T E C H N I C A L
R E P O R T

0478

WHAT PIECES OF LIMIT ORDER BOOK INFORMATION ARE INFORMATIVE?

PASCUAL, R. and D. VEREDAS

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

CORE DISCUSSION PAPER
2004/33

What pieces of limit order book information are informative?*

Roberto Pascual[1] and David Veredas[2]

May 2004

Abstract

This paper studies the importance of different pieces of limit order book information in characterizing order aggressiveness and the timing of trades, order submissions and cancellations. Using limit order book information on a representative sample of Spanish stocks, we evidence that most of the explanatory power of the book concentrates on the best quotes. However, the book beyond the best quotes also matters in explaining the aggressiveness of traders. In particular, liquidity providers (limit-order traders) benefit more from an increased degree of pre-trade transparency than liquidity consumers (market-order traders). Finally, no piece of book information matters in explaining the timing of orders.

Keywords: Open limit order book, order aggressiveness, durations, pre-trade transparency, order driven markets.

JEL classification: C41, C35, G14

[1]Department of Business, Universidad de las Islas Baleares
[2]Department of Econometrics & OR, Tilburg University and CORE.

* We wish to thank David Abad, Sonia Falconieri, Pierre Giot, Angelo Ranaldo, Andreu Sansó, José Vidal, and Klaus F. Zimmermann for their useful suggestions. We acknowledge the insightful comments of the participants at seminars at Universidad Carlos III de Madrid (Spain), Catholique University of Louvain (Belgium), and Sociedad de Bolsas (Spain), and participants at the Namur's Workshop on Automated Auction Markets (Belgium). David Veredas acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, [MICFINMA]. Roberto Pascual also acknowledges the financial sponsorship of the Fulbright Grant and the Spanish Ministry of Education, Culture and Sports, and of the Spanish DGICYT project BEC2001-2552-C03-03. The first version of this paper was finished while Roberto Pascual was a Visiting Scholar at the New York University Salomon Center. Corresponding author: Roberto Pascual, Departamento de Economía de la Empresa, Universidad de las Islas Baleares, Ctra. de Valldemossa Km. 7.5, 07122 Palma de Mallorca, Baleares, Spain. Phone: +34-971-17-13-29, Fax: +34-971-17-23-89, e-mail: rpascual@uib.es.
This text presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1. Introduction

Some recent trends in market design are towards the supply of limit order book information in real time, the introduction of competing order driven venues in traditional dealer markets, and the creation of new pure electronic limit order book systems (e.g., Domowitz and Wang, 1994). Together with the increasing availability of limit order book data, these trends generate a renovated interest in the microstructure of order-driven markets based on open limit order books (LOBs). A characteristic feature of these trading platforms is their high degree of pre-trade transparency (i.e. the ability of market participants to observe the content of the LOB).

In this paper, we analyze the relevance of pre-trade transparency in determining the trading strategies of the participants in one of these platforms: the Stock Exchange Interconnection System (henceforth SIBE) of the Spanish Stock Exchange (SSE). In this particular electronic platform, real-time information about the five best bids and offers on the book is widely disseminated through a computer-assisted information system. We use six months of limit order book data of the SIBE to evaluate the information content of the book and, particularly, whether it influences the behavior of incoming traders.

This paper is not, of course, the first to address such issue. Biais, Hillion and Spatt (1995), Griffiths *et al* (2000), Coppejans and Domowitz (2002) and Ranaldo (2004), among others, provide evidence in the affirmative. The value added of this paper is that we evaluate what particular pieces of all the book information do really matter in characterizing the upcoming order flow. We distinguish between two large sets of information: the best bid and offer quotes and the second to fifth bid and offer quotes. This distinction is not arbitrary since even some of the most pre-trade opaque markets provide information about the best quotes.

Our main goals are, first, to provide a measurement of the information value of the limit order book beyond that of the best bid and offer quotes and, second, to infer what traders do really benefit from the additional information in limit order book data. This questions are of great interest not only to economists, in terms of modeling pure order driven markets and characterizing the traders behavior, but also to policy makers since our findings shed some light on some basic questions of the pre-trade transparency literature: how publicly revealing information about the book does affect trading strategies and who benefits from a pre-trade transparent venue.

We characterize the upcoming order flow using a categorization of order aggressiveness similar to that proposed by Biais et *al* (1995), where the degree of impatience of the trader is approximated by the type of order he/she submits to the system. We also typify the degree of aggressiveness of market participants by the time between consecutive market orders (trades), limit order submissions, and cancellations. Our methodology merges the Griffiths *et al* (2000) and Ranaldo (2004) approach of using ordered probit models to study order aggressiveness and the Coppejans and Domowitz (2002) approach of using the family of autoregressive conditional duration (ACD) models (Engle and Russell, 1998) to analyze the time between consecutive events of some kind.

We report that the whole limit order book matters in explaining the aggressiveness of traders, although the best quotes account for the most important part of the useful information. The book beyond the best quotes is particularly relevant in explaining the aggressiveness of an upcoming liquidity provider (limit-order trader) but it does not affect the strategic decision of an upcoming liquidity consumer (market-order trader). Finally, neither the best quotes nor the book beyond the best quotes provides noteworthy information in determining the timing of orders. In general, our findings suggest that pre-trade transparency does not benefit the same to all market participants, and that it provides additional information only in the very short-run.

The paper proceeds as follows. In the next section, we review the pertinent literature. In section 3, we describe the data and the market. In section 4, we analyze what pieces of limit order book information matter in determining order aggressiveness. In section 5, we analyze what pieces of book information are important in explaining the timing or trades, order submissions and cancellations. Finally, we conclude in section 6.

2. Literature review

Three topics in market microstructure literature are essential to set the limits of this paper: the issue of whether the limit order book contains information about future price movements and trading decisions, the study of the determinants of order aggressiveness, and the discussion about the benefits and inconveniences of pre-trade transparent trading systems. In this section, we briefly review these three lines of research so as to make clear our contribution to the literature.

The informativeness of the LOB has been the subject of recent theoretical and empirical research. Handa and Schwartz (1996), Parlour (1998) and Foucault (1999), among others, argue that the state of the LOB influences the forthcoming order flow. These models suggest that an unbalanced limit order book reflects the market sentiment. Similarly, Huang and Stoll (1994) assert that an unbalanced LOB is a sign of asymmetric information. Contrary to the usual claim, Seppi (1997) and Kaniel and Liu (2001) theoretically show that limit orders stored on the LOB may be information motivated since informed traders may prefer to submit limit orders rather than market orders under certain conditions. Empirically, Harris and Panchapagesan (2003) and Madhavan and Panchapagesan (2000) conclude that the privileged access of the NYSE specialist to the book turns out to be an informative advantage about short-run market movements. In a similar vein, Corwin and Lipson (2000) evidence that the repositioning of limit orders on the book during trading halts is informative about market movements when trading resumes. Irvine, Benston and Kandel (2000) show that a liquidity measure computed using limit order book data is more informative about subsequent order flow than other traditional liquidity measures based on the best bid and offer quotes. Finally, Coppejans and Domowitz (2002) evidence that the information gleaned from the book substantially affects the timing of trades, order submissions and cancellations. In opposition to the previous papers, Franke and Hess (2000) observe that the book is informative only during periods of low information intensity.

All previous empirical and theoretical works look at the LOB as a whole. Contrarily, in this paper we understand the LOB as a set of individual pieces of information.

Consequently, we aim to evaluate what part of the whole pack provides brand new information and what part is just information redundant.

Biais *et al* (1995) proposed a categorization of the aggressiveness of traders based on the particularities of the orders submitted. From the most to the less aggressive category:

- C7: Buy (sell) orders that demand more volume than is available at the best prevailing ask (bid) and are allowed to walk up (down) the book.
- C6: Buy (sell) orders that demand more volume than is available at the best ask (bid) but are not allowed to walk up (down) the book.
- C5: Buy (sell) orders that demand less volume than is available at the best ask (bid).
- C4: Orders with prices lying between the best bid and offer.
- C3: Buy (sell) orders that have prices equal to the best bid (ask).
- C2: Buy (sell) orders that have prices above (below) the best bid (ask).
- C1: Cancellations.

Categories C5 to C7 imply total or partial immediate execution of the order. Categories C1 to C4 imply non-immediate execution.

The theoretical models of Parlour (1998), Foucault (1999), and Handa, Schwartz and Tiwari (2004) predict that the LOB conditions the aggressiveness of traders. These models suggest that variables like the imbalance between potential buyers and sellers and the volatility of the asset determine the non-execution risk of a limit order and, hence, the mix between market (C2-C4) and limit (C5-C7) orders. In general, the lower the non-execution risk the less aggressive the order flow. Empirically, Ahn, Bae and Chan (2001) and Daníelson and Payne (2001) observe that investors become less aggressive buyers (sellers) when liquidity driven volatility rises from the offer (bid) side of the book. Griffiths *et al* (2000) and Ranaldo (2004) report that traders become more aggressive when their own (opposite) side book is thicker (thinner), the spread narrower, and temporary volatility increases.

Once again, neither of the previous papers distinguishes the effect of different pieces of book information on the aggressiveness of traders. In addition, these papers model the incoming trader's choice between providing or consuming liquidity (limit vs. market order) and the ex-post liquidity provider/consumer's level of aggressiveness, as a (one-step) simultaneous decision. This paper, however, seeks to discern whether active traders (liquidity consumers) rely more on LOB data when decide about the aggressiveness of their market orders than passive traders (liquidity providers) when decide about the aggressiveness of their limit orders.

Madhavan (2000, pg. 234) defines pre-trade transparency as "the wide dissemination of current bid and ask quotations, depths, and possibly also information about limit orders away from the best prices, as well as other pertinent trade related information such as the existence of large order imbalances". Electronic limit order markets are usually characterized as highly pre-trade transparent since they normally offer real-time information about the LOB.

Up to now, the empirical research on the supposed benefits of a pre-trade transparent venue has been inconclusive. Bloomfield and O'Hara (1999) and Flood *et al* (1999) develop independent laboratory experiments to evaluate the influence of quote information disclosure in multi-dealer settings reporting mixed findings. Madhavan, Porter and Weaver (2000), for the Toronto Stock Exchange, and Boehmer, Saar and Yu (2003), for the NYSE, investigate the impact of an exogenous increase in the level of public information about the LOB. These papers report varied conclusions regarding the effect of greater transparency on displayed liquidity. Nonetheless, Boehmer *et al* detect that greater transparency improves informational efficiency. Finally, Harris (1996) argues that pre-trade transparency increases the exposure-risk of limit order traders.

Our empirical study does not provide additional insights about the beneficial or pervasive effects of pre-trade transparency, but it sheds some light on who benefits from an open LOB in an electronic order-driven market and provides a measurement of how much valuable is the book information beyond the best bid and offer quotes.

In a recent independent unpublished paper, Cao, Hansh and Wand (2003) also evaluate the informativeness of the Australian Stock Exchange LOB beyond its first step. Although both papers inevitably overlap at some point, the focus is different and there are remarkable methodological differences. Cao *et al* focus on the value of the book information in determining the true value of the stock. Using depth-weighted estimators of the efficient price from both the best quotes and the complete book, these authors estimate an error correction model for 5-minute snapshots of the book. They find that the best quotes lead the whole book and provide a better estimator of the true value. The averaged Hasbrouck (1995) information share bounds attribute a 70% of the price discovery to the best quotes and a 30% to the rest of the book. Nonetheless, the lower information share bound for the book away from the best quotes is always very close to zero. These authors evidence that the liquidity information derived from the secondary steps of the book has also some "marginal" explanatory power on future returns.

Cao *et al* provide some insight about the issue this paper focus on performing a probit analysis of order aggressiveness. However, they do not provide a measurement of the value added of the book beyond the best quotes and do not study patient and impatient traders' behavior independently. In addition, we believe that those traders that continuously monitor the market may be able to infer about the state of the book beyond the best quotes by closely studying the evolution of the best quotes. Consequently, some of the information derived from the secondary quotes may be redundant. We take this possibility into account by considering summary measures of the information on the secondary levels of the book that are unconnected with the information on the best quotes.

3. Market background and data

The SIBE is an electronic platform that connects the four stock exchanges that constitute the Spanish Stock Exchange (SSE), located at Barcelona, Bilbao, Madrid and Valencia. Since 1995, this electronic system holds the trading activity of all the Spanish stocks that achieve pre-determined minimum levels of trading frequency and liquidity.[3] Every order submitted to the system in any of the four markets is electronically routed to a centralized LOB to proceed with its immediate execution or storage. The matching of

[3] Illiquid or infrequently traded stocks are negotiated through an auction-based trading system called Fixing.

orders is, therefore, computerized. It is a highly transparent market that provides real-time information on the LOB and trades through the Computerized Dissemination Information System (IDS). The status of the book is updated instantaneously on broker's screens each time there is a cancellation, execution, modification or new submission.

The SIBE is organized as a pure order-driven market with a daily continuous trading session from 9:00 a.m. to 5:30 p.m. and two call auctions that determine the opening and closing prices. In this paper, we discard data from the daily auctions. During the continuous trading session, orders are submitted, modified or cancelled. A trade takes place whenever a counterpart order hits the quotes. The market is governed by a strict price-time priority rule. However, an order may lose priority if modified.

Stocks are quoted in euros. The minimum price variation (tick) equals 0.01 for prices below 50 and 0.05 for prices above 50. The minimum trade size is one share. There are no market makers; there is no floor trading, and price-improvements are not possible either. At the Block Market and at the Off-Hours Market (5:40 to 8:00 p.m.), the brokers can execute pre-arranged trades or "applications", although they are subject to restrictive price and minimum size/value conditions. In this paper, however, only trades from the ordinary market are considered.

There are three basic types of orders: market, limit and market-to-limit. Market orders are executed against the best prices on the opposite side of the book. Any excess that cannot be executed at the best bid or ask quote is executed at less favorable prices by walking down (up) the book until the order is fulfilled. Market orders belong to the C5 to C7 categories of aggressiveness previously defined. Market-to-limit orders do not specify a limit price but are limited to the best opposite-side price on the book at the time of entry. Any excess that cannot be executed at that price is converted into a limit order at that price. Therefore, these are C6 orders. Finally, limit orders are to be executed at the limit price or better. Any unexecuted part of the order is stored in front of the book at the limit price. These are C2 to C4 orders.[4]

By default, orders expire at the end of the session. Nonetheless, the broker can enter a specific expiration date, with a maximum of 90 calendar days. For all type of orders, brokers may specify special conditions, like "immediate execution or elimination", "minimum execution" and "fill or kill". In practice, orders with these conditions are not distinguishable from some of the seven categories defined above. The SIBE also allows partially undisclosed limit orders, known as "iceberg" orders.[5]

Our database consists on all the movements of the 5 best bid and offer quotes of the LOB and all trades executed from July to December 2000 (124 trading days) during the continuous trading session. We use the first five months of data to perform estimations and in-sample analyses and the last month to carry out out-of-sample analyses. The LOB data includes quotes, disclosed depth and the number of orders supporting each quote. All the movements of the book are time stamped at the nearest hundredth of a second. We have the same LOB information as the SSE brokers have (in real time) during the session. The trading data details the price, the size and the counterparties of each trade. We have developed a foolproof algorithm that perfectly matches trade and quote data. Using this matched data, it is straightforward to classify all the movements of the LOB into one of the seven categories of aggressiveness formerly defined. Since price-improvement is not possible, buyer and seller initiated trades are easily identified.

[4] Notice that limit orders at a price equal to the best opposite-side quote and for a smaller (larger) quantity than that available at that quote cannot be distinguished in practice from C4-market (C5-market to limit) orders. Therefore, we pool these two categories as market (market to limit) orders. Similarly, we put together limit orders that walk up or down the book and become totally fulfilled with C7-market orders. Limit orders that walk up or down the book but become partially executed are very unusual in the SSE. They represent less than 0.3% of all orders submitted. These orders are also considered C7.

[5] In this paper, we take into account the presence of undisclosed depth when determining the aggressiveness of an order. Thus, a market order with size larger than the disclosed depth at the best-opposite quote on the book is classified as C7 only if it exhausts all the available depth, disclosed plus undisclosed, at that quote. See Pardo and Pascual (2004) for a study on the usage and impact of hidden limit orders at the SSE.

We consider data on the 36 stocks that were included in the IBEX-35 index sometime through the year.[6] One stock is excluded because of a merger. Table I provides some descriptive statistics, including information about the book and daily trading measures. Even though these are the most frequently traded and liquid stocks of the SSE, there are huge differences between them in terms of immediacy costs, depth and activity.

[Table I]

Table II provides summary statistics about order aggressiveness. We classify each update of the LOB into the 7 categories of aggressiveness defined earlier. The most frequent category is small market orders (C5) with an average of 38.74% followed by limit orders within the best offer and bid quotes (C4). The less frequent category is that of large market orders -or alike- (C7). On average, 38.42% of the orders submitted provide liquidity and 61.58% either consume or withdraw liquidity.

[Table II]

Table III provides additional descriptive statistics about durations of trades, limit orders and cancellations. We disaggregate the data into offer and bid flow. These series are highly autocorrelated (Spearman's rho statistics are significantly different from zero) and overdispersed (the standard deviation is bigger than the mean). Notice also that durations have very long right tails (compare the differences between the 25% and 75% percentiles with respect to the median). A differential characteristic of our dataset is the absence of zero-durations because of the high precision at which book updates are time stamped. Another common feature of these series is a strong intra-daily seasonality. All these aspects are revisited in section 5.

[Table III]

We consider two large sets of book information. The first piece consists on the *best quotes* (BQ). Even traditionally opaque markets, like the NYSE, have always provided

[6] The IBEX-35 index is computed as a cross-stock average trade price weighted by market capitalization. It is composed of the 35 most liquid and active SIBE-listed stocks during the most recent six-month control period. The composition is ordinarily revised twice a year, but extraordinary revisions are possible due to major events like mergers or new stock issues.

this information to the market. We summarize this piece of book information into the following variables, all of them defined with respect to the incoming order:

- *SPR:* Bid-ask spread.
- DS^1 (DO^1): Pending number of shares or depth on the same (opposite) side of the market.
- NS^1 (NO^1): Number of orders on the same (opposite) side of the market.

Since the correlation between DS^1 and NS^1 and between DO^1 and NO^1 is very high, we consider two alternative BQ sets: $BQ_1 = (SPR, DS^1, DO^1)$ and $BQ_2 = (SPR, NS^1, NO^1)$.

The second piece of book information consists of the *additional* four levels of *quotes* (AQ) publicly available at the SSE. We summarize this second set of quotes using:

- DS^{25} (NS^{25}): Accumulated depth (number of orders) on the same side of the market.
- DO^{25} (NO^{25}): Accumulated depth (number of orders) on the opposite side of the market.
- LS^{12} (LS^{25}): Distance –ticks- between the best and the second best (the second best and fifth best) quotes on the same side of the market.
- LO^{12} (LO^{25}): Distance between the best and the second best (the second best and fifth best) quotes on the opposite side of the market.

These latter "length" measures are less frequent in microstructure research and some justification is in order. On the one hand, these measures capture the expected price impact of large market orders. Thus, LS^{12} and LO^{12} measure de incremental cost of consuming more than the depth available at the best quotes. This cost may influence the aggressiveness of traders and the timing of orders. On the other hand, the length of the book may signal the consensus among traders about the true value of the stock; it may be informative about future price changes or it may indicate the presence of informed traders; it may also be interpreted as a measure of the willingness of traders to provide liquidity on a given side of the LOB.

We will also consider two alternative AQ sets: $AQ_1 = (DS^{25}, DO^{25}, LS^{12}, LO^{12}, LS^{25}, LO^{25})$ and $AQ_2 = (NS^{25}, NO^{25}, LS^{12}, LO^{12}, LS^{25}, LO^{25})$. Finally, the variables in the *AQ* set are defined as the residuals of a linear regression of each of its components on the variables in the pertinent *BQ* set, so that the *AQ* sets have no redundant information with respect to the *BQ* sets.[7]

4. Aggressiveness

Suppose that the degree of aggressiveness (impatience) of a given trader i is a (linear) function of a variety of factors X_i^k, k = 1,…, K. The LOB information, we presume, is included among these aggressiveness-inducing factors. Hence, the aggressiveness index A_i^* can be represented as,

$$A_i^* = \sum_{k=1}^{K} \beta_k X_i^k + \varepsilon_i = Z_i + \varepsilon_i, \qquad [1]$$

where β_k is the coefficient associated with the k^{th} factor. The error term ε_i indicates that the relationship in [1] is not an exact one. The aggressiveness index A_i^* is difficult, if not impossible, to observe. Therefore, equation [1] is a latent regression. However, we can infer about the degree of aggressiveness of trader i by observing the specific order submitted by that trader. The seven categories of order aggressiveness (C1 to C7) previously described represent a partition of the state space that allows mapping the latent degree of aggressiveness into observable discrete values. Let A_i be a ordinal response variable such that,

$$A_i = \begin{cases} 1 & \text{if } A_i^* \le \delta_1 \\ m & \text{if } \delta_{m-1} < A_i^* \le \delta_m,\ m = 2,\ldots,6 \\ 7 & \text{if } A_i^* > \delta_6 \end{cases} \qquad [2]$$

with δ_m being unknown thresholds, to be estimated along with the β_k parameters in [1]. If $A_i^* > \delta_6$, the trader is extremely aggressive and submits C7-market orders ($A_i = 7$); if

[7] That is, AQ_1 contains the residuals of a linear regression of each of its components (DS^{25}, DO^{25}, LS^{12}, LO^{12}, LS^{25} and LO^{25}) on the variables in the BQ_1 set (SPR, DS^1 and DO^1), and AQ_2 contains the residuals of a linear regression of each of its components (NS^{25}, NO^{25}, LS^{12}, LO^{12}, LS^{25} and LO^{25}) on the variables in the BQ_2 set (SPR, NS^1 and NO^1). This guarantees that any information contain attributed to the secondary levels of the book (AQ sets) is (linearly) unconnected with the information on the best quotes.

$\delta_1 < A_i^* \le \delta_2$, the trader is highly patient and submits C2-limit orders ($A_i = 2$), and so on. Assuming that the probability distribution of the error terms ε_i is normal, equations [1]-[2] define an ordered probit model. The probability of A_i taking value m is,

$$\Pr(A_i = m) = \Pr(\delta_{m-1} - Z_i < \varepsilon_i \le \delta_m - Z_i) = \Phi(\delta_m - Z_i) - \Phi(\delta_{m-1} - Z_i)\,, \quad [3]$$

with $\delta_0 = -\infty$, $\delta_7 = +\infty$ and $\Phi(.)$ being the normal cumulative distribution function.[8] The log-likelihood function is,

$$L(\beta_1, \ldots \beta_K, \delta_1, \ldots \delta_m) = \sum_{i=1}^{N} \sum_{m=1}^{6} d_{im} \log\left[\Phi(\delta_m - Z_i) - \Phi(\delta_{m-1} - Z_i)\right]$$

where d_{im}, $i=1,\ldots,m$ is an indicator variable that equals 1 if $A_i = m$ and 0 otherwise.

In order to measure the information content of the LOB beyond that of the best bid and ask quotes, we estimate three alternative models. The "Baseline" model (BM) only includes the first lag of the dependent variable in the explanatory variable set; the "Best Quotes" model (BQM) adds the BQ set of explanatory variables to the BM model; the "Complete Book" model (CBM) adds the AQ set of explanatory variables to the BQM. Therefore, the BM assumes that the information gleaned from the book is not relevant in explaining order aggressiveness; the BQM presumes that only the best quotes of the LOB provide valuable information and, finally, the CBM is based on the notion that the book provides relevant information away from the best quotes. We evaluate the relative performance of each model both in-sample and out-of-sample. For the in-sample analysis, we use all orders submitted from July to November 2000. For the out-of-sample analysis, we use the in-sample estimated coefficients on the December 2000 data.

Table IV summarizes the estimation of the CBM for the 36 stocks. We consider two alternative specifications: model M1 includes BQ_1 and AQ_1 as exogenous variables and

[8] Alternatively, we can assume that the error terms are logistically distributed. In such a case, equations [1] and [2] define an ordered *logit* model. There is no theoretical reason to prefer a priori a normal or a logistic distribution. The difference between both distributions is in the tails, much heavier in the case of the logistic distribution. Generally, either model will give identical substantive conclusions. In case of large number of observations and a heavy concentration of observations in the tails of the distribution, however, the estimates may differ substantially (e.g., Liao, 1994). Since both Griffiths *et al* (2000) and Ranaldo (2004) consider the case of normality, we also base our analysis on the ordered probit model. We have not found, however, remarkable differences using the ordered logit model.

model M2 includes BQ_2 and AQ_2 instead. We also distinguish between buyer-initiated-orders and seller-initiated-orders. The estimates of β_k and δ_m (not reported) are obtained by maximum likelihood (ML). Griffiths *et al* (2000) for the Toronto Stock Exchange and Ranaldo (2004) for the Swiss Stock Exchange estimate similar ordered probit models. These papers, however, do not distinguish the effect of different pieces of book information.

[Table IV]

The estimated coefficients for the variables in the *BQ* sets are consistent with the hypotheses H1 to H3 discussed and tested by Ranaldo (2004). A wider spread reduces the aggressiveness of traders (H1), consistent with Foucault (1999). As argued by Parlour (1989) and Handa et al. (2002), the thicker the book on the buy (sell) side, the more aggressive the incoming buyer (seller) (H2) and the thicker the book on the sell (buy) side, the less aggressive the incoming buyer (seller) (H3). These relationships are stronger in model M2 (number of orders) than in model M1 (quoted depth). We also find a strong first order positive autocorrelation in order aggressiveness, an expected result given the "diagonal effect" reported by Biais et al. (1995).

The results for the number of orders (NS^{25}, NO^{25}) and depth (DS^{25}, DO^{25}) apart from the best quotes generally support H2 and H3, particularly in model M2, but they are far less convincing. Regarding the length measures, we obtain a weak but clearly negative effect of LS^{12} and a strong positive effect of LO^{12} on order aggressiveness. A small value of LS^{12} may signal a tight or crowded book on the same side of the market as the incoming trader. In this situation, gaining precedence by price might be difficult and hitting the best quotes would bring a longer-than-average time to execution. Consequently, patient traders could become more aggressive and submit market orders. On the other hand, a higher dispersion on the offer (bid) quotes may be associated with a lower probability of execution of a limit order to buy (sell), inducing the incoming trader to be more aggressive. The results for the other variables are weak or inconclusive.

Table V shows the relative in-sample and out-of-sample performance of each of the three models estimated. We provide four alternative goodness-of-fit measures that correspond to the in-sample (adjusted) pseudo-R^2s of McFadden (1973, p.121), Maddala

(1983, p.39) with the Cragg and Uhler (1970) correction, Aldrich and Nelson (1984) with the Veall and Zimmermann (1992) correction, and McKelvey and Zavoina (1975).[9] Namely, Table V contains the median psuedo-R^2s for the BM model, the increase in the BQM pseudo-R^2s with respect to the BM and the increase of the CBM pseudo-R^2s with respect to the BQM. We observe that the in-sample fit improves on the BM a median 211.18% (289.45%) for sellers and 359.61% (446.54%) for buyers with the M1 (M2) specification when the variables from the best quotes of the book are added to the model. In addition, the fit improves on the BQM a median 47.33% (18.22%) for sellers and 78.86% (27.11%) for buyers when the whole LOB is taken into account. This increasing pattern indicates that the state of the book determines, at least partially, the aggressiveness of traders. Most of the explanatory power concentrates on the best quotes. However, traders examine not only the information available at the best quotes but also the less aggressive quotes.

A similar conclusion can be derived from the out-of-sample McKelvey-Zavoina (adjusted) pseudo-R^2. The predictive capacity of the BQM outperforms that of the BM model by a median 263.21% (156.46%) for sellers and 424.8% (157.69%) for buyers. When the complete book is considered, there is an additional improvement of 53.79% (37.36%) for sellers and 47.24 (70.8%) for buyers.

As an alternative to the previous point measures of goodness-of-fit, we have also performed an additional experiment. Using the in-sample estimated coefficients, we have computed the one-step-ahead probability for each of the 7 categories of aggressiveness and for each out-of-sample observation. We have compared the predicted probabilities for the actual event with a constant probability given by the in-sample relative

[9] No one of these measures is universally accepted or employed. The values between zero and one have no natural interpretation, though it has been suggested that the pseudo-R^2 value increases as the fit of the model improves. In a comparative analysis performed by Veall and Zimmermann (1996), these authors conclude that, for the particular case of the ordered probit model, the pseudo-R^2 due to McKelvey-Zavoina outperforms the other measures and has a strong numerical relationship to the OLS-R^2 in the latent variable. The Veall-Zimmermann and the Cragg-Uhler's measures also perform reasonably well. We include the McFadden's pseudo-R^2 because it is the most common in statistical packages. For a review of all these goodness-of-fit measures see Veall and Zimmermann (1996). For a definition see Appendix A. As in standard regression analysis, we use adjusted versions of these measures to take into account the change in degrees of freedom.

frequency.[10] The results are summarized in the last section of Table V. The CBM usually outperforms the BQM and the BM on the basis that it allocates a higher-than-its-relative-frequency probability to the actual event more often than the other two models do. For example, the CBM is the best model against the relative frequency rule for 80.56% of the stocks for the sellers-M2 specification; only for 13.89% of the stocks the BQM outperforms the other two models. Table V also provides a direct comparison between models. For example, the CBM does better than the BQM for 94.44% of the stocks in the buyers-M1 model in the sense that the CBM usually allocates a higher probability to the true event than the BQM. Both book models usually improve on the BM.

[Table V]

The former ordered probit model does not allow studying the relevance of the different pieces of book information in the trading decisions of passive and active traders independently. The submission of an order, however, can be though as a sequential process with two steps. In the first step, the trader chooses between a cancellation, a limit order or a market order. In the second stage, the patient trader places the limit order either away from, at or within the best quotes, and the impatient trader fixes the size of the order: less volume than available at the best opposite quote, a market to limit order or more volume than available at the best opposite quote. These stages in the decision process conform a sequential ordered probit analysis (e.g., Liao, 1994), which consists in estimating an ordered probit model at each stage of the sequence.

Table VI summarizes the estimation of the second stage of the sequential ordered probit CBM for the entire sample. We do not report the first stage since the results are similar to those in Table IV. Table VI shows that patient traders submit more aggressive limit orders as the spread increases. This is consistent with Biais et al. (1995) conclusion that in pure order driven markets the traders provide liquidity when it is valuable for the marketplace. Large impatient traders are also more frequent when the spread is large, probably because the high immediacy costs discourage the small investor.

[10] Since all the categories of aggressiveness are not equally frequent, the predicted probability for the most frequent category (small market orders) is always the largest, independently of the specification of the ordered probit model. However, the expected probabilities for each category and each observation do are model-specific.

The aggressiveness of the patient trader increases with DS^1 or NS^1, but it decreases with the length of their side of the market (LS^{12} and LS^{25}). All these variables may proxy for the proportion of traders with a similar valuation. In order to gain precedence when the book is thick the patient trader has to submit orders within the best quotes. When the book is long, the patient trader will demand a larger compensation so as to provide liquidity. The impatient trader is also more aggressive the more crowded the book on her side of the market (in terms of DS^1, LS^{12} and LS^{25}) and the more disperse the book in the opposite side of the market (in terms of DO^1, NO^1, LO^{12} and LO^{25}). An increase in LO^{12}, for example, means a larger cost of submitting C5-market orders, which reduces the aggressiveness of the impatient trader. [11]

[Table VI]

Table VII reports the relative in-sample and out-of-sample performance of the sequential ordered probit models CBM, BQM and BM. We only report the results for the second step. The in-sample analysis shows that passive traders' strategic decisions clearly depend on the book information. There is a median improvement of 211.61% (134.48%) for sellers and 250.42% (177.09%) for buyers with the M1 (M2) specification when the best quotes of the book are considered. More important, the CBM improves on the BQM by a median of 174.81% (263.19%) for sellers and 153.95% (210.51%) for buyers with the M1 (M2) models, which means that order submissions by liquidity providers are (at least partially) based on an examination of the state of the whole LOB. From this point of view, liquidity traders undoubtedly benefit from an increased degree of pre-trade transparency. The out-of-sample (adjusted) pseudo-R^2 leads to the same conclusion showing a similar increasing pattern from BM to BQM and from BQM to BCM. In addition, the CBM obtains the best scores against the relative frequency rule. It always outperforms the BQM: for all the stocks in the sample, the CBM allocates higher probabilities to the actual event than the BQM and the BM.

The results for the active traders (liquidity consumers) are remarkably different. The strategic decision of active traders at this second step is to choose the size of their market

[11] The strong negative effect of the number of orders on the same side of the book on the decision of the impatient trader reported in Table VI – Panel B suggests that what matters is neither the depth nor the number of orders but the number of large orders on the same side of the book. The larger the average order size supporting the best quotes the larger the aggressiveness of the incoming impatient trader.

order. Table VII shows that this decision strongly depends on the best offer and bid quotes. There is a median in-sample fit improvement of 1512% (1100%) for sellers and 2690% (991%) for buyers with the M1 (M2) specification when the best quotes are added to the BM. However, the CBM improves on the BQM only by a median 12.97% (23.75%) for sellers and 3.71% (34.03%) for buyers. This means that the most aggressive traders in the market barely base their strategic decisions on the state of the LOB beyond the best quotes. The out-of-sample predictive performances support this conclusion: the pseudo-R^2 for sellers-M1 (buyers-M1), for example, increases a negligible 0.36% (0.71%) from the BQM to the CBM. Moreover, the book-based models rarely do better than the relative frequency rule and the BQM probabilities outperform those of the CBM as many times as the CBM outperforms the BQM.

Table II showed that 80.5% of the orders that are executed instantaneously (C5 to C7) in the SSE are small market orders (C5) and 11.4% are market-to-limit orders (C6). This suggests that the Spanish active traders tend to adjust the size of their orders to the available depth on the best quote of the opposite side of the market. Consequently, an increase in the pre-trade transparency would have only a marginal impact on the order submission strategy of liquidity consumers.

5. The timing of cancellations, limit orders and market orders

In this section, we analyze what pieces of book information are important in explaining the time between two consecutive trades, limit order submissions and cancellations on the same side of the market. These three types of orders coincide with the three levels of aggressiveness in the first step of the sequential ordered probit model estimated in the previous section. The analysis of durations is performed using the logarithmic version of the autoregressive conditional duration (ACD) model (Engle and Russell, 1998) introduced by Bauwens and Giot (2000). The $Log-ACD(p,q)$ model for the duration d_i is defined as,

$$d_i = \Psi_i \varepsilon_i = \exp(\psi_i)\varepsilon_i$$
$$\psi_{\mathrm{i}} = w + \sum_{j=1}^{p} \alpha_j \ln d_{i-j} + \sum_{j=1}^{q} \beta_j \psi_{i-j}$$

where ε_i (for $i = 1, \ldots, n$) are *iid* innovations with $E[\varepsilon_i] = \mu$, such that $E[d_i | I_{i-1}] = \Psi_i \mu_i$. An alternative, and convenient, specification is $d_i = \Psi_i \eta_i$, where $\eta_i = \mu^{-1}\varepsilon_i$ and hence $E[\eta_i] = 1$ and $E[d_i | I_{i-1}] = \Psi_i$. The conditional duration (Ψ_i) is specified as a linear function of the previous *p* durations and *q* conditional durations. As in the GARCH literature, numerous studies have shown that the $Log-ACD(1,1)$ captures correctly the dynamics of a very general class of models (e.g., Bauwens et al., 2003). Therefore, we restrict our attention to the $Log-ACD(1,1)$ case, $\psi_i = w + \alpha \ln d_{i-1} + \beta \psi_{i-1}$.

We use a $Log-ACD(1,1)$ instead an $ACD(1,1)$ model because when additional explanatory variables, implied by microstructure theory, are added linearly in the conditional expectation negative slope coefficients may be expected for some of those variables. Since no sign restrictions are needed on the parameters to ensure the positivity of the conditional duration, we can linearly introduce any set of exogenous variables, regardless of the expected sign of its accompanied parameter,

$$\psi_i = \omega + \alpha \ln d_{i-1} + \beta \psi_{i-1} + x_i^{'} \delta\,, \qquad [6]$$

where *x* is a row vector of dimension *s* including the exogenous variables.

A parametric model is obtained when the distribution of ε_i is specified up to a finite number of parameters. Given the particularities of duration data (see Table III), we use a three-parameter distribution, the generalized gamma.[12] Let ε_i follow a generalized

[12] Engle and Russell (1998) proposed the standard exponential distribution and, as an extension, the Weibull distribution. However, as documented by Bauwens and Veredas (2003) and Grammig and Maurer (2000), the Weibull distribution may not be flexible enough for duration processes with high intensity. This is our case, where orders, trades and cancellations arrive at a high rate and extreme events (very short and very long durations) are often observed. The generalized gamma distribution nests the exponential and Weibull as particular cases.

gamma distribution, i.e. $\varepsilon_i \approx GG(1,\upsilon,\gamma)$ where υ and γ are the shape parameters. Then $d_i \approx GG(1/\exp(\psi_i),\upsilon,\gamma)$ with density function,

$$f_{GG}(d|I_{i-1};\theta) = \frac{\gamma}{\exp(\psi_i)^{\upsilon\gamma}\Gamma(\upsilon)} d^{\upsilon\gamma-1} \exp\left[-\left(\frac{d}{\exp(\psi_i)}\right)^{\gamma}\right],$$

where $\theta = (\omega,\alpha,\beta,\gamma,\upsilon)$ is the parameter set. It is estimated by maximizing the likelihood function,

$$L(\theta) = \sum_{i=1}^{n} \log f_{GG}(d|I_{i-1};\theta).$$

Prior to estimation, we adjust the observed durations by intra daily seasonality.[13] Let

$$d_i = D_i / \phi(t_i'),$$

where D_i is the original duration, d_i is the adjusted duration, $\phi(t_i')$ is the time-of-day effect, and t_i' is a bounded random variable that measures the number of accumulated seconds since the opening. It is obtained from the arrival times using,

$$t_i' = \begin{cases} t_i - \left\lfloor \dfrac{t_i - t_0}{t_c - t_0} \right\rfloor (t_c - t_0) & if \quad t_i > t_c \\ t_i & otherwise \end{cases},$$

where $\lfloor x \rfloor$ is the integer part of *x*, and t_o and t_c stand respectively for the market opening and closing (in hundredths of a second). This gives a sequence of arrival times that are, everyday, monotonically increasing from t_o to t_c. They are hence bounded, the whole process looking like a toothed sequence.

The estimated seasonal pattern $\widehat{\phi}(t_i')$ is computed by a nonparametric regression of the observed duration on the time of the day, a methodology introduced by Veredas *et al* (2001). The result is the Nadaraya-Watson estimator,

[13] The durations can be thought of as consisting of two parts: a stochastic component to be explained by the Log-ACD model, and a deterministic part, namely the seasonal intra-daily pattern. This effect arises from the systematic variation of the market activity during each trading day.

$$\widehat{\phi}\left(t_o^{'}\right)=\frac{\frac{1}{nh}\sum_{i=0}^{n}K\left(\frac{t_0^{'}-t_i^{i}}{h}\right)d_i}{\frac{1}{nh}\sum_{i=0}^{n}K\left(\frac{t_0^{'}-t_i^{i}}{h}\right)},$$

with the function $K(\cdot)$ being a kernel estimator and *h* the bandwidth. The kernel chosen is the quartic and the bandwidth is $2.78\sigma n^{-1/5}$, where σ is the sample standard deviation and *n* is the number of observations.[14]

For each stock in the sample, we compute six different time-of-day adjusted durations: trades, limit orders, and cancellations, distinguishing between buyer and seller-initiated orders. As in the ordered probit analysis, we estimate three alternative models for each duration: the $Log-ACD(1,1)$ is our baseline model (BM) in this case; the BQM-$Log-ACD(1,1)$ adds the BQ set of explanatory variables to the BM model; the CBM-$Log-ACD(1,1)$ adds the AQ set of explanatory variables to the BQM. Finally, for the BQM and CBM models we consider two alternative specifications: model M1 includes BQ_1 and AQ_1 as exogenous variables and model M2 includes BQ_2 and AQ_2 instead. We evaluate the relative performance of each model both in sample (July to November 2000) and out-of-sample (December 2000).

The estimation results are not reported because of space limitations.[15] However, only the bid-ask spread (*SPR*) shows a strongly significant effect on all six durations and for all the model specifications. As the *SPR* increases, the time between consecutive trades (either buyer or seller-initiated) increases and the time between consecutive cancellations and limit order submissions decreases in both sides of the LOB. This result is consistent with the former evidence from the ordered probit models: a wide bid-ask spread decreases order aggressiveness as the increase in immediacy costs discourages market order traders.

[14] We also observe intra-daily deterministic patterns in some of the exogenous variables, in particular in the variables that correspond to the best quotes of the book. All these explanatory variables have been time-of-day adjusted using the same nonparametric regression as for the durations.

[15] The estimation results are available upon request from the authors.

Table VIII summarizes the relative in-sample and out-of-sample performance of the log-ACD models BM, BQM and CBM. We provide four alternative in-sample and out-of-sample goodness-of-fit measures for each model. The adjusted pseudo-$R^2(1)$ is a function of the mean square error (MSE), and the adjusted pseudo-$R^2(2)$ is a function of the sample correlation coefficient between the actual (d_i) and the fitted values of the durations ($\hat{\Psi}_i$). We also provide the Akaike (AIC) and the Schwarz Bayesian (SBC) information criteria computed from the residuals $\hat{\varepsilon}_i = d_i / \hat{\Psi}_i$. See the Appendix for a definition of all these measures. Table VIII contains the median psuedo-R^2s, AIC and BIC for the BM model, the percent increase in the BQM measures with respect to the BM and the percent increase of the CBM measures with respect to the BQM.

[Table VIII]

The in-sample analysis shows that the timing of orders barely depend on the book information. The goodness-of-fit adjusted-pseudo-R^2 measures improve, in median, a 1.3% for cancellations, 0.87% for limit order submissions and a 1.4% for trades when the best quotes of the book are considered. More important, the CBM improves on the BQM by a median of 1.2% for cancellations, 1.11% for limit order submissions and 0.73% for trades. We report a similar (or even more negligible) decreasing pattern in the AIC and BIC information criteria going from BM to BQM and from BQM to BCM. No remarkable differences are observed between seller and buyer-initiated orders. The results for the out-sample analysis strongly reject that either piece of book information is relevant in explaining the timing of cancellations, limit orders, or market orders. Many adjusted pseudo-R^2 measures decrease and the AIC and BIC information criteria increase as we add extra exogenous variables in the log-ACD model. In summary, Table VIII evidences that no piece of book information matters in explaining the timing of orders.[16]

This evidence may seem contradictory with Coppejans and Domowitz's (2002) findings. These authors use a Generalized ACD o GACD model (Lunde, 1999) to conclude that the information gleaned from the electronic LOB substantially affects

[16] We have performed some robustness tests. In particular, we have considered the exponential distribution for ε_i instead of the generalized gamma and we have also used cubic splines (e.g., Engle and Russell, 1998) instead of the Veredas *et al* (2001) methodology to estimate the seasonal component of the durations. The results in Table VIII are invariant to these alternative specifications.

trader behavior. Using unadjusted pseudo-R^2 measures, they compare the goodness of fit of a GACD model, which includes book and order flow information, with a simple ACD(1,1) model.[17] The ACD and the GACD are not nested models. On the contrary, the ACD model assumes that the exogenous variables are time-invariant, i.e. they do not change within the durations. As a consequence, it is impossible to discern if, for example, the reported 293% (in-sample) and 149% (out-of-sample) increase in the pseudo-R^2 for seller-initiated trade durations (see Table 4 in Coppejans and Domowitz, 2002) is due to the book information, to the order flow information, or simply to the fact that the GACD is a more rich and complex model than the ACD(1,1) used as a reference. Our findings in Table VIII suggest that Coppejans and Domowitz's results would not vary if the book information were dropped from their GACD model.

The results in Table VIII also refine to some extent the results of the ordered probit analysis in the previous section. The order aggressiveness analysis concluded that the LOB information was relevant in explaining the aggressiveness of an incoming order. Namely, the best quotes on the LOB were important in explaining the strategic decisions of any incoming trader and at any stage of the decision process; the book information beyond the best quotes, however, was only relevant in explaining the strategic decisions of an incoming limit order trader. In contrast, the analysis of cancellation, limit order, and trade durations in this section indicates that neither piece of book information, apart from the bid-ask spread, matters in explaining the particular submission time of an incoming order of a similar level of aggressiveness than the most recent order submitted.

Notice that in the order aggressiveness analysis we study the capacity of the book to provide information about an event that is going to occur almost instantaneously: the next order to be submitted. In the order duration analysis, however, we evaluate the capacity of the book to explain an event that may take a longer time to be accomplished: the next order to be submitted with a given level of aggressiveness. Therefore, a possible interpretation of our mixed findings is that the book information has explanatory power only in the very short run.

[17] For any two consecutive events of the same kind at time t-1 and t, the GACD model includes book information measured at t-1, namely the bid-ask spread and accumulated depth measures, and information about the order flow in that interval.

In addition, Franke and Hess (2000) show that the information value provided by the insight into the LOB in an electronic trading system declines when the intensity of private and public information arrival increases. The basic idea behind this result is that in times of low information intensity there are only a few updates in the state of the LOB and, consequently, the insight into the book may provide valuable information. In times of high information intensity, however, the order flow increases and the book updates continuously. Consequently, the snapshots of the book have little value. This feature of the book information may be more relevant in our duration analysis than in our aggressiveness analysis. When the order flow between two consecutive events of the same type at time t-1 and t is severe, the information on the most recent orders submitted may be more relevant in explaining the length of the duration than the book at time t-1. The information intensity has no impact on the order aggressiveness analysis because it always considers the most recent update of the LOB. An interesting extension of this study would be to evaluate whether the results in Table VIII improve when the analysis focuses on low information intensity periods, for example by separating the intermediate intervals of the trading sessions from the, usually more trading-intensive, opening and closing intervals.

6. Summary and conclusions

This paper has studied the importance of different pieces of LOB information in characterizing the strategic decisions of traders. Two basic pieces of book information have been considered: the best bid and offer quotes and the second to fifth bid and offer quotes on the book. Two related aspects of the order flow have been analyzed: the aggressiveness of traders, and the timing of trades, order submissions and cancellations.

Order aggressiveness has been modeled using (sequential) ordered probit models and the timing of orders has been modeled using log-ACD models. We have evaluated the relative improvement on the in-sample and out-of-sample goodness-of-fit performance of these models when the different pieces of book information are added sequentially as exogenous variables. The data for this empirical experiment consisted on six months of high-frequency book and trade files on the most frequently traded and liquid stocks of the Spanish Stock Exchange in 2000.

We have shown that the state of the book determines, at least partially, the aggressiveness of traders. Most of the explanatory power of the book concentrates on the best quotes. Nevertheless, our results suggest that traders also examine the less aggressive quotes. The model specification that includes all pieces of book information usually outperforms the model specification that only includes the best quotes. We have also analyzed the relevance of the different pieces of book information in the trading decisions of passive traders (liquidity providers) and active traders (liquidity consumers) independently. The aggressiveness of the passive traders clearly depends on the examination of the state of the whole LOB. The aggressiveness of the active traders, however, strongly depends on the best quotes of the book but it barely depends the state of the LOB beyond the best quotes.

The analysis of the duration of trades, limit orders, and cancellations assigned a less remarkable role to the LOB information. Only the bid-ask spread shows some remarkable explanatory power in explaining the timing of orders. Indeed, the goodness-of-fit analysis of durations evidences that no piece of book information matters in explaining the timing of orders.

We suggest that these apparently inconsistent findings could be reconciled. The inability of the book to explain the timing of orders may be well-matched with this information being useful in the very short-run only or with the informativeness of the book varying with the information intensity, as measured by the order flow. Neither the short-term nature of the book information nor their dependence on the information intensity should have an effect on the analysis of order aggressiveness but they may seriously affect the analysis of durations.

In summary, our empirical findings are mixed. The informativeness of the LOB largely concentrates on the best quotes. However, all traders seem to benefit from the additional quotes of the book at some stage in the decision-making process. For example, these extra quotes partially explain the decision of whether to submit a cancellation, a limit order or a market order. Then, if the incoming trader is a liquidity provider, the whole LOB matters in fixing the price of the limit order. However, if the incoming trader is a liquidity consumer, the book beyond the best quotes hardly counts in fixing the size of the market order. From this point of view, passive traders undoubtedly benefit from an increased degree of pre-trade transparency. However, an increase in the pre-trade transparency would have only a marginal impact on the order submission strategy of active traders. Moreover, pre-trade transparency does not help in explaining the timing of orders, although this result might change in periods of low information intensity.

References

Ahn, Hee-Joon, Kee-Hong Bae, and Kalok Chan, 2001, Limit orders, depth and volatility: evidence from the stock exchange of Hong Kong, *Journal of Finance*, 56, 2, 769-790.

Aldrich, J.H., and F.D. Nelson, 1984, *Linear Probability, Logit, and Probit Models*, Sage University Press.

Bauwens, Luc, and P. Giot, 2000, The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks, *Annales d'Economie et de Statistique*, 60, 117-149

Bauwens, Luc, Pierre Giot, Joaquim Grammig, and David Veredas, D., 2003, A comparison of financial duration models via density forecast, *International Journal of Forecasting*, forthcoming.

Bauwens, Luc, and David Veredas, 2003, The Stochastic Conditional Duration Model: A latent factor model for the analysis of financial durations, *Journal of Econometrics*, forthcoming.

Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *The Journal of Finance*, 50, 1655-1689.

Bloomfield, Robert, and Maureen O'Hara, 1999, Market transparency: who wins and who loses? *Review of Financial Studies*, 12, 1, 5-35.

Boehmer, Ekkehart, Gideon Saar, and Lei Yu, 2003, Lifting the Veil: An analysis of pre-trade transparency at the NYSE, *Working Paper*, New York Stock Exchange.

Cao, Charles, Oliver Hansch, and Xiaoxin Wang, 2003, The informational content of an open limit order book, *Working Paper*, Pennsylvania State University.

Coppejans, Mark, Ian Domowitz, 2002, An empirical analysis of trades, orders, and cancellations in a limit order market, *Working Paper*, Duke University.

Corwin, Shane A., and Marc L. Lipson, 2000, Order flow and liquidity around NYSE trading halts, *Journal of Finance*, 55, 4, 1771-1801.

Gragg, J.G., and R. Uhler, 1970, The demand for automobiles, *Canadian Journal of Economics*, 3, 386-406.

Daníelsson, Jón, and Richard Payne, 2001, Measuring and explaining liquidity on an electronic limit order book: evidence from Reuters D200-2, *Working Paper*, London School of Economics.

Domowitz, Ian, and Jianxin Wang, 1994, Auctions as algorithms: Computerized trade execution and price discovery, *Journal of Economic Dynamics and Control*, 18, 29-60.

Engle, Robert F., and J.R. Russell, 1998, Autoregressive conditional duration: a new approach for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, *Journal of Financial Markets*, 99-134.

Flood, Mark D., Ronald Huisman, Kees G. Koedijk, and Ronald J. Mahieu, 1999, Quote disclosure and price discovery in multiple-dealers financial markets, *Review of Financial Studies*, 12, 1, 37-59.

Franke, Günter, and Dieter Hess, 2000, Information diffusion in electronic and floor trading, *Journal of Empirical Finance*, 7, 455-478.

Gourieroux, C., and A. Monfort, 1995, *Statistics and Econometric Models*, Vol. II, Cambridge University Press.

Grammig, Joaquim, and K.O. Maurer, 1999, Non-monotonic hazard functions and the autoregressive conditional duration model, *The Econometrics Journal*, 3, 16-38.

Griffiths, Mark D., Brian F. Smith, D. Alasdair S. Turnbull, and Robert W. White, 2000, The costs and determinants of order aggressiveness, *Journal of Financial Economics*, 56, 65-88.

Handa, Puneet, and Robert A. Schwartz, 1996, Limit order trading, *The Journal of Finance* 51, 1835-1861.

Handa, Puneet, Robert A. Schwartz, and Ashish Tiwari, 2004, Quote setting and price formation in an order driven market, *Journal of Financial Markets,* forthcoming.

Harris, Lawrence, 1996, Does a minimum price variation encourages order exposure? *Working Paper*, Marshall School of Business.

Harris, Lawrence E., and Venkatesh Panchapagesan, 2003, The information-content of the limit order book: evidence from NYSE specialist actions, *Working Paper*, Marshall School of Business, University of Southern California.

Huang, Roger D., and Hans R. Stoll, 1994, Market microstructure and stock return predictions, *Review of Financial Studies*, 7, 1, 179-213.

Irvine, Paul, Benston, George, and Eugene Kandel, 2000, Liquidity beyond the inside spread: measuring and using information in the limit order book, *Working Paper*, Emory University.

Kaniel, Ron, and Hong Liu, 2001, So what orders do informed traders use?, *Working Paper*, University of Texas at Austin.

Liao, Tim F, 1994, Interpreting probability models. Logit, probit, and other generalized linear models, *Quantitative Applications in the Social Sciences*, 07-101, Sage University Paper.

Maddala, G.S., 1983, *Limited-dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.

Madhavan, Ananth, 2000, Market microstructure: a survey, *Journal of Financial Markets*, 3, 205-258.

Madhavan, Ananth, and Venkatesh Pachapagesan, 2000, Price discovery in auction markets: a look inside the black box, *Review of Financial Studies*, 13, 3, 627-658.

Madhavan, Ananth, David Porter, and Daniel Weaver, 2000, Should securities markets be transparent? *Working Paper*, University of Southern California.

McFadden, D., 1973, Conditional logit analysis of qualitative choice behavior, in Zarembka, P, (Ed.) *Frontiers in Econometrics*, pp. 105-142, Acadamic Press, New York.

McKelvey, R., and W. Zavoina, 1975, A statistical model for the analysis of ordinal level dependent variables, *Journal of Mathematical Sociology*, 4, 103-120.

Pardo, Angel, and Roberto Pascual, 2004, On the hidden side of liquidity, *Working Paper Series 2004*, FIN-04-004, Leonard N. Stern School of Business, New York University.

Parlour, Christine A., 1998, Price dynamics and limit order markets, *Review of Financial Studies*, 789-816.

Ranaldo, Angelo, 2004, Order aggressiveness in limit order book markets, *Journal of Financial Markets*, 7, 53-74.

Seppi, Duane, 1997, Liquidity provision with limit orders and a strategic specialist, *Review of Financial Studies*, 10, 103–150.

Veall, Michael R., and Klaus F. Zimmermann, 1992, Pseudo-R^2's in the ordinal probit model, *Journal of Mathematical Sociology*, 16, 332-342.

Veall, Michael R., and Klaus F. Zimmermann, 1996, Pseudo-R^2 measures for some common limited dependent variable models, *Journal of Economic Surveys*, 10, 3, 241-259.

Verbeek, M., 2000, *A Modern Guide to Econometrics*, John Wiley and Sons.

Veredas, David, Juan Rodriguez-Poo, and Antoni Espasa, 2001, On the (intra-daily) seasonality of a financial point process, *Working Paper*, #01-33 Statistics and Econometrics Series, Universidad Carlos III de Madrid.

TABLE I
Sample Statistics: Book and Activity

This table provides some descriptive statistics about the 36 stocks in the sample, averaged for July-December 2000. "Quote midpoint" is the average between the best offer and bid quotes. "Spread" is the distance, in number of ticks, between the best ask and bid quotes. "Ask (Bid) depth" is the accumulated number of shares offered at the five best ask (bid) quotes on the LOB. "Ask (Bid) orders" is the accumulated number of limit orders supporting the best ask (bid) quotes on the book. Ask_1-Ask_5 (Bid_1-Bid_5) is the distance, in number of ticks, between the first and the fifth ask (bid) quotes on the book. "Daily Vol. (Trades)" is the daily average share volume (number of trades). The tick is 0.01 euros for all stocks and during all the sample period.

	Quote Midpoint	Spread (# ticks)	Ask Depth (Book)	Bid Depth (Book)	Ask Orders (Book)	Bid Orders (Book)	Ask_1-Ask_5 (# ticks)	Bid_1-Bid_5 (# ticks)	Daily Vol./1000	Daily Trades
ACR	9.30	3.17	10571.1	14832.1	18.36	11.84	5.83	5.78	277.85	370.48
ACS	27.11	13.70	4028.2	4392.9	7.01	7.69	19.22	18.79	133.83	249.15
ACX	32.09	13.07	3701.3	3926.0	8.03	7.40	18.91	19.57	175.34	328.65
ALB	27.55	19.03	3083.9	3232.3	6.31	6.44	23.59	21.30	136.75	188.77
ALT	16.31	4.05	12583.4	11611.8	8.60	12.94	8.05	9.07	1131.46	773.14
AGS	14.19	6.63	5316.6	5719.1	7.47	7.75	11.24	11.29	171.98	232.81
AMS	10.23	2.84	14416.6	16484.7	17.03	12.66	6.33	6.14	1177.81	924.65
ANA	38.48	14.68	2955.1	2883.1	7.20	7.01	22.24	21.73	125.52	288.21
AUM	16.63	8.90	6913.5	6626.0	7.69	9.65	12.75	18.70	102.09	76.23
BBV	16.07	1.71	84451.5	53025.4	13.90	39.14	4.84	4.96	6974.52	2023.90
BKT	44.19	11.73	3694.6	3734.3	9.17	7.99	18.67	16.34	224.68	559.02
CAN	20.97	12.06	8585.1	5853.9	6.45	8.07	15.58	18.93	146.01	148.24
CTG	18.94	7.29	6700.7	6860.1	8.53	8.99	10.93	10.97	372.38	369.15
DRC	9.94	3.60	18047.6	11783.3	9.62	15.56	6.69	7.81	624.81	440.19
ELE	20.80	2.45	24505.2	22732.6	10.07	12.75	6.34	6.34	3099.08	1496.06
FCC	19.56	8.63	5168.4	5081.2	8.01	8.35	13.52	13.90	183.95	322.99
FER	13.92	4.96	5974.4	6629.5	9.79	10.07	8.90	8.74	191.16	396.99
GPP	4.40	1.71	42577.5	40534.4	20.56	20.38	4.45	4.50	1107.38	620.10
IBE	13.80	2.57	31215.0	27039.0	11.22	15.85	5.74	6.20	2112.56	758.00
IDR	17.81	5.87	5819.6	5662.0	9.01	9.49	10.07	9.64	290.02	530.90
MAP	18.27	12.12	6629.1	5302.8	7.10	10.03	16.58	21.56	130.70	116.61
NHH	13.09	6.26	11189.7	9085.4	7.43	9.29	10.02	11.47	357.15	232.24
POP	34.72	8.57	8096.0	5174.1	7.43	12.55	13.66	14.05	395.66	463.60
PRS	23.68	7.95	5207.0	5825.4	8.76	11.72	12.52	14.06	382.35	554.59
REE	10.66	4.49	7034.9	9465.6	10.03	9.66	7.75	7.69	160.62	279.86
REP	20.32	2.41	25296.6	24654.2	13.76	12.37	6.14	5.99	3528.20	1655.46
SCH	11.45	1.34	178610.4	115052.2	28.81	97.43	4.33	4.31	9719.32	2962.41
SGC	33.37	12.24	3056.1	3041.5	8.58	7.21	18.59	17.10	202.47	545.57
SOL	10.92	4.70	7633.1	8448.2	9.47	8.42	8.27	7.89	280.60	292.52
TEF	21.90	1.56	65666.8	56251.7	23.83	19.05	4.90	4.67	19714.10	6608.87
TPZ	4.92	1.60	50389.1	68614.1	40.60	20.30	4.45	4.38	1433.03	847.68
TRR	34.53	4.75	8983.2	8675.3	15.11	11.48	9.59	8.37	2935.99	4596.82
TPI	8.83	2.41	14769.5	17494.5	17.24	12.05	5.61	5.39	1133.28	1008.90
UNF	20.47	4.40	14094.7	19545.4	8.21	9.20	8.89	7.72	768.06	413.85
VAL	6.71	3.01	14156.4	12739.4	10.09	11.35	5.95	6.49	334.15	253.67
ZEL	34.95	4.25	6826.7	5900.9	10.71	11.15	7.78	7.52	865.26	1776.36
Average	19.47	6.41	20220.79	17608.74	11.98	14.26	10.53	10.81	1697.23	936.30
Std. Dev.	(10.07)	(4.53)	(32720.8)	(23084.8)	(7.13)	(15.41)	(5.46)	(5.69)	(3677.81)	(1324.33)

TABLE II
Statistics: Aggressiveness

This table describes the distribution of orders in terms of the 7 categories of order aggressiveness (C1 to C7) defined in section 3. "M.O." means market orders. "L.O." means limit orders. C1 are cancellations. C2 are limit orders to sell (buy) above (below) the best ask (bid) quotes. C3 are limit orders to sell (buy) hitting the best ask (bid) quote. C4 are limit orders within the best quotes. C5 are market orders for a lower size than the depth available at the best quote on the opposite side of the book. C6 are totally or partially executed orders that consume (only) the best quote on the opposite side of the book. C7 are totally or partially executed orders that consume more than one level of quotes on the opposite side of the book. For each statistic, the proportion of observations belonging to each category is provided.

		Limit Orders			Small	Market	Large	
	Cancellat.	Ab./Bel.	At	Within	M.O.	to Limit	M.O.	
	(C1)	(C2)	(C3)	(C4)	(C5)	(C6)	(C7)	Obs.
Mean	13.51	12.56	10.63	15.23	38.74	5.51	3.82	221344
Std. Dev.	2.65	2.09	1.98	5.57	5.33	0.99	0.73	285811
Median	12.89	12.23	10.61	16.14	38.65	5.31	3.87	120505
Max.	20.26	19.51	16.55	25.80	53.17	7.21	5.36	1414546
Min.	8.18	9.59	7.78	4.02	28.65	3.38	2.16	20203

TABLE III
Statistics: Durations

This table provides average statistics about durations of trades, limit orders and cancellations. We separate the bid flow from the offer flow and buyer-initiated trades from seller-initiated trades. Durations are measured in seconds. All values are averages of the individual statistics for the 36 stocks in the sample. "rho(0,-*k*)" stands for the Spearman's rho *k*th order autocorrelation coefficient.

Statistic	Cancellation Ask	Cancellation Bid	Limit order Ask (to sell)	Limit order Bid (to buy)	Trade Ask (buys)	Trade Bid (sells)
Mean	431.86	368.09	210.91	171.95	210.20	172.52
Std. Dev.	774.77	697.17	346.52	283.48	375.68	284.40
Pctile. 25%	31.49	25.19	24.94	19.95	18.51	20.40
Median	146.83	118.17	86.75	70.06	72.32	70.10
Pctile. 75%	487.89	396.67	249.45	202.38	238.95	204.35
Pctile. 95%	1819.47	1582.49	832.91	680.67	873.21	684.07
Min.	0.01	0.04	0.01	0.01	0.01	0.01
Max.	11455.19	12088.52	6829.70	5521.76	8158.09	6449.56
Obs.	15467	16745	31675	46222	55081	50886
rho (0,-1)	0.2815	0.2743	0.2919	0.2872	0.3416	0.2779
rho (0,-5)	0.1630	0.1509	0.1927	0.1877	0.2347	0.1744

TABLE IV
Aggressiveness: Ordered Probit Models - Estimates

This table summarizes the results of estimating the ordered probit model [1]-[3]. The dependent variable is the aggressiveness of the order, ranked from the least to the most aggressive type of order. Therefore, a positive estimated coefficient means that the associated explanatory variable is positively related to order aggressiveness. We report the median estimated coefficient for the 36 stocks in the sample, the percentage of statistically significant coefficients and the percentage of statistically significant and positive coefficients. We provide separated results for buyer initiated orders and seller initiated orders. We also provide separated results for a model with depth measures (model M1 -Panel A) and a model with number-of-orders measures instead (model M2 –Panel B). The exogenous variables defined with respect to an incoming order are: (1) Computed using the best ask and bid quotes: the bid-ask spread (*SPR*), the depth on the same side of the market (DS^1), the depth on the opposite side of the market (DO^1), the number of orders on the same side of the market (NS^1), the number of orders on the opposite side of the market (NO^1). (2) Computed using from the 2[nd] to the 5[th] best quotes: the accumulated depth (number of orders) on the same side of the market DS^{25} (NS^{25}), the accumulated depth (number of orders) on the opposite side of the market DO^{25} (NO^{25}). (3) The distance between the best and the second best (the second best and fifth best) quotes on the same side of the market LS^{12} (LS^{25}), the distance between the best and the second best (the second best and fifth best) quotes on the opposite side of the market LO^{12} (LO^{25}). All models include one lag of the dependent variable.

Panel A: Model M1

Sellers	Agr(-1)	*SPR*	DS^1	DO^1	DS^{25}	DO^{25}	LS^{12}	LO^{12}	LS^{25}	LO^{25}
Median	0.0314	-0.0193	9.10E-06	-1.47E-06	2.77E-06	-7.20E-07	-0.0163	0.0167	0.0013	0.0017
% Signif.	97.22	100	83.33	38.89	47.22	58.33	69.44	97.22	47.22	44.44
% Positive	97.22	8.33	77.78	8.33	33.33	11.11	8.33	97.22	25	25
Buyers										
Median	0.0290	-0.0222	9.33E-06	-2.57E-06	3.09E-06	5.30E-07	-0.0281	0.0177	-0.0033	0.0017
% Signif.	91.67	97.22	66.67	38.89	61.11	55.56	58.33	88.89	50	44.44
% Positive	88.89	5.56	61.11	8.33	47.22	30.56	2.78	86.11	16.67	25

Panel B: Model M2

Sellers	Agr(-1)	*SPR*	NS^1	NO^1	NS^{25}	NO^{25}	LS^{12}	LO^{12}	LS^{25}	LO^{25}
Median	0.0321	-0.0199	0.0284	-0.0184	0.0039	-0.0005	-0.0170	0.0173	0.0014	0.0015
% Signif.	97	100	100	72.22	77.78	44.44	66.67	97.22	47.22	52.78
% Positive	97	8.33	100	2.78	72.22	16.67	5.56	97.22	25	27.78
Buyers										
Median	0.0252	-0.0220	0.0357	-0.0114	0.0052	-0.0002	-0.0270	0.0186	-0.0049	0.0013
% Signif.	97.22	100	100	86.11	75	55.56	61.11	91.67	47.22	50
% Positive	94.44	5.56	91.67	2.78	63.89	22.22	2.78	88.89	19.44	27.78

TABLE V

Aggressiveness: Ordered Probit Models - Performance.

This table summarizes the results of an in-sample goodness-of-fit analysis and an out-of-sample predictive-ability analysis of three alternative specifications of the ordered probit model [1]-[2] of order aggressiveness: the Baseline Model (BM) only includes the first lag of the dependent variable in the set of explanatory variables; the Best Quotes Model (BQM) adds variables computed from the best quotes of the book to the BM model; the Complete Book Model (CBM) adds variables computed using from the 2nd to the 5th best quotes of the book to the BQM. The variables computed using the best quotes are: the bid-ask spread (*SPR*), the depth on the same side of the market (DS^1), the depth on the opposite side of the market (DO^1), the number of orders on the same side of the market (NS^1), the number of orders on the opposite side of the market (NO^1). The variables computed using the additional 4 quotes available of the book are: the accumulated depth (number of orders) on the same side of the market DS^{25} (NS^{25}), the accumulated depth (number of orders) on the opposite side of the market DO^{25} (NO^{25}), the distance between the best and the second best (the second best and fifth best) quotes on the same side of the market LS^{12} (LS^{25}), the distance between the best and the second best (the second best and fifth best) quotes on the opposite side of the market LO^{12} (LO^{25}). An "M1" model includes depth measures but do not include number-of-orders measures as explanatory variables. An "M2" model includes number-of-orders measures but do not include depth measures. The table provides separated results for buyers and sellers. The in-sample analysis uses data from July to November 2000 and the out-of-sample analysis uses data from December 2000. The table reports the in-sample adjusted pseudo-R^2s of McFadden (1973), Maddala (1983) with the Cragg-Uhler (1970) correction, Aldrich-Nelson (1984) with the Veall-Zimmermann (1992) correction, and McKelvey-Zavoina (1975). The out-of-sample adjusted McKelvey-Zavoina pseudo-R^2 is also provided. Finally, the table provides: (1) the percentage of stocks for which a given model is the best against using relative frequencies to predict the aggressiveness of the incoming order (2) the percentage of stocks for which the CBM and the BQM outperform the BM on the basis that they usually allocate higher probabilities to the actual event than BM, and (3) the percentage of stocks for which the CBM outperforms the BQM on the basis of the same prediction rule.

	Ordered Probit Model			
In-sample Ajusted Pseudo-R^2s	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2
BM Model Pseudo-R^2				
McF	0.0011	0.0008	0.0011	0.0008
MadCU	0.0037	0.0028	0.0037	0.0028
AN	0.0047	0.0035	0.0047	0.0035
MZ	0.0039	0.0030	0.0039	0.0030
BQM Model (% increase over BM)				
McF	211.88	361.51	290.38	453.46
MadCU	210.49	358.43	288.52	444.94
AN	208.92	356.69	287.03	439.85
MZ	212.10	360.40	291.50	448.14
CBM Model (% increase over BQM)				
McF	39.10	78.17	11.95	24.62
MadCU	46.96	79.39	17.93	27.56
AN	48.33	78.32	18.63	28.64
MZ	47.70	83.94	18.52	26.66
Out-of-sample Adjusted Pseudo-R^2 (MZ)	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2
BM Model Pseudo-R^2	0.0040	0.0029	0.0076	0.0068
BQM Model (% increase over BM)	263.21	424.80	156.46	157.69
CBM Model (% increase over BQM)	51.79	47.24	37.36	70.80

Additional out-of-sample analysis (% stocks)	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM
Best model against the relative frequency	50.00	30.56	19.44	55.56	41.67	2.78	80.56	13.89	5.56	38.89	22.22	38.89
Better perf. than BM	100	100		100	100		97.22	97.22		97.22	97.22	
Better perf. than BQM	94.44			94.44			88.89			91.67		

TABLE VI

Aggressiveness: Sequential Ordered Probit Models - Estimates

This table summarizes the results of estimating a sequential ordered probit model with two steps. In the first step, the dependent variable has three levels of aggressiveness: cancellations (C1), limit orders (C2-C4) and market orders (C5-C7). In the second stage, the trader that has chosen to submit a limit order has to decide whether to place the limit order away from the best quotes (C2), at the best quotes (C3) or within the best quotes (C4); the trader that chooses to submit a market order has to decide the size of his/her order: less volume than available at the best quote on the opposite side of the book (C5), a market to limit order (C6) or more volume than available at the best quote on the opposite side of the book (C7). A positive estimated coefficient means that the associated explanatory variable is positively related to order aggressiveness. We report the median estimated coefficient for the 36 stocks in the sample, the percentage of statistically significant coefficients and the percentage of statistically significant and positive coefficients. We provide separated results for buyer initiated orders and seller initiated orders. We also provide separated results for a model with depth measures (model M1 -Panel A) and a model with number-of-orders measures instead (model M2 –Panel B). For a definition of the exogenous variables see Table V. All models include one lag of the dependent variable.

Panel A: Model M1

	Agr(-1)	*SPR*	DS^{1}	DO^{1}	DS^{25}	DO^{25}	LS^{12}	LO^{12}	LS^{25}	LO^{25}
Sellers	2nd Step - Liquidity Providers (Passive Traders)									
Median	0.0911	0.0195	6.35E-05	-1.09E-05	-7.74E-06	-7.66E-07	-0.0858	-0.0105	-0.0072	0.0096
% Signif.	94.44	91.67	94.44	25.00	86.11	22.22	94.44	16.67	63.89	13.89
% Positive	94.44	91.67	94.44	0	0	5.56	0	0	0	11.11
Buyers										
Median	0.0881	0.0170	7.44E-05	-9.60E-06	-6.59E-06	-3.13E-06	-0.0900	-0.0174	-0.0065	0.0070
% Signif.	97.22	91.67	94.44	38.89	77.78	30.56	97.22	27.78	55.56	30.56
% Positive	97.22	91.67	94.44	0	0	0	0	5.56	0	16.67
Sellers	2nd Step - Liquidity Consumers (Active Traders)									
Median	0.1288	0.0469	9.87E-06	-2.06E-04	-4.29E-07	6.42E-06	0.0132	-0.0194	0.0207	-0.0106
% Signif.	97.22	55.56	61.11	100	38.89	63.89	30.56	55.56	61.11	52.78
% Positive	97.22	55.56	55.56	0	13.89	61.11	27.78	2.78	58.33	8.33
Buyers										
Median	0.1304	0.0282	7.54E-06	-2.86E-04	-4.20E-06	4.74E-06	0.0190	-0.0275	0.0092	-0.0100
% Signif.	86.11	72.22	36.11	91.67	36.11	52.78	50.00	66.67	66.67	33.33
% Positive	86.11	72.22	33.33	0	5.56	50	44.44	0	58.33	2.78

Panel B: Model M2

	Agr(-1)	*SPR*	NS^{1}	NO^{1}	NS^{25}	NO^{25}	LS^{12}	LO^{12}	LS^{25}	LO^{25}
Sellers	2nd Step - Liquidity Providers (Passive Traders)									
Median	0.1017	0.0201	0.0733	0.0032	-0.0063	0.0059	-0.0838	-0.0074	-0.0081	0.0099
% Signif.	94.44	88.89	88.89	8.33	91.67	22.22	94.44	16.67	66.67	19.44
% Positive	94.44	88.89	88.89	8.33	0	22.22	0	2.78	2.78	13.89
Buyers										
Median	0.0966	0.0167	0.0934	-0.0015	-0.0076	0.0030	-0.0888	-0.0149	-0.0062	0.0105
% Signif.	97.22	91.67	97.22	5.56	94.44	33.33	97.22	30.56	52.78	30.56
% Positive	97.22	91.67	97.22	2.78	0	27.78	0	5.56	0	16.67
Sellers	2nd Step - Liquidity Consumers (Active Traders)									
Median	0.1144	0.0424	-0.0099	-0.2878	-0.0020	0.0078	0.0123	-0.0212	0.0178	-0.0106
% Signif.	97.22	66.67	27.78	100	58.33	63.89	36.11	63.89	66.67	55.56
% Positive	97.22	66.67	2.78	0	8.33	61.11	33.33	0	63.89	5.56
Buyers										
Median	0.1079	0.0304	-0.0306	-0.2401	-0.0058	0.0079	0.0220	-0.0269	0.0145	-0.0076
% Signif.	94.44	86.11	38.89	97.22	77.78	66.67	58.33	80.56	61.11	27.78
% Positive	91.67	86.11	5.56	0	5.56	63.89	52.78	0	58.33	2.78

TABLE VII

Aggressiveness: Sequential Ordered Probit Models – Performance

This table summarizes the results of an in-sample goodness-of-fit analysis and an out-of-sample predictive-ability analysis of three alternative specifications of a sequential ordered probit model of order aggressiveness with two steps. In the first step, the dependent variable has three levels of aggressiveness: cancellations, limit orders and market orders. In the second stage, the trader that has chosen to submit a limit order has to decide whether to place the order away from the best quotes, at the best quotes or within the best quotes; the trader that chooses to submit a market order has to decide the size of his/her order: less volume than available at the best quote on the opposite side of the book, a market to limit order or more volume than available at the best quote on the opposite side of the book. The three alternative specifications are: the Baseline Model (BM), which only includes the first lag of the dependent variable in the set of explanatory variables; the Best Quotes Model (BQM), which adds variables computed from the best quotes of the book to the BM model; the Complete Book Model (CBM), which adds variables computed using from the 2nd to the 5th best quotes of the book to the BQM. For a description of the explanatory variables see Table V. An "M1" model includes depth measures but do not include number-of-orders measures as explanatory variables. An "M2" model includes number-of-orders measures but do not include depth measures. Separated results are provided for buyers and sellers. The in-sample analysis uses data from July to November 2000 and the out-of-sample analysis uses data from December 2000. The table reports the in-sample adjusted pseudo-R^2s of McFadden (1973), Maddala (1983) with the Cragg-Uhler (1970) correction, Aldrich-Nelson (1984) with the Veall-Zimmermann (1992) correction, and McKelvey-Zavoina (1975). The out-of-sample adjusted McKelvey-Zavoina pseudo-R^2 is also provided. Finally, the table provides: (1) the percentage of stocks for which a given model is the best against using relative frequencies to predict the aggressiveness of the incoming order; (2) the percentage of stocks for which the CBM and the BQM outperform the BM on the basis that they usually allocate higher probabilities to the actual event than BM, and (3) the percentage of stocks for which the CBM outperforms the BQM on the basis of the same prediction rule. The results for the first step of the model are not reported but they are available upon request from the authors.

	Sequential Ordered Probit Model 2nd Step: Passive Traders (Liquidity Providers)				Sequential Ordered Probit Model 2nd Step: Active Traders (Liquidity Conumers)			
In-sample Ajusted Pseudo-R^2s	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2
BM Model Pseudo-R^2								
McF	0.0037	0.0030	0.0037	0.0031	0.0027	0.0022	0.0019	0.0022
MadCU	0.0092	0.0075	0.0092	0.0078	0.0049	0.0043	0.0033	0.0041
AN	0.0119	0.0096	0.0119	0.0100	0.0063	0.0055	0.0043	0.0054
MZ	0.0107	0.0085	0.0107	0.0087	0.0061	0.0056	0.0043	0.0053
BQM Model (% increase over BM)								
McF	212.26	251.51	131.04	177.32	1542.76	2745.48	1095.97	1005.21
MadCU	210.95	249.34	134.62	176.86	1481.25	2636.37	1104.73	976.90
AN	208.00	246.94	134.34	175.29	1429.09	2555.81	1088.45	956.08
MZ	225.64	280.88	143.11	197.91	10155.57	13550.25	3739.82	3128.25
CBM Model (% increase over BQM)								
McF	181.17	155.92	272.94	221.91	14.25	5.37	25.78	34.43
MadCU	173.01	151.98	259.05	213.73	13.37	4.35	24.14	34.78
AN	167.27	149.30	250.48	207.29	12.57	3.07	23.36	33.62
MZ	176.61	156.44	267.34	202.79	0.27	-6.13	1.85	13.54
Out-of-sample Adjusted Pseudo-R^2 (MZ)	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2
BM Model Pseudo-R^2	0.0107	0.0082	0.0107	0.0086	0.0054	0.0044	0.0054	0.0772
BQM Model (% increase over BM)	218.17	252.06	126.73	152.47	12689.56	13437.40	2097.78	1258.31
CBM Model (% increase over BQM)	275.91	213.41	334.57	325.53	0.36	0.71	7.82	21.88

Additional out-of-sample analysis (% stocks)	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM	BM	CBM	BQM
Best model against the relative frequency	97.22	2.78	0	86.11	13.89	0	91.67	8.33	0	88.89	11.11	0	2.78	11.11	86.11	8.33	2.78	88.89	38.89	22.22	38.89	27.78	19.44
Better perf. than BM	100	97.22		100	100		100	97.22		100	100		80.56	63.89		66.67	69.44		44.44	22.22		47.22	27.78
Better perf. than BQM	100			100			100			100			58.33			69.44			38.89			50	

TABLE VIII
Durations: ACD Models - Performance.

This table summarizes the results of an in-sample goodness-of-fit analysis and an out-of-sample predictive-ability analysis of three alternative specifications of the log-ACD(1,1) model [5]-[6] for three different durations: cancellations, limit orders and trades. The Baseline Model (BM) is the log-ACD(1,1) model; the Best Quotes Model (BQM) adds variables computed from the best quotes of the book to the BM model; the Complete Book Model (CBM) adds variables computed using from the 2nd to the 5th best quotes of the book to the BQM. For a definition of the explanatory variables see Table VII. An "M1" model includes depth measures but do not include number-of-orders measures as explanatory variables. An "M2" model includes number-of-orders measures but do not include depth measures. The table provides separated results for buyer-initiated and seller-initiated orders. The in-sample analysis uses data from July to November 2000 and the out-of-sample analysis uses data from December 2000. The table reports two adjusted pseudo-R^2s measures, the AIC and the SBC information criteria (see Appendix).

		In-sample goodness-of-fit				Out-of-sample goodness-of-fit			
		Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2	Sellers-M1	Buyers-M1	Sellers-M2	Buyers-M2
A. Cancellations									
BM Model:	Psd-R2(1)	0.0843	0.0848	0.0843	0.0848	0.0648	0.0809	0.0648	0.0809
	Psd-R2(2)	0.0923	0.0914	0.0923	0.0914	0.0736	0.0851	0.0736	0.0851
	AIC	5242.1	5961.5	5242.1	5961.5	1070.3	1157.9	1070.3	1157.9
	SBC	5259.4	5979.2	5259.4	5979.2	1087.9	1175.7	1087.9	1175.7
BQM Model:	Psd-R2(1)	2.3119	2.2433	0.9778	1.7113	0.0000	-1.0305	-0.3568	-0.4529
(% variation over BM)	Psd-R2(2)	1.1790	1.3091	0.8329	1.2957	-0.1331	-1.5525	-1.0643	-1.5549
	AIC	-0.2318	-0.2449	-0.1866	-0.1669	0.1094	0.1323	0.1189	0.1390
	SBC	-0.0686	-0.0798	-0.0149	-0.0371	0.8150	1.0322	0.9567	1.0684
CBM Model:	Psd-R2(1)	1.9413	1.3043	1.5303	0.9292	-1.0008	-3.1051	-0.9432	-2.8101
(% variation over BQM)	Psd-R2(2)	1.7204	1.0138	1.0871	1.0986	-3.7291	-2.9211	-3.2010	-2.2377
	AIC	-0.1298	-0.0899	-0.1093	-0.1161	0.6262	0.4918	0.6183	0.4757
	SBC	0.1967	0.1718	0.1996	0.2019	2.4392	2.1667	2.4947	2.2097
B. Limit orders									
BM Model:	Psd-R2(1)	0.1281	0.1253	0.1281	0.1253	0.1181	0.1096	0.1181	0.1096
	Psd-R2(2)	0.1293	0.1267	0.1293	0.1267	0.1202	0.1096	0.1202	0.1096
	AIC	13392.8	17750.8	13392.8	17750.8	2511.9	2913.9	2511.9	2913.9
	SBC	13412.3	17770.8	13412.3	17770.8	2531.0	2933.8	2531.0	2933.8
BQM Model:	Psd-R2(1)	1.4969	0.4501	1.6432	0.2802	0.8004	-0.0078	1.1383	0.1824
(% variation over BM)	Psd-R2(2)	1.2574	0.5024	1.5463	0.2836	0.7141	0.0132	0.9099	0.0711
	AIC	-0.2698	-0.1784	-0.2513	-0.1735	0.0124	-0.0379	-0.0505	-0.0406
	SBC	-0.1652	-0.0816	-0.1509	-0.0810	0.3039	0.2877	0.2733	0.2388
CBM Model:	Psd-R2(1)	1.4611	0.6170	1.5529	0.6671	-1.4848	-1.4341	-1.6028	-1.7654
(% variation over BQM)	Psd-R2(2)	1.5139	0.6319	1.5035	0.7672	-1.5402	-1.5913	-1.3685	-1.1511
	AIC	-0.1943	-0.0848	-0.1750	-0.1090	0.2713	0.2482	0.2463	0.1653
	SBC	-0.0166	0.0294	-0.0010	-0.0034	1.1163	1.1564	1.0765	1.0054
C. Market orders (trades)									
BM Model:	Psd-R2(1)	0.1105	0.0978	0.1105	0.0978	0.1029	0.0752	0.1029	0.0752
	Psd-R2(2)	0.1126	0.0987	0.1126	0.0987	0.1093	0.0781	0.1093	0.0781
	AIC	14192.9	19548.8	14192.9	19548.8	2358.5	2993.0	2358.5	2993.0
	SBC	14212.4	19569.1	14212.4	19569.1	2378.1	3012.9	2378.1	3012.9
BQM Model:	Psd-R2(1)	2.1170	0.8118	2.1260	1.0303	0.2785	0.3301	0.3060	0.5361
(% variation over BM)	Psd-R2(2)	1.8801	0.8607	1.9892	0.8954	0.7721	0.8032	0.7087	0.6423
	AIC	-0.2314	-0.1184	-0.2195	-0.1068	-0.0263	-0.0076	-0.0311	-0.0227
	SBC	-0.1403	-0.0590	-0.1388	-0.0342	0.3165	0.2915	0.2860	0.2593
CBM Model:	Psd-R2(1)	0.6861	0.7563	0.8798	0.7840	-0.9634	-0.4859	-0.4908	0.0917
(% variation over BQM)	Psd-R2(2)	0.6249	0.7157	0.7476	0.5919	-0.7707	-0.6589	-0.8263	-0.4101
	AIC	-0.0577	-0.0636	-0.1035	-0.0705	0.2694	0.1682	0.1790	0.1493
	SBC	0.0576	0.0472	0.0541	0.0313	1.1946	0.8733	1.1465	0.8749

APPENDIX

Diagnosis Measures

This table contains the computational details of the goodness-of-fit measures used in this paper. Ordered probit models: l_M is the log-likelihood value of the unrestricted model; l_0 is the log-likelihood function of the restricted model (the coefficients of all the exogenous variables equal to zero); l_{MAX} is the maximum possible likelihood (i.e., perfect fit); $\hat{y}_i^* = x_i'\beta$, evaluated at maximum likelihood estimates of the model; $\bar{y}^*$ is the sample average of $\hat{y}_i^*$. ACD models: $\hat{\sigma}_d^2$ is the sample variance of the duration process; $corr^2(d, \exp(\psi))$ is the squared sample correlation coefficient between the actual and fitted values. The Akaike (AIC) and Schwarz Bayesian (SBC) information criteria are based on the residuals of the model ($\hat{\varepsilon}_i$). Finally, k is the number of parameters and N is the sample size. All the pseudo-R^2 measures are adjusted in the standard way: Adj-R2=1-((N-1)/(N-k))(1-R^2).

Measure	Additional definitions	Reference
Ordered Probit Models		
$R^2_{VZ} = \dfrac{2(l_M - l_0)}{2(l_M - l_0) + N} \Big/ \dfrac{-2l_0}{N - 2l_0}$		Aldrich and Nelson (1984) and Veall and Zimmermann (1992)
$R^2_{McF} = 1 - \dfrac{l_M}{l_0}$		McFadden (1973)
$R^2_{CU} = \dfrac{1 - (l_0/l_M)^{2/N}}{1 - (l_0/l_M)^{2/N}}$		Maddala (1983) and Cragg and Uhler (1970)
$R^2_{MZ} = \dfrac{\sum_{i=1}^{N} (\hat{y}_i^* - \bar{y}^*)^2}{\sum_{i=1}^{N} (\hat{y}_i^* - \bar{y}^*) + N}$		McKelvey and Zavoina (1975)
ACD Models		
$Adj - \bar{R}^2(1) = 1 - \dfrac{N-1}{N-K} \dfrac{MSE}{\hat{\sigma}_d^2}$	$MSE = N^{-1} \sum_{i=!}^{N} (d_i - \exp(\psi_i))^2$	Verbeek (2000)
$Adj. - \bar{R}^2(2) = 1 - \dfrac{N-1}{N-k} corr^2(d, \exp(\psi))$		Verbeek (2000)
$AIC = \log \hat{\sigma} + 2\dfrac{k}{N}$ $SBC = \log \hat{\sigma} + 2\dfrac{k}{N} \log N$	$\hat{\sigma}^2 = Var(\hat{\varepsilon}_i)$ $\hat{\varepsilon}_i = d_i / \hat{\Psi}_i$ $\hat{\Psi}_i = \exp(\bar{\psi}_i) = \exp(\varpi + \alpha \ln d_{i-1} + \beta \bar{\psi}_i)$	Gourieroux and Monfort (1995)