The Role of Time in Price Discovery: Ultra-high frequency trading in a Limit Order Book Market

Article

2 authors:


Bernard Mualuke Ben Sita
Lebanese American University


P. Joakim Westerholm
University of Sydney

The Role of Time in Price Discovery: Ultra-high frequency trading in a Limit Order Book Market

Bernard Ben Sita

and

P. Joakim Westerholm ♣

Draft 25 May, 2006

Abstract

In this paper we investigate the price effects of trading intensity. Extending on the Madhavan et al. (1997) model, we split the intensity effect into liquidity and information effects. We provide a measure of market quality that is the ratio of the covariance bias to the variance bias. Analyzing about 6 years of tick by tick data, we find that the bid-ask spread in a pure limit order book market contains a risk component associated with managing the time to trade, and this component accounts for roughly 19.6% of the implied bid-ask spread. Extending our model to investigate intraday patterns, we find that the adverse selection cost exhibits a U-shaped pattern reflecting uncertainty at market openings in the Helsinki Stock Exchange (HEX) and in the New York Stock Exchange (NYSE). The results emphasize the importance of managing time in limit order book markets.

JEL classification: C30; C41; G14

Keywords: Trading Intensity; Market Microstructure; Price Impact

---

We are grateful to SIRCA, Sydney, Australia for providing access to data from Reuters Group
♣ Corresponding author: P J Westerholm, H60 School of Business, UNIVERSITY OF SYDNEY, NSW 2006, Australia. Email: j.westerholm@econ.usyd.edu.au Phone +612 9351 6454 Fax +612 9351 6461

The Role of Time in Price Discovery: Ultra-high frequency trading in a Limit Order Book Market

Abstract

In this paper we investigate the price effects of trading intensity. Extending on the Madhavan et al. (1997) model, we split the intensity effect into liquidity and information effects. We provide a measure of market quality that is the ratio of the covariance bias to the variance bias. Analyzing about 6 years of tick by tick data, we find that the bid-ask spread in a pure limit order book market contains a risk component associated with managing the time to trade, and this component accounts for roughly 19.6% of the implied bid-ask spread. Extending our model to investigate intraday patterns, we find that the adverse selection cost exhibits a U-shaped pattern reflecting uncertainty at market openings in the Helsinki Stock Exchange (HEX) and in the New York Stock Exchange (NYSE). The results emphasize the importance of managing time in limit order book markets.

JEL classification: C30; C41; G14

Keywords: Trading Intensity; Market Microstructure; Price Impact

1. Introduction

The purpose of this paper is to investigate the components of the bid-ask spread in a pure limit order book market[1] taking into account the duration[2] between consecutive trades. The literature on the components of the bid-ask spread is rich and the subject is theoretically and empirically settled from many aspects. The theory in Glosten and Milgrom (1985) and Kyle (1985) describe the agents, the actions and the process by which prices change to incorporate new information. Time is not expressively modelled in these models. However, since informed traders randomize between trading and waiting, the bid-ask spread is also in these models a function of the arrival rate of bid and ask prices. In Glosten and Milgrom (1985) informed traders transact intensively as if there will not be another opportunity to trade, while in Kyle (1985) informed traders are patient choosing to trade gradually. While there is no indication in Glosten and Milgrom (1985) and Kyle (1985) that time affects the trading behaviour of traders, there is a time dimension simply because prices converge at different rates for uninformed and informed traders in these models. Comparing the Kyle (1985) model with the Glosten and Milgrom (1985) model, Back and Baruck (2004) demonstrate that informed traders trade with a lower trading rate that corresponds to one time unit, whereas uninformed traders trade with a higher trading rate that corresponds to the square root of one time unit.

The importance of time in price discovery emerges clearly in the Easley and O'Hara (1992) model. In this model, the information flow is not continuous because informed traders choose not to trade from time to time. The model predicts that market makers

[1] With a pure limit order book market we refer to an exchange that has consolidated all its trading to one limit order book with no dealers with affirmative obligations and no size and liquidity determined market segments.

[2] Time elapsed between consecutive trades.

interpret the absence of informed traders in the market as no new information. However, this is true only under the assumption that liquidity is not a hurdle for informed traders to trade at their convenience, and that informed traders are not patient in the sense of Kyle (1985). In her presidential address, O'Hara (2003) points out that liquidity is an important factor for price discovery. Hence, the absence of trades is not necessarily an indication that there is no new information. In addition to illiquidity another reason that might keep informed traders out of the market is trading halts. Market regulators might voluntarily suspend informed trading when the new information is expected to have an extreme impact on price as in Diamond and Verrecchia (1987). A third reason is that informed traders may stay out of the market for purely strategic reasons. For example, Back and Baruck (2004) assert that when informed traders realize that their aggregate profits increase more when they trade gradually than when they trade intensively, they will adopt a low trading rate strategy. Modelling the risk of trading in limit order book markets, Foucault, Kanda and Kendal (2003) show that patient traders will require a compensation for providing liquidity to traders that choose to trade intensively as defined by the Glosten and Milgrom (1985) model.

In empirical studies duration is respectively considered as a measure of trading intensity, a measure of liquidity and a measure of risk. E.g. Jasiak and Ghysels (1998), Engle and Russell (1997), Engle and Russell (1998), Engle (2000), Renault and Werker (2002), Manganelli (2002), and Spierdijk (2004)) consider duration as a measure of trading intensity. These studies show that duration is inversely proportional to the expected return variance. E.g. Engle (2000) shows that variation in duration and variation in return variance are linked to the same news events. In

Gouriéroux, Jasiak and Le fol (1999), Dufour and Engle (2000), Engle and Lange (2001), Engle and Lunde (2004)) duration is considered as a measure of liquidity. The general result in these studies is that liquidity is an important determinant of the bid-ask spread, and that it can be estimated with trading activity-based measures. E.g., Dufour and Engle (2000) show that informed traders are more active during periods when the number of informed transactions can be maximized. In (Renault and Werker (2002), and Ghysels, Gouriéroux, and Jasiak (2004)) duration is considered as a measure of risk

Madhavan et al. (1997) extending on Glosten and Milgrom (1985) show that the price change is a function of the order flow and the autocorrelation in the order flow. We extend the Madhavan et al. (1997) model by inserting trade duration with the aim to capture the speed by which liquidity is created and information is disseminated. Motivated by the Dufour and Engle (2000) and Saar and Hasbrouck (2002) findings that spreads and volatilities are higher when traders observe short durations, we examine the impact of durations on price changes. In that our study is closely related to Dufour and Engle (2000) while we base our model of price impact of duration on the Hasbrouck (1991) vector autoregressive (VAR) model. In contrast to Dufour and Engle (2000) we split the trade duration into a transitory and a permanent component. Following Renault and Werker (2002), we relate the transitory component of the trade duration to the expected duration, and the permanent component of the trade duration to innovations in the trading intensity.

Liquidity is a critical measure in markets operating without market makers. Foucault, Kanda and Kendal (2003) assert that patient traders providing liquidity in limit order book markets require a compensation for that, and this compensation is a function of waiting times. They predict that the bid-ask spread is lower when patient traders observe short durations. Hence, "managing the time to trade" is important in pure limit order book markets because this is the only way to increase the execution likelihood, and to minimize the risk to be picked off. However, in traditional spread models investors are not compensated for managing time, rather for managing inventory. The time of trade can be managed in a similar way to inventory in the sense that patient traders adjust their prices to reflect the trading intensity just as market makers adjust their quotes to balance their inventory.

We estimate our model on transactions data for Nokia. The Nokia stock is the most traded stock in the Helsinki Stock Exchange (HEX), and one of the most active American Depository Receipts (ADR) traded in the New York stock Exchange (NYSE). We focus on the trading on the main market (HEX) and our data sample spans six years of transactions with a total of 6,753,243 observations Our primary findings are as follows. Firstly we find that prices are persistent under both liquidity and informed trading. Liquidity is perceived as risky as predicted by Foucault et al. (2003). We estimate that 19.6% of the implied bid-ask spread is attributable to liquidity costs (duration effects), 23.8% to adverse selection costs, and 56.6% to order-handling costs. We cross-validate our result by estimating the Madhavan et al. (1997), the DeJong et al. (1996), and the Dufour and Engle (2000) model. Secondly we find that trades in Nokia are persistent, consistent with previous studies (e.g. Hasbrouck (1991), and Ahn et al. (2002)). Thirdly we find that the first order autocorrelation of the trade process is negative and decreases when we control for trading frictions and microstructure effects. Fourthly investigating the intraday patterns of the components of the spread, we find that the adverse selection cost exhibits a U-shaped pattern consistent with Ahn et al. (2002). This pattern shows that traders adjust the price of Nokia close to end of the trading day. Hence, since Nokia is traded both in Helsinki and in New York, the U-shaped pattern is consistent with information models predicting a decrease in information asymmetry toward the end of the trading day. With two or several markets opening at different times, the U-shaped pattern expresses the information flow across exchanges. Overall, this study shows the importance of managing time in a pure limit order book market because time affects the component of the bid-ask spread.

The rest of the paper proceeds as follows. In section 2, we present a simple structural model to examine the price effect of trade durations. In section 3, we present descriptive statistics and the intraday patterns of the trading intensity. In section 4, we report the empirical results both on transactions and intraday data. The last section concludes the paper.

2. The empirical model

2.1 The structural model

A large number of studies examine the components of the bid-ask spread in organized securities markets (e.g. Glosten (1987), Glosten and Harris (1988), Stoll (1989), George, Kaul and Nimalendran (1991), Lin, Sanger and Booth (1995), DeJong, Nijman and Röell (1996), Madhavan, Richardson and Roomans (1997), and Huang and Stoll (1997)). In this section we present a simple model that decomposes the bid-ask spread into order-handling, liquidity and adverse selection costs based on if the trade is buyer or seller intitated including a duration variable in the model. There are at least three statical approaches to decompose the bid-ask spread. The first approach by Roll (1984) is to infer the components of the bid-ask spread from the serial covariance properties of observed prices. The second approach by Hasbrouck (1988) and Glosten and Harris (1988) is to infer the components of the bid-ask spread from the trade indicator variable. The third by Huang and Stoll (1997) is to infer the components of the bid-ask spread from a trade direction indicator and trading volume. We decompose the spread by a trade direction indicator and trade duration:

$$m_t = m_{t-1} + \left[\phi_1 + \left(\phi_2 + \rho_q \vartheta_1\right)T_t\right]q_t + e_t, \tag{1}$$

where m_t is the unobservable price, T_t is the trade time, q_t is the trade direction indicator variable (taking -1 for seller-initiated trades, +1 for buyer-initiated trades, and 0 for trades occurring within the bid-ask spread), and e_t is the serially uncorrelated error term. Eq. (1) decomposes the effects on m_t into four trading parameters. A change in m_t reflects the cost of processing a trade and this is captured by ϕ_1. A change in m_t reflects the speed by which transactions occur and this captured by ϕ_2. A change in m_t reflects private information revealed by the order flow and the trading intensity and this is captured by ϑ_1. It is assumed that informed traders observe short durations when present in the market (Easley and O'Hara (1992)). A change in m_t depends on the temporal dependence in the order flow ρ_q that results from either informed traders in their efforts to consume creeping liquidity or noise traders acting on past information. While we do not observe m_t, we do observe the trade price p_t that is measured with an error. The first difference of Eq. (1) after we have decomposed the duration variable is given by

$$r_t = \lfloor \phi_1 \Delta q_t + (\phi_2 \psi_t + \rho_q \theta_1 C_t) q_t \rfloor + (\varepsilon_t + (n_t - n_{t-1})), \tag{2}$$

where $r_t = 100 * \ln(p_t / p_{t-1})$ the return in percent, $\psi_t = a_0 + a_1 x_{t-1} + a_2 \psi_{t-1}$ is the expected duration process truncated at lag 1, $x_t = T_t - T_{t-1}$ is the duration, $C_t = (x_t/\psi_t)((x_t/\psi_t) - 1)$ is the innovation in trading intensity, $\Delta q_t = q_t - q_{t-1}$ is the residual of the order flow, $\varepsilon_t = m_t - m_{t-1}$ is the residual of the unobservable prices, which is serially uncorrelated, and $\Delta n_t = n_t - n_{t-1}$ is the residual of the price errors, which is serially correlated. Eq. (2) shows that change in r_t reflects information shocks through ε_t and trading friction shocks through n_t associated with p_t, which is observed with error according to $p_t = p_t^* + n_t$.

In Dufour and Engle (2000) the association between r_t and x_t is related to short durations, but not to long durations (Easley and O'Hara (1992)). Eq. (2) differs from that of Dufour and Engle (2000) in that x_t is split into a temporal and a permanent component. The temporal component is associated with ψ_t capturing liquidity effects and the permanent component is associated with C_t capturing information effects. We estimate an Autoregressive Conditional Duration ACD(1,1) model of Engle and Russell (1998) to obtain ψ_t and C_t. In this model, all the temporal dependence in durations is captured by ψ_t so that the duration residuals $d_t = \tilde{x}_t / \psi_t$ are by assumption identically and independently distributed. Eq. (2) decomposes also the error term into a fundamental and a transitory component. The fundamental error induces permanent price changes related to the fundamental variance consisting of seemingly random price changes that do not revert. The transitory component induces transitory price changes related to the transitory variance that reverts from time to time. In Hasbrouck (1993), the transitory variance is taken as an implicit transaction cost and is a measure of the market quality. Based on Eq. (2), we propose the following standardized measure of the market quality,

$$\rho = \frac{-\sigma_n^2 + \left(\phi_1^2 + \phi_2^2 + \rho_q \vartheta_1^2 + 2\phi_1\phi_2\right)\rho_q}{\left(\sigma_\varepsilon^2 + 2\sigma_n^2\right) + \left(\phi_1^2 + \phi_2^2 + \vartheta_1^2 + 2\theta_1\phi_2\rho_q\right)}, \tag{3}$$

where ρ is the ratio of the covariance bias to the variance bias (see Appendix for the full derivation). The ratio of Eq. (3) is simply the first order autocorrelation of the return process, and is interpreted as a measure of the market quality. In Hasbrouck (1993), the pricing error determines the quality of the stock, in Eq. (3) not only the pricing error is accounted for, but also trading frictions and divers microstructure effects. It follows that this ratio considers the entire price process, and is useful for investors and market regulators to distinguish between the fundamental and the transitory volatility when dealing with excess volatility in a stock or a market.

2.2 Model estimation

Eq. (2) is estimated with the generalized method of moments (GMM) estimator with mild distribution assumptions. The population moments are encapsulated in $g_N\left(\beta^0\right)$ and its counterpart sample moments in $g_N\left(\hat{\beta}\right)$, where the subscript N denotes the sample size. The expectations of the following four population moments are zero,

$$\begin{bmatrix} \varepsilon_t = r_t - \left(\phi_1 \Delta q_t + \left(\phi_2 \psi_t + \rho_q \vartheta_1 C_t\right) q_t\right) \\ q_t q_{t-1} - \rho_q q_t^2 \\ \varepsilon_{t-1}\varepsilon_t + \sigma_n^2 \\ \varepsilon_t^2 - \left(\sigma_\varepsilon^2 + 2\sigma_n^2\right) \end{bmatrix} = 0 . \tag{4}$$

Eq. (4) is a system of linear equations whose expectations equal 0. The first equation captures the effect of revision in belief that grows rapidly when traders observe short durations and slowly when traders observe long durations (Easley and O'Hara (1992)). The second linear equation captures the lagged effects of noise traders on price change by allowing for the autocorrelation in the order flow. The third and the fourth equation capture the transitory and the fundamental variance of returns respectively. As the instrumental variable, we use the square of residuals obtained from regressing q_t and q_{t-1} on r_t. In GMM, $g_N\left(\beta^0\right)$ is chosen through $g_N\left(\hat{\beta}\right)$ by minimizing the objective function, $J(\beta) = \bar{g}_N(\beta)' W_N \bar{g}_N(\beta)$, where W_N is a symmetric non-singular weighting matrix that satisfies $W_N \rightarrow W$ almost surely. Under regularity conditions $g_N\left(\hat{\beta}\right)$ is a consistent estimator of $g_N\left(\beta^0\right)$ with an asymptotic variance covariance matrix, Σ_W, that depends on the limiting weighting matrix W (Hansen (1982)).

2.3 Alternative empirical models

There is now a set of competitive models for the estimation of the implied bid-ask spread both on specialist and on pure limit order book markets. By its structure, the Madhavan et al. (1997) model is closely related to our empirical model. Estimated on the New York Stock Exchange (NYSE) stocks, they find that about 60% of the total variance is attributable to the transitory variance. Similarly, Dufour and Engle (2000) estimate their VAR model on NYSE stocks and find a negative relationship between the midquote return and the signed duration. We only estimate the first equation of their bivariate VAR model that is relevant to our study. DeJong et al. (1996) extending on the Glosten (1994) model provide estimates on data from the Paris Bourse showing that the price impact increases with the transaction size. We cross-validate our results based on Eq. (2) by estimating the following empirical models,

$$r_t = \theta(q_t - \rho_q q_{t-1}) + \phi(q_t - q_{t-1}) + v_{1t}, \tag{5}$$

$$r_t = (\gamma_0 q_t + \gamma_1 q_{t-1}) + [\delta_0 \ln(x_t) q_t + \delta_1 \ln(x_{t-1}) q_{t-1}] + v_{2t}, \tag{6}$$

$$r_t = a_0 + (R_0 \Delta q_t + R_1 \Delta q_t z_t) + (e_0 q_{t-1} + e_1 q_t z_{t-1}) + v_{3t}, \tag{7}$$

where v_{1t}, v_{1t}, and v_{1t} are random error terms from the Madhavan et al. (1997), the Dufour and Engle (2000), and the DeJong et al. (1996) model respectively, and z_t is the trading volume. The Madhavan et al. (1997) equation suggests that the order processing cost is captured by ϕ, and the adverse selection cost by θ. In the Dufour and Engle (2000) equation, the order processing costs are given by $(\gamma_0 - \gamma_1)q$ and the adverse selection costs by $(\delta_0 - \delta_1)\ln(x)q$. The coefficients in DeJong et al. (1996) are more complicate. They derive the order handling cost by establishing the following relation, $c_0 + c_1 z$, where c_0 is $R_0 - e_0 - e_1\alpha$, c_1 is $R_1 - 0.5e_1$ and α is the median of trade size divided by the logarithm of 2, and the adverse selection cost by $(R_0 - c_0) - (R_1 - c_1)z$. We estimate Eq. (5) by the GMM estimator, and Eq. (6) and (7) by the OLS estimator with robust errors. The OLS error might be heteroskedastic and autocorrelated, therefore we use White's heteroskedasticity consistent standard errors to compute t-statistics.

3. Data, summary statistics and intraday patterns

3.1 Data

We use tick-by-tick data for the common stock of the telecommunications corporation Nokia, the most actively traded stock listed on HEX[3], and one of the most active ADRs listed on the NYSE. The trade and quote (highest bid and lowest ask order in the limit order book) data is obtained from Reuters covers April 12, 1999 through December 30, 2004. Trading on HEX is organized as an electronic limit order book market in three main trading sessions: The opening session from 8:30am to 9:45am, under which authorized broker-dealers are allowed to enter their publicly invisible sell and buy orders into the system and the opening price is determined in an opening call. The continuous trading session is held from 10:00am to 6:20pm.[4] during which orders are submitted with price and time priority, transactions are matched automatically, transactions arranged upstairs are reported to the trading system without delay, and submitted quotes are binding until there is a match or the end of the trading day. The after market session is held from 6:40pm to 7:00pm. During this session, transactions are matched under the conditions prevailing during the continuous trading session. HEX has no market makers with affirmative obligations except for in recent years when liquidity providers have been introduces in certain sponsored stocks. This does not concern the Nokia stock under investigation here.

3.2 Data preparation and weighted duration measures

---

[3] HEX is since 2003 operated by OMX Exchange which is a division of OMX a listed company headquartered in Stockholm, Sweden that owns and operates the largest integrated securities market in Northern Europe and is a provider of marketplace services and solutions for the financial and energy markets.

[4] The continuous (trading) session schedule has been modified several times during the sample period. The modification of the trading session has mostly concerned the time of the close of the continuous trading session. The time of the opening call session has invariably been kept at 10:00am. For simplicity, we fix the closing time at 6:00pm for the entire sample. Since September 2004, the continuous session is held from 10:00am to 6:20pm.

For a total of about 6 years and of 1363 trading days, there are 6,753,243 recorded transactions. Of these transactions only a fraction affect the price indicating that only a part of the trades contain information. Many subsequent transactions are traded at the same price as the previous trade without price adjustment. To focus on the trades that contribute to price discovery, we thin the trade price process. We ignore price at time t-1 if this price equals the price at time t and in such a case adding the trading volume of the price at time t-1 to the trading volume at time t. This way we obtain a series of transactions that is known as price based durations (e.g. Engle and Russell (1997)). This data set is used to estimate the ACD, the GMM and the OLS model for 127 different sub-sets. The number of trading days included in each sub-set is about 10 trading days, which is a comparable sample to previous studies (e.g. Engle (2000)).

<Insert Table 1 about here>

We also adjust for the time of day effects on durations and volumes. Seasonal patterns in the rate of arrival of quotes and the realization of prices have been examined in several papers (e.g. Bollerslev and Andersen (1997), and Engle (2000)). Since the intensity of trading is periodic in financial markets, we divide the trading day into four different trading periods, each of 2 hours long. We include only transactions that occur during the continuous trading session (from 10.00am to 6.00pm), the first trading period is from 10.00 to 12.00am, the second trading period from 12.01am to 2.00pm, the third trading period from 2.01 to 4.00pm and the fourth trading period from 4.01 to 6.00pm. We use a piecewise function of the times of day to adjust the trade duration and the trading volume for the time of day effects,

$$\tilde{x}_t = x_t / E[x_t \mid f(t)], \tag{8}$$

$$\widetilde{z}_t = z_t / E[z_t \mid f(t)], \qquad (9)$$

where

$$\begin{aligned} f(t) = \beta_0 + \beta_1(t-k_1) + \beta_2(t-k_2;(t>k_2);0) \\ +\beta_3(t-k_3;(t>k_3);0) + \beta_4(t-k_4;(t>k_4);0) \end{aligned}, \qquad (10)$$

where k_i are nodes fixed at 43200 (12.00am), 50400 (2.00pm), 57600 (4.00 pm) and 64800 (6.00pm) seconds since midnight, and β_k are consistent but inefficient OLS estimates as we regress x_t and z_t on $f(t)$, respectively. Furthermore, in order to explore and to investigate the intraday trading activity in the Nokia stock, we transform the durations into weighted duration measures using a normal density kernel:

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right), \qquad (11)$$

where u is defined for $u_1 = (x_{t-1} - \bar{x})/h$ trade durations and $u_2 = (x_{t-1}^z - \bar{x}^z)/h$ volume durations, h is the bandwidth parameter fixed at 30 seconds, $\bar{x}$ the trade duration average and $\bar{x}^z$ the volume duration average. The weighted duration measures are obtained as

$$\Lambda_p^x = N^{-1} \sum_{j=1}^{N} \frac{1}{h} K(u_1), \qquad (12)$$

$$\Lambda_p^v = N^{-1} \sum_{j=1}^{N} \frac{1}{h} K(u_2), \qquad (13)$$

where N is the number of trades during a given trading period, the subscript p stands for 1, 2 and 3 trading period, respectively, Λ_p^x is the weighted trade duration, and Λ_p^v the weighted trading volume duration for a given trading period, respectively. In Gouriéroux et al. (1999), the volume duration is defined for a given trading volume. We define the volume duration for transactions over 10,000 shares. We use the weighted duration measures to explore the "aggregate" U-shaped pattern over the sample period of about 6 years long, and to investigate the intraday patterns of the price effects of the trading intensity.

3.3 Descriptive statistics

Table 2 presents overall statistics of the tick-by-tick variables and of the estimated variables. Since the data sample is large, the sample is divided in 127 sub-samples from which primary mean statistics are obtained, which are then used to compute the statistics presented in Table 2. The ACD(1,1) model is estimated on each of the sub-samples.

<Insert Table 2 about here>

Table 2 shows that the return have a platokurtic distribution with a kurtosis coefficient much lower than 3. Negative skewness is an indication that the probability to observe a large negative jump in price for Nokia was greater than the probability of a positive jump during the sample period. The adjusted variables are all close to 1, as expected. Dealing with adjusted variables allows us to investigate the effects of excess dispersion in returns. Table 2 shows that Nokia is actively traded in HEX showing a mean duration about 15 seconds. The trade indicator variable is not zero as expected under the null hypothesis that a trade at buy (ask) is immediately followed by a trade at ask (buy). The ACD coefficients show that the model captures trading momentum in Nokia. Similarly, ψ_t and C_t deviate strongly from the expected mean of 1 and 0, respectively.

3.4 The Intraday pattern across the sample

Having access to about 6 years of transactions data gives us the opportunity to examine the long term trading dynamic in the Nokia stock. The U-shaped pattern characterizes trading momentum in three different trading periods throughout the trading day. Figures 1, 2 and 3 display the aggregate intraday trading patterns of volatility, trade weighted durations, and volume weighted durations, respectively

<Insert Figures 1, 2 and 3 about here>

Figure 1 shows interesting volatility patterns in a range where the highest volatility is about 39.7%, and the lowest volatility is about 14.2% in consistency with previous studies on Nokia. The highest volatility occurs invariably towards the market close and the closing period volatility is the quickest to converge to a normal-shaped distribution. Volatility appears to converge toward the end of the sixth year at about 20.5%. Volatility during the period around noon is the lowest and slowly converges to a normal shape, indicating a slow down in trading around noon time. Figures 2 and 3 show that the trading intensity is lowest at noon and highest at close, consistently with earlier evidence that transactions are concentrated at open and close. Comparing the three time periods, the graphs are mostly parallel, suggesting that the three intra-day periods evolve independently. However, transactions are more concentrated on low trading volumes (Figure 2) than on high trading volumes (Figure 3).

4. Empirical results

4.1 GMM estimates on tick-by-tick data

Table 3 reports the summary statistics of the GMM estimates on tick-by-tick data of the model of Eq. 2 in Panel A and of the model of Eq. 5 due to Madhavan, Richardson and Roomans (1997) in Panel B. The difference between the model of Eq. 2 and the Madhavan et al. (1997) model is that the model of Eq. 2 examines the price effect of the trade duration. Following Renault and Werker (2002), we split the duration effect into liquidity and information effects. Table 3 reports the summary statistics across the 127 sub-sets of the sample period.

<Insert Table 3 about here>

The results of Table 3 are interesting in number of ways. First, there is weak evidence, looking at $\hat{\phi}_2$ and $\hat{\vartheta}_1$, that duration exerts a great influence on prices. The two coefficients are significant in 29% of the 127 estimated sub-sets. Averaging across the sub-sets, Table 3 shows that both duration expectations and innovations in trading intensity are positively related to returns. Second, the order-handling coefficient, $\hat{\phi}_1$, is negative and significantly related to the return. The negative sign on $\hat{\phi}_1$ means that the price of Nokia felt by roughly €0.07 immediately after a sale had occurred. The autocorrelation of the order flow ρ_q is on average 0.14. From this coefficient, we estimate the conditional probability as $1-\lambda = 0.571$, where λ is the probability that a trade is executed at the bid (ask) is immediately followed by another trade at bid (ask) and the autocorrelation structure of the order flow is given by $1-2\lambda=\rho_q$. This finding holds for different market platforms (e.g. Hasbrouck (1991), and Madhavan et al. (1997) for the NYSE, Biais et al. (1995) for the Paris Bourse, and Ahn et al. (2002) for the Tokyo Stock Exchange). The implied bid-ask spread is €0.224. The proportion of the adverse selection cost in the implied bid ask spread is 23.8% and of this number the order-handling cost is 19.6%. Table 3 shows that trading frictions increase the variance bias. The proportion of variance attributable to pricing errors is 31.3%, whereas the proportion of variance attributable to pricing errors and trading frictions is 53.6%. Sixth, the ratio of the covariance to the variance bias indicates that the first order autocorrelation decreases the intensity effects are controlled. The autocorrelation is only -0.02 when theses effects are controlled, but -0.25 when these effects are disregarded.

4.2 Empirical results on the alternative models

We estimate the DeJong et al. (1996) model including trading size, and the Dufour and Engle (2000) model including the trade duration to cross-validate the findings of Table 3.

<Insert Table 4 about here>

Table 4 presents the OLS estimates of the DeJong et al. (1996) model in Panel A and the OLS estimates of the Dufour and Engle (2000) model in Panel B. The proportion of the adverse selection in the DeJong et al. (1996) model is 16.6% and in the Dufour and Engle model 17.6%. Both models show significant coefficients relating the trading intensity to the return. In particular, the Dufour and Engle (2000) model shows that duration is negatively related to return, suggesting that prices increase when traders observe shorter durations. The contrast between the models of Table 3 and Table 4 appears in the estimate of the implied spread. The implied bid-ask spread is €0.038 for the DeJong et al. model and €0.093 for the Dufour and Engle (2000) model, which are much lower than the implied bid-ask spread of Eq. 2 and Eq. 5.

4.3 Intraday patterns

Hasbrouck (1991), Madhavan et al. (1997), and Ahn et al. (2000) estimate their model on intraday day, as information models predict that price uncertainty should decrease toward the end of the trading day. We construct intraday data on the basis of tick-by-tick data. For each trading day, we obtain 4 observations. Using the average trade and volume durations, which we obtain according to Eq. (12) and Eq. (13), we obtain the return variable by aggregating the tick-by-tick returns across the trading period, and the trade indicator variable in the following way:

$$Q_j = \left[+1 \middle| \sum_{i=1}^{N} q_i > 0 \right] \text{ or } Q_j = \left[-1 \middle| \sum_{i=1}^{N} q_i < 0 \right], \tag{14}$$

where Q_j is the aggregate order flow at intraday time interval, j = 1, 2, 3 and 4 for each trading day. Since q_i is positively auto-correlated, Q_j will be wandering far away from zero. The aggregate order flow Q_j presents some advantages over the observed order flow, q_t. One advantage is that Q_j can be viewed as the periodic expected order flow, representing the underlying information during a given trading period. Another advantage is that Q_j is robust to microstructure effects due to multiple transactions on one unit information.

<Insert Table 5 about here>

Table 5 reports the intraday GMM estimates. The M1 model is estimated on intraday data from 10.00am to 6.00pm, the M2 model on data from 10.00am to 12.00am, the M3 model on data from 12.00am to 2.00pm, the M4 model on data from 2.00pm to 4.00pm, and the M5 model on data from 4.00pm to 6.00pm. The 5 models are estimated for the model of Eq. 2, which estimates are reported in Panel A, and for the model of Eq. 5, which estimates are reported in Panel B. Following results can be highlighted. In consistency with Dufour and Engle (2000), the association between duration and return is significant, with both positive liquidity and negative information effects on returns. The information effect is decreasing toward the end of the trading day looking at the model of Eq. 2. Trades are more persistent in the middle of trading day than at other times of the trading day. The ratio of the covariance bias to the variance bias is positive for the model of Eq. 2, but negative for the model of Eq.5.

<Insert Figure 4 about here>

Consistent with Ahn et al. (2002) who analyse data from the Tokyo Stock Exchange, the adverse selection and the order-handling component are highest at openings and closures, but lowest at the middle of the trading day. Figure 4 displays the intraday pattern of the adverse selection component. The U-shaped pattern is stronger for the MRR graph than for the BSW graph. Table (5) and Figure (4) show that the information effect is increasing in the duration effect, and that the price effect of duration consists of liquidity and information effects exerting an influence on the autocorrelation of the order flow and the variance process.

5. Conclusions

The thrust of this paper is to examine the role of time in a pure limit order book market. Building on the Madhavan et al. (1997) model, we include duration in a simple empirical model to investigate the effect of time on returns. In their empirical study, Dufour and Engle (2000) document that durations affect the trading behaviour of market makers in the New York Stock Exchange (NYSE). We extend their study by splitting the duration effect into liquidity and information effects. Our simple structural model allows us to derive a measure of the quality of the market for a stock. We obtain the following results. Firstly, duration effects on returns are stronger when they are aggregated, and durations exert an influence on the return process in consistency with Dufour and Engle (2000). Secondly, the variance bias due to pricing errors and trading frictions lies between 53.6% and 59.8%. Interestingly, we find that the ratio of the covariance bias to the variance bias is lower in the return model of trading intensity. Thirdly, we find that adverse selection cost exhibits a U-shaped pattern. A plausible explanation for the U-shaped pattern is that traders in Helsinki revise their prices in Nokia during the time when HEX is closing and NYSE is opening. Overall, our study provide evidence for that durations influence returns, the order flow and the variance process, suggesting that managing time is an important aspect of trading on pure limit order book markets such as the Helsinki Stock Exchange.

Appendix

The derivation of the second moment of the return:

$$
\begin{aligned}
E[r_t]^2 &= \underbrace{E\varepsilon_t^2}_{\sigma_\varepsilon^2} + \underbrace{E2\varepsilon_t(n_t - n_{t-1})}_{=0} + \underbrace{E(n_t - n_{t-1})^2}_{2\sigma_n^2} + \underbrace{E2\rho_q\theta_1\varepsilon_t q_t}_{=0} + \underbrace{E2\phi_2 c_t\varepsilon_t q_t}_{=0} \\
&+ \underbrace{E2\rho_q\theta_1 q_t(n_t - n_{t-1})}_{=0} + \underbrace{E2\phi_2 c_t q_t(n_t - n_{t-1})}_{=0} + \underbrace{\rho_q\theta_1^2 q_t^2}_{\rho_q\theta_1^2} + \underbrace{E2\phi_2\rho_q\theta_1 c_t q_t^2}_{=0} \\
&+ \underbrace{E\phi_2^2 c_t^2 q_t^2}_{\phi_2^2} + \underbrace{E2\phi_1\varepsilon_t q_t\psi_t}_{=0} + \underbrace{E2\phi_1 q_t\psi_t(n_t - n_{t-1})}_{=0} + \underbrace{E2\phi_1\rho_q\theta_1 q_t^2\psi_t}_{2\phi_1\rho_q\theta_1} \\
&+ \underbrace{E\phi_1\rho_q\theta_1 c_t q_t^2\psi_t}_{=0} + \underbrace{E\phi_1^2 q_t^2\psi_t^2}_{\phi_1^2} \\
&= \sigma_\varepsilon^2 + 2\sigma_n^2 + \rho_q\theta_1^2 + \phi_2^2 + 2\phi_1\rho_q\theta_1 + \phi_1^2
\end{aligned}
$$

This variance expression is evaluated under $E\psi_t = E\psi_t^2 = 1$, $Ec_t = 0$ and $Ec_t^2 = 1$, $Eq_t = 0$ and $Eq_t^2 = 1$, $E\varepsilon_t = 0$ and $E\varepsilon_t^2 = \sigma_\varepsilon^2$, and $En_t = 0$ and $En_t^2 = \sigma_n^2$. See Engle and Russell (1998) for the properties of ψ_t.

The derivation of the first order covariance:

$$
\begin{aligned}
&E[r_t, r_{t-1}] \\
&= \underbrace{E\phi_1^2 q_t q_{t-1}}_{\phi_1^2\rho_q} + \underbrace{E\phi_1\phi_2 q_t q_{t-1}\psi_{t-1}}_{=\phi_1\phi_2\rho_q} + \underbrace{E\phi_1\rho_q\theta_1 q_t q_{t-1} c_{t-1}}_{=0} + \underbrace{E\phi_1 q_t(n_{t-1} - n_{t-2})}_{=0} + \underbrace{E\phi_1 q_t\varepsilon_{t-1}}_{=0} \\
&+ \underbrace{E\phi_2\phi_1\psi_t q_t q_{t-1}}_{\phi_1\phi_2\rho_q} + \underbrace{E\phi_2^2 q_t q_{t-1}\psi_t\psi_{t-1}}_{\phi_2^2\rho_q} + \underbrace{E\phi_2\rho_q\theta_1\psi_t q_t q_{t-1} c_{t-1}}_{=0} + \underbrace{E\phi_2\psi_t q_t(n_{t-1} - n_{t-2})}_{=0} \\
&+ \underbrace{E\phi_2\psi_t q_t\varepsilon_{t-1}}_{=0} + \underbrace{E\rho_q\theta_1\phi_1 q_{t-1} q_t c_t}_{=0} + \underbrace{E\rho_q\theta_1\phi_2 c_t\psi_{t-1} q_t q_{t-1}}_{=0} + \underbrace{E\rho_q\theta_1^2 c_t c_{t-1} q_t q_{t-1}}_{\rho_q\theta_1^2\rho_q} \\
&+ \underbrace{E\rho_q\theta_1 c_t q_t(n_{t-1} - n_{t-2})}_{=0} + \underbrace{E\rho_q\theta_1 c_t q_t\varepsilon_{t-1}}_{=0} + \underbrace{E(n_t - n_{t-1})\phi_1 q_{t-1}}_{=0} + \underbrace{E(n_t - n_{t-1})\rho_q\phi_2\psi_{t-1} q_{t-1}}_{=0} \\
&+ \underbrace{E(n_t - n_{t-1})\rho_q\theta_1 c_{t-1} q_{t-1}}_{=0} + \underbrace{E(n_t - n_{t-1})(n_{t-1} - n_{t-2})}_{-\sigma_n^2} + \underbrace{E(n_t - n_{t-1})\varepsilon_{t-1}}_{=0} \\
&+ \underbrace{E\varepsilon_t\phi_1 q_{t-1}}_{=0} + \underbrace{E\varepsilon_t\phi_2\psi_{t-1} q_{t-1}}_{=0} + \underbrace{E\varepsilon_t\rho_q\theta_1 c_{t-1} q_{t-1}}_{=0} + \underbrace{E\varepsilon_t(n_{t-1} - n_{t-2})}_{=0} + \underbrace{E\varepsilon_t\varepsilon_{t-1}}_{=0} \\
&= \phi_1^2\rho_q + \phi_1\phi_2\rho_q + \rho\theta_1^2\rho_q + \phi_1\phi_2\rho_q + \phi_2^2\rho_q - \sigma_n^2
\end{aligned}
$$

The covariance expression is evaluated under $E\psi_t = E\psi_t^2 = 1$ and $E[\psi_t\psi_{t-1}] = 0$, $Ec_t = 0$, $Ec_t^2 = 1$ and $E[c_t c_{t-1}] = 0$, $Eq_t = 0$, $Eq_t^2 = 1$ and $E[q_t q_{t-1}] = \rho_q$, $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma_\varepsilon^2$ and $E[\varepsilon_t\varepsilon_{t-1}] = 0$, and $En_t = 0$, $En_t^2 = \sigma_n^2$ and $E[n_t n_{t-1}] = 0$. See Engle and Russell (1998) for the properties of ψ_t.

References

Admati, A., and P. Pfleiderer (1988), 'A theory of intraday patterns: Volume and price variability,' The Review of Financial Studies, 1, 3-40.

Ahn, H-J., J. Cai, Y. Hamao, R. Y. K. Ho (2002), 'The components of the bid-ask spread in a limit-order market: evidence from the Tokyo Stock Exchange,' Journal of Empirical Finance, 9, 399-430.

Andersen, T.G. and T. Bollerslev (1997), 'Intraday periodicity and volatility persistence in financial markets,' Journal of Empirical Finance, 1997, 115-158.

Back, K. and S. Baruch (2004), 'Information in securities markets: Kyle meets Glosten and Milgrom,' Econometrica, 2, 433-467.

Biais, B., Hillion, P., C. Spatt (1995), 'An empirical analysis of the limit order book and the order flow in the Paris Bourse,' Journal of Finance 50, 1655-1689.

De Jong, F., Nijman, T. and A. Röell (1996), 'Price effects of trading and components of the bid-ask spread on the Paris Bourse,' Journal of Empirical Finance 3, 193-213.

Diamond, D. and R.E. Verrecchia (1987), 'Constraints on short-selling and asset price adjustments to private information,' Journal of Financial Economics 18, 277-311.

Dufour, A. and R.F. Engle (2000), 'Time and the Price Impact of a Trade,' Journal of Finance 55, 2467-2498.

Easley, D. and M. O'Hara (1992), 'Time and the process of security price adjustment,' Journal of Finance 47, 1992, 557-605.

Engle, R.F. (2000), 'The econometric of ultra-high frequency data,' Econometrica 68, 1-22.

Engle, R.F. and A. Lunde (2003), 'Trades and quotes: A bivariate point process,' Journal of Financial Econometrics 1, 159-188.

Engle, R.F. and J. Lange (2001), 'Predicting VNET: A model of the dynamics of market depth,' Journal of Financial Markets 4, 113-142.

Engle, R.F. and J.R. Russell (1997), 'Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model,' Journal of Empirical Finance, 4, 187-212.

Engle, R.F. and J.R. Russell (1998), 'Autoregressive conditional duration: a new model for irregularly spaced transactions data,' Econometrica, 66, 1127-1162.

Foucault, T., O. Kadan, and E. Kandel (2003), 'Limit order book as a market for liquidity' Working Paper, HEC School of Management.

George, T. J., G. Kaul and M. Nimalendran (1991), 'Estimation of the bid-ask spread and its components: A new approach,' Review of Financial Studies 4, 623-656.

Ghysels, E. and J. Jasiak (1998), 'GARCH for irregularly spaced financial data: The ACD-GARCH model,' Studies on Nonlinear Dynamics and Econometrics, 2, 133-149.

Ghysels, E., C. Gouriéroux and J. Jasiak (2004), 'Stochastic volatility duration models,' Journal of Econometrics, 119, 413-433.

Glosten, L. and P. Milgrom (1985), 'Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,' Journal of Financial Economics, 14, 71-100.

Glosten, L. R. (1987), 'Components of the bid-ask spread and the statistical properties of transaction prices,' Journal of Finance 42, 1293-1307.

Glosten, L.R. and L.E. Harris (1988), 'Estimating the components of the bid-ask spread,' Journal of Financial Economics, 21, 123-142.

Gouriéroux, C., J. Jasiak, and G. Le Fol (1999), 'Intraday trading activity,' Journal of Financial Markets, 2, 1999, 193-226.

Hansen, L. (1982), 'Large sample properties of Generalized Method of Moments Estimators,' Econometrica, 50, 1029-1084.

Harris, L. (1990), 'Estimation of stock price variances and serial covariances from discrete observations,' Journal of Financial and Quantitative Analysis, 25, 291-306.

Hasbrouck, J. (1988), 'Trades, quotes, investors, and liquidity,' Journal of Financial Economics 22, 229-252.

Hasbrouck, J. (1991), 'Measuring the information content of stock trades,' Journal of Finance 46, 179-207.

Hasbrouck, J. (1993), 'Assessing the quality of a security market: A new approach to transaction-cost measurement,' Review of Financial Studies, 6, 191-212.

Huang, R.D. and H.R. Stoll (1997), 'The Components of the bid-ask spread: A general approach,' Review of Financial Studies 10, 995-1034.

Lin, J-C., G. C. Sanger and G. G. Booth (1995), 'Trade size and the components of the bid-ask spread,' Review of Financial Studies 8, 1153-1183.

Madhavan A., M. Richardson and M. Roomans (1997), 'Why do security prices change? A transaction-level analysis of NYSE stocks,' Review of Financial Studies 10, 1035-1064.

Manganelli, S. (2002), 'Duration, volume and the price impact of trades,' Working paper, European Central Bank Nr 125.

Renault, E. and B. J. M., Werker (2002), 'Stochastic volatility with transaction time risk,' Discussion Paper, University of Montreal.

Roll, R. (1984), 'A simple implicit measure of effective bid-ask spread in an efficient market,' Journal of Finance, 39, 1127-1139.

Saar, G. and J. Hasbrouck (2002), 'Limit orders and volatility in a hybrid market: The island ECN,' Working Paper, New York University.

Spierdijk, L. (2004), 'An empirical analysis of the role of the trading intensity in information dissemination on the NYSE,' Journal of Empirical Finance, 11, 2004, 163-184.

Stoll, H. R., (1989), 'Inferring the components of the bid-ask spread: Theory and empirical tests, Journal of Finance 33, 1133-1151.

Table 1
Data preparation
Table 1 reports the different data preparation levels. The data set is obtained from Reuters, and runs from April 12, 1999 to December 30, 2004. First, the tick-by-tick data is divided in 202 sub-sets of about 60,000 observations. Second, by thinning the price process some redundant observations are eliminated. Third, observations falling outside the trading session from 10.30am to 6.00pm are deleted. Finally, the intraday data are constructed on the time adjusted data.

Preparation Level	Actions undertaken to prepare the data set	Observations	Sub-sets
I. Tick-by-tick data	As it is. Including transactions for the three different trading sessions in the HEX.	6,753,243	202
II. Price duration	Thinning process as following. If price at i equals price at i+1, then price at i+1 is selected, price at i ignored, and trade volume at i summed with volume at i+1.	2,309,717	127
III. Estimation data	Transactions outside the trading session disregarded and volumes at the same time aggregated. The ACD and the GMM model are estimated with this data.	1,738,462	127
IV. Intraday data	Created on estimation data. Used a normal density kernel to obtain weighted variables at four regular spaced time intervals of the trading day.	5,452	1

Table 2
Summary statistics on tick-by-tick data and the ACD model
Table 2 reports the overall mean of the observed and estimated variables. These summary statistics are on the mean of each sub-set of the sample. There are in total 127 sub-sets, for each denoted by the subscript i, thus i runs from 1 to 127. r_i is the mean return in percent, s_i is the relative mean spread, $\tilde{x}_i$ is the adjusted duration mean, x_i is the duration mean, $\tilde{z}_i$ is the adjusted trading volume mean, z_i is the trading volume mean, and q_i is the trade indicator mean. ψ_i is the expected duration mean from an ACD(1,1) model of Engle and Russell (1998) according to $\psi_t = a_0 + a_1\tilde{x}_{t-1} + a_2\psi_{t-1}$, where a_{i0}, a_{i1} and a_{i2} are positive coefficients. C_i is the signed duration innovation mean, and y_i is the signed expected duration mean.

	Mean	Median	Std Dev	Kurtosis	Skewness	Minimum	Maximum
r_i	5.E-06	1.E-05	0.0003	0.930	-0.470	-0.001	0.001
s_i	1.E-03	0.0010	0.0003	2.763	1.271	0.001	0.002
$\tilde{x}_i$	1.0001	0.9999	0.0013	57.050	6.683	0.998	1.012
x_i	8.0445	7.6803	2.1134	0.937	0.707	2.972	14.70
$\tilde{z}_i$	1.0024	1.0003	0.0089	51.055	6.469	0.992	1.081
z_i	6 618	6 472	2 883	1.487	0.721	1 139	15 685
ψ_i	1.0084	1.0065	0.0126	21.669	2.801	0.957	1.092
q_i	0.0360	0.0167	0.1085	3.167	1.493	-0.160	0.456
C_i	0.0831	0.0716	0.5672	34.602	2.443	-2.987	4.457
y_i	0.0361	0.0122	0.1062	2.851	1.377	-0.183	0.409
a_{i0}	0.0287	0.0245	0.0190	4.388	1.813	0.004	0.109
a_{i1}	0.0739	0.0740	0.0192	-0.146	0.311	0.037	0.135
a_{i2}	0.8983	0.9016	0.0310	1.902	-0.989	0.765	0.949

Table 3
Summary statistics over the GMM estimates on tick-by-tick data
Panel A reports the summary statistics of the estimates of our model for trading intensity: $\varepsilon_t = r_t - \lfloor \phi_1 \Delta q_t + (\phi_2 \psi_t + \rho_q \theta_1 C_t) q_t + n_t \rfloor$, where r_t is the return, q_t is the trade indicator variable, $\Delta q_t = q_t - q_{t-1}$, ψ_t is the expected duration , C_t is innovation in trading intensity, ε_t is the random error, and n_t is the pricing error.
Panel B reports the summary statistics of the estimates of the Madhavan et al. (1997) model: $\varepsilon_t = r_t - \lfloor \theta(q_t - \rho_q q_{t-1}) + \phi \Delta q_t + n_t \rfloor$,
S is the implied spread, ϑ_1 is the proportion of adverse selection costs in S and ϑ_2 is the proportion of liquidity costs in S.
π_1 is the proportion of variance bias due to pricing errors.
π_2 is the proportion of variance due to pricing errors and trading frictions.
ρ is the first autocorrelation of the trade process.

	Mean	Std Error	Median	Std Dev	Minimum	Maximum
	Panel A: Descriptive Statistics of the GMM Estimates for Model 1					
ϕ_1	-0.0687	0.0389	-0.0369	0.4381	-4.7968	1.1745
ϕ_2	0.0123	0.0274	0.0018	0.3084	-1.5924	2.7462
θ_1	0.0212	0.0284	0.0000	0.3196	-0.4979	3.5041
ρ_q	0.1421	0.0079	0.1169	0.0891	-0.0001	0.6503
σ_ε^2	0.3164	0.2291	0.0030	2.5814	0.0004	27.6500
σ_n^2	0.0508	0.0367	0.0006	0.4135	-0.1924	3.5691
S	0.2242	0.0819	0.0841	0.9234	0.0187	9.4234
ϑ_1	0.2383	0.0178	0.1817	0.2009	0.0000	0.8277
ϑ_2	0.1964	0.0161	0.1402	0.1818	0.0008	0.9414
π_1	0.3126	0.0141	0.3223	0.1584	0.0026	0.8566
π_2	0.5355	0.0130	0.5474	0.1465	0.0638	0.9206
ρ	-0.0203	0.0026	-0.0184	0.0292	-0.1058	0.1611
	Panel B: Descriptive Statistics of the GMM estimates for Model 2					
ϕ	-0.0461	0.0053	-0.0397	0.0601	-0.5092	0.2219
θ	0.0292	0.0253	0.0018	0.2851	-0.3103	3.0892
ρ_q	0.1415	0.0071	0.1178	0.0797	0.0613	0.5104
σ_ε^2	0.0123	0.0067	0.0026	0.0756	-0.2048	0.7968
σ_n^2	0.0022	0.0026	0.0005	0.0290	-0.1720	0.2675
ϑ_1	0.2232	0.0159	0.1681	0.1796	0.0071	0.9900
π_1	0.3511	0.0146	0.3418	0.1644	0.0035	0.8661
π_2	0.5984	0.0160	0.6110	0.1801	0.0951	0.9900
S	0.1045	0.0260	0.0536	0.2926	0.0043	3.1203
ρ	-0.2510	0.0087	-0.2697	0.0983	-0.4363	0.0484

Table 4
Descriptive statistics upon the 127 OLS model estimates on tick-by-tick data
Panel A reports the summary statistics of the OLS estimates of the De Jong et al. (1996) model: $r_t = a_0 + \left(R_0 \Delta q_t + R_1 \Delta q_t \tilde{z}_t\right) + \left(e_0 q_{t-1} + e_1 q_t \tilde{z}_{t-1}\right) + v_t$, where $\tilde{z}_t$ is the adjusted trading volume, q_t is the trade indicator variable and v_t is the error term. In this model, the order-handling cost is given by $\varphi = c_0 + c_1 \tilde{z}_i$, where $\tilde{z}_i$ is the mean of the adjusted trading volume for i = 1, …, 127 the total number of estimated models, $c_0 = R_0 - e_0 - e_1 \alpha$, $\alpha = \tilde{z}_i / \ln(2)$, $c_1 = R_1 - (1/2) e_1$ and the adverse selection cost is given by $\vartheta = (R_0 - c_0) + (R_1 - c_1)\tilde{z}_i$.
Panel B reports the summary statistics of the OLS estimates of the Dufour and Engle (2000) model: $r_t = (\gamma_0 q_t + \gamma_1 q_{t-1}) + [\delta_0 \ln(x_t) q_t + \delta_1 \ln(x_{t-1}) q_{t-1}] + v_t$, where x_t is the trade duration. In this model, the order-handling cost is given by $\varphi = |(\gamma_0 - \gamma_1)| q$ for $q = 1$, and the adverse selection cost by $\vartheta = |(\delta_0 - \delta_0)| \ln(x_i)$. The implied spread is given by $S = 2(\vartheta + \varphi)$, and the proportion of the adverse selection cost in spread by $\vartheta_1 = \vartheta / (\varphi + \vartheta)$.

	Mean	Std Error	Median	Std Dev	Minimum	Maximum
	Panel A : The De Jong, Nijman and Roell (1996) Model					
a_0	0.0003	0.0001	0.0001	0.0009	-0.0014	0.0040
R_0	-0.0202	0.0007	-0.0188	0.0076	-0.0537	-0.0095
R_1	-0.0006	0.0000	-0.0005	0.0005	-0.0037	0.0002
e_0	-0.0063	0.0006	-0.0070	0.0066	-0.0285	0.0145
e_1	0.0005	0.0000	0.0005	0.0002	0.0000	0.0013
c_0	-0.0146	0.0005	-0.0133	0.0062	-0.0358	-0.0061
c_1	-0.0009	0.0000	-0.0008	0.0005	-0.0042	0.0000
α	1.4462	0.0011	1.4431	0.0128	1.4309	1.5594
φ	-0.0159	0.0006	-0.0140	0.0064	-0.0376	-0.0079
ϑ	0.0023	0.0002	0.0028	0.0028	-0.0066	0.0081
ϑ_1	0.1658	0.0056	0.1668	0.0630	0.0139	0.3244
S	0.0381	0.0013	0.0351	0.0151	0.0173	0.0884
	Panel B: The Dufour and Engle (2000) Model					
γ_0	-0.0250	0.0008	-0.0231	0.0093	-0.0707	-0.0133
γ_1	0.0170	0.0006	0.0149	0.0070	0.0081	0.0401
δ_0	0.0030	0.0001	0.0028	0.0015	0.0007	0.0106
δ_1	-0.0015	0.0001	-0.0014	0.0007	-0.0037	0.0000
φ	0.0420	0.0013	0.0384	0.0149	0.0223	0.1043
ϑ	0.0090	0.0004	0.0085	0.0040	0.0023	0.0241
ϑ_1	0.1738	0.0035	0.1687	0.0395	0.0732	0.2698
S	0.0931	0.0029	0.0857	0.0332	0.0493	0.2327

Table 5: GMM Estimates of the intraday models

The model of Panel A: $\varepsilon_j = R_j - \left[\left(\phi_1 \Delta Q_j + \left(\phi_2 \Lambda_j^x + \hat{\rho}_q \theta_1 \Lambda_j^v\right) Q_j\right) + n_j\right]$, where Q_j is the aggregate order flow, Λ_j^x is the weighted trade duration, Λ_j^v is the weighted volume duration, R_j is the aggregate return, ε_j is the error term, and n_j is the pricing error. $J(\beta)$ is the J-statistic, and $J(\beta) - PV$ is the P-value of $J(\beta)$. The model of Panel B: $\varepsilon_j = R_j - \left[\left(\phi \Delta Q_j + \theta\left(Q_j - \rho_Q Q_{j-1}\right)\right) + n_j\right]$. S is the implied spread, ϑ_1 is the proportion of adverse selection costs in S, ϑ_2 is the proportion of liquidity costs in S, and ϑ_3 is the proportion of adverse selection costs in S, π_1 is the proportion of the variance bias due to pricing errors, and ρ is the first order autocorrelation of the trade process (return). one asterisk (*) means that the coefficient is statistically significant at 5% level, at least.

Model	M1:10.00am – 6.00pm	M2:10.00am – 12.00pm	M3:12.00am – 2.00pm	M4:2.00pm – 4.00pm	M5:4.00pm – 6.00pm
	Panel A: The Intraday Models: The Intensity Trading Intensity Approach				
ϕ_1	-0.0390*	-0.0078*	-0.0054*	-0.0072*	-0.0124
ϕ_2	0.0145*	-0.0230*	0.0211*	0.0096*	0.0173
θ_1	-0.0493*	0.2563*	-0.0805*	-0.0440*	-0.0965
ρ_q	0.2481*	0.1289*	0.2868*	0.2536*	0.1674*
σ_ε^2	0.0005*	0.0012*	0.0002*	0.0002*	0.0007*
σ_n^2	6.E-04*	-8.E-04	-1.E-05	3.E-05*	-8.E-05
S	0.2055	0.5742	0.2140	0.1217	0.2523
ϑ_1	0.2398	0.4463	0.3761	0.3613	0.3825
ϑ_2	0.0706	0.0401	0.0986	0.0792	0.0684
ϑ_3	0.6897	0.5136	0.5253	0.5595	0.5491
π_1	0.7057	0.5583	0.0907	0.2756	0.1854
ρ	0.1088	0.1324	0.1169	0.0788	0.1129
$J(\beta)$	19.880	296.348	26.972	13.995	15.531
$J(\beta) - PV$	0.1340	0.0000	0.0795	0.4501	0.6252
	Panel B: The Intraday Models: The Madhavan et al. (1997) Model				
ϕ	-0.0319*	-0.0035*	-0.0050*	-0.0074*	0.0050
θ	0.0038*	-0.0039*	0.0014	0.0017*	0.0045
ρ_q	0.2483*	0.1698*	0.3427*	0.2481*	0.1638*
σ_ε^2	0.0002*	0.0004*	0.0002*	0.0002*	0.0005*
σ_n^2	0.0006*	-0.0001	-7.E-06	3.E-05*	-4.E-06
S	0.0713	0.0148	0.0128	0.0181	0.0190
ϑ_1	0.0530	0.2657	0.1085	0.0919	0.2373
ϑ_3	0.9470	0.7343	0.8915	0.9081	0.7627
π_1	0.8583	0.2090	0.0718	0.3004	0.0184
ρ	-0.4021	-0.1264	-0.0621	-0.1955	-0.0679
$J(\beta)$	24.4055	25.6082	19.6674	14.2800	14.9792
$J(\beta) - PV$	0.0585	0.0423	0.1851	0.5044	0.4529

Figure 1
Intraday Volatility Patterns

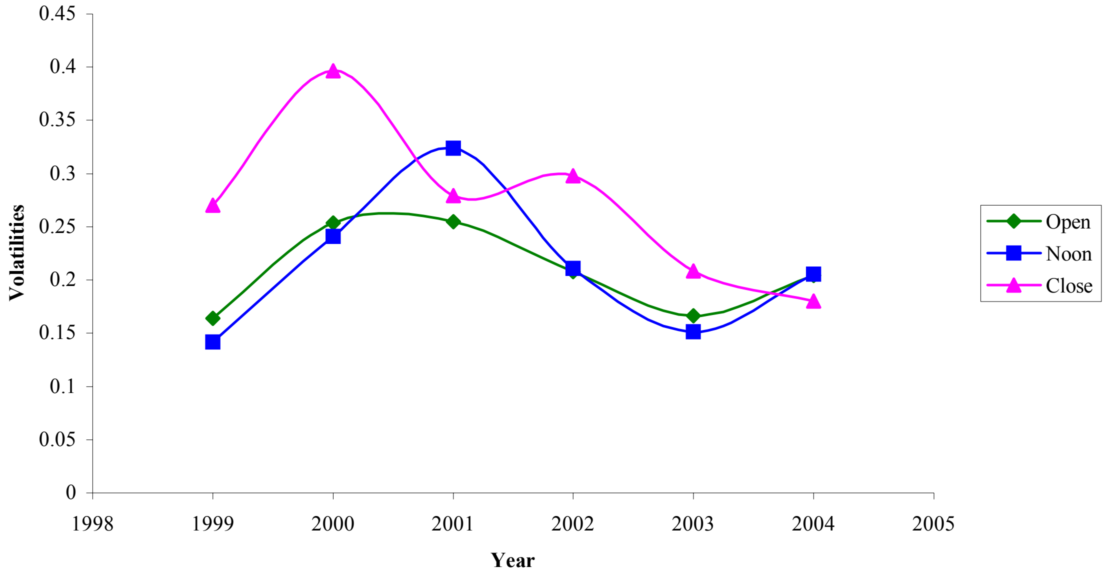

Figure 1 reports the volatility pattern across the six years sample. Volatilities are obtained in the following manners. First, the logarithmic returns are computed at transactions-data level. Second, the logarithmic returns are squared. Third, the squared returns are summed for three different trading periods of each trading day of the sample. The X-axis represents the six years of the entire data set. The Y-axis are volatilities for the trading period after the open; the trading period around noon denoted, and the trading period towards the close, respectively.

Figure 2
Trade-based Intensity Measures

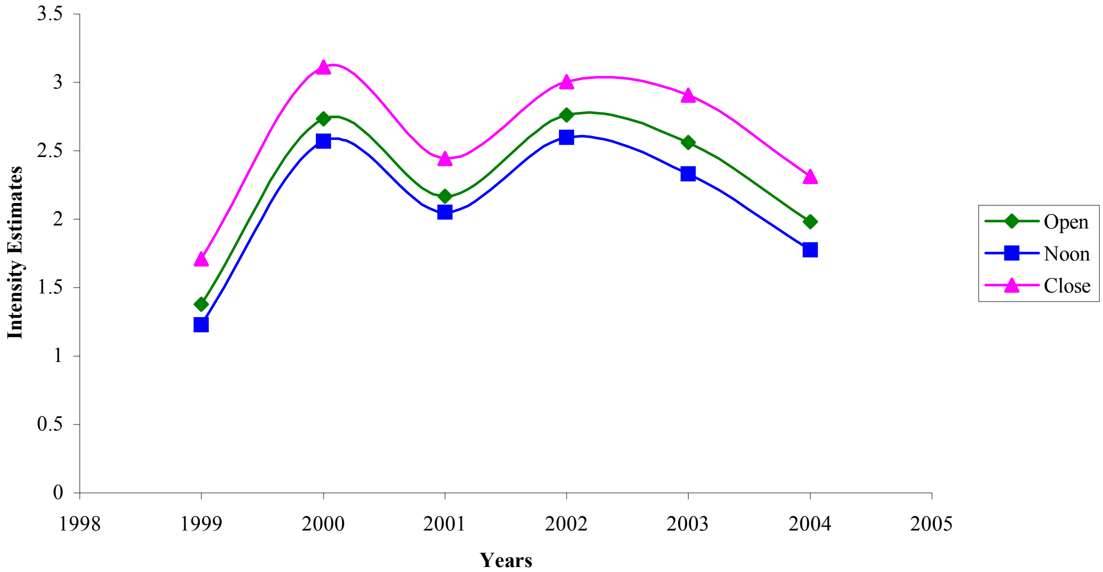

Figure 2 reports trade-based intensity measures. The intensity measures are obtained in the following manners. First, trade-durations are computed as the difference in seconds between consecutive transaction times. Second, a normal density kernel is used to weight durations for each trading period of the trading day. The weighted duration measures represent the intensity measures and are computed following Equation (16). Third, the intensity measures are summed for each trading period of the trading day of the sample and weighted using as scaling factor the number of transactions per trading period. Fourth, the measures are summed trading period for trading period over the years. The X-axis represents the 6 years of the sample. The Y-axis represents the intensity measures for the open trading period denoted as 1, the noon trading period denoted as 2, and the close trading period denoted as 3, respectively. The lowest intensity value is about 1.23 and the highest intensity value is about 3.11.

Figure 3
Volume-based Intensity Measures

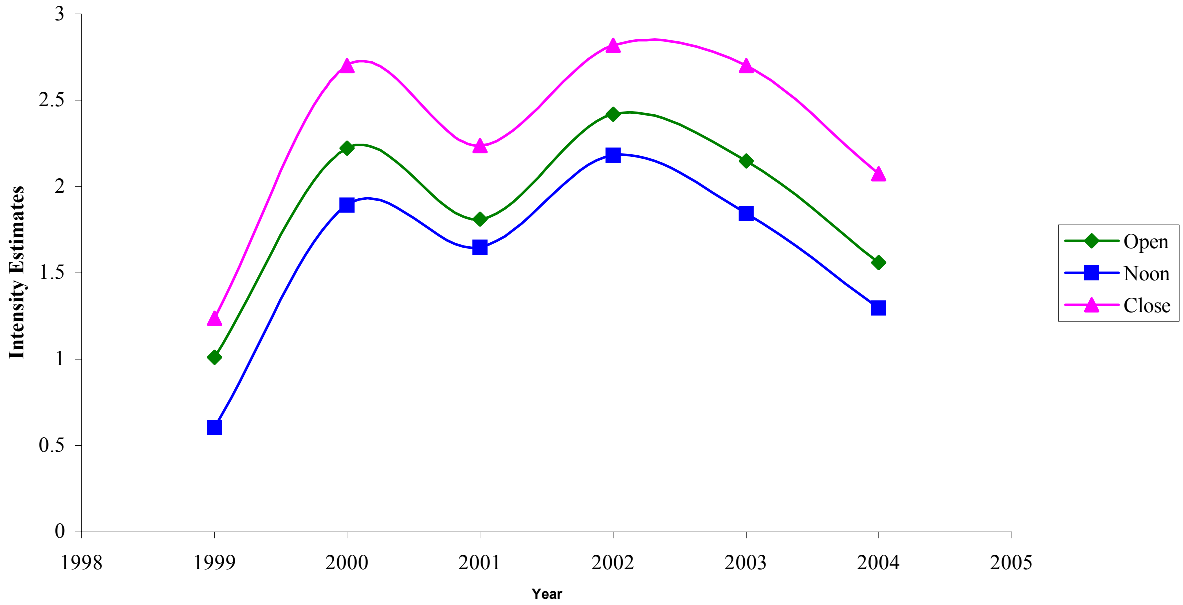

Figure 3 reports the Kernel-based intensity measures. The intensity measures are obtained in the following manners. First, volume-durations are computed as the difference in seconds between consecutive transaction times associated with volume over 10,000 numbers of shares traded. Second, a normal density kernel is used to weight durations for each trading period of the trading day. The weighted duration measures represent the intensity measures and are computed following Equation (17). Third, the intensity measures are summed for each trading period of the trading day of the sample and weighted using as scaling factor the number of transactions per trading period. Fourth, the measures are summed trading period for trading period over the years. The X-axis represents the 6 years of the sample. The Y-axis represents the intensity measures for the open trading period denoted as 1, the noon trading period denoted as 2, and the close trading period denoted as 3, respectively. The lowest intensity value is 0.60 and the highest intensity value is about 2.70.

Figure 4
The intraday pattern of the adverse selection component

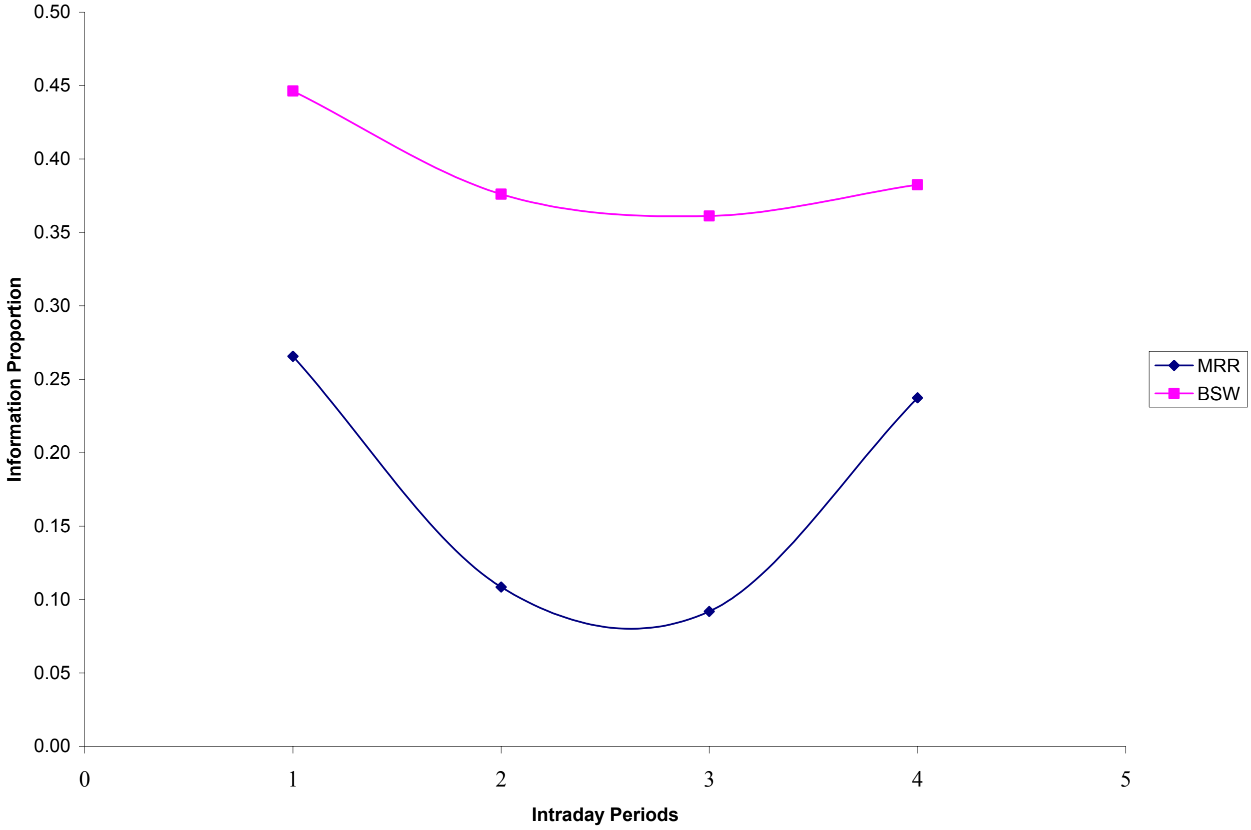

Figure 4 exhibits the intraday adverse selection cost patterns. The Y-axis displays the proportion of the adverse selection cost in the implied bid-ask spread. The X-axis displays the intraday periods. The MRR is the graph on GMM estimates of the model of Eq. 5 due to Madhavan, Richardson and Roomans (1997) (MRR). The BSW is the graph on the intraday GMM estimates of our model of Eq. 2.