121**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BMO-C1	**-0.014**	-0.008	-0.002	0.010	**0.069**	**0.046**
	(0.00)	(0.18)	(0.62)	(0.61)	(0.00)	(0.00)
BMO-C2	**-0.010**	**-0.016**	**-0.011**	**0.051**	**0.082**	**0.067**
	(0.02)	(0.02)	(0.00)	(0.02)	(0.00)	(0.00)
BMO-C3	**0.009**	0.008	0.004	0.024	**0.020**	0.017
	(0.00)	(0.08)	(0.16)	(0.09)	(0.05)	(0.16)
BLO-C4	0.003	**0.011**	0.001	**0.049**	**0.016**	**0.026**
	(0.21)	(0.00)	(0.62)	(0.00)	(0.02)	(0.00)
BLO-C5	**0.005**	**0.009**	**-0.004**	0.009	0.011	**0.022**
	(0.01)	(0.01)	(0.01)	(0.29)	(0.08)	(0.01)
BLO-C6	-0.002	-0.004	0.003	-0.001	0.002	**0.022**
	(0.19)	(0.07)	(0.06)	(0.90)	(0.71)	(0.00)
SMO-C1	**0.071**	0.004	**0.016**	**0.094**	-0.005	**0.060**
	(0.00)	(0.61)	(0.00)	(0.00)	(0.77)	(0.00)
SMO-C2	**0.022**	**0.035**	0.008	0.014	-0.003	-0.007
	(0.00)	(0.00)	(0.08)	(0.57)	(0.87)	(0.72)
SMO-C3	0.009	-0.002	**0.032**	-0.005	0.022	0.018
	(0.07)	(0.84)	(0.00)	(0.83)	(0.18)	(0.33)
SLO-C4	**0.013**	**0.036**	**0.012**	**0.132**	0.015	**0.025**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.01)
SLO-C5	**0.009**	**0.007**	**0.007**	**0.053**	**0.103**	**0.052**
	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
SLO-C6	0.002	**0.006**	**0.006**	**0.040**	**0.037**	**0.123**
	(0.29)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)
β	**0.682**	**0.632**	**0.677**	**0.559**	**0.676**	**0.512**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Bfact1	**0.008**	-0.003	**0.035**	4.08E-4	0.001	-0.002
	(0.01)	(0.37)	(0.00)	(0.85)	(0.68)	(0.45)
Diffslope	0.006	**0.037**	2.05E-4	**0.022**	0.006	**-0.019**
	(0.46)	(0.00)	(0.98)	(0.00)	(0.34)	(0.00)
Bfact3	0.004	-0.023	**0.072**	-0.013	0.010	**-0.025**
	(0.82)	(0.19)	(0.00)	(0.29)	(0.42)	(0.03)
Sfact1	**0.009**	**0.014**	-1.93E-4	**0.007**	**0.011**	**-0.014**
	(0.00)	(0.00)	(0.95)	(0.00)	(0.00)	(0.00)
Sfact3	-0.026	0.001	**0.049**	0.005	-0.016	0.007
	(0.09)	(0.94)	(0.00)	(0.64)	(0.21)	(0.52)
Spread	**-1.349**	**-2.467**	**-1.902**	-0.456	-0.274	-0.378
	(0.00)	(0.00)	(0.00)	(0.07)	(0.28)	(0.13)
Bidvol	**-3.37E-5**	**-1.865E-5**	**3.86E-5**	**5.82E-6**	5.24E-6	1.68E-6
	(0.00)	(0.00)	(0.00)	(0.13)	(0.17)	(0.68)
Askvol	**1.39E-5**	**1.07E-5**	**1.31E-5**	**1.67E-5**	**-9.53E-6**	-1.96E-6
	(0.00)	(0.03)	(0.01)	(0.00)	(0.00)	(0.64)
Volat	**-0.493**	**-0.334**	-0.142	**-0.120**	**0.431**	0.487
	(0.01)	(0.05)	(0.21)	(0.01)	(0.02)	(0.10)
Disp	**1.388**	**1.195**	**1.402**	**0.652**	**0.773**	**0.775**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$W(\psi' s = 0)$	**43.14**	**37.18**	**12.87**	**14.10**	2.12	**90.53**
	(0.00)	(0.00)	(0.01)	(0.01)	(0.71)	(0.00)
Var(ϵ_t)	1.69	1.48	1.61	1.01	1.03	1.02
Log likelihood	-5516.6	-6952.0	-5497.3	-16220.2	-13507.1	-13613.9

Table 6: **Estimation results of MDACP models: cancelations**

Parameters	**BMO**	**SMO**
ω	**0.034**	**0.057**
	(0.00)	(0.00)
α_1	**0.086**	0.002
	(0.00)	(0.63)
α_2	**0.020**	0.003
	(0.00)	(0.31)
α_3	0.012	**0.089**
	(0.07)	(0.00)
α_4	-0.006	**0.033**
	(0.09)	(0.00)
BCANC1-2	**-0.020**	-0.007
	(0.01)	(0.39)
BCANC3-5	-0.013	-0.007
	(0.08)	(0.37)
BCANC6-ss	-0.009	0.010
	(0.10)	(0.09)
SCANC1-2	**-0.016**	**-0.031**
	(0.03)	(0.00)
SCANC3-5	0.014	-0.006
	(0.06)	(0.46)
SCANC6-ss	**0.012**	0.004
	(0.04)	(0.53)
β	**0.848**	**0.698**
	(0.00)	(0.00)
$W(\psi's = 0)$	**24.01**	**15.42**
	(0.00)	(0.00)
Var(ϵ_t)	1.02	1.16
Log likelihood	-14656.7	-12042.91

Table 7: **Estimation results of MDACP models: cancelations**

Parameters	**BLO-C4**	**BLO-C5**	**BLO-C6**	**SLO-C4**	**SLO-C5**	**SLO-C6**
ω	**0.133**	**0.041**	**0.057**	**0.137**	**0.048**	**0.109**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BLO-C4	**0.143**	**0.035**	-0.007	**0.065**	**0.045**	**0.044**
	(0.00)	(0.00)	(0.14)	(0.00)	(0.00)	(0.00)
BLO-C5	**0.069**	**0.109**	0.005	0.003	-0.002	**0.015**
	(0.00)	(0.00)	(0.47)	(0.74)	(0.76)	(0.01)
BLO-C6	**0.040**	**0.077**	**0.091**	-0.009	**-0.014**	0.012
	(0.00)	(0.00)	(0.00)	(0.39)	(0.04)	(0.19)
SLO-C4	**0.058**	**0.041**	**0.043**	**0.155**	**0.024**	0.004
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.52)
SLO-C5	**-0.023**	-4.91E-4	-0.002	**0.064**	**0.115**	**0.019**
	(0.01)	(0.94)	(0.71)	(0.00)	(0.00)	(0.01)
SLO-C6	**-0.036**	-0.010	0.010	**0.046**	**0.052**	**0.087**
	(0.01)	(0.21)	(0.22)	(0.00)	(0.00)	(0.00)
BCANC1-2	**-0.037**	**-0.036**	**0.053**	0.012	0.005	0.015
	(0.00)	(0.01)	(0.00)	(0.39)	(0.57)	(0.18)
BCANC3-5	**-0.021**	**-0.037**	**0.093**	0.009	**0.027**	**0.024**
	(0.00)	(0.00)	(0.00)	(0.50)	(0.01)	(0.05)
BCANC6-ss	**-0.027**	0.006	**0.025**	0.011	**0.017**	-0.007
	(0.01)	(0.49)	(0.00)	(0.34)	(0.04)	(0.45)
SCANC1-2	-1.49E-5	-0.006	-0.005	-0.014	**-0.039**	**0.089**
	(0.99)	(0.51)	(0.62)	(0.32)	(0.00)	(0.00)
SCANC3-5	**0.061**	0.011	0.016	-0.028	**-0.044**	**0.103**
	(0.00)	(0.30)	(0.13)	(0.06)	(0.00)	(0.00)
SCANC6-ss	**0.029**	**0.049**	**0.017**	0.007	**0.016**	0.013
	(0.01)	(0.00)	(0.03)	(0.52)	(0.02)	(0.17)
β	**0.680**	**0.713**	**0.685**	**0.581**	**0.707**	**0.533**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$W(\psi's = 0)$	**32.88**	**24.58**	**25.51**	**14.92**	0.81	**56.89**
	(0.00)	(0.00)	(0.01)	(0.01)	(0.94)	(0.00)
Var(ϵ_t)	0.98	1.00	1.02	1.01	1.03	1.03
Log likelihood	-17525.1	-14027.4	-13866.3	-16242.7	-13519.5	-13601.6

Table 8: **Estimation results of MDACP models: cancelations and disaggregated category 6**

Parameters	BC6-1-2	BC6-3-5	BC6-6-ss	SC6-1-2	SC6-3-5	SC6-6-ss
ω	**0.019**	**0.014**	**0.025**	**0.038**	**0.021**	**0.032**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
BC6-1-2	**0.058**	0.005	**0.011**	-0.004	**0.012**	**0.013**
	(0.00)	(0.34)	(0.00)	(0.48)	(0.01)	(0.00)
BC6-3-5	**0.022**	**0.074**	**0.025**	0.011	0.001	**-0.012**
	(0.00)	(0.00)	(0.00)	(0.11)	(0.92)	(0.00)
BC6-6-ss	0.001	**0.015**	**0.084**	**0.012**	-0.001	-0.002
	(0.82)	(0.00)	(0.00)	(0.05)	(0.89)	(0.48)
SC6-1-2	0.006	**0.011**	0.005	**0.079**	**0.011**	**-0.008**
	(0.27)	(0.01)	(0.11)	(0.00)	(0.03)	(0.01)
SC6-3-5	0.006	0.001	0.006	**0.021**	**0.046**	**0.013**
	(0.38)	(0.90)	(0.14)	(0.01)	(0.00)	(0.00)
SC6-6-ss	0.003	**0.011**	**-0.013**	**-0.014**	-0.001	**0.087**
	(0.63)	(0.04)	(0.00)	(0.02)	(0.79)	(0.00)
BCANC1-2	**0.046**	**0.023**	**-0.009**	**0.025**	**0.012**	**0.008**
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)
BCANC3-5	**0.029**	**0.060**	**-0.009**	**0.012**	**0.014**	**0.009**
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)
BCANC6-ss	**0.007**	-0.005	**0.027**	0.002	-0.001	**0.0097**
	(0.04)	(0.13)	(0.00)	(0.65)	(0.70)	(0.00)
SCANC1-2	**0.011**	0.005	-0.004	**0.084**	**0.016**	**-0.007**
	(0.01)	(0.21)	(0.12)	(0.00)	(0.00)	(0.02)
SCANC3-5	**0.014**	0.006	**0.008**	**0.021**	**0.063**	**0.007**
	(0.00)	(0.15)	(0.02)	(0.00)	(0.00)	(0.03)
SCANC6-ss	**0.020**	3.36E-4	**0.012**	0.003	-0.003	**0.023**
	(0.00)	(0.89)	(0.00)	(0.37)	(0.23)	(0.00)
β	**0.684**	**0.716**	**0.768**	**0.558**	**0.692**	**0.663**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$W(\psi's = 0)$	**25.65**	**11.83**	**62.81**	**32.29**	**16.67**	**92.29**
	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
Var(ϵ_t)	1.30	1.32	1.36	1.36	1.30	1.57
Log likelihood	-8615.7	-7683	-6701.9	-8560.2	-7364.3	-6238.8