

Screen Information, Trader Activity, and Bid-Ask Spreads in a Limit Order Market

Mark Coppejans
Department of Economics
Duke University
Durham, NC 27708
mtc@econ.duke.edu

Ian Domowitz
Department of Finance
Smeal College of Business Administration
Pennsylvania State University
University Park, PA 16802
domowitz@psu.edu

August 1999

We are grateful to Lester Loops for making the data available and providing useful insights with respect to the market structure examined here. Some help with the database was provided by Hao Zhou. We thank Bruno Biais, Burton Hollifield, Bruce Lehman, Steven Slezak, and Chester Spatt for helpful discussion and comments, as well as participants of the Paris 1998 Microstructure and High Frequency Data Conference and the 1999 Western Finance Association Meetings.

Abstract

Screen Information, Trader Activity, and Bid-Ask Spreads in a Limit Order Market

A key focus of empirical work on limit order markets is the relative importance of individual pieces of information in characterizing order submission and trade execution. We enlarge this focus to include an examination of pricing behavior, using data on index futures trading in a pure electronic limit order book market. A theoretical link between order, trade, and cancellation arrival rates, and the distribution of bid-ask spreads is empirically implemented. Evaluation of models across different information sets is based on relative ability to predict market activity and pricing out-of-sample. A main finding of the paper is the importance and superiority of information embodied in continuous individual traders' actions in characterizing order submission behavior and the structure of pricing. The book information on characteristics of resting orders alone cannot explain subsequent order submission, trade, or pricing behavior, and has little impact on the shape of the spread distribution.

1. Introduction

A clear trend in securities exchange design is towards the introduction of electronic limit order book systems for equities, derivatives, and bonds. This development has generated a growing literature on order flow and its contribution to market activity and price formation.¹ We analyze the links between information observed through the system by traders, order placement behavior, and the probability structure of the bid-ask spread in this paper. We are particularly interested in whether information on the characteristics of resting orders on the book influences trader behavior and pricing, over and above information available through the flow of trader activity. The question is relevant in debates over disclosure of order information in markets more generally.² The investigation here proceeds in three steps, using intraday data on stock index futures trading in a pure limit order book market.

We first exploit advances in conditional duration analysis for the modeling of trader behavior. Duration models are estimated for the arrival of bids, offers, transactions, and order cancellations. Market activity studies, such as that by Biais, Hillion, and Spatt (1995), generally consist of an examination of a large number of individual pieces of information garnered from order book trading. In a multivariate model-based setting, we instead aggregate such pieces into interpretable blocks. The order book represents a stock of price, volume, and other liquidity information, augmented by the flow of various actions by individual traders. Our fundamental

¹ Examples include Biais, Hillion, and Spatt (1995), Foucalt (1993), Glosten (1994), and Hollifield, Miller, and Sandas (1996). Recent developments in electronic limit order book markets and references to other work on the institution are contained in Domowitz and Steil (1999).

² The New York Stock Exchange has debated the merits of permitting complete disclosure of the specialist's book, for example, a proposal opposed by the specialists themselves. See Rickard and Lupien (1996) for discussion of the merits of new opaque electronic systems.

differentiation is between these stocks and flows. Engle and Russell (1998) find that information in the form of durations between market events also helps to explain the evolution of transactions behavior. Our models are therefore differentiated by four information sets conditioning the price processes. Abstracting from time-of-day effects, the information sets include (i) lagged event durations and autonomous dynamics in the duration process; (ii) durations augmented by observed order book information pertaining to available pricing and liquidity; (iii) durations supplemented by information on traders' activity, as distinguished from the state of the order book; and (iv) models combining elements (i)-(iii). Models of market activity also are benchmarked to a no-information case, represented by Poisson arrivals of events.

In the second step, we establish a link between data observed by traders, the evolution of the market, and pricing, through a framework that connects activity in terms of event arrivals to the probability distribution of the bid-ask spread. In a model of the mechanics of the limit order book, the stock of book information and market events influence the arrival rates of orders, trades, and cancellations at various prices. Through the mechanics of the market structure, these conditional arrival rates determine the probability of observing particular bid-ask spreads. Complementing the new literature on duration modeling, the technique provides a structural interpretation of durations of trading activity in terms of pricing.³

Using the probability model, information sets nested within the alternative duration models are compared against each other by their usefulness in describing pricing behavior. We present summary statistics with respect to the goodness-of-fit of

³ See also Russell and Engle (1998), who present a statistical model of trade durations and price changes, without a structural modeling interpretation and with no link to limit order book activity. We share an advantage with their technique, namely a statistical structure capable of handling discrete irregularly spaced data. Gouriéroux, Le Fol, and Meyer (1996) present a

the spread distributions predicted by the models. To the extent that the model probabilities match the data, the exercise also reinforces the hypothesized links between duration of order placement, transactions, and cancellations, and the probability structure of the order book.

The third step consists of an examination of changes in the spread distribution as market circumstances change, as exhibited by book information and the flow of trader activity. The goal is to determine what changes in information most affect the shape of the distribution, thereby influencing average spreads and the volatility of trading costs represented by the spread. Percentage changes in the spread probabilities are calculated, given changes in each piece of information embodied in the duration and probability models. An aggregate measure of the impact of changes within different information sets is introduced, and supplemented by data on movements of probability mass for specific points of support.

A traditional way to evaluate observable information in duration and spread models is to use inference based on in-sample asymptotic tests of statistical significance for individual variables or groups thereof. This practice presents interpretation problems across the 30 models estimated here. Instead, we use out-of-sample evaluation techniques to judge relative adequacy in describing behavior. Out-of-sample criteria mitigate not only problems with "data-snooping" (e.g., Lo and MacKinlay (1990)), but also those associated with potential overfitting and testing of models within sample.⁴ An analysis of out-of-sample forecasts of trade and order

factor model of the probabilities of bid and ask prices based on a queuing model of the book, in which the spread influences subsequent bids and offers.

⁴ Such an approach has been advocated more generally. See, for example, Granger, King, and White (1995), who note the difficulties in using standard in-sample hypothesis tests for choosing between competing models, and recommend out-of-sample criteria.

placement durations, and the probability distribution of the bid-ask spread, permits us to distinguish between information sets in terms of characterizing market activity.

We also minimize biases by splitting the data into three parts. All “experiments” are performed on an initial three week period of trading data. This includes specification of summary statistics reported, selection of information sets for the conditional duration analysis, and the functional form of models used for the latter. Estimation results reported in the paper are based on nine weeks of trading data, with a period of four weeks separating the “trial” period from the estimation period. The third data period consists of another three weeks of activity, chronologically separated from the estimation period by two weeks. We refer to this data as the “out-of-sample” period. The duration models, probability specification analysis, and comparison of information sets are evaluated based on out-of-sample calculations. The chronological difference between the trial period and the data used for model evaluation is then 15 weeks, the maximum that could be feasibly achieved with our data set.

The data are obtained from trading activity in stock index futures contracts from the consolidated limit order book operated by OM Stockholm and the London Securities and Derivatives Exchange. The data include a complete record of transactions activity and the five best bids and offers on the book, with volume available at each price. Data also are available from an integrated odd-lot book and for block trade crossing activity. In addition to being one of the few complete limit order book datasets in existence, our data have two advantages.⁵ First, the use of the future on a broad stock index mitigates asymmetric information problems relative to

⁵ The most commonly used is the Paris Bourse data, described by Biais, Hillion, and Spatt (1995). Hollifield, Miller, and Sandas (1996) also use OM data, for a selection of 10 stocks

trading in individual equities.⁶ This is particularly useful in an analysis of the importance of screen information, which is available to all traders. Second, the examination of trading activity on the index eliminates the need to average results across individual equities, as has been done elsewhere. We deal with only a single order book, as opposed to, say, the 40 on the Paris Bourse system.

We find that trading activity consisting of submissions to the book (bids, offers, and cancellations) substantively depends only on the flow information. The state of the order book, the stock information, cannot predict the evolution of the market. Any correlation between lagged stocks and flows provides negligible improvement in describing limit order book activity. This does not mean that the order book is irrelevant; except for transactions, flow information is garnered from changes in book data. It does imply, however, that liquidity information on the book is secondary to trading flows in determining order placement.

The results on the distribution of the bid-ask spread are consistent with those obtained from the duration models. The usefulness of the flow information in describing arrival activity translates directly into superior pricing forecasts. This provides some support for the hypothesized link between arrival rates and price distributions. There is little gain in the ability to characterize pricing structure from the addition of stock information to that provided by simple lagged durations of activity. A full information model, exploiting the correlations between stock and flow data, exhibits no improvement in descriptions of pricing, relative to conditioning only on the flow of individual trader actions.

traded on the equities order book. Data also are available for trading on the Australian SEATS automated system.

⁶ We do not observe customer order flow to dealers operating on the system. In index futures trading, the existence of private information is most likely in such order flow, similar to the interbank foreign exchange market.

We also find that the probability distribution of the spread is little affected by changes in book information. In contrast, changes in trader activity shift the distribution considerably, especially with respect to the aggressiveness of order submission.

We describe the data and institutional form of the trading structure in the next section. Summary statistics relevant to the modeling exercise to follow also are presented therein. Section 3 is devoted to the specification and analysis of duration models differentiated by information sets. We delineate the link between durations and the probability structure of bids and offers on the limit order book in Section 4. Estimates of such probabilities are computed and compared to the empirical distribution of bid-ask spreads out-of-sample, differentiating between information sets. A comparison of the relative impact of different information sets on the shape of the spread distribution is undertaken in Section 5. Some concluding remarks are contained in Section 6.

2. Trading Institutions and Data

2.1 The Market Institution

Trading in stock index futures contracts was launched by OM Stockholm (OMX) in 1985. OMLX, the London Securities and Derivatives Exchange, was established in 1989. We refer to the overall market as OMX, since the order book and trading activity are integrated.⁷ Trading in Swedish index futures is based on a consolidated book that includes activity from Sweden, the U.K., Denmark, and the Netherlands. The trading day is six hours, beginning at 9:00 a.m. and ending at 3:00 p.m. London local time.

⁷ Clearing is conducted on a local basis.

The basic trading institution is that of a continuous “pure” limit order market. Trading is anonymous. Orders are prioritized for execution in terms of price and time. Trades are executed automatically, under one of two circumstances. Counterparty prices may match on the book, or a trader may “hit the bid” or “lift the offer,” by executing a single keystroke and submitting desired volume.⁸ After a trade, unexecuted volume at the trade price remains on the order book, unless explicitly cancelled. Cancellation of orders may be done at any time. There is no opening algorithm or batch auction at the beginning of the trading day.

Trading on the order book is in round lots of 10 contracts. There is a facility for a small amount of odd-lot trading, but it is integrated with the main book. For example, two odd lots of 5 contracts will automatically match with a round lot of 10. Block trades are allowed, in the form of “crosses.” They appear as separate transactions, but the two sides are not listed on the order book. There is no interference with a cross from activity on the limit order market, unlike the practice on, say, the automated Swiss SOFFEX derivatives system.

Price and volume information are distributed directly from the trading system. The data include a record of transactions activity and the five best bids and offers on the system, with aggregate volume available at each price. No “indicative” prices are distributed. A trader may choose to view the information through the OM Click interface, or to accept a real-time data feed allowing customized screens, data summaries, and direct processing for any trading strategies chosen.

2.2 *The Data*

⁸ If an entered bid price is higher than an offer price on the book, there is no “locked” market. A transaction occurs based on time priority, at the offer price in this example.

Our data are obtained from a trading house that chose the real-time feed option, which permits the archiving of data for analysis.⁹ The data are restricted to trading of the OMX futures contract on the Swedish stock index. Prices are denominated in Swedish currency (SEK), and volume is given in number of contracts. Information is time-stamped to the second.

Transactions files and separate records containing order book information are matched. The order book itself is reconstructed from the raw data and fully consistent with transactions reported.¹⁰ Odd-lot trades are few, but identifiable. Similarly, cross trades are broken out, and matched in terms of time with all limit book trading activity.

The complete span of the data set runs from 7/31/95 through 2/23/96. We choose periods in which only a single contract (i.e., a single expiration) is traded, in order to eliminate liquidity effects stemming from lack of trading at the end of expiration cycles, due to rollover effects.¹¹ The trial period is chosen to be 7/31/95 through 8/18/95. The estimation period is from 9/25/95 through 12/15/95, representing trading in three contracts differentiated by expiration, but with 10 day gaps to account for any rollover activity.¹² Finally, the out-of-sample period is 1/29/96 through 2/16/96. This choice eliminates an entire contract cycle before the evaluation period. Sample sizes during the estimation period include 30,866 orders, 2876 trades, and 21,070 cancellations of orders.

⁹ We are grateful to Lester Loops, who not only provided the raw numbers, but also helped with data problems and commented on “traders’ rationale” for some regularities observed in our own analysis.

¹⁰ Some irregularities, about one percent of activity, initially were uncovered, but all were reconciled with the assistance of the trading house providing the data.

¹¹ This procedure also eliminates the need for averaging activity over different books devoted to different contract cycles, with potentially different liquidity characteristics, in this case depending on expiration cycle and rollovers.

¹² The gap period was originally suggested by conversations with traders in this market, but a check of the data firmly confirms their intuition as well.

2.3 *Prices, Order Placement, and Transaction Activity*

The average bid-ask spread in the market is between 7 and 8 ticks and the average percentage spread is around 0.14 percent.¹³ About 70 percent of spreads are observed in the range of 4 to 8 ticks. Some tick multiples are clearly too small relative to price, while very large spreads are simply uncommon, suggesting good liquidity provision. The order book is empty on the bid and offer side only for 60 and 66 seconds a day, respectively.

Table I contains daily averages of activity.¹⁴ There are 1,856 orders and 178 trades per day. The order cancellation rate is 67 percent. The total numbers of orders, trades, and cancellations per day are split virtually evenly between bid and offer activity. This symmetry in orders and cancellations also is reflected by statistics that disaggregate the activity into activity at the best quote, worse than the quote, or better than the best price on the book.

Incoming bids and offers at the current best quote account for 8.2 percent of orders. An order betters the existing best price over 18 percent of the time. Although bids and offers below the market are dominant, the percentage of such orders is the same as the percentage of cancellations of orders below the quote (73.1 percent versus 72.9 percent). Duration analysis suggests that such orders below the best price are killed off very quickly, relative to order arrival rates. Large volume transactions account for 79 percent of all trades based on limit orders, and there also are an additional 19 crosses per day at larger volumes. Odd-lots account for only about 3 percent of trades, averaging less than 5 contracts per trade.

¹³ There is little discreteness in prices. The tick size, relative to transactions price, is only 0.02 percent, an order of magnitude below that observed in many markets.

¹⁴ Figures in all summary tables are computed based on the data used for model estimation in the next section. Numbers from the trial and out-of-sample periods are quite similar and not reported.

Table II documents a small addition to the large body of evidence on U-shaped trading activity across time of day. The empirical regularity is striking with respect to the symmetry of the effect across the day and across trading behavior disaggregated by orders, transactions of different sizes, cancellations, and crossing activity. Beyond noting the desirability of including time of day effects in model estimation, we have little else to add, except to note that U-shaped activity does not seem to be a figment of market structure, *per se*.

Selected average trade and order information across the sample is given in Table III. Average prices are quite similar across all forms of activity, including orders, trades, and cancellations. There is a 0.35 percent discount on crosses, but this discount is not statistically different from zero. Percentage returns over this period are slightly negative, averaging less than 1 percent in absolute value. There is no identifiable trend in price levels.

Average trade and order sizes are just below 20 contracts, i.e., 2 round lots. Although orders, trades, and cancellations on the offer side are larger than on the bid, the differences are small and not statistically significantly different from zero. The averages conceal a great deal of variation. The standard deviation of traded volume is 10.6, compared to a mean of 16.8, for example. Most variation is related to the prevalence of large trades, although transactions of single round lots are common.

Average duration information complements the activity reported in Table I. The total of 1856 orders per day translates into a time between orders of just over 11 seconds.¹⁵ There is no statistical difference between bid and offer arrival rates. Trades occur once every 2 minutes, on average. The time between cancellations is 48 percent higher than observed for orders, in part reflecting intervening transaction

activity. There is no duration asymmetry between cancellation activity at the bid and at the offer. Crosses and odd-lot trades are relatively rare, occurring once every 17 and 44 minutes, respectively.

The models in section 4 relate the distribution of bid-ask spreads to order and trade arrival rates. We check the correlation between market activity and spreads using a simple Poisson formulation incorporating time-of-day effects. There is a statistically significant negative relationship between the spread and order, cancellation, and trade durations, supporting the analysis to follow.

Since there is no opening algorithm in the market, it might be suspected that the first hour of trading might be somewhat special. All statistics are recomputed based on the first hour and the remainder of the day. They differ little from those reported here, except that overall activity is slightly higher in the first hour, consistent with the time of day statistics in Table II.

3. Information Sets and Associated Conditional Duration Models

There are many “events” that might be used to characterize behavior. We restrict the analysis to the arrival of bids, offers, trades at the bid, trades at the offer, cancellations at the bid, and cancellations at the offer. These might be considered the most obvious events of interest in any case, but are chosen because they comprise inputs to the analysis of bid-ask spreads which follows.

Similarly, there are many “information sets” that could be examined. The choice of a discrete information set is just one way of cutting the data. There is no unique optimal way to do so, and the issue is interpretation. We concentrate on cases of no information, that obtained through autonomous dynamics of the duration

¹⁵ As one basis for comparison, quote arrivals in the foreign exchange data studied by Engle and Russell (1995) are every 15 seconds.

process, book information, and information available through traders' actions. The information sets are described and interpreted in Section 3.2 below.

There are also many ways in which to model conditional duration.¹⁶ We use the autoregressive conditional duration (ACD) model of Engle and Russell (1998). We are more interested in the composition of information relevant to trading activity and spreads than in the precise specification of the autonomous dynamics of duration. The ACD model provides one way to account for the latter, and there is no specification search over the extremely wide range of alternatives in this respect. Within the ACD class, there is a variety of distributional and functional form assumptions possible. We clarify those used here in Section 3.1.

Even with all these restrictions, 30 models are estimated and analyzed. We begin with a brief description of the statistical framework upon which the results are conditioned.

3.1 *The Statistical Model*

In the ACD model, the arrival times of the data are treated as continuous time-varying stochastic point processes, an alternative to fixed interval discrete time analysis. The model is derived and justified in several papers.¹⁷ We describe only the parametric form and pseudo-likelihood function used in the estimation.

For any given set of events, e.g., bid arrivals, the underlying stochastic process is manifested by a sequence of times $\{t_1, t_2, \dots, t_n, \dots\}$. The elements are arrival times of the point process. Corresponding to the arrival times is a counting process, $N(t)$, which is the number of events which have occurred by time t . The ACD model is a

¹⁶ This literature goes back at least to Cox and Lewis (1966). For more recent treatments, see for example Engle (1996), Gouriéroux, Jasiak, and Le Fol (1996), Ghysels, Gouriéroux, and Jasiak (1998), and references therein.

description of such point processes with intertemporally correlated events, for some selected probability density for event i occurring at time t conditioned on past arrival times.

Let the i^{th} duration between events be defined as $x_i = t_i - t_{i-1}$. Time dependence within the ACD framework is summarized by a function \mathbf{y} , which is the expected duration conditioned on past information. The functional form used in this paper corresponds to the ACD (1,1) model of Engle and Russell (1998). The formulation is analogous to GARCH (1,1) models, and shares many of the same basic properties. The conditional expectation is written as

$$\mathbf{y}_i = \mathbf{w} + \mathbf{a}x_{i-1} + \mathbf{b}\mathbf{y}_{i-1} + \exp(\mathbf{g}'z_{i-1}) \quad (1)$$

for all $i = 1, \dots, N$. The composition of the vector of variables z varies with the specification of information sets. Regularity conditions for ACD models require positivity with respect to the determinants of durations. This is particularly important in our case, since we analyze out-of-sample forecasts. The restriction is imposed on the z 's by taking an exponential transformation. The case of “no information” is modeled by setting all coefficients save the constant term to zero, i.e., constant event arrival rates.

The form of the underlying density of arrivals is taken to be exponential. The possibilities are legion, and we do not claim an exact distributional assumption. As in Engle (1996), the exponential form is used simply to compute quasi-maximum likelihood estimates (QMLE) of the parameters, a common practice in other contexts. The pseudo-log-likelihood is given as

¹⁷ The original presentation is Engle and Russell (1998). Summaries at alternative levels of probabilistic rigor appear in Engle and Russell (1995), which also contains applications to the foreign exchange market, and Engle (1996).

$$\sum_{i=1}^{N(T)} -\ln(\mathbf{y}_i) - \left(\frac{x_i}{\mathbf{y}_i} \right). \quad (2)$$

All parameter estimates are based on maximization of this statistical criterion.

Conditional on the functional form for the conditional duration, Theorem 1 of Engle (1996) ensures estimated parameter consistency and asymptotic normality under weak conditions. The form of the robust QMLE variance-covariance estimator for the coefficients described therein is used here.

3.2 *Information Sets*

We consider four types of information, which are nested in different ways across models. An additional benchmark is provided by the case of no information. The data comprising each set is summarized in Table IV. All data are directly observable by traders through the trading system.¹⁸

In the context of duration models, the case of no information translates simply into Poisson arrivals, for which arrival rates are constant. The Poisson model is provided only as a benchmark, but is a special case of the ACD formulation. Information in the ACD model includes lagged and expected durations. The model does not necessarily presume that traders ignore all other observables, rather that other information is completely summarized by expected and lagged duration in activity.

The ACD model is nested within the other three information sets, in the sense that all include lagged duration and expected duration. We call the third information set “state of the order book” or simply the book. Book information consists of resting order volumes disaggregated by bid and offer, the “length” of the book on bid and

¹⁸ An exception may appear to be expected conditional duration, i.e., the autonomous dynamic. This may be computed in real time from observables, however, in the same way as some real

offer sides, measured as the number of prices on each side of the book, whether or not the order book is empty, and a small spread indicator. Volume, lack of resting orders, and spreads have been used by Biais, Hillion, and Spatt (1995), for example, as liquidity characteristics. A theoretical link between the number of prices on the limit order book and execution risk is given in Domowitz and Wang (1994). Conversations with limit order traders also suggest that a decision to place an order depends on the perceived ability to exit the trade within short periods of time. They have some notion of reservation prices for exit, and look at the range of prices, as well as volumes, available at the time of order submission. The combination yields information on desirability of submission at some price.¹⁹

The state of the book is not the sole purveyor of liquidity information. Arrival rates and other actions by traders also contribute. The dichotomy between book and trader actions is naturally interpreted as stock versus flow. The book represents a stock of price, volume, and other liquidity information (empty book and spread), augmented by the flow of trader actions.

We construct the fourth information set to represent this flow in terms of observable characteristics. Transaction information is disaggregated by side of market (trade initiation), and further by size of transaction, discretized into large and small trades. Following Biais, Hillion, and Spatt (1995), large trades are those that remove all volume at the best quote on the opposite side of the book. New orders and cancellation activity are split between bids and offers. Behavior may differ depending on the aggressiveness of orders, also as documented by Biais, Hillion, and Spatt

time options trading systems calculate recursive estimates of GARCH processes, stochastic volatility, or volatility smiles.

¹⁹ We examined this intuition, using statistics from the trial sample. Durations between orders depend greatly on whether one observes combinations of large volume/large length, small volume/large length, large volume/small length, and small volume/small length.

(1995) for the Paris Bourse. We split bid and offer flows according to whether orders are submitted at the best quote, worse than the quote, or improving the quote, as separate variables. Similarly, cancellations are disaggregated, depending on whether orders are cancelled at prices equal to the best quote or lower down in the book hierarchy by price. The trial data suggest that odd lot trading has no effect on flows of market activity. Cross trades at larger volumes are included, however.

The state of book and trader action information sets are not nested. We therefore estimate a fifth model that includes a combination of book and trader data. There are other ways of combining the data described above, and several were investigated using the trial sample, based on simple correlations with durations. On the other hand, no included variables were dropped from the formal duration models on grounds of “statistical significance.”

3.3 Comparisons Across Information Sets

The models are estimated using the functional forms and QMLE techniques described in section 3.1. The probability model of the next section requires that arrival rates of orders, cancellations, and trades be conditioned on the price at which they occur, measured as the number of ticks away from the best offer price in the market at the time of submission. For example, if the best bid is 1350, and an offer arrives at a price of 1425, this measure would be 3 ticks. These variables are included in all models, but coefficient estimates are not reported in Tables V through VIII, which contain other estimates for all 30 models. Time of day effects in the form of hourly indicator variables are included, but also are not reported in the tables.

Out-of-sample summary statistics for forecasts of durations across information sets are contained in Table IX. We include the root mean-square error of forecasts and an out-of-sample pseudo R^2 . The R^2 is constructed by subtracting the ratio of

forecast mean square error to the variance of duration from one. Comparisons below are with respect to the out-of-sample R^2 , the most natural measure of relative model performance.

Lagged duration information vastly improves predictions of all activity relative to the case of constant arrival rates. The out-of-sample fit improves using the ACD formulation by almost 500 percent for bids.²⁰ Trade duration exhibits improvements of 388 percent on the bid side and 127 percent on the offer. Bid cancellation predictions are 966 percent better using the ACD model.²¹

The out-of-sample predictive ability of simple order and cancellation duration information is dominated by use of data on the flow of traders' actions. Order and cancellation duration predictions are 266 to 600 percent better using the larger flow information set.

Trading activity consisting of arrivals of bids, offers, and cancellations substantially depends only on the flow of individual traders' actions. The use of stock information embodied in the state of the order book does little to improve predictive content with respect to subsequent activity. The percentage improvement of trader flow information over book information in describing behavior ranges from 266 to 700 percent.

Correlation between lagged stock and flow information in the full information model provides negligible improvement in predicting market activity in the form of submissions to the book. Addition of book information to trader actions improves predictability by only 0 to 5 percent.

²⁰ The out-of-sample pseudo R^2 is negative for Poisson offer and offer cancellation arrivals, indicating that the model introduces extraneous noise relative to the variance of the process.

²¹ These results reinforce those in other studies comparing conditional duration models with constant arrival rates, based on in-sample statistics and measures of overdispersion. See, for example, Engle and Russell (1998).

The introduction of book and/or trader information generally fails to improve trade duration forecasts relative to the ACD formulation. The decision to trade immediately, as opposed to submitting a limit order, is implemented through hitting the bid or lifting the offer. This is somewhat analogous to submitting a market order, and may not depend on observable market factors directly available through the system, besides price. Alternatively, only a small subset of non-price market information is relevant, while the remainder introduces extraneous noise. If one relies on the in-sample precision and size of coefficient estimates in the trade models, there is some support for this hypothesis. In terms of stock information, the arrival of trades is conditioned by liquidity in terms of volume and depth of prices on the book. With respect to flow information, previous trades on the same side increase new trade activity, as do cancellations away from the best quotes. Finally, it may be that all information is adequately summarized for the purposes of trades by lagged duration. Lagged expected duration plays no role once extra information is included, but the coefficient on expected duration is large in the pure ACD formulations for trade duration.

4. The Probability Structure of the Bid-Ask Spread

We now consider the contribution of screen information to the characterization of limit order market activity, in terms of pricing. We begin with the link between durations and the probability structure of bid-ask spreads.

4.1. Order Arrival Rates and Price Distributions

Conditional duration analysis is simply another means of describing event arrival rates. The links between expected duration and intensity functions describing arrival rates are explicated in Cox and Lewis (1966), and for ACD models in particular by Engle and Russell (1998). Domowitz and Wang (1994) present a model in which

various stationary distributions pertaining to limit order book trading are derived from conditional intensity functions describing, order, trade, and cancellation activity. In this section, we summarize a subset of results, specializing the analysis to the distribution of the bid-ask spread.²²

The set of possible limit prices is denoted by an increasing sequence, $\{p_k\}$, in which each price is separated by one tick. Let A^m and B^m be the best offer and best bid prices on the book. The bid-ask spread is $S = A^m - B^m$. The theory in Domowitz and Wang (1994) is phrased in terms of conditional arrival rates over all possible prices, conditional on the best price on the opposite side of the market. As a practical matter, given the small tick size in this market, there are too many prices to maintain such disaggregation. We impose a symmetry condition here, in order to make the empirical analysis tractable.

Let $s = |k - j|$, the difference in ticks between two prices p_k and p_j . Arrival rates conditioned on price are functions of s and some information set Ω . The arrival rate of offers at a price s ticks away from the best bid price is $I^a(s, \Omega)$. The arrival rate of bids at a price s ticks away from the best offer price is $I^b(s, \Omega)$. The same interpretation applies to arrivals of trades on the offer and bid side, $g^a(s, \Omega)$ and $g^b(s, \Omega)$, respectively. Finally, $m^f(s, \Omega)$ denotes the arrival rate of a cancellation of an offer at a price s ticks away from the best bid, and similarly on the bid side, $m^b(s, \Omega)$.

A specialization of Theorem 1 of Domowitz and Wang (1994) then yields the implied probability distribution of the bid-ask spread, given by

²² Elaboration of notation required for regularity conditions, and statements of the latter, are omitted. In particular, some stationarity in the conditioning sets must be assumed, as well as conditions pertaining to the relative inflows and outflows of orders, covered in Domowitz and Wang (1994).

$$\Pr(S = s) = \frac{I^b(s, \Omega)[1 - d^b(s, \Omega)]}{g^a(s, \Omega)} \quad (3)$$

over all points of support given by the range of values of s . The quantity $d^b(s, \Omega)$ relates to cancellation rates, given by

$$d^b(s, \Omega) = \frac{m^b(s, \Omega)E(Y_s)}{I^b(s, \Omega)} \quad (4)$$

where Y_s is the number of bids on the book at a price s ticks away from the best offer.²³

4.2 Comparisons of Implied Distributions Across Information Sets

The out-of-sample duration predictions are the primary inputs to the spread probability model. The models of Section 3 are inverted to obtain corresponding conditional average arrival rates of orders, trades, and cancellations, differentiated by information set. There is no theoretical guidance for the modeling of Y_s . A linear function of s and observed volume is used in the results reported here. Conditional probabilities thus obtained are averaged across the out-of-sample period.²⁴ All coefficients used, including those characterizing Y_s , are those based on the estimation period, and data input is solely from the out-of-sample period.²⁵

²³ Due to the symmetry condition, an alternative calculation uses offer information, but is otherwise the same.

²⁴ Alternatively, one could average the data in estimating the arrival rates, or average the arrival rates across the sample to produce the probabilities. The first is common in static conditional logit analysis, for example. We prefer to average the probabilities themselves to approximate the stationary distribution. This corresponds, for example, to the asymptotic equivalence of sample frequencies and derived stationary distributions from conditioning on the transition matrix in a Markov chain model.

²⁵ There is no natural way to impose summing up constraints on the probabilities, given the parameterization of the arrival models. Affine scaling factors were estimated based on the same sample as used for the estimation of the duration models, and employed in the out-of-sample probability forecasts. We also computed these factors using out-of-sample data as a check, and found that the scaling factors are virtually the same across sample periods. Positivity constraints are not imposed.

Out-of-sample comparisons of results across information sets are summarized in Panel A of Table X using mean square error (MSE) and weighted MSE statistics.²⁶ The MSE is computed relative to the empirical distribution for each point of support, and then averaging. Weighted MSE statistics use the empirical distribution as weights in the averaging process. The predicted distribution based on the full information model is reported alongside the empirical distribution in Panel B. The points of support are separated by one tick, i.e., 25 SEK, 50 SEK, and so on. We report the first ten points of support in the table, accounting for 80 percent of the probability mass of the cumulative distribution.

The usefulness of the flow information in characterizing market activity translates directly into a superior description of pricing behavior. The MSE for the pricing model using the ACD arrivals is over 30 percent higher than for the model incorporating trader activity. Flow information reduces forecast error relative to the stock of information embodied in the limit order book by about 18 percent.

Book information improves the spread probability prediction by only about 6 percent relative to the use only of duration information. Consistent with our earlier results on event durations, the addition of book information to trader flows does not help to describe pricing behavior. The predictive accuracy of the full information model is exactly equal to that of the probability model incorporating flow information.

Flow information does not improve the description of pricing behavior as much as it does the characterization of market activity in terms of arrival rates. In addition to order and cancellation processes, the probability model incorporates trade duration data. Trade durations are not well explained by the addition of observable screen or

²⁶ More formal comparisons of discrete distributions in settings such as ours are available; see, for example, Bollerslev, Domowitz, and Wang (1997). A comparison of asymptotic statistical

trader activity information, relative to the exclusive use of lagged durations. A proxy for the number of orders resting on the book at various prices away from the best offer also is required. Estimates of the latter are highly imprecise, and the problem is exacerbated in an out-of-sample comparison. Finally, the pricing model is an approximation to the underlying spread probability process, and specification error may play a role.

Particular points of support at which the model fails are evident by comparing the model-based distribution with the empirical distribution. There is clustering of spreads at 4 and 8 ticks in the empirical distribution. Over 50 percent of the spread probability mass is concentrated at the points of support for these spread sizes. Although the model does well in capturing the probability mass at 8 ticks, it predicts only an 8.8 percent probability of observing a spread of 4 ticks, compared with an average probability of over 22 percent. Clustering at 4 ticks appears to be a phenomenon unrelated to screen information, while a jump to the next tick cluster is well predicted by the model. This one error accounts for approximately two thirds of the overprediction of the cumulative distribution after the point of support that includes the clustering at 4 ticks.²⁷ The model does well at describing the probability of smaller spreads and the sharp increase in the spread from 4 to 8 ticks.

5. The Impact of Trading and Book Information on Spread Probabilities

The results thusfar suggest that trading activity describes trader behavior and market spreads better than book information. This finding is consistent with a smaller economic impact of changes in book information relative to traders' actions.

significance levels across information sets that are not always nested adds little insight, however.

²⁷ The calculation excludes the point of support at 8 ticks, which is slightly underestimated by the model.

In order to investigate this hypothesis, we calculate the average percentage changes in the spread probabilities, given each market event embodied in the probability models. Let j denote an event, such as the arrival of a bid at the market quote, with the variable k_j taking on the value of 1 if j occurs, and zero otherwise. The average is taken over all N events in the sample. The change is calculated as

$$\frac{1}{N} \sum_{i=1}^N [\Pr(S_i = s | k_j = 1, \Omega_{i-1}) - \Pr(S_i = s | k_j = 0, \Omega_{i-1})] / \Pr(S_i = s | k_j = 0, \Omega_{i-1}) \quad (5)$$

where the probabilities are conditional on Ω_{i-1} , the state of the market just prior to the observation of the spread, excluding the event j . The result is an "average derivative" relative to the occurrence of a discrete event. Averaging the differences yields the change in the probability of a particular spread due to the event across a variety of accompanying market conditions. The full information model of the last section is used as the basis for the computations.

The changes are reported in percentage terms in Table XI. There is a large number of average derivatives involved, and only a subset of the results is illustrated. Trade and quote information is reported only for the bid side, since the qualitative results are generally similar for trades or quotes at the offer. We omit cancellation activity.²⁸ The mean spread is between 7 and 8 ticks, and we report percentage changes averaged across spreads of 1 through 7 ticks and 8 through 10 ticks.²⁹ Variation across individual spread sizes is illustrated by spreads of 4 and 8 ticks, since much of the activity is concentrated at these points of support.

We also report an aggregate measure of the relative impact of trader activity and book information on the spread distribution, termed the trader/book variation. For

²⁸ Cancellations shift the spread distribution to the right, widening spreads on average. This is a purely mechanical result, and says little about pricing in terms of the spread.

²⁹ The averages are weighted by the empirical probability for each spread size.

each point of support in the spread distribution, let M be the number of possible events, j , within our classification of trader activity. Trader variation is given by

$$\frac{1}{NM} \sum_{j=1}^M \sum_{i=1}^N |\Pr(S_i = s | k_j = 1, \Omega_{i-1}) - \Pr(S_i = s | k_j = 0, \Omega_{i-1})| / \Pr(S_i = s | k_j = 0, \Omega_{i-1})|. \quad (6)$$

Book variation is computed the same way, with $j=1, \dots, M$ now indexing the possible events within the our classification of book information. The ratio of trader variation to book variation yields the relative absolute magnitude of the impact of trader flow information relative to the stock of information in the limit order book.

Trader/book variation suggests that the impact of trader activity on the spread distribution is far larger than that of book information. Averaging the variation over all spread sizes indicates a three-fold difference between the two types of information. The ratio can be lower for individual points of support in the distribution, but the difference in impact remains substantial. At 4 ticks, for example, the variation is just over 2, and at 8 ticks, it is about 1.8. The variation is over 73 percent bigger at large tick sizes (8 to 10 ticks) relative to the points of support for smaller spreads. Investigation of the numerator and denominator of the trader/book ratio suggests that the impact of trader activity increases, while that of book information decreases on average, for points of support for the smaller spreads. This suggests that the book is less relevant to pricing behavior when spreads are small.

Individual entries in Table XI suggest the source of these results. For example, large volume on the book decreases the probability of a spread at 4 ticks by about 5 percent and raises the probability of observing an 8 tick spread by only 0.7 percent. The effect of an empty book on one side of the market is negligible, on average. In comparison, large trades affect those probabilities by 66 to 80 percent in absolute value.

Trader aggressiveness is an especially important factor for larger spreads, illustrated by the results for spreads of 8 to 10 ticks. The impact of the arrival of a bid that betters the market quote is four times that predicted when the new bid is at the market quote. The difference rises to 10 times the effect when comparing a new bid worse than the market with a quote that betters the existing best bid. Such differences are nonexistent for points of support for spreads of 1 to 4 ticks.

We cannot provide a precise interpretation of the signs of the average derivatives. The average changes reported here are in the spirit of partial derivatives for each individual point of support. A negative sign for an impact on the distribution's probability mass for a spread of 8 ticks, for example, indicates that an event reduces the probability of observing that spread. It follows that the distribution shifts, relative to market circumstances in which the event does not occur. It is not generally possible to determine, however, to where the probability mass actually shifts, e.g., to a higher versus a lower spread size. This would require a theoretical model yielding the dynamics of the spread distribution, which is beyond our scope.³⁰

6. Concluding Remarks

In a limit order market, the state of the order book and information events influence order flow. Through the mechanics of the electronic market structure, the order flow updates the book, and the combination influences trading activity. The state of the book changes, providing additional information to the market and leading to new order submission. This cycle is empirically investigated in depth through descriptive statistics in Biais, Hillion, and Spatt (1995), for example. The emphasis in such work is on the relevance of individual pieces of observable activity on the evolution of the market.

We enlarge this focus through a joint investigation of market activity and pricing in an electronic market, using intraday data on stock index futures trading in a pure limit order book setting. Disparate information describing the market is aggregated into interpretable blocks. Duration models of order submission and trading behavior, differentiated by these blocks of information, are nested within a model describing the probability distribution of the bid-ask spread. We examine cases of no information, autonomous dynamics in arrival processes, stock information, and flow information. The last depends on order and trade activity on the part of individual traders. The stock information is embodied in the state of the order book just prior to observed market events.

The usefulness of our combination of data, information aggregation, characteristics of behavior, and evaluation techniques is illustrated through the consistency and sharpness of results. The main finding of the paper is the importance and superiority of flow information embodied in individual traders' actions in characterizing order submission behavior and the structure of pricing. The state of the book, the stock information, cannot explain subsequent order submission, trade, or pricing behavior based on out-of-sample predictive performance. Any correlation between stocks and flows provides negligible improvement in predictive accuracy relative to information based only on the flows of trader activity. The latter result again holds for market activity and spread distributions. Finally, the probability distribution of the bid-ask spread is little affected by changes in book information. In contrast, changes in trader activity shift the distribution considerably, especially with respect to the aggressiveness of order submission.

³⁰ We know of no such model, except for *ad hoc* specifications of the parameters of some specific distribution and purely statistical Markov chain formulations.

The nature of the results suggests an interesting conjecture. Despite some differences between electronic and floor markets, the rules governing the double auction that forms the basis for price discovery are fundamentally the same. Market making activity also exists in electronic venues, including the one examined here, similar to that which characterizes trading floors. With few exceptions, a floor trader observes much the same data that we have characterized as flow information, but cannot refer to a limit order book. We find that such flow information drives activity and pricing in a setting in which the book is available. This suggests that trader behavior may not be very different across market structures. Although we cannot verify this conjecture given the nature of our data, some support is available from other studies, based on observable market characteristics.³¹

There is one type of activity for which neither stock nor flow information appears to be important, however, namely trade execution. Market activity summarized by the speed with which trading is taking place appears to be sufficient to describe future trading activity. Trade execution is largely implemented through the device of hitting the bid or lifting the offer in this market, a separate type of “order submission” from bids and offers. Limit orders entail the problem of a “free option” that a firm quote presents to the market, permitting the possibility of trades at a sure profit that disadvantage the liquidity provider. The flow of trader activity may permit a better evaluation of the potential benefits of order submission, while the exercise of the option, the trade, does not require screen information. The fact that lagged trade duration information does provide explanatory power for trade execution suggests a heuristic story in our conversations with traders. When the market is moving up or down quickly, real-time traders simply want to buy or sell into the trend, however

³¹ See, for example, Coppejans and Domowitz (1999). A review of the results of empirical

short lived. This activity is summarized quite well by the autonomous dynamics of trade execution. The order submission decision appears to be more calculated, depending on the arrivals of others to the market, and the speed at which others are buying the security.

studies comparing electronic and floor markets is given in Domowitz and Steil (1999).

References

- Biais, Bruno, Pierre Hillion, and Chester Spatt, 1995, An empirical analysis of the limit order book and the order flow in the Paris Bourse, *Journal of Finance* 50, 1655-1689.
- Bollerslev, Tim, Ian Domowitz, and Jianxin Wang, 1997, Order flow and the bid-ask spread: An empirical probability model of screen-based trading, *Journal of Economic Dynamics and Control* 21, 1471-1491.
- Coppejans, Mark, and Ian Domowitz, 1999, Pricing behavior in an off-hours computerized market, *Journal of Empirical Finance*, forthcoming.
- Cox, David R. and Paul A.W. Lewis, 1966, *The statistical analysis of series of events*, New York: John Wiley & Sons.
- Domowitz, Ian, and Benn Steil, 1999, Automation, trading costs, and the structure of the securities trading industry, *Brookings-Wharton Papers on Financial Services*, 33-82.
- 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., 1996, The econometrics of ultra-high frequency data, *Econometrica*, forthcoming.
- Engle, Robert F., and Jeffrey Russell, 1998, Autoregressive conditional duration: A new model of irregularly spaced data, *Econometrica* 66, 1127-1162.
- Foucart, Thierry, 1993, Price formation in a dynamic limit order market, Working paper, HEC.
- Ghysels, Eric, Christian Gouriéroux, and Joanna Jasiak, 1998, Stochastic volatility duration models, Working paper, Pennsylvania State University.

- Glosten, Lawrence, 1994, Equilibrium in an electronic open limit order book, *Journal of Finance* 49, 1127-1161.
- Gourieroux, Christian, Joanna Jasiak, and G. Le Fol, 1996, Intra-day market activity, Discussion paper CREST, Paris.
- Gourieroux, Christian, G. Le Fol, and B. Meyer, 1996, Analysis of order queues, Discussion paper, CREST, Paris.
- Granger, Clive W.J., Maxwell L. King, and Halbert White, 1995, Comments on testing economic theories and the use of model selection criteria, *Journal of Econometrics* 67, 173-187.
- Hollifield, Burton, Robert A. Miller, and Patrik Sandas, 1996, An empirical analysis of a pure limit order market, Working paper, GSIA, Carnegie Mellon University.
- Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431-467.
- Rickard, John T. and William A. Lupien, 1996, Optimal market structure based on mutual satisfaction, *Proceedings of the 30th Asilomar Conference on Signals, Systems, and Computers*.
- Russell, Jeffrey R. and Robert F. Engle, 1998, Econometric analysis of discrete-valued irregularly-spaced financial transactions data using a new autoregressive conditional multinomial model, Working Paper, University of Chicago.

TABLE I**Summary of Daily Activity**

This table contains daily order, trade, and cancellation activity, for the period 9/25/95 through 12/15/95, averaged over 30,866 orders, 2876 trades, and 21,070 cancellations of orders. The designations "at quote," "> quote," and "< quote" mean that the order arrived at a price equal to the best price (bid or offer) on the book, greater than that price, and less than that price, respectively. "Large trade" denotes a trade which removed all volume at the best quote on the book, and "small trade" includes all other trades.

	orders	trades	Cancels
Total	1856	178	1245
Bid side	921	90	626
Offer side	935	88	619
Bid side at quote	75	-----	176
Offer side at quote	78	-----	161
Bid side > quote	176	-----	-----
Offer side < quote	171	-----	-----
Bid side < quote	670	-----	451
Offer side > quote	686	-----	457
Large trade at bid	-----	67	-----
Large trade at offer	-----	73	-----
Small trade at bid	-----	23	-----
Small trade at offer	-----	15	-----

TABLE II**Percentage of Daily Activity by Time of Day**

This table contains percentages of daily activity by hour of the trading day, reported in decimal form. There are six hours in each trading day, and no special opening protocol precedes the first hour of trading. The daily averages are computed over the period 9/25/95 through 12/15/95, based on 30,866 orders, 2876 trades, and 21,070 cancellations of orders. "Large trades" denotes trades which remove all volume at the best quote on the book, and "small trade" includes all other trades. "Crosses" are trades arranged away from the market, i.e., do not involve bids and offers on the order book at the time of reporting through the system.

	Hour 1	Hour 2	Hour 3	Hour 4	Hour 5	Hour 6
Orders	0.182	0.158	0.141	0.150	0.175	0.189
Large trades	0.204	0.149	0.119	0.136	0.176	0.213
Small trades	0.183	0.145	0.122	0.145	0.177	0.221
Cancellations	0.173	0.162	0.147	0.156	0.173	0.189
Crosses	0.231	0.145	0.106	0.129	0.175	0.214

TABLE III**Average Trade, Order, and Duration Information**

This table contains summary statistics on price, volume, and time between market events listed in the left column (duration). Duration is reported in seconds. Volume is denominated in number of contracts. Prices are in hundreds of Swedish Kroner, with the exception of the spread, which is the difference between the best offer and best bid on the book. An odd lot is a trade for less than ten contracts.

		Price	Volume	Duration
Orders		1394	18.05	11.44
Order at bid		1382	17.78	23.03
Order at offer		1405	18.32	22.69
Trades		1389	16.79	118.0
Trade at bid		1387	15.68	231.8
Trade at offer		1390	17.92	237.5
Cancel		1384	16.31	16.97
Cancel bid		1382	16.11	33.63
Cancel offer		1387	16.52	34.02
Crosses		1384	24.04	1034
Odd lots		1382	4.710	2653
Spread		198.4	-----	-----

TABLE IV
Information Sets

This table contains lists of variables constituting the different information sets labeled by the titles in the first row of the table. The term, "durations," denotes both expected and actual duration. Large volume is volume above the sample median, and small spread is a bid-ask spread below the sample median. Large offer (bid) length denotes greater than 3 prices appearing on the offer (bid) side of the book. "Large trade" denotes a trade which removed all volume at the best quote on the book, and "small trade" includes all other trades. "Crosses" are trades arranged away from the market, not involving bids and offers on the order book at the time of reporting through the system. "Book empty" means that there are no quotes on either the bid or offer side of the limit order book.

Poisson	ACD	Book	Traders	All
none	durations	durations	durations	durations
		large bid volume		large bid volume
		large offer volume		large offer volume
		large bid length		large bid length
		large offer length		large offer length
		book empty		book empty
		small spread		small spread
			large trade at bid	large trade at bid
			large trade at offer	large trade at offer
			small trade at bid	small trade at bid
			small trade at offer	small trade at offer
			new bid at quote	new bid at quote
			new offer at quote	new offer at quote
			new bid betters quote	new bid betters quote
			new offer betters quote	new offer betters quote
			new bid worse than quote	new bid worse than quote
			new offer worse than quote	new offer worse than quote
			cancel bid at quote	cancel bid at quote

			cancel offer at quote	cancel offer at quote
			cancel bid below quote	cancel bid below quote
			cancel offer above quote	cancel offer above quote
			cross trade	cross trade

TABLE V**Coefficient Estimates: Poisson and ACD Models**

This table contains coefficient estimates and standard errors adjusted for quasi-maximum likelihood estimation (in parentheses) for models of the form

$$y_i = w + ax_{i-1} + by_{i-1}$$

where the i^{th} duration between events is $x_i = t_i - t_{i-1}$, and y is the expected duration between events conditioned on past information. The quasi-maximum likelihood estimation is based on a pseudo-likelihood function in which the density of arrivals is exponential. Models are separately estimated for events consisting of the arrival of trades, orders, and cancellations as given in the first row of the tables in Panels A and B. In Panel A, the coefficients on duration and expected duration are set equal to zero.

Panel A: Poisson models

	bids	offers	trade at bid	trade at offer	cancel bid	cancel offer
constant	20.75	18.96	167.4	142.2	30.63	28.97
	(0.353)	(0.324)	(7.967)	(6.493)	(0.627)	(0.615)
sample	14835	16013	1439	1437	10051	11019

Panel B: ACD models

	bids	offers	trade at bid	trade at offer	cancel bid	cancel offer
constant	0.591	0.000	22.46	29.77	1.133	0.100
	(0.098)	(0.111)	(4.001)	(5.222)	(0.190)	(0.040)
x_{t-1}	0.0761	0.072	0.086	0.127	0.092	0.068
	(0.004)	(0.003)	(0.012)	(0.019)	(0.005)	(0.003)
y_{t-1}	0.879	0.908	0.815	0.692	0.863	0.920
	(0.006)	(0.005)	(0.026)	(0.041)	(0.008)	(0.004)
sample	14835	16013	1439	1437	10051	11019

TABLE VI

Coefficient Estimates: State of Order Book

This table contains coefficient estimates and standard errors adjusted for quasi-maximum likelihood estimation (in parentheses) for models of the form

$$y_i = w + ax_{i-1} + by_{i-1} + \exp(\mathbf{g}'z_{i-1})$$

where the i^{th} duration between events is $x_i = t_i - t_{i-1}$, and y is the expected duration between events conditioned on past information. The variables, z , are indicator variables in the book information set. The quasi-maximum likelihood estimation is based on a pseudo-likelihood function in which the density of arrivals is exponential. Models are separately estimated for events consisting of the arrival of trades, orders, and cancellations given in the first row of the table.

	bids	offers	trade at bid	trade at offer	cancel bid	cancel offer
constant	1.006 (0.178)	0.000 (0.093)	9.048 (11.13)	10.42 (9.749)	0.884 (0.434)	0.081 (0.036)
x_{i-1}	0.109 (0.005)	0.073 (0.003)	0.105 (0.022)	0.074 (0.014)	0.124 (0.007)	0.074 (0.003)
y_{i-1}	0.782 (0.010)	0.902 (0.005)	0.691 (0.058)	0.761 (0.050)	0.766 (0.012)	0.906 (0.004)
large bid volume	0.389 (0.078)	0.821 (0.217)	0.201 (0.169)	-0.132 (0.302)	0.069 (0.075)	2.965 (0.305)
large offer volume	0.029 (0.072)	0.480 (0.189)	0.017 (0.139)	0.211 (0.278)	0.019 (0.076)	-0.849 (0.312)
large bid length	-0.211 (0.179)	-0.436 (0.245)	0.702 (0.386)	0.659 (1.241)	-0.037 (0.153)	-1.323 (0.253)
large offer length	0.771 (0.163)	-0.122 (0.364)	0.755 (0.327)	1.902 (3.067)	0.323 (0.159)	0.134 (0.319)
book empty	0.418 (0.146)	-0.227 (0.258)	-----	-----	0.763 (0.133)	0.066 (0.285)
small spread	4.126 (0.230)	4.280 (0.330)	1.744 (0.371)	0.480 (2.196)	2.441 (0.705)	10.01 (1.749)
sample	14835	16013	1439	1437	10051	11019

TABLE VII

Coefficient Estimates: Trader Information

This table contains coefficient estimates and standard errors adjusted for quasi-maximum likelihood estimation (in parentheses) for models of the form

$$y_i = w + ax_{i-1} + by_{i-1} + \exp(\mathbf{g}'z_{i-1})$$

where the i^{th} duration between events is $x_i = t_i - t_{i-1}$, and y is the expected duration between events conditioned on past information. The variables, z , are indicator variables in the trader information set. The quasi-maximum likelihood estimation is based on a pseudo-likelihood function in which the density of arrivals is exponential. Models are separately estimated for events consisting of the arrival of trades, orders, and cancellations given in the first row of the table.

	bids	offers	trade at bid	trade at offer	cancel bid	cancel offer
constant	8.175 (0.306)	6.972 (0.272)	0.000 (17.30)	0.000 (17.28)	6.501 (0.381)	6.397 (0.369)
x_{i-1}	0.109 (0.007)	0.108 (0.007)	0.128 (0.026)	0.082 (0.019)	0.099 (0.008)	0.085 (0.008)
y_{i-1}	0.000 (0.009)	0.024 (0.009)	0.131 (0.082)	0.093 (0.089)	0.004 (0.009)	0.000 (0.009)
large trade bid	3.415 (0.092)	3.519 (0.081)	-80.70 (1.819)	5.263 (0.560)	4.200 (0.073)	3.948 (0.089)
large trade offer	3.809 (0.091)	3.566 (0.140)	5.259 (0.495)	-2345 (0.000)	4.249 (0.097)	4.245 (0.086)
small trade bid	3.792 (0.128)	3.379 (0.116)	-78.10 (2.061)	5.436 (0.301)	3.854 (0.142)	3.865 (0.117)
small trade offer	3.511 (0.118)	3.763 (0.106)	5.939 (0.355)	-63.04 (4.803)	3.947 (0.125)	3.864 (0.119)
bid at quote	-102.3 (2.209)	3.667 (0.061)	4.375 (0.378)	5.384 (0.219)	3.928 (0.060)	3.934 (0.066)
offer at quote	3.562 (0.057)	-132.1 (39.31)	5.167 (0.207)	4.476 (0.297)	3.980 (0.065)	3.769 (0.071)
bid > quote	-129.6 (3.547)	3.745 (0.042)	4.749 (0.155)	5.281 (0.126)	3.781 (0.049)	3.900 (0.043)
offer < quote	3.670 (0.047)	-114.5 (4.414)	5.181 (0.153)	4.614 (0.160)	3.894 (0.044)	3.718 (0.054)
bid < quote	-114.6 (3.146)	3.569 (0.042)	5.068 (0.146)	5.531 (0.150)	3.788 (0.046)	3.871 (0.044)
offer > quote	3.496 (0.043)	-102.1 (15.90)	5.115 (0.152)	5.386 (0.143)	3.922 (0.043)	3.663 (0.047)
cancel bid at quote	3.275 (0.060)	3.602 (0.060)	5.004 (0.236)	5.127 (0.274)	-101.3 (2.095)	3.942 (0.068)
cancel offer at quote	3.485 (0.068)	3.279 (0.074)	5.471 (0.218)	5.246 (0.196)	3.977 (0.072)	-98.08 (2.779)
cancel bid < quote	3.188	3.577	5.067	5.490	-134.3	3.829

	(0.053)	(0.046)	(0.163)	(0.204)	(2.239)	(0.051)
cancel offer>quote	3.561	3.245	5.495	5.293	3.918	-136.2
	(0.048)	(0.051)	(0.153)	(0.211)	(0.055)	(2.006)
cross trade	3.513	3.348	5.250	5.149	3.429	3.512
	(0.121)	(0.146)	(0.338)	(0.419)	(0.130)	(0.156)
sample	14835	16013	1439	1437	10051	11019

TABLE VIII

Coefficient Estimates: All Information

This table contains coefficient estimates and standard errors adjusted for quasi-maximum likelihood estimation (in parentheses) for models of the form

$$y_i = w + ax_{i-1} + by_{i-1} + \exp(g'z_{i-1})$$

where the i^{th} duration between events is $x_i = t_i - t_{i-1}$, and y is the expected duration between events conditioned on past information. The variables, z , are indicator variables in the full information set. The quasi-maximum likelihood estimation is based on a pseudo-likelihood function in which the density of arrivals is exponential. Models are separately estimated for events consisting of the arrival of trades, orders, and cancellations given in the first row of the table.

	bids	offers	trade at bid	trade at offer	cancel bid	cancel offer
constant	8.146 (0.304)	6.972 (0.0.275)	0.000 (18.73)	0.000 (15.45)	6.588 (0.388)	6.482 (0.368)
x_{i-1}	0.107 (0.007)	0.106 (0.007)	0.121 (0.026)	0.080 (0.019)	0.095 (0.008)	0.085 (0.007)
y_{i-1}	0.007 (0.009)	0.030 (0.009)	0.029 (0.078)	0.046 (0.078)	0.010 (0.008)	0.000 (0.008)
large trade bid	3.145 (0.149)	3.298 (0.122)	-91.19 (2.688)	-0.269 (0.766)	3.022 (0.177)	2.543 (0.213)
large trade offer	3.593 (0.132)	3.312 (0.160)	3.582 (0.787)	-63.77 (1.874)	2.977 (0.191)	3.016 (0.200)
small trade bid	3.437 (0.164)	3.068 (0.150)	-61.98 (2.671)	0.061 (0.655)	2.554 (0.212)	2.381 (0.229)
small trade offer	3.176 (0.157)	3.381 (0.143)	4.340 (0.726)	-110.7 (5.795)	2.638 (0.206)	2.396 (0.228)
bid at quote	-103.9 (2.377)	3.354 (0.114)	2.876 (0.774)	-0.091 (0.620)	2.640 (0.183)	2.476 (0.215)
offer at quote	3.267 (0.121)	-84.87 (31.81)	3.530 (0.763)	-0.955 (0.414)	2.689 (0.183)	2.313 (0.215)
bid > quote	-129.8 (3.507)	3.428 (0.107)	3.245 (0.704)	-0.197 (0.610)	2.475 (0.177)	2.442 (0.206)
offer < quote	3.347 (0.117)	-84.05 (3.097)	3.674 (0.694)	-0.836 (0.615)	2.606 (0.173)	2.264 (0.209)
bid < quote	-112.9 (3.209)	3.259 (0.106)	3.434 (0.715)	0.029 (0.608)	2.439 (0.181)	2.397 (0.208)
offer > quote	3.210 (0.116)	-120.6 (29.99)	3.536 (0.722)	-0.103 (0.609)	2.631 (0.176)	2.137 (0.214)
cancel bid at quote	3.008 (0.115)	3.373 (0.115)	3.342 (0.763)	-0.419 (0.630)	-84.74 (1.678)	2.571 (0.206)
cancel offer at quote	3.272 (0.118)	3.013 (0.115)	3.758 (0.758)	-0.224 (0.618)	2.761 (0.175)	-90.95 (2.688)
cancel bid < quote	2.892	3.292	3.481	0.040	-214.1	2.368

	(0.117)	(0.106)	(0.725)	(0.624)	(0.000)	(0.206)
cancel offer>quote	3.275	2.936	3.836	-0.213	2.622	-157.5
	(0.113)	(0.108)	(0.706)	(0.619)	(0.175)	(2.745)
cross trade	3.192	3.038	3.676	-0.335	2.137	2.062
	(0.158)	(0.172)	(0.763)	(0.716)	(0.211)	(0.249)
large bid volume	0.113	0.132	0.284	-0.003	0.091	0.175
	(0.028)	(0.027)	(0.078)	(0.089)	(0.026)	(0.027)
large offer volume	0.032	0.042	0.072	0.133	0.059	0.064
	(0.028)	(0.027)	(0.068)	(0.097)	(0.027)	(0.028)
large bid length	-0.221	0.043	0.994	-0.275	0.584	0.231
	(0.056)	(0.064)	(0.374)	(0.197)	(0.085)	(0.086)
large offer length	0.111	-0.118	-0.006	5.553	0.180	0.608
	(0.0.079)	(0.065)	(0.280)	(0.737)	(0.092)	(0.114)
book empty	0.352	0.299	-----	-----	0.369	0.384
	(0.034)	(0.034)			(0.040)	(0.039)
small spread	1.375	1.059	0.427	0.137	0.974	-1.092
	(0.232)	(0.322)	(0.303)	(0.337)	(0.424)	(0.869)
sample	14835	16013	1439	1437	10051	11019

TABLE IX**Out-of-Sample Statistics for Duration Models**

This table contains out-of-sample summary statistics for forecasts of durations between events given in the first row of the table. Models for each event are differentiated by information sets labeled in the first column. The duration models are estimated over the period 9/25/95 through 12/15/95, and the forecasts are produced and evaluated over the period 1/29/96 through 2/16/96. RMSE is the root mean-square error of forecast. R^2 is an out-of-sample pseudo R^2 , constructed by subtracting the ratio of forecast mean square error to the variance of duration from one.

model	statistic	bids	offers	trade at bid	trade at offer	cancel bid	cancel offer
Poisson							
	RMSE	36.52	33.76	387.5	363.3	51.07	46.78
	R^2	0.006	-0.003	0.018	0.033	0.003	-0.014
ACD							
	RMSE	35.98	33.18	373.5	355.3	50.33	46.03
	R^2	0.035	0.031	0.088	0.075	0.032	0.018
Traders							
	RMSE	34.23	31.51	377.5	357.4	47.52	43.44
	R^2	0.128	0.122	0.068	0.064	0.137	0.126
Book							
	RMSE	35.97	33.18	375.7	354.8	50.26	46.13
	R^2	0.035	0.031	0.077	0.078	0.034	0.014
All							
	RMSE	34.21	31.55	378.6	356.8	47.33	43.34
	R^2	0.128	0.124	0.063	0.067	0.144	0.130

TABLE X

Out-of-Sample Statistics for Probability Models

This table contains out-of-sample statistics for forecasts of the probability distribution of the bid-ask spread, given by

$$\Pr(S = s) = \frac{I^b(s, \Omega)[1 - d^b(s, \Omega)]}{g^a(s, \Omega)}$$

over all points of support given by the range of values of s , in which

$$d^b(s, \Omega) = \frac{m^b(s, \Omega)E(Y_s)}{I^b(s, \Omega)}$$

where $s = |k - j|$, the difference in ticks between two prices p_k and p_j . The arrival rate of bids at a price s ticks away from the best offer price is $I^b(s, \Omega)$. The arrival of trades on the offer side is $g^a(s, \Omega)$; $m^b(s, \Omega)$ denotes the arrival rate of a cancellation of an offer at a price s ticks away from the best offer. Y_s is the number of bids on the book at a price s ticks away from the best offer. Models for each event are differentiated by information sets, Ω , labeled in the first column. Arrival rates are constructed by inverting forecasts of the conditional duration models. The duration models are estimated over the period 9/25/95 through 12/15/95, and the forecasts are produced and evaluated over the period 1/29/96 through 2/16/96. In Panel A, the MSE is computed relative to the empirical distribution for each point of support, and then averaged. Weighted MSE statistics use the empirical distribution as weights in the averaging process. The predicted distribution based on the full information model is reported with the out-of-sample empirical distribution in Panel B. The points of support are separated by one tick. The first ten points of support reported in the table account for 80 percent of the probability mass of the cumulative distribution.

Panel A: Mean Squared Error of Forecast

model		Mean Squared Error		Weighted MSE
ACD		0.231		0.099
Traders		0.177		0.078
Book		0.217		0.094
All		0.177		0.078

(Table X continued)

Panel B: Bid-Ask Spread Probabilities

Spread in ticks		Full Information Model		Empirical Distribution
1		-0.030		0.003
2		0.030		0.028
3		0.011		0.014
4		0.088		0.223
5		0.026		0.017
6		0.128		0.106
7		0.123		0.041
8		0.260		0.291
9		0.048		0.013
10		0.104		0.053

TABLE XI

Selected Magnitudes of the Impact of Changes on Spread Probabilities

This table contains the average percentage changes in the spread distribution, given market events listed in the first column. The change for any event j , for each point of support, is computed according to

$$\frac{1}{N} \sum_{i=1}^N [\Pr(S_i = s | k_j = 1, \Omega_{i-1}) - \Pr(S_i = s | k_j = 0, \Omega_{i-1})] / \Pr(S_i = s | k_j = 0, \Omega_{i-1})$$

where the variable k_j takes on the value of 1 if j occurs, and zero otherwise. The average is taken over all N events in the sample, and the probabilities are conditional on O_{i-1} , the prior state of the market, excluding the event j . The numerator and denominator of average trader/book variation are computed as

$$\frac{1}{NM} \sum_{j=1}^M \sum_{i=1}^N | [\Pr(S_i = s | k_j = 1, \Omega_{i-1}) - \Pr(S_i = s | k_j = 0, \Omega_{i-1})] / \Pr(S_i = s | k_j = 0, \Omega_{i-1}) |$$

where M is the number of possible events j within the classifications of trader (numerator) and book (denominator) activity. The average derivatives over 1-7 (8-10) ticks are computed by averaging the percentage changes over the points of support of the distribution covering spread sizes of 1 through 7 (8 through 10) ticks.

	average derivatives 1-7 ticks	average derivatives 8-10 ticks	derivatives 4 ticks	derivatives 8 ticks
average trader/book variation	2.749	4.758	2.011	1.779
large trade bid	-27.22	-30.76	-79.58	-65.89
small trade bid	-23.13	32.52	-77.62	-36.71
bid at quote	52.06	320.2	134.4	449.9
bid >quote	54.50	1287	126.8	492.2
bid <quote	59.96	122.3	159.3	755.9
cross trade	-36.50	-85.98	-98.77	-141.6
large bid volume	-1.161	2.380	-4.821	0.680
large offer volume	2.878	7.606	8.046	13.55
large bid length	9.643	-121.7	15.34	-397.9
large offer length	45.66	10.70	158.1	131.2
book empty	0.470	-5.739	-3.504	37.60

