The Components of the Bid-Ask Spread in a Limit-Order Market:
Evidence from the Tokyo Stock Exchange

Hee-Joon Ahn
College of Economics and Commerce
Sookmyung Women's University
Seoul, Korea

Jun Cai
Department of Economics and Finance
City University of Hong Kong
Tat Chee Avenue
Kowloon, Hong Kong

Yasushi Hamao*
Department of Finance and Business Economics
Marshall School of Business
University of Southern California
Los Angeles, CA 90089-1427

Richard Y. K. Ho
Department of Economics and Finance
City University of Hong Kong
Tat Chee Avenue
Kowloon, Hong Kong

Forthcoming in *Journal of Empirical Finance*

The Components of the Bid-Ask Spread in a Limit-Order Market: Evidence from the Tokyo Stock Exchange

Abstract

This paper analyzes the components of the bid-ask spread in the limit-order book of the Tokyo Stock Exchange (TSE). While the behavior of spread components in U.S. markets has been extensively studied, little is known about the spread components in a pure limit-order market. We find that both the adverse selection and order handling cost components of the TSE exhibit U-shape patterns independently, in contrast to the findings of Madhavan, Richardson, and Roomans (1997) for U.S. stocks. On the TSE, there does not exist an upstairs market that allows large trades to be prenegotiated or certified as on the New York Stock Exchange (NYSE). This feature of the TSE provides a valuable opportunity to examine the relationship between trade size and spread components. Our results show that the adverse selection cost increases with trade size while order handling cost decreases with it.

Key words: Bid-ask spread, adverse selection cost, order-processing cost, trade size.

JEL classification: C14; G15

1. Introduction

Over the past years, limit-order trading has received growing attention as more exchanges implement electronic public limit-order books and open up the market-making process. A number of studies have examined various aspects of the limit-order market. In particular, Glosten (1994), Handa and Schwartz (1996), Rock (1996), Seppi (1997), Viswanathan and Wang (1998), and Foucault (1999) offer a variety of equilibrium models on limit-order trading.[1] Biais, Hillion, and Spatt (1995) offer an empirical analysis of the supply and demand of liquidity and interaction between the order book and order flow in the Paris Bourse. Harris and Hasbrouck (1996) investigate the relative importance of market and limit orders. Ahn, Bae, and Chan (2001) analyze the interaction between transitory volatility and order flow composition in a limit-order market. Chung, Van Ness, and Van Ness (1999) and Kavajecz (1999) examine whether quoted spreads reflect the trading interest of specialists or limit-order traders.[2]

The purpose of this paper is to examine the components of the bid-ask spread in a limit-order market. Existing market microstructure theories on the components of the bid-ask spread are largely developed within the framework of quote-driven single (multiple) dealer markets. In addition to the order-processing costs, the bid-ask spread must cover the following two components: the inventory and information costs in a dealer market.[3] However, the bid-ask spread is not unique to the dealer markets. Cohen, Maier, Schwartz, and Whitcomb (1981) establish the existence of the bid-ask spread in a limit-order market when investors face transaction costs of assessing information, monitoring market, and conveying orders to the market. Glosten (1994) shows that the limit-order market will have a positive

---

[1] See Seppi (1997) for a more detailed summary of other equilibrium models of limit-order market.

[2] Other empirical studies of limit-order market include Frino and McCorry (1995) and Hollifield, Miller, and Sandås (1999).

[3] Demsetz (1968) and Tinic (1972) identify the order-processing costs incurred by the providers of market liquidity. Stoll (1978), Ho and Stoll (1983), and Amihud and Mendelson (1980), emphasize the inventory holding costs. Copeland and Galai (1983), Easley and O'Hara (1987), and Glosten and Milgrom (1985) concentrate on the information costs faced by liquidity suppliers when trading with informed traders.

bid-ask spread arising from the possibility of trading on private information. Nevertheless, empirical evidence on the bid-ask component in a limit-order market has been extremely limited.[4]

We examine the bid-ask component in a limit-order book of the Tokyo Stock Exchange (TSE). On the TSE, there are no designated market makers with an obligation to take positions in the market. Every transaction is executed by a *saitori* who maintains each offer to buy or sell in an order book, which is open to all exchange members on the floor. All liquidity is supplied by traders who submit limit or market orders. In this sense, the TSE may be better described as a market where a multiple number of dealers provide market-making at their own discretion (Takagi, 1993).

A number of earlier papers explored various aspects of the TSE. Amihud and Mendelson (1989, 1991, and 1993) examine liquidity provision and price discovery on the TSE. Hamao (1992) and Takagi (1993) present an overview of the institutional features of the TSE trading. Lindsey and Schaede (1992) compare the role of a *saitori* with that of the specialist of the NYSE. George and Hwang (1995) and Kim and Rhee (1997) investigate the effectiveness of the TSE price limit rules. Two recent studies examine the intertemporal behavior of the market microstructure on the TSE. Lehmann and Modest (1994) study the size of the bid-ask spread and its cross-sectional and intraday stability, among other issues. They report U-shape intraday patterns in bid-ask spreads, return volatility, and trading volume. Hamao and Hasbrouck (1995) investigate the properties of intraday trades and quotes on the TSE for a number of representative firms. They find evidence consistent with asymmetric information effects within the limit-order book.

Our primary findings are as follows. First, we found that both the adverse selection and order-processing cost components exhibit U-shape patterns. This contrasts with the finding by Madhavan, Richardson, and Roomans (1997) that on the NYSE the adverse selection component declines and order-process component increases during the day. The TSE evidence of an increase in adverse selection costs around the end of the trading day suggests that transactions around this period convey private information

---

[4] Brockman and Chung (1999) and Chan (2000) study the bid-ask components on the Stock Exchange of Hong Kong. De Jong, Nijman, and Roell (1996) examine the bid-ask component on the Paris Bourse.

2

that would otherwise be released during the non-trading hours that follow the exchange close. Second, we also find that adverse selection costs increase with trade size while order-processing costs decrease with it. This result is compared with the evidence on the NYSE that medium trades contain more information than large trades, as reported by Barclay and Warner (1994) and Huang and Stoll (1997). We believe that this difference in the relation between trade size and information content of a trade between the two exchanges comes from how the two exchanges treat large order flows. On the NYSE, large block orders are sent to the upstairs, where transactions are made through a search-brokerage mechanism. On the TSE, there is no separate venue for block trading. All trades, large or small, are consolidated into the central electronic order book, where investor identity is not revealed as in upstairs trading on the NYSE. Our evidence that large trades convey more information provides strong support for the theoretical prediction given by Easley and O'Hara (1987) that informed investors prefer to trade large volume. The majority of the world's stock markets take the form of the limit order system. Most of the extant studies, however, focus on quote driven markets or a hybrid system that adopts both limit order book and dealer system. Our study implies that the process of how information is incorporated into stock price through trading on a limit order market is different from the existing evidence reported based on the quote driven or hybrid system.

The remainder of the paper is organized as follows. Section 2 describes the tick-by-tick transaction data, the process of filtering the data, and summary statistics. Section 3 presents a simple characterization of the bid-ask component model and estimation procedure. Section 4 presents the empirical results for the cross-sectional difference and the intraday pattern of the bid-ask components. The relation between trade size and bid-ask components is also examined. Section 5 concludes the paper.

2. Data sources, filtering, and summary statistics

2.1 Data sources

We obtain the real-time TSE trades and quotes data from Nikkei Economic Electronic Database System (NEEDS) historical tick data. The database is time-stamped to the nearest minute and includes

the information on all quotes and trades in both price and quantity.[5] It also has detailed flags indicating the conditions of each trade and quote. These flags include the opening/closing trade indicators, buy/sell indictors, and special and warning quote indictors, among others.[6] The NEEDS database essentially reflects all the trade and quote information broadcast to TSE members by the TSE. The database is the most detailed and extensive among the known data sets on the Japanese stock market.

Our initial data set includes all transactions and quotes recorded for the 225 stocks that constitute the Nikkei Price Average Index. The Nikkei 225 component stocks, selected from the first section of the TSE, are highly liquid and widely known as a good representation of the Japanese stock market. The sample period is from January 5, 2000 to March 31, 2000, for a total of 60 trading days. The first trading day of the year, January 4, 2000, is excluded since trading was conducted only during the morning session of the day.

2.2 Data filtering process

2.2.1 Tick size

We use only the transactions and quotes on the TSE in our analysis.[7] On the TSE, the tick size is a step function of stock prices. For example, during the sample period used in this study, stocks priced less than or equal to ¥2,000 have a tick size of ¥ 1; stocks priced between ¥2,001 and ¥3,000 have a tick size of ¥5; and stocks priced between ¥3,001 and ¥30,000 have a tick size of ¥10. While some stocks are priced above ¥30,000 and traded on greater tick sizes, most stocks trade at prices below ¥30,000. In our initial sample of Nikkei 225 constituent stocks, six are priced above ¥30,000 and therefore excluded from the sample. Exclusion of these six firms is to avoid possible confounding effects on the spread due to significantly larger tick sizes. We then partition the remaining 219 firms into three groups based on individual tick sizes (i.e. ¥1, ¥5, and ¥10).

---

[5] Even if time-stamped to the minute, the data are recorded to preserve the sequence of events within each minute.
[6] On August 24, 1998, the TSE officially abolished warning quotes. However, the TSE order matching system is not updated until the end of year 2000 and the system occasionally shows warning quotes for some stocks.

2.2.2 Classification of buy and sell transactions

    To ensure a clean data set, we apply several filters to individual transactions and quotes for each stock in the sample. Each transaction during the morning and afternoon sessions is classified as a buyer or a seller-initiated trade by the price location flags available from the dataset. These flags indicate the location of each trade at the bid, ask, between the bid and ask, or outside the bid-ask range. Almost all of the trades for our sample firms are made either at the bid or ask. However, there is a small number of trades that are made either within or outside the spread. For an average firm in our sample, of the total 19,682 transactions made during the regular sessions for the entire three-month period, 0.12 percent (24 trades) are made within the spread while less than 0.01 percent (2 trades) are made outside the spread. Since the TSE is a pure order-driven market without responsible market makers, under normal conditions of a double auction, a trade should always hit either the bid or ask. Therefore, we exclude all trades made within or outside the spread.

2.2.3 Opening and closing transactions in the morning and afternoon sessions

    The trading day on the TSE is divided into morning (9:00-11:00 a.m. local time) and afternoon (12:30-3:00 p.m.) sessions. All observations recorded before 9:00 a.m. or after 3:00 p.m. are excluded from the sample. Therefore, we do not include transactions made during the off-hours trading session, which often deals with negotiated large block trades through a system called ToSTNet (Tokyo Stock Exchange Trading Network System).[8] To examine the intraday patterns, we divide each trading day into nine 30-minute intervals, four in the morning session and five in the afternoon session.

---

[7] There are five regional exchanges in Japan: Osaka, Nagoya, Hiroshima, Fukuoka, and Sapporo Stock Exchanges. Some of the Nikkei constituent stocks also trade in regional exchanges. However, regional exchanges are dominated by the TSE in terms trading volume. Trade-throughs are not allowed.
[8] According to Tokyo Stock Exchange Fact Book (2000), off-hour trading was introduced for block orders on November 14, 1997. ToSTNet-1 was established on June 29, 1998 to deal with off-hours trading for single issue orders and basket cross orders. TosTNet-2 was established on August 7, 1998 to allow orders be executed at the TSE's closing prices.

On the TSE, the opening and closing transactions in each session are made under a batch clearing process called *itayose* while the rest of the trades during the day are made under a continuous double auction called *zaraba*. Since the model for our empirical test applies to normal trading conditions under a continuous double auction, we only need transactions made under *zaraba*. Thus, we exclude all *itayose* trades. With the opening and closing trade and quote indicators, the data set enables us to accurately identify all trades made at *itayose*.

2.2.4 Price limit on the TSE

One important feature of the TSE is the price limit rule. The TSE uses price limits to trigger indicative quote dissemination, to halt trade temporarily, and to allow continued price discovery through quote adjustment in some circumstances.[9] For a stock trading between ¥1,001 and ¥1,500, the tick size is ¥1, the maximum price variation between trades without trade being halted is ¥20, and the daily price limit is ¥200. The daily price limits are quite large, ranging from 10 to 30 percent for most stocks priced below ¥30,000. Consequently, these limits are rarely hit. For 156 of our 219 sample stocks, the price never hits the daily price limit during the entire three-month sample period. In contrast, the maximum price variation allowed between trades is on the order of 0.5 to 1.5 percent of the stock price, and thus the limit is hit more frequently. For example, for 96 out of the 219 sample firms, the stock price reaches the intraday price limit during the sample period. However, the average number of hits per day is low at 0.62 times (0.19 percent of the total number of transactions on an average day).

It is possible that these price-limit rules affect our estimation process based on order arrival sequences. For example, the price limits can hinder the natural price discovery process by suppressing part of the orders that are informative and, thus, would reduce the amount of information asymmetry that would otherwise be present in the market. This is especially true for the case of daily price limits since trading is allowed only within the daily limit during the entire day. Therefore, our estimates of spread components

---

[9] It is a debatable issue whether these rules help price continuity. See Kim and Rhee (1997) for an empirical study of the effects of daily price limit on TSE.

could be smaller than the amount of the information asymmetry that would appear in the absence of the price limits.[10]

2.2.5 Special quotes

The TSE requires the *saitori* to post a special quote when a major order imbalance between buy and sell orders arises. Once posted, the special quote remains until a counter order arrives and a new equilibrium is established, at which point the quote is withdrawn. In the absence of the arrival of any counter order, the quote can be renewed at an interval of every five minutes within the variation limit set forth by the exchange. The display mechanism of a special quote has several distinct features. A special quote is always accompanied by a null quote (zero price and zero quantity) on the other side. The quantity displayed with a special quote is the net buying and selling order quantities rather than the quantity available at the quote. In addition, the price location flag for the trade following a special quote does not indicate the usual 'transaction at ask' or 'transaction at bid' but the arrival and execution of a buy order or a sell order. Since a trade or trades matched with special quotes cannot be seen as part of a successive normal order flow, we exclude any trades following special quotes during the regular trading hours.[11]

2.2.6 Partial execution of large market orders

Sometimes an incoming market order is too large to be absorbed by the current bid or ask quotes. In this case, the *saitori* partially fills the order up to the size of the current quote. Then the remaining portion is converted into a limit order at the current quote.[12] For example, suppose the current ask is ¥500 for 1,000 shares. If a market buy order of 1,200 shares arrives, it can only be partially executed against the current ask up to 1,000 shares. The remaining 200 shares will be converted into a limit buy order (an

---

[10] As an alternative approach we apply a filter that excludes the entire 30-minute interval if a trade hits either the daily or intraday price limits. However, our results are not sensitive to this filtering method.
[11] Issuances of special quotes could be related with severe information asymmetry in the market. In a separate analysis, we include the transactions executed against special quotes. However, the estimation results are virtually identical to those estimated without special quotes.

ask quote of 200 shares at ¥500). Two outcomes can follow: (1) an incoming market sell order or limit sell order priced at ¥500 or below hits this ask quote or, (2) when there is no immediate arrival of either type of orders, the 200 share buy order walks up the book and hits the limit sell order at the next best price.[13]

Particular attention on buy-sell classification is required when a trade is executed under the first situation (i.e. an incoming order hits the limit order originated from partial execution of a market order). The trade, in the above example, is classified as a sell transaction in the original dataset. However, since this 200 share trade originates from the market buy order of 1,200 shares, it is more reasonable to treat it as a buy transaction instead of a sell transaction. A large market order (and often a series of large market orders on the same side) indicates the arrival of new information about the security. A series of transactions initiated on the same side will be more consistent with the information arrival. Therefore, we reclassified all trades that result from partial execution of large market orders. We also repeat all following analyses using the original buy-sell classification provided by the data set. Our reclassification rule, however, generally results in better, if marginal, convergence in the estimation of spread components. We report the results based on the reclassification. [14]

2.2.7 Final sample

It also happens that the tick size of some stocks changes in our sample. In this case, we exclude the half-hour section if the tick size changes (for example, from ¥1 to ¥5 or vice versa) during the interval. To ensure enough observations within each half-hour interval, we impose the condition that there are at least five valid transactions during the interval. Then, we drop the entire trading day if it has fewer than

---

[12] The saitori used to display a warning quote briefly to invite liquidity in this situation before the TSE's rule change on warning quotes in 1998. Since the rule change, the quote is treated as part of general quotes.
[13] The saitori uses discretion on the exposure duration of the limit order converted from a partially executed market order.
[14] The way to obtain immediacy of a large order in the TSE's institutional setting might be not to submit a market order but to submit an opposing limit order with a price that guarantees an immediate and full trade execution. Indeed, a limit sell (buy) order submitted in this fashion should be regarded as a market buy (sell) order in disguise of a limit order. Hence, it would be more relevant for our purpose to treat these limit orders as opposing market

nine half-hour intervals. Finally, we make sure that, for every stock, there are more than 30 valid trading

days available. Otherwise, we drop the stock. This final filtering process excludes 15 firms, leaving 204

firms in our final sample. These 204 stocks include 171 stocks within the price range of less than ¥2,000,

11 within the price range of ¥2,000 to ¥2,999, and 22 within the range of ¥3,000 to ¥29,999. For easy

exposition, we call them group 1 (under ¥2,000), group 2 (¥2,000 to ¥2,999), and group 3 (¥3,000 to

¥29,999) stocks.

2.3 Descriptive statistics

    Table 1 presents basic summary statistics including the cross-sectional mean, median, and standard

deviation for a number of market variables. For group 1 stocks, the median transaction price is ¥471.

Trading is very active with the median number of transactions being 249 per day. The median trade size

is 4,630 shares. The median spreads are ¥1.85 and 0.45 percent of the bid-ask midpoint. For group 2

stocks, the median transaction price is ¥2,460. Group 2 stocks are even more actively traded with the

median number of daily transactions being 361. The median trade size is 3,530 shares. The median

spreads are ¥7.77 and 0.31 percent of the bid-ask midpoint. For group 3 stocks, the median price is much

larger at ¥5,861. The median number of daily trades is 333. Trade size measured by number of shares is

smaller, with a median of 2,200 shares; trade size measured in yen is much larger. The median spreads

are ¥21.05 and 0.33 percent of the bid-ask midpoint. Finally, the market capitalization is much larger for

group 2 stocks than for group 1 stocks. The median market capitalization of group 2 stocks is almost

seven times that of group 1 stocks. The median market capitalization is the largest for group 3 stocks,

more than nine times that of group 1 stocks.

    Table 2 provides the cross-sectional mean estimates of the market variables for the nine half-hour

intraday intervals during the morning and afternoon sessions on the TSE. For all three groups of stocks,

intraday volatility, number of transactions, yen volume, and yen trade size in general exhibit the familiar

U-shape over the course of the trading day. The increase in volatility, number of trades, trading volume,

---

orders and reclassify the buy/sell signs of such trades accordingly. Our data, however, do not provide specific order

and trade size is more pronounced towards the end of the trading day than in the opening half-hour. There is evidence that the lunch break also induces higher volatility, more transactions, higher trading volume, and larger trade size, but at a smaller magnitude compared to the opening and closing half-hours.

Table 3 summarizes the intraday patterns of bid-ask spread for stocks sorted by the three price categories. For group 1 stocks in Panel A for example, the spread averages ¥2.89 in the opening half-hour, drops to ¥2.12 in the middle of the day, then rises to ¥2.41 towards the closing half-hour. This is equivalent to 0.66, 0.54, and 0.60 percent of the bid-ask midpoint. The spread in yen terms is much larger for high-priced stocks in Panels B and C. The corresponding spread in percentage terms is lower, but the intraday patterns are similar. The standard errors also suggest that the mean estimates are reliable, despite the small sample size for medium- and high-priced stocks. There is also some evidence of a slight increase in the spread around the lunch break, indicating that both morning and afternoon sessions have their own U-shape pattern. The spreads during the last half hour in the morning session and the first half hour of the afternoon session are slightly greater than the spreads in the neighboring half hour intervals. The intraday pattern of the spread reported in Table 3 is consistent with those reported by Lehmann and Modest (1994) and Hamao and Hasbrouck (1995).

3. Estimating the components of the bid-ask spread

3.1 The model

Theoretical and empirical work on the components of bid-ask spread is extensive. In general, there are two classes of statistical models. The first one relies on the serial covariance properties of the observed transaction prices (see Roll (1984), Choi, Salandro, and Shastri (1984), George, Kaul, and Nimalendran (1991), Stoll (1989), Lin (1992), Huang and Stoll (1994), Lin, Sanger, and Booth (1995)). The second class of models is based on the trade initiation indicator variable. The work by Glosten and Harris (1988) and Madhavan and Smidt (1991) fall into this category.[15] More recently, Huang and Stoll

---

flow information and allow us of such reclassification.

[15] Other related works include Hasbrouck (1991a, 1991b, 1993) and Hamao and Hasbrouck (1995), who apply a vector autoregression framework to make inferences about the sources of spread.

(1997) develop a general indicator model within which the various existing models may be reconciled. This model allows a three-way decomposition of the adverse selection, inventory-carrying, and order-processing components. At the same time, Madhavan, Richardson, and Roomans (1997) extend the indicator model of Glosten and Milgrom (1985) to allow the order flow to be correlated.

The trade indicator models are appropriate for our purpose to investigate the intraday variations in spread components since they can be easily used to estimate different components for different times of the trading day by introducing time indicator variables. In particular, the framework developed by Madhavan, Richardson, and Roomans (1997) or MRR model can be readily applied to the study of bid-ask components in an order driven market such as the TSE. Although they do not explicitly model the limit-order book, a limit-order trader can be interpreted as another market maker.[16] In Huang and Stoll's model, the inventory cost component of the spread has an important role and is estimated separately. However, the inventory component is ignored in the model developed by Mahavan, Richardson, and Roomans. Instead, the model decomposes the spread into the permanent component due to information (adverse selection) and the transitory component. This approach has an attractive feature since it is believed that inventory costs are of a less important concern for limit order traders, unlike the case of the NYSE specialist who is obliged at all times to maintain inventory positions on both sides of the market. We use the general framework of the MRR model for our analysis of spread decomposition. However, estimation results might be sensitive to the specifications used by different market microstructure models. Hence, we use other models such as de Jong, Nijman, Roell (1997) model, the Glosten and Harris model (1988), and the Hasbrouck's VAR model (1988, 1991a, b) to cross-validate our estimation results. These alternative models are discussed later in Section 4.4.

Now we briefly describe the MRR model for estimating the components of bid-ask spread on the TSE. We denote the price of transaction at time t as $P_t.$, and define Q_t to be the buy-sell trade indicator variable for the transaction price. $Q_t = +1$ if the transaction is buyer initiated and -1 if the trade is seller initiated. The change in transaction price can be described as:

$$\Delta P_t = \alpha(Q_t - \rho Q_{t-1}) + \beta(Q_t - Q_{t-1}) + u_t. \tag{1}$$

The first term captures the effect of revision in belief, where $\alpha \geq 0$ measures the possible asymmetric

information revealed by the trade at time t. The second term captures the effect of bid-ask bounce, where

$\beta \geq 0$ denotes the liquidity suppliers' cost per share for supplying liquidity. Equation (1) extends the

approach in Glosten and Milgrom (1985) and allows for the autocorrelation in order flows. Let π denote

the probability of a trade continuation on the same side of the bid or ask. Then, the conditional

expectation of a trader indicator at t given Q_{t-1} is $E(Q_t|Q_{t-1}) = (1-2\pi)Q_{t-1} \equiv \rho Q_{t-1},$ where ρ is the first-order

autocorrelation of the trade initiation variable. In the absence of adverse selection ($\alpha = 0$) and inventory

and ordering processing costs ($\beta = 0$), the model reduces to the classical case where prices follow a

random walk. When the autocorrelation of trade indicators variable is zero ($\rho = 0$), the model reduces to

the setting in Glosten and Milgrom (1985).

3.2 Estimation procedure

The three parameters (α, β, ρ) governing the behavior of transaction prices and quotes in Equation (1)

can be estimated using generalized method of moments (GMM), which imposes very weak distribution

assumptions. This is especially important because the error term includes rounding errors due to

discreteness of stock prices. The GMM procedure also easily accounts for the presence of conditional

heteroskedasticity of an unknown form.[17] Specifically, the expectation of the following four population

moments is zero :

---

[16] Greene (1996) extends the model in this direction.

[17] We can think of an OLS regression as an alternative to the GMM procedure for parameter estimation. However, the OLS cannot identify each parameter when there is a correlation term, ρ, as in (1). There are three parameters, α, β, and ρ, to be estimated in model (1), but the OLS will generate only two coefficients corresponding to Q_t and Q_{t+1}.

$$E[f(\Delta P_t, Q_t, Q_{t-1}, \alpha, \beta, \rho)] = 0,$$

$$\text{where } f(\Delta P_t, Q_t, Q_{t-1}, \alpha, \beta, \rho) = \begin{pmatrix} Q_t Q_{t-1} - \rho Q_{t-1}^2 \\ u_t - u_0 \\ (u_t - u_0)Q_t \\ (u_t - u_0)Q_{t-1} \end{pmatrix}, \tag{2}$$

where u_t is as defined in (1), or $\Delta P_t - \alpha(Q_t - \rho Q_{t-1}) - \beta(Q_t - Q_{t-1})$ and u_0 is a constant drift. The first

equation is the definition of the autocorrelation in the trade initiation variable. The second equation

defines the constant term u_0 as the average pricing error. The last three equations are the normalizing

equations. The instrumental variables are the lagged trade initiation variable Q_{t-1} plus a constant. It is

worth noting that since all of the transactions included in the final sample occur on either side of the bid

and ask quotes, the probability of a transaction occurring within the bid ask quotes is zero. Therefore, we

do not need to estimate that probability as in Madhavan, Richardson, and Roomans (1997).

4. Parameter estimates

4.1 Cross-sectional patterns

Table 4 presents the summary statistics on the individual parameter estimates for three groups of

stocks. The summary statistics include the mean coefficient estimate, mean standard error, standard

deviation of the estimates, and median estimates for adverse selection cost (α), order-processing cost (β),

autocorrelation of order flow (ρ), the implied spread, $2(\alpha + \beta)$, and the proportion of the adverse selection

component in the implied spread ($\gamma=\alpha/(\alpha+\beta)$).[18] The adverse selection cost, order-processing cost, and

the implied spread in percentage terms is calculated relative to the average prices over the sample period.

The drift term μ_0 is essentially zero and therefore not reported.

4.1.1 Adverse selection (α) and order-processing (β) components

---

[18] The stock price is given by $P_t = V_{t-1} + \alpha(Q_t - \rho Q_{t-1}) + \beta Q_t + \varepsilon_t + \eta_t$, replace Q_t by $+1$ and -1 to obtain the bid-ask spread $2(\alpha + \beta)$, where V_{t-1} is the unobserved fundamental value of the stock.

13

The mean values of adverse selection component α are ¥0.460, ¥1.416, and ¥6.268 for three groups of stocks. In percentage terms, the adverse selection components are 0.091, 0.057, and 0.071 percent respectively. The lowest-priced group has the largest percentage adverse selection component, which is significantly different from median- and highest-priced stocks based on the non-parametric Wilcoxon signed rank test. The results are consistent with the evidence from the market capitalization in Table 1. The lowest-priced stocks have a significantly smaller market value than higher-priced stocks. Therefore, the lowest-priced stocks tend to be less widely followed in the market and are subject to a greater degree of asymmetric information. Table 2 also provides the standard errors for the estimates. In all cases, the standard errors are small.[19]

The median values of order-processing cost β are ¥0.395, ¥1.657, and ¥4.001, respectively for the three groups of stocks. These translate into 0.141, 0.067, and 0.052 percent of the bid-ask midpoint. The order-processing component is significantly lower for the highest-priced stocks, as indicated by the p-values from the Wilcoxon signed rank tests. For the lowest- and median-priced stocks, order-processing costs are similar in percentage terms.

4.1.2 Auto-correlation of order flow (ρ)

The autocorrelation of order flow ρ are 0.315, 0.235, and 0.284, respectively for three groups of stocks. The differences are highly significant between either pair of groups. Since by definition, $(1-2\pi) \equiv \rho$, we can infer from the estimates of $\rho = 0.315$ that $1- \pi = 0.658$, which is the probability of a transaction continuing on the same side of the bid (ask) for the lowest-price stocks. Similarly, the probability of continuation is $1 - \pi = 0.618$ for median-priced stocks, and $1 - \pi = 0.642$ for highest-priced stocks. A positive autocorrelation in order arrivals has been found for U.S. stocks by Hasbrouck and Ho (1987), Choi, Salandro, and Shastri (1988), Hasbrouck (1991a), and Lin, Sanger, and Booth (1995). For limit-

---

[19] The estimates are obtained from a sample including the half-hour session when the price limit is violated. We also estimate the model with a sample excluding the half-hour session when the price limit is violated. The results are essentially the same.

order markets such as the Paris Bourse and TSE, the order flow is also found to be positively correlated (Biais, Hillion, and Spatt, 1995; Hamao and Hasbrouck, 1995). Biais, Hillion, and Spatt (1995) offer several hypotheses for the positive autocorrelation of order flow including strategic order splitting, traders imitating or piggy-backing each other, and traders reacting similarly, but successively, to the same events.

4.1.3 The implied spread and the proportion of adverse selection in spread

The implied spread at time t is a random variable with mean $2(\alpha + \beta)$. The median implied spreads are ¥1.45, ¥6.26, and ¥15.81 respectively. The median quoted spreads from Table 1 are ¥1.85, ¥7.77, and ¥21.05. In percentage terms, the implied spreads are 0.33, 0.25, and 0.25 percent respectively. The sample spreads are 0.45, 0.31, and 0.33 percent. The estimated spread is lower than the sample mean spread by a factor of 20-30 percent for the group 1 stocks that account for a majority of the sample. The discrepancy is slightly smaller in magnitude to Madhavan, Richardson, and Roomans (1997) for a sample of U.S. stocks. For U.S. stocks, the discrepancy can be readily attributed to transactions occurring between the quotes, which happen much more often than for Japanese stocks.[20] There are two pontential explanations for underestimation on the TSE. The sample spread in Table 1 is the average of all quoted spreads. The implied spread, on the other hand, is based on the effective spread, i.e. conditioned on actual trades. When the quoted spread is large, it is less likely that the quotes are hit. Therefore, even after eliminating all trades that occur inside the spreads on the TSE, the implied spread remains smaller than the average quoted spread. Alternatively, trades may take place at the side of the market offering the highest liquidity. Suppose liquidity is plentiful on one side of the market while it is not on the other side. The asymmetry in liquidity supply in this situation can cause an asymmetry in the spread: the spread on the side with plentiful liquidity can be smaller than the spread on the other side. In other words, the spread of the side with greater liquidity is less than one half of the quoted spread. Trades are likely to hit

---

[20] According to NYSE reported statistics, about 30 percent of NYSE transactions occur at the midpoint, but in transactions where the spread is larger than 1/8, the figure increases to 70 percent.

the side with plentiful liquidity more often. As our implied spread measure reflects the half spread of the side where transactions take place, it can be smaller than the quoted spread.[21]

If we define γ as the ratio of the adverse selection component of the spread (2α) to the total implied spread, the estimate of γ has a mean of $\alpha/(\alpha + \beta)$. The estimates are 0.498, 0.448, and 0.571 respectively. The adverse selection component accounts for a bigger portion of the implied spread for the highest-priced stocks. We also perform a non-parametric test to examine the significance of the differences between the three groups. The difference is particularly significant between the highest-priced stocks and the lowest-priced stocks.

4.2 Intraday patterns

Perhaps the best known stylized fact about the intraday patterns in bid-ask spread (and volume, and volatility) is that there is a broad U-shaped pattern in virtually all equity markets, whether they are quote-driven or order-driven.[22] Chung, Van Ness, and Van Ness (1999) recently report that the U-shape pattern of spreads on 144 stocks on the NYSE largely reflects the intraday variation in spreads established by limit-order traders. The intraday variation in dealers' spreads is different; the spreads are widest at the open, narrow until late morning, and then level off. Madhavan, Richardson, and Roomans (1997) decompose the adverse selection and order-process parts of the spread for a sample of 274 of NYSE stocks. Their major finding is that the adverse selection component declines steadily throughout the day, while the dealer cost component increases over the course of the day, resulting in the U-shape pattern of the implied spread. Their work on NYSE stocks can be refined by classifying the quotes and transactions according to whether they originate from the dealers or limit-order traders. On the TSE, all the orders are supplied through the limit-order book. In this regard, the evidence from TSE will increase our understanding of the forces behind the observed U-shaped pattern.

---

[21] We thank an anonymous referee for suggesting the explanation.
[22] See Goodhart and O'Hara (1997) for a survey of theoretical models and empirical evidence.

Table 5 presents the intraday pattern for the estimates. For all three groups of stocks, the adverse selection component, both in yen and percentage terms, exhibit a clear U-shape by itself.[23] The mean value of α is highest at the opening and closing half-hours, and lowest in the middle of the day. The U-shape pattern is also pronounced for the order-processing component. The mean standard errors in parentheses indicate that the estimates are reliable across each intraday interval. The U-shape patterns in both components necessarily imply a U-shape in the implied spread. Moreover, the proportion of adverse selection cost to the implied spread also displays a broad U-shape over the course of a trading day. The last column reports the F-statistics for testing the hypotheses that the variables are equal across eleven intraday intervals. The results strongly reject the null hypotheses.

Our analysis is based on the stock groups formed by different tick sizes. However, the spread may be a function of other characteristics that are closely related with the liquidity of the stock, such as trading volume or return volatility. To check the robustness of our finding, we regroup the sample firms into three subgroups based on daily trading volume. Table 6 reports the estimated results of the spread components for different half-hour intervals of the trading day. The results are qualitatively identical to those found from the samples based on tick size. All three subgroups based on trading volume exhibit strong U-shape patterns in the adverse selection component, the order processing component, the implied spread, and the adverse selection portion of the implied spread.[24]

Since α represents the magnitude of the revision in limit-order traders' beliefs concerning the fundamental value of the stock induced by the order flow, a large α in the opening half-hour is consistent with the degree of asymmetric information being highest when information is accumulated overnight. The gradual decline of α is consistent with the market learning about the fundamental asset value through trading (Madhavan, Richardson, and Roomans, 1997). But learning or price discovery cannot explain why α rises again towards the end of the trading day. The pattern of a rise in α towards the close clearly

---

[23] There is also a pattern of a small increase in the adverse selection around the lunch break. The differences, though, are statistically indistinguishable when the estimates during the intervals of half hour before and after the lunch break are compared with the estimates from the neighboring intervals.

suggests that information contents of trading increases in the last 30-minute session of trading, and that limit-order traders reflect this increase in informed trading in their revision of beliefs.

The striking pattern of a rise in the adverse selection cost toward the end of the trading day can be explained in the framework of Admati and Pfleiderer (1988), who predict covariations of trading volume, the variance of price changes, and adverse selection costs. In the equilibrium of the Admati and Pfleiderer model, trading by discretionary liquidity traders is relatively more concentrated in periods closer to the realization of their demands. This concentration of liquidity trading in turn attracts privately informed traders, who camouflage their informed trades among liquidity trades. Therefore, prices are more informed in these periods. The last 30-minute trading session is characterized by the fact that it falls just before the exchange is closed, after which it is difficult to trade. Therefore, discretionary liquidity traders will concentrate their trading around this period. Also, the settlement rules will cause a further concentration of trading around this time. The increase in trading around the end of the trading day is well supported by the intraday trading pattern on the TSE. In fact, the daily number of transactions is highest during the closing 30 minutes for all three groups of stocks (Table 2). As liquidity trading increases at the end of the trading day, so does informed trading. Besides, informed traders have added motivations to trade before the market close than wait until the next market opening. First, the value of the private information these investors have will decline during the overnight non-trading period. Further, the overnight non-trading period adds uncertainty that the price moves unfavorably to the informed traders. As a result, the end of the trading day will better reflect the private information that would otherwise be released during the non-trading hours that follow the exchange close.[25]

We further investigate whether the significant increase in the adverse selection cost during the last half hour of the day is mainly due to the final minutes of trading right before the exchange close. We divide the final half hour period of the day into six five-minute intervals and estimate the spread

---

[24] We also partition the sample into three groups based on return volatility and find the similar U-shape patterns. The results are available upon request.

[25] It will be interesting to see if the introduction of off-hour trading for large block trades through the ToSTNet system affects the adverse selection component in any significant way.

components. The results are reported in Table 7. The adverse selection component displays a dramatic

increase in the final five-minutes. For example, the average α of Group 1 stocks is 0.107 percent from

ten minutes to five minutes before the close. Then, it escalates to 0.165 percent in the final five minutes.

Even though not as dramatic as for Group 1, a similar pattern is found for both Group 2 and 3 stocks. The

evidence suggests that the adverse selection component shifts up just before the close. The order

handling component also exhibits a similar pattern but its pattern of increase is rather gradual.

4.3 Trade size

Recent research has begun to examine the relation of trade size and various components of bid-ask

spread. Trade size introduces an adverse selection problem because, given that they wish to trade,

informed investors prefer to trade large amounts at any given price (Easley and O'Hara, 1987). This

implies that large trades tend to convey more information than small trades. Empirically, Lin, Sanger,

and Booth (1995) show that the adverse selection component increases with trade size and the order-

processing cost decreases with trade size for a sample of 150 U.S. firms in 1988. However, using a

sample of 20 component stocks in the Major Market Index in 1992, Huang and Stoll (1997) report

conflicting evidence that the adverse selection component is smaller for sequences of two trades that end

in a large trade than for sequences that end in a medium or small trade. Huang and Stoll (1997) argue that

the results could be due to the upstairs trading of large block trades.[26] Since there is no upstairs trading

for block trades on the TSE, all trades have to go through the trading system. The data from TSE offers

an opportunity to look further into the issue.

We generalize Equation (1) to allow different coefficient estimates by trade size category. We choose

three trade size categories, although any number of categories is possible. Consider the following

specification:

---

[26] In upstairs markets, large transactions are accomplished through a search-brokerage mechanism where an
intermediary or broker locates counterparts to a trade before sending it to the downstairs market for execution.
Keim and Madhavan (1996) provide a theoretical and empirical analysis of the upstairs market. See also Madhavan
and Cheng (1997).

$$\Delta P_t^{ij} = \alpha^{ij}(Q_t - \rho^{ij}Q_{t-1}^j) + \beta^{ij}(Q_t - Q_{t-1}^j) + u_t^{ij} \qquad (3)$$

where i, j = small, medium, large, referring to the trade size categories at time t-2 and t-1. α^{ij}, β^{ij}, and

ρ^{ij} are the adverse selection component, order-processing component, and order-flow correlation when

trade size moves from i category at time t-2 to j category at time t-1. The model in Equation (3) is fairly

complex for it requires nine different values of α, β, and ρ for each of the nine trade-size transitions

between t-2 and t-1. Therefore, we use data points to classify each possible transition.

In Japan, trading takes place in units of a specified number of shares, or minimum trading unit (MTU).

The common unit is 1,000 shares.[27] For the 204 stocks in our final sample, the majority (196) have a

trading unit of 1,000 shares. Only eight firms have a trading unit of 100 shares. We define a trade to be

small if the number of shares traded is equal to the MTU. A medium-size transaction trades between the

MTU and 10 times the MTU. A large-size transaction trades at more than 10 times the MTU. Using the

196 firms with a MTU of 1,000 shares as an example, a small-size trade involves 1,000 shares. A

medium-size trade involves between 1,001 and 9,999 shares. A large-size trade involves at least 10,000

shares. We calculate the distribution of all transactions over the three-month period for each firm in our

sample. The cross-sectional average distribution is 40, 45, and 14 percent for small, medium, and large

trades.[28]

Table 8 summarizes the empirical results.[29] The column categories refer to the transaction size in

time t-1 and the row categories refer to the transaction size in t-2. Looking across the columns, the

estimates of α are larger for sequences of two trades that end in a large trade than for sequences of two

---

[27] For stocks listed on the TSE, the Exchange recommends units of 1,000, 500, 100, 50, or 10 shares. There are some cases when trading units are 2,000 shares or 1 share.

[28] The cross-sectional averages of the correlations between different-size trades, normalized for daily number of trades, are 0.10 between small and medium trades, -0.04 between small and large trades, and 0.12 between the medium and large trades.

[29] The cross-sectional average of observations in the GMM estimation is 3,105 and 1,116, respectively, for small-to-small and large-to-large trade sequences. Trade sequences that begin or end with small or medium transactions typically have more observations than sequences that begin or end with large transactions. We require both α and β to be positive and a minimum of 30 observations for a particular trade sequence to be included in Table 6. The

trades that end in a medium or small trade. There is in fact a monotonic increase in the adverse selection

component regardless of the initiating trade size at time t-2. The corresponding F-value indicates that the

difference is highly significant. Looking down the rows in the first two columns, trade sequences

initiated by a large trade at time t-2 always have a larger estimate of α than trade sequences initiated by a

medium or small trade at time t-2. The estimate for α is 0.059 for a small-to-small trade sequence versus

0.127 for a large-to-large trade sequence. These results strongly indicate that large trades on the TSE

contain more information about the security price and are consistent with the way in which TSE handles

large trades.[30]

When we look across the columns in Panel B, the estimates of order-processing cost β are smaller for

sequences of two trades that end in a large trade than for sequences that end in a medium or small trade.

The difference is monotonic and highly significant for the first two rows. The overall pattern is that

order-processing cost decreases with trade size. For example, the estimate for β is 0.158 for a small-to-

small trade sequence versus 0.105 for a large-to-large trade sequence. But the decrease in the order-

processing cost is less dramatic than the increase in the adverse selection cost as trade size increases. Note

that the negative relation between the order processing component and trade size has an implication on

inventory costs on the TSE. Our order processing component estimates include both inventory costs and

usual order handling costs. While order handling costs are expected to decrease with trade size due to

economies of scale[31], inventory costs are likely to increase with it. The fact that our estimates that

combine these two cost factors decrease with trade size implies the possibility that the effect of order

---

average number of stocks included is 203 and 157, respectively, for small-to-small and large-to-large trade
sequences.
[30] We also test whether there is any significant difference in the magnitudes of the adverse selection as well as the
order processing components between large-to-large trades at the same side of the market and large-to-large trades
at the different side of the market. For stocks in Group 2 and Group 3, large-to-large trades on the same side of the
market, a large buy followed by another large buy for an example, turn out to have significantly greater adverse
selection (smaller order processing) costs than large-to-large trades at the different side of the market. Group 1
stocks on the other hand do not display any distinct difference.
[31] For example, we can think of brokerage fees as part of order handling costs for limit order traders. The brokerage
fees on the TSE monotonically decrease in trade size.

handling costs dominates the effect of inventory costs. This is consistent with the notion that inventory costs are of a less important concern to limit order traders on the TSE.

4.4 Estimation by alternative models

Spread component estimates are known to be sensitive to model specifications (de Jong, Nijman, and Roell (1996)). In this section, we apply alternative spread decomposition models and demonstrate that our findings are not simply driven by methodological differences. Specifically, we compare our results based on the Madhavan, Richardson, and Roomans (1997) or MRR model of spread decomposition with those based on the De Jong, Nijman, and Roell (1996) model and the Glosten and Harris (1988) model. Later, we introduce the VAR model developed by Hasbrouk (1988, 1991a, b) and examine whether different models produce comparable spread estimates.

De Jong, Nijman, and Roell (1996) extend the Glosten (1994) model and analyze the intraday price effects on the Paris Bourse. The De Jong et al. (1996) model is specified as follows:

$$\Delta P_t = \mu + R_0 \Delta Q_t + R_1 \Delta(q_t Q_t) + e_0 Q_{t-1} + e_1 q_t Q_{t-1} + u_t, \qquad (4)$$

where P_t is the transaction price at time t, Q_t is the buy-sell trade indicator variable, q_t is the size of trade measured or number of shares traded measured in multiples of Minimum Trading Unit (MTU). In this model, the order-processing component is $c_0 + c_1 q_t$, where c_0 is $R_0 - e_0 - e_1 \cdot \alpha$, c_1 is $R_1 - 0.5 \cdot e_1$, and α is the median of trade size divided by log(2). The adverse selection component is given by $(R_0 - c_0) + (R_1 - c_1)q_t$.

Meanwhile, we can specify the Glosten and Harris (1988) model in the context of De Jong, Nijman, and Roell (1986) as follows:

$$\Delta P_t = \mu + c_0 \Delta Q_t + c_1 \Delta(q_t Q_t) + z_0 Q_t + z_1 q_t Q_t + u_t, \qquad (5)$$

22

where P_t, Q_t, and q_t are defined as in Model (4). In the above model, $z_0+z_1q_t$ is the post-trade expectation to the order flow, or the adverse selection cost component of the spread. $c_0+c_1q_t$ is the order handling cost component. The two models are estimated using ordinary least squares (OLS), where the standard errors have been adjusted using the Newey and West (1987) method. In actual estimation, the trades size is censored so that transactions larger than 99.5 percentile of all trades are excluded. This is to mitigate the effect of very large trades on the estimates (Hausman, Lo and MacKinlay, 1992).

The estimation results by both models are presented in Table 9. For ease of comparison we also report the spread component estimates form the MRR model in the same table. Panel A reports the mean parameter estimates and mean standard errors (in parenthesis) from the alternative models. Panel B compares the adverse selection component, the order-processing component, and the proportion of adverse selection component of the spread from the three models. Since spread components depend on trade size, the spread components for each individual stock are first calculated when trade size q_t is set equal to 1 MTU, median MTUs, 10 MTUs, and 99 percentile MTUs of all trades, respectively. Then, the cross-sectional mean spread components are reported.

As expected, the spread component estimation is sensitive to model specifications. The estimates of the adverse selection (order handling) component from the two alternative models are generally lower (greater) than our estimate from the MRR model. In the De Jong et at. (1996) model, the adverse selection component is $(R_0-c_0)+(R_1-c_1)q_t$. Using Group 1 stocks as an example, when q_t is set at the median trade size, the cross-sectional average of the adverse selection is 0.301. In the Glosten and Harris (1988) model, the adverse selection cost component is $z_0+z_1q_t$. Again, for Group 1 with the median trade size, the average estimate of the adverse selection component is 0.280. These two numbers are compared with our mean MRR estimate of the adverse selection component for the same group, 0.460. The order handling component in the De Jong et at. (1996) framework is $c_0+c_1q_t$. For Group 1, the average order handling component estimate is 0.541 with q_t set at the median trade size. The order handling component in the Glosten and Harris model, also $c_0+c_1q_t$, is 0.562. The MRR counterpart of these numbers is 0.395. The cross-sectional average of the proportion of the adverse selection component in the implied spread is

0.332 under the De Jong et al. model and 0.306 under the Glosten and Harris model, compared with 0.395 for the MRR model.[32] A similar pattern of discrepancies in the estimates by the alternative models can be found from Group 2 and Group 3.

In the previous subsection, we demonstrated that the informativeness of trading increases with trade size on the TSE. Panel B of Table 9 shows that the alternative models, despite the slight discrepancies in the estimates, confirm the same positive relation between the adverse selection component estimates and trade size. The adverse selection component rises monotonically as trade size increases. Based on the De Jong et at. (1996) model, using Group 1 stocks as an example, the cross-sectional average of the adverse selection is 0.295 when q_t is set to 1 MTU. When q_t is set at the median trade size, the average adverse selection component is 0.301. With q_t set at 10 MTUs, it is 0.349. For the Glosten and Harris (1988) model, the average estimate of the adverse selection component for the same stock group is 0.269 when q_t is 1 MTU. It then increases to 0.280 and to 0.375 as q_t increases to the median trade size and to 10 MTUs. The proportion of the adverse selection component in the implied spread also confirms the pattern. For both models, the proportion of the spread attributable to informed trading increases monotonically as the trade size increases. The same pattern is observed from Group 2 and Group 3.

The overall negative association of order-processing costs with trade size observed in Table 8 is not found in the estimated results by the De Jong et al. (1996) model. However, the estimation results by the Glosten and Harris (1988) model indicates that order processing costs decrease monotonically as the trade size increases.

The trade indicator models of Glosten and Harris (1988), de Jong et al. (1996), and MRR assume that the information content of a trade is immediately and fully revealed after each transaction. No lagging price effect is assumed. However, the results on Table 8 indicate that trades may have lagged information effects as well. Further, the models assume that the trading pattern is exogenous. However, if the trading pattern is endogenously determined and lagged trades have some effects on the current trades, the

---

[32] Under the De Jong et al. (1996) model, the proportion of the adverse selection component in the implied spread is given as $((R_0 - c_0) + (R_1 - c_1)q_t)/(R_0 + R_1 q_t)$.

coefficients of the regressions based on the above structural models might be biased.[33] To address these problems, we apply the Vector Auto Regressive (VAR) model introduced by Hasbrouck (1988, 1991a, b). In Hasbrouck's model, prices and trade patterns are jointly modeled by a system of equations. Specifically, the Hasbrouck VAR model is specified as follows:

$$x_t = [\Delta P_t \quad Q_t \quad q_t Q_t]',$$

$$x_t = A_0 x_t + A_1 x_{t-1} + \cdots + A_5 x_{t-5} + v_t,$$ (6)

$$A_0 = \begin{pmatrix} 0 & a_{01} & a_{02} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where P_t is the transaction price at time t, Q_t is the buy-sell trade indicator variable, q_t is the size of trade measured or number of shares traded measured in multiples of Minimum Trading Unit (MTU). The inclusion of A_0 term allows the current price changes to depend on the current buy-sell indicator Q_t and signed trade size $q_t Q_t$. We consider lagged price effects from transactions for up to five lags.

   Table 10 presents the estimation results based on the Hasbrouck's VAR model. Panel A reports the mean estimates and mean standard errors (in parenthesis) for a_{01} and a_{02} from the VAR model. Panel B reports the estimated half-spread from the VAR model along with the estimated half-spreads based on the other alternative models. In the Glosten and Harris (1988) model, the half-spread is given by $(c_0+z_0)+$ $(c_1+z_1)q_t$. In the De Jong et al. (1996) model, the half-spread is calculated as $R_0+R_1 q_t$. In the Hasbrouck VAR model, the half-spread is computed as $a_{01}+a_{02}q_t$. The half-spread from the MRR model is calculated as $\alpha+\beta$. Since half-spreads depend on trade size, half-spreads for each individual stock are first calculated by setting trade size q_t to be equal to 1 MTU, median MTUs, 10 MTUs, and 99 percentile MTUs of all trades, respectively. Then the cross-sectional means and mean standard errors (in parentheses) of half-spreads are reported in Panel B. The panel reveals that the Hasbrouk's VAR model provides half spread estimates very close to those produced by the MRR and the other alternative models. For example, in the

---

[33] We thank an anonymous referee to point out these issues.

case of Group 1 stocks, when the trade size is a medium one, the average half spread by the Hasbrouk's model is 0.833. This figure is very close to 0.855 by the MRR model or 0.842 by the de Jong et al. and Glosten and Harris models.

5. Conclusions

This paper examines the components of the bid-ask spread in the TSE limit-order book. We analyze unique tick-by-tick data for Nikkei 225 constituent stocks over a three-month period from January to March 2000. Our analysis reveals a number of findings that are different from those for other major exchanges such as the NYSE. In particular, both the adverse selection and order-processing components exhibit U-shape patterns independently, which necessarily implies a U-shape pattern in the implied spread. This is in striking contrast to the finding by Madhavan, Richardson, and Roomans (1997) for NYSE stocks that the adverse selection component declines and dealer costs increase over the trading day, resulting in an overall U-shape of the spread. Second, small transactions on the TSE typically take place at the minimum trading unit. These trades account for 40 percent of all transactions. Large trades take place at 10 times the minimum trading unit and account for 14 percent of all transactions. These large trades contain more information than medium or small trades on the TSE. The adverse selection component increases with trade size and order processing cost decreases with trade size. The evidence is consistent with the practice of handling large trades on the TSE. Upstairs trading does not exist on the TSE. In contrast, large trades are often prenegotiated through the upstairs markets, and this often leads to smaller adverse selection costs associated with large trades. Our findings of the striking U-shape pattern of the adverse selection component of the spread and the positive relation between trade size and informativeness of a trade are robust to the choice of different market microstructure models.

References

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

Ahn, H. J., K., Bae, and K. Chan, 2001, Limit orders, depth, and volatility: Evidence from the Stock Exchange of Hong Kong, *Journal of Finance* 54, 767-788.

Amihud, Y., and H. Mendelson, 1980, Dealership market: Market-making with inventory, *Journal of Financial Economics* 8, 31-53.

Amihud, Y., and H. Mendelson, 1989, Market microstructure and price discovery on the Tokyo Stock Exchange, *Japan and the World Economy* 1, 341-370.

Amihud, Y., and H. Mendelson, 1991, Volatility, efficiency, and trading: Evidence from the Japanese stock market, *Journal of Finance* 46, 1765-1790.

Amihud, Y., and H. Mendelson, 1993, Liquidity, trading costs and asset prices: Implications for Japan, in I. Walter and T. Hiraki (eds.), *Restructuring Japan's Financial Markets*, Business One, Irwin, Homewood, Ill.

Biais, B., P. Hillion, and 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.

Brockman, P., and D. Y. Chung, 1999, Bid-ask spread components in an order-driven environment, *Journal of Financial Research* 22, 227-246.

Chan, Y. C., 2000, The price impact of trading on the Stock Exchange of Hong Kong, *Journal of Financial Markets* 3, 1-16.

Choi, J. Y., D. Salandro, and K. Shastri, 1988, On the estimation of bid-ask spreads: Theory and evidence, *Journal of Financial and Quantitative Analysis* 23, 219-230.

Chung, K. H., F. Van Ness, and R. Van Ness, 1999, Limit orders and the bid-ask spread, *Journal of Financial Economics* 53, 255-287.

Cohen, K., S. Maier, R. Schwartz, and D. Whitcomb, 1981, Transaction costs, order placement strategy, and existence of the bid-ask spread, *Journal of Political Economy* 89, 287-305.

Copleland, T. and D. Galai, 1983, Information effects on the bid-ask spread, *Journal of Finance* 38, 1457-1469.

De Jong, F., T. Nijman, and A. Roell, 1996, Price effects of trading and components of the bid-ask spread on the Paris Bourse, *Journal of Empirical Finance* 3, 193-213.

Demsetz, H., 1968, The costs of transacting, *Quarterly Journal of Economics* 82, 33-53.

Easley, D., and M. O'Hara, 1987, Price, trade size and information in securities markets, *Journal of Financial Economics* 19, 69-90.

Foucault, T., 1999, Order flow composition and trading costs in a dynamic limit-order market, *Journal of Financial Markets* 2, 193-226.

George, T., 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.

George, T., and C. Y. Hwang, 1995, Transitory price changes and price-limit rules: Evidence from the Tokyo Stock Exchange, *Journal of Financial and Quantitative Analysis* 30, 313-327.

Glosten, L., 1994, Is the electronic open limit-order book inevitable? *Journal of Finance* 49, 1127-1161.

Glosten, L., and L. Harris, 1988, Estimating the components of the bid-ask spread, *Journal of Financial Economics* 21, 123-142..

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

Goodhart, C., and M. O'Hara, 1997, High frequency data in financial markets: Issues and applications, *Journal of Empirical Finance* 4, 73-114.

Greene, J., 1996, The impact of limit-order executions on trading costs in New York Stock Exchange stocks, Working paper, Georgia State University.

Hamao, Y., 1992, Tokyo Stock Exchange, in P. Newman, M. Milgate, and J. Eatwell (eds.), *The New Palgrave Dictionary of Money and Finance*, Macmillan Press, London.

Hamao, Y., and J. Hasbrouck, 1995, Securities trading in the absence of dealers: Trades and quotes in the Tokyo Stock Exchange, *Review of Financial Studies* 8, 849-878.

Handa, P., and R. Schwartz, 1996, Limit-order trading, *Journal of Finance* 51, 1835-1861.

Harris L., and J. Hasbrouck, 1996, Market vs. limit orders: the SuperDot evidence on order submission strategy, *Journal of Financial and Quantitative Analysis* 31, 213-231.

Hasbrouck, J., 1991a, Measuring the information content of stock trade, *Journal of Finance* 46, 179-207.

Hasbrouck, J., 1991b, The summary informativeness of stock trades: An econometric analysis, *Review of Financial Studies* 4, 571-595.

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

Hasbrouck J., and T. Ho, 1987, Order arrival, quote behavior and the return generating process, *Journal of Finance* 42, 1035-1048.

Ho, T., and H. Stoll, 1983, The dynamics of dealer markets under competition, *Journal of Finance* 38, 1053-1074.

Hollifield, B., A. Miller, and P. Sandås, 1999, An empirical analysis of limit-order markets, Working paper, Carnegie Mellon University, Pittsburgh, PA.

Huang, R., and H. Stoll, 1994, Market microstructure and stock return predications, *Review of Financial Studies* 7, 179-213.

Huang, R., and H. Stoll, 1997, The components of the bid-ask spread: A general approach, *Review of Financial Studies* 10, 995-1034.

Kavajecz, K., 1999, A specialist's quoted depth and the limit-order book, *Journal of Finance* 54, 747-771.

Keim, D., and A. Madhavan, 1996, The upstairs market for large-block transactions: Analysis and measurement of price effects, *Review of Financial Studies* 9, 1-36.

Kim, K., and S. Ghon Rhee, 1997, Price limit performance: Evidence from the Tokyo Stock Exchange, *Journal of Finance* 52, 885-901.

Lehmann, B., and D. Modest, 1994, Trading and liquidity on the Tokyo Stock Exchange: A bird's eye view, *Journal of Finance* 48, 1595-1628.

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

Lindsey, R., and U. Schaede, 1990, Specialist vs. saitori: market making in New York and Tokyo, *Financial Analysts Journal* 48, 48-57.

Madhavan, A, and M. Cheng, 1997, In search of liquidity: Block trades in the upstairs and downstairs markets, *Review of Financial Studies* 10, 175-203.

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.

Madhavan, A., and S. Smidt, 1993, A Bayesian model of intraday specialist pricing, *Journal of Financial Economics* 30, 99-134.

Rock, K., 1996, The specialist's order book and price anomalies, forthcoming in *Review of Financial Studies*.

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

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

Stoll, H., 1978, The supply of dealer services in securities markets, *Journal of Finance* 33, 1133-1151.

Takagi, S., 1993, The Japanese equity market: past and present, *Journal of Banking and Finance* 13, 537-570.

Tinic, S., 1972, The economics of liquidity services, *Quarterly Journal of Economics* 66, 79-97.

Viswanathan, S., and J. D. Wang, 1998, Market architecture: limit-order books versus dealership markets, Working Paper, Duke University

Table 1. Descriptive Statistics of the Sample

This table presents descriptive statistics for 204 TSE stocks used in this study. The summary includes stock price, market value, variance of transaction price changes, average number of trades per day, average share volume per day, average yen volume per day, average trade size in number of shares, average trade size in yen, average spread in yen, and average spread in percentage terms. Group 1 includes the stocks whose average prices are less than or equal to ¥2,000. Group 2 is for the stocks whose average prices range between ¥2,001 and ¥3,000. Group 3 is for the stocks whose average prices are between ¥3,001 and ¥30,000. The sample period is three months from January 5 to March 30, 2000. The market values in ¥ are as of March 31, 2000.

	Group 1 Price $\leq$ ¥2,000 (N = 171)			Group 2 ¥2,000 $<$ Price $\leq$ ¥3,000 (N = 11)			Group 3 ¥3,000 $<$ Price $\leq$ ¥30,000 (N = 22)		
	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median	Mean	Std. Dev.	Median
Price (¥)	595	454	471	2500	221	2460	9473	6649	5861
Market Value in ¥ ($\times 10^9$)	626	944	285	2,369	551	1,951	3,981	1,013	2,596
Variance of ΔPrice	2.41	2.68	1.34	24.83	9.96	24.33	424.02	537.92	177.2
Daily No. of Trades	296	176	249	473	275	361	540	570	333
Daily Share Volume ($\times 10^3$)	1,760	1,793	1,190	1,860	1,436	1,165	1,114	1,238	740
Daily Yen Volume ($\times 10^6$)	1,073	1,507	550	4,686	3,720	2,920	8,505	10,721	4,876
Trade Size in Shares ($\times 10^3$)	5.19	3.06	4.63	3.57	1.25	3.53	2.26	1.46	2.20
Trade Size in ¥ ($\times 10^6$)	2.63	2.01	2.06	8.95	3.39	8.39	14.72	5.93	14.19
Spread (¥)	2.22	1.16	1.85	7.76	2.08	7.77	29.07	20.6	21.05
Spread (%)	0.56	0.38	0.45	0.31	0.08	0.31	0.34	0.12	0.33

Table 2. Mean Estimates of Variances and Trading Activity Variables by 30-Minute Intraday Trading Intervals

This table presents the cross-sectional mean estimates of the variance of transaction price changes, average number of trades, average yen volume, and average trade size in yen during 30-minute intraday trading intervals for 204 TSE stocks used in this study. Panels A, B, and C provide the mean estimates for firms whose average prices are less than or equal to ¥2,000, between ¥2,001 and ¥3,000, and between ¥3,001 and ¥30,000 respectively.

Panel A. Group 1: Price ≤ ¥2,000 (N = 171)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
Variance of ΔPrice	4.49	2.32	2.16	2.20	2.10	1.85	1.90	1.92	3.00
Daily No. of Trades	35	34	35	35	30	29	29	31	39
Daily Yen Volume ($\times 10^6$)	122	118	116	125	114	104	105	116	152
Trade Size in ¥ ($\times 10^6$)	2.51	2.53	2.37	2.51	2.81	2.62	2.63	2.64	2.74

Panel B. Group 2: ¥2,000 < Price ≤ ¥3,000 (N = 11)

	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
Variance of ΔPrice	43.76	24.51	21.38	22.01	22.19	18.21	18.58	21.03	34.01
Daily No. of Trades	59	58	60	55	44	43	44	49	62
Daily Yen Volume ($\times 10^6$)	584	566	581	530	470	419	437	488	610
Trade Size in ¥ ($\times 10^6$)	8.52	8.55	8.38	8.53	9.54	8.95	9.01	8.91	9.02

Panel C. Group 3: ¥3,000 < Price ≤ ¥30,000 (N = 22)

	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
Variance of ΔPrice	794.79	491.07	422.35	398.31	361.84	305.17	273.10	300.11	559.70
Daily No. of Trades	60	63	63	64	53	52	53	59	74
Daily Yen Volume ($\times 10^6$)	999	955	923	990	914	801	785	907	1231
Trade Size in ¥ ($\times 10^6$)	14.58	14.06	13.87	14.26	15.78	14.70	14.41	14.56	14.80

Table 3. Mean Yen and Percent Spreads by 30-Minute Intraday Trading Intervals

This table presents the cross-sectional mean estimates of the spread in yen and percent terms during 30-minute intraday trading intervals for 204 TSE stocks used in this study. The standard errors are in parentheses. Panels A, B, and C provide the mean estimates for the firms whose average prices are less than or equal to ¥2,000, between ¥2,001 and ¥3,000, and between ¥3,001 and ¥30,000 respectively.

Panel A. Group 1: Price ≤ ¥2,000 (N = 171)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
Spread in ¥	2.89 (0.14)	2.21 (0.09)	2.10 (0.08)	2.15 (0.09)	2.12 (0.08)	2.00 (0.08)	1.97 (0.07)	2.01 (0.07)	2.41 (0.09)
Spread in %	0.66 (0.03)	0.55 (0.03)	0.53 (0.03)	0.54 (0.03)	0.54 (0.03)	0.52 (0.03)	0.52 (0.03)	0.53 (0.03)	0.60 (0.03)

Panel B. Group 2: ¥2,000 < Price ≤ ¥3,000 (N = 11)

	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
Spread in ¥	9.96 (1.12)	7.75 (0.67)	7.19 (0.45)	7.56 (0.55)	7.42 (0.59)	6.90 (0.49)	6.78 (0.40)	7.00 (0.48)	8.67 (0.83)
Spread in %	0.40 (0.04)	0.31 (0.03)	0.29 (0.02)	0.30 (0.02)	0.30 (0.02)	0.28 (0.02)	0.27 (0.02)	0.28 (0.02)	0.35 (0.03)

Panel C. Group 3: ¥3,000 < Price ≤ ¥30,000 (N = 22)

	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
Spread in ¥	40.26 (6.69)	30.45 (4.94)	27.86 (4.51)	27.52 (4.02)	27.60 (3.89)	25.22 (3.51)	24.69 (3.72)	24.98 (3.51)	32.05 (4.84)
Spread in %	0.45 (0.04)	0.35 (0.03)	0.32 (0.02)	0.33 (0.02)	0.33 (0.02)	0.30 (0.02)	0.30 (0.02)	0.30 (0.02)	0.38 (0.03)

Table 4. GMM Estimates of Spread Components, Autocorrelation of Trade Directions, and Implied Spreads

This table presents the GMM model estimates of the parameters and the spread implied by the parameters for 204 TSE stocks used in this study. The parameters of interest are the adverse selection cost component of the spread (α), the order processing cost component of the spread (β), and autocorrelation of trade directions (ρ). The implied spread is defined as 2 ($\alpha + \beta$). The proportion of the adverse selection in the implied spread is denoted as γ. Panels A, B, and C provide the estimates for firms whose average prices are less than or equal to ¥2,000, between ¥2,001 and ¥3,000, and between ¥3,001 and ¥30,000 respectively. Panel D presents the statistical test results for the null hypothesis that there is no difference in the estimates between different price groups.

Panel A. Group 1: Price ≤ ¥2,000 (N = 171)

	Mean	Mean S.E.	Std. Dev.	Median
α (¥)	0.460	0.022	0.287	0.416
β (¥)	0.395	0.010	0.132	0.370
α (% of Price)	0.091	0.003	0.035	0.085
β (% of Price)	0.141	0.013	0.172	0.072
ρ	0.315	0.005	0.061	0.316
γ	0.498	0.013	0.175	0.528
Implied Spread (¥)	1.710	0.053	0.693	1.450
Implied Spread (%)	0.464	0.028	0.368	0.330

Panel B. Group 2: ¥2,000 < Price ≤ ¥3,000 (N = 11)

	Mean	Mean S.E.	Std. Dev.	Median
α (¥)	1.416	0.176	0.584	1.367
β (¥)	1.657	0.092	0.306	1.640
α (% of Price)	0.057	0.007	0.023	0.058
β (% of Price)	0.067	0.004	0.015	0.066
ρ	0.235	0.017	0.056	0.256
γ	0.448	0.043	0.143	0.467
Implied Spread (¥)	6.146	0.224	0.744	6.264
Implied Spread (%)	0.248	0.012	0.038	0.248

**Table 4. GMM Estimates of Spread Components, Autocorrelation of
Trade Directions, and Implied Spreads (Continued)**

Panel C. Group 3: ¥3,000 < Price ≤ ¥30,000 (N = 22)

α (¥)	6.268	0.944	4.427	5.195
β (¥)	4.001	0.371	1.738	3.177
α (% of Price)	0.071	0.006	0.026	0.074
β (% of Price)	0.052	0.004	0.020	0.047
ρ	0.284	0.013	0.060	0.292
γ	0.571	0.029	0.138	0.587
Implied Spread (¥)	20.539	2.415	11.329	15.805
Implied Spread (%)	0.246	0.013	0.060	0.254

Panel D. : *p*-values from Wilcoxon Signed Rank Test for the Difference
　　　in Median Estimates

	Between Groups 1 and 2	between Groups 1 and 3	between Groups 2 and 3
α (% of Price)	0.001	0.018	0.151
β (% of Price)	0.505	0.001	0.025
ρ	0.000	0.054	0.031
γ	0.240	0.045	0.024
Implied Spread (%)	0.009	0.000	0.932

Table 5. GMM Model Parameters and Implied Spreads by 30-Minute Intraday Intervals

This table presents the cross-sectional mean and mean standard error (in parentheses) of the GMM model estimates of the parameters and the implied spread by 30-minute intraday trading intervals for 204 TSE stocks used in this study. The parameters of interest are the adverse selection cost component of the spread (α), order processing cost component of the spread (β), and autocorrelation of trade directions (ρ). The implied spread is defined as $2(\alpha + \beta)$. The proportion of the adverse selection in the implied spread is denoted as γ. Panels A, B, and C provide the estimates for the firms whose average prices are less than or equal to ¥2,000, between ¥2,001 and ¥3,000, and between ¥3,001 and ¥30,000 respectively.

Panel A. Group 1: Price ≤ ¥2,000 (N = 171)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
α (¥)	0.570	0.470	0.428	0.444	0.429	0.403	0.408	0.409	0.544
	(0.029)	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)	(0.020)	(0.019)	(0.026)
β (¥)	0.481	0.359	0.400	0.392	0.376	0.373	0.369	0.372	0.409
	(0.020)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.011)
α (% of Price)	0.112	0.094	0.085	0.087	0.085	0.079	0.083	0.083	0.108
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
β (% of Price)	0.151	0.132	0.142	0.141	0.138	0.140	0.138	0.138	0.142
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
ρ	0.338	0.303	0.303	0.310	0.363	0.314	0.310	0.303	0.304
	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.004)
γ	0.508	0.530	0.478	0.490	0.494	0.478	0.487	0.486	0.528
	(0.014)	(0.016)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.014)	(0.013)
Implied Spread (¥)	2.102	1.658	1.655	1.672	1.610	1.552	1.555	1.562	1.907
	(0.080)	(0.053)	(0.053)	(0.053)	(0.048)	(0.045)	(0.044)	(0.044)	(0.061)
Implied Spread (% of Price)	0.526	0.451	0.455	0.456	0.446	0.438	0.441	0.441	0.501
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)	(0.029)	(0.028)

Table 5. GMM Model Parameters and Implied Spreads by 30-Minute Intraday Intervals (Continued)

Panel B. Group 2: ¥2,000 < Price ≤ ¥3,000 (N = 11)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
α (¥)	1.882	1.477	1.193	1.330	1.295	1.159	1.226	1.254	1.769
	(0.259)	(0.190)	(0.126)	(0.174)	(0.153)	(0.143)	(0.150)	(0.176)	(0.245)
β (¥)	1.659	1.582	1.718	1.620	1.587	1.623	1.628	1.530	1.800
	(0.156)	(0.121)	(0.094)	(0.092)	(0.086)	(0.094)	(0.101)	(0.141)	(0.058)
α (% of Price)	0.075	0.059	0.048	0.053	0.052	0.047	0.049	0.050	0.071
	(0.010)	(0.007)	(0.005)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)
β (% of Price)	0.068	0.064	0.069	0.065	0.064	0.066	0.066	0.062	0.073
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.003)
ρ	0.253	0.221	0.219	0.225	0.268	0.238	0.237	0.229	0.237
	(0.024)	(0.017)	(0.018)	(0.019)	(0.015)	(0.018)	(0.020)	(0.015)	(0.021)
γ	0.515	0.471	0.404	0.438	0.439	0.409	0.421	0.443	0.473
	(0.053)	(0.048)	(0.037)	(0.045)	(0.040)	(0.042)	(0.043)	(0.057)	(0.042)
Implied Spread (¥)	7.082	6.119	5.822	5.900	5.764	5.563	5.710	5.569	7.138
	(0.387)	(0.241)	(0.186)	(0.209)	(0.181)	(0.165)	(0.212)	(0.163)	(0.384)
Implied Spread (% of Price)	0.285	0.246	0.235	0.238	0.233	0.225	0.231	0.225	0.287
	(0.018)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.012)	(0.010)	(0.016)

Table 5. GMM Model Parameters and Implied Spreads by 30-Minute Intraday Intervals (Continued)

Panel C. Group 3: ¥3,000 < Price ≤ ¥30,000 (N = 22)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00
α (¥)	7.783	6.592	5.910	6.364	5.994	5.539	5.178	5.631	7.052
	(1.111)	(1.012)	(0.945)	(1.027)	(0.847)	(0.861)	(0.845)	(0.893)	(0.996)
β (¥)	5.013	4.050	3.972	3.674	3.634	3.585	3.596	3.465	4.655
	(0.669)	(0.501)	(0.409)	(0.368)	(0.375)	(0.313)	(0.258)	(0.310)	(0.395)
α (% of Price)	0.089	0.074	0.066	0.072	0.069	0.062	0.058	0.063	0.081
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
β (% of Price)	0.057	0.050	0.051	0.048	0.046	0.048	0.050	0.049	0.061
	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)
ρ	0.293	0.278	0.275	0.283	0.319	0.292	0.282	0.282	0.272
	(0.016)	(0.015)	(0.014)	(0.015)	(0.013)	(0.014)	(0.011)	(0.014)	(0.013)
γ	0.594	0.590	0.557	0.589	0.589	0.562	0.537	0.566	0.565
	(0.026)	(0.030)	(0.032)	(0.035)	(0.032)	(0.035)	(0.033)	(0.038)	(0.026)
Implied Spread (¥)	25.593	21.283	19.763	20.077	19.255	18.248	17.547	18.191	23.414
	(3.268)	(2.753)	(2.534)	(2.517)	(2.124)	(2.113)	(1.943)	(1.948)	(2.632)
Implied Spread (% of Price)	0.292	0.248	0.234	0.241	0.232	0.220	0.216	0.224	0.284
	(0.013)	(0.013)	(0.013)	(0.014)	(0.012)	(0.012)	(0.014)	(0.013)	(0.015)

Table 6. GMM Model Parameters and Implied Spreads for Stock Groups based on Trading Volume

This table presents the cross-sectional means and mean standard errors (in parentheses) of the GMM model estimates of the parameters and the implied spread by 30-minute intraday trading intervals as well as for the entire day for the 204 TSE stocks used in this study. The parameters of interest are the adverse selection cost component of the spread (α) and order processing cost component of the spread (β), both in percentages of stock price. The implied spread is defined as 2 ($\alpha + \beta$) divided by the stock price. The proportion of the adverse selection in the implied spread is denoted as γ.

Panel A. Mean Daily Share Volume<810,000 (N: 68)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00	Day
α	0.135	0.112	0.103	0.108	0.102	0.096	0.098	0.101	0.124	0.109
	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
β	0.131	0.112	0.124	0.123	0.120	0.121	0.121	0.118	0.126	0.123
	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)
Implied spread	0.532	0.448	0.455	0.462	0.443	0.435	0.438	0.437	0.500	0.464
	(0.045)	(0.043)	(0.044)	(0.044)	(0.043)	(0.044)	(0.045)	(0.045)	(0.045)	(0.044)
γ	0.581	0.601	0.555	0.561	0.560	0.551	0.561	0.574	0.595	0.570
	(0.023)	(0.024)	(0.022)	(0.020)	(0.023)	(0.023)	(0.023)	(0.023)	(0.020)	(0.020)

Table 6. GMM Model Parameters and Implied Spreads for Stock Groups based on Trading Volume (Continued)

Panel B. 820,000 ≤ Mean Daily Share Volume < 1,511,000 (N: 68)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00	Day
α	0.110	0.093	0.082	0.084	0.085	0.077	0.082	0.082	0.109	0.090
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
β	0.137	0.117	0.128	0.126	0.123	0.126	0.123	0.124	0.130	0.126
	(0.019)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Implied spread	0.494	0.419	0.421	0.420	0.416	0.405	0.409	0.412	0.478	0.433
	(0.040)	(0.040)	(0.041)	(0.040)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
γ	0.522	0.543	0.481	0.498	0.506	0.478	0.489	0.489	0.535	0.505
	(0.019)	(0.021)	(0.019)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.018)	(0.018)

Panel C. Mean Daily Share Volume ≥ 1,511,000 (N: 68)

Trading Interval	9:00 – 9:30	9:31 – 10:00	10:01 – 10:30	10:31 – 11:00	12:30 – 13:00	13:01 – 13:30	13:31 – 14:00	14:01 – 14:30	14:31 – 15:00	Day
α	0.079	0.065	0.057	0.059	0.058	0.054	0.054	0.054	0.076	0.063
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)
β	0.140	0.128	0.133	0.131	0.128	0.130	0.130	0.131	0.134	0.132
	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)
Implied spread	0.438	0.387	0.380	0.382	0.374	0.368	0.369	0.37	0.421	0.389
	(0.041)	(0.041)	(0.042)	(0.042)	(0.041)	(0.041)	(0.042)	(0.042)	(0.041)	(0.041)
γ	0.451	0.455	0.411	0.436	0.438	0.422	0.415	0.413	0.458	0.434
	(0.020)	(0.023)	(0.021)	(0.023)	(0.023)	(0.024)	(0.022)	(0.022)	(0.020)	(0.021)

Table 7. GMM Model Parameters and Implied Spreads During the Last 30-Minute Period of Trading Day

This table presents the cross-sectional means and mean standard errors (in parentheses) of the GMM model estimates of the parameters and the implied spread by 5-minute trading intervals during the last half hour period of trading day for the 204 TSE stocks used in this study. The parameters of interest are the adverse selection cost component of the spread (α) and order processing cost component of the spread (β), both in percentages of stock price. The implied spread is defined as $2 (\alpha + \beta)$ divided by the stock price. The proportion of the adverse selection in the implied spread is denoted as γ.

Panel A. Group 1: Price $\leq$ ¥2,000 (N = 171)

Trading Interval	14:31 – 14:35	14:36 – 14:40	14:41 – 14:45	14:46 – 14:50	14:51 – 14:55	14:56 – 15:00
α	0.102	0.095	0.093	0.101	0.107	0.165
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.006)
β	0.125	0.133	0.136	0.136	0.148	0.148
	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)
Implied spread	0.454	0.456	0.457	0.474	0.511	0.627
	(0.029)	(0.028)	(0.028)	(0.028)	(0.029)	(0.030)
γ	0.557	0.523	0.515	0.534	0.526	0.591
	(0.019)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)

Table 7. GMM Model Parameters and Implied Spreads During the Last 30-Minute Period of Trading Day (Continued)

Panel B. Group 2: ¥2,000 < Price ≤ ¥3,000 (N = 11)

Trading Interval	14:31 – 14:35	14:36 – 14:40	14:41 – 14:45	14:46 – 14:50	14:51 – 14:55	14:56 – 15:00
α	0.053	0.052	0.052	0.073	0.079	0.119
	(0.008)	(0.008)	(0.007)	(0.011)	(0.011)	(0.019)
β	0.074	0.071	0.070	0.063	0.069	0.087
	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.010)
Implied spread	0.254	0.246	0.244	0.272	0.296	0.412
	(0.018)	(0.013)	(0.012)	(0.016)	(0.015)	(0.032)
γ	0.398	0.407	0.411	0.511	0.510	0.549
	(0.046)	(0.046)	(0.042)	(0.055)	(0.049)	(0.052)

Panel C. Group 3: ¥3,000 < Price ≤ ¥30,000 (N = 22)

Trading Interval	14:31 – 14:35	14:36 – 14:40	14:41 – 14:45	14:46 – 14:50	14:51 – 14:55	14:56 – 15:00
α	0.061	0.067	0.067	0.074	0.093	0.121
	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)	(0.009)
β	0.058	0.051	0.059	0.062	0.063	0.070
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.007)
Implied spread	0.237	0.235	0.252	0.273	0.312	0.383
	(0.014)	(0.013)	(0.014)	(0.014)	(0.020)	(0.022)
γ	0.512	0.567	0.528	0.542	0.590	0.628
	(0.032)	(0.033)	(0.039)	(0.026)	(0.033)	(0.031)

Table 8. GMM Estimates of Spread Components by Trade Size

This table presents the GMM model estimates of the parameters and standard errors (in parentheses) of the adverse selection cost component of the spread (α) and the order processing cost component of the spread (β) sorted by trade size categories for 204 TSE stocks used in this study. Also reported are the F-test results on the equality of the estimates across each row and column. t-1 and t-2 represent the time of the trades with respect to the current quote at t. * and ** indicate significance at the 5% and 1% levels respectively.

Panel A. Adverse Selection Cost (α) by Trade Size

t-2 \ t-1	Small	Medium	Large	F-value
Small	0.059 (0.002)	0.090 (0.003)	0.127 (0.006)	84.23**
Medium	0.068 (0.002)	0.092 (0.003)	0.125 (0.005)	65.59**
Large	0.078 (0.003)	0.106 (0.004)	0.127 (0.005)	34.86**
F-value	14.40**	7.05**	0.04	

Panel B. Order Processing Cost (β) by Trade Size

t-2 \ t-1	Small	Medium	Large	F-value
Small	0.158 (0.012)	0.124 (0.012)	0.091 (0.011)	8.30**
Medium	0.147 (0.012)	0.122 (0.012)	0.093 (0.011)	5.13**
Large	0.133 (0.011)	0.104 (0.011)	0.105 (0.012)	2.14
F-value	1.09	0.85	0.43	

Table 9 Estimated Spread Components by De Jong, Nijman, Roell Model and Glosten and Harris Model

This tables compares the De Jong, Nijman, and Roell (1996) model, the Glosten and Harris (1988) model, and the Madhavan, Richardson, and Roomans (1997) or MRR model of spread components. All three models are estimated for 204 TSE stocks used in this study. The sample period is from January 5 to March 30, 2000.

Panel A: Mean parameter estimates from alternative models

	De Jong, Nijman, and Roell (1996) model				Glosten and Harris (1988) model			
	R_0-c_0	R_1-c_1	c_0	c_1	z_0	z_1	c_0	c_1
Group 1	0.290	0.006	0.540	-0.001	0.257	0.012	0.573	-0.006
(N=171)	(0.014)	(0.001)	(0.015)	(0.001)	(0.013)	(0.001)	(0.015)	(0.001)
Group 2	0.967	0.043	1.921	0.011	0.745	0.086	2.142	-0.032
(N=11)	(0.108)	(0.009)	(0.057)	(0.004)	(0.078)	(0.019)	(0.044)	(0.008)
Group 3	4.232	0.180	5.519	0.075	3.549	0.361	6.202	-0.105
(N=22)	(0.627)	(0.054)	(0.584)	(0.029)	(0.532)	(0.108)	(0.643)	(0.028)

Panel B: Adverse selection component, order-processing component, and proportion of adverse selection

	De Jong, Nijman, and Roell (1996) model				Glosten and Harris (1988) model				MRR model
	$q = 1$	q = median	$q = 10$	q=99% MTU	$q = 1$	q = median	$q = 10$	q=99% MTU	
Group 1 (N=171)									
Adverse selection	0.295	0.301	0.349	0.472	0.269	0.280	0.375	0.622	0.460
Order-processing	0.540	0.541	0.537	0.553	0.567	0.562	0.510	0.403	0.395
Proportion	0.328	0.332	0.366	0.443	0.295	0.306	0.394	0.596	0.498
Group 2 (N=11)									
Adverse selection	1.010	1.043	1.397	1.943	0.832	0.899	1.607	2.699	1.416
Order-processing	1.932	1.941	2.032	2.146	2.110	2.086	1.823	1.390	1.657
Proportion	0.335	0.341	0.389	0.465	0.276	0.294	0.444	0.650	0.448
Group 3 (N=22)									
Adverse selection	4.412	4.469	6.035	6.333	3.909	4.022	7.155	7.751	6.268
Order-processing	5.594	5.608	6.268	6.217	6.097	6.055	5.148	4.799	4.001
Proportion	0.417	0.421	0.458	0.494	0.361	0.374	0.527	0.624	0.571

Table 10 Half Spread Estimates by Hasbrouk's VAR Model

This tables presents the estimation results of the Hasbrouck vector autoregression (VAR) model (1988, 1991a, b). Panel A reports the mean parameter estimates and their standard errors (in parentheses) of the VAR coefficients of the model. Panel B compares the estimated half-spread from the Glosten and Harris (1988) and De Jong, Nijman, and Roell (1996) models, the Hasbrouck vector autoregression (VAR) model (1988, 1991a, b), and the MRR model. All models are estimated for 204 TSE stocks used in this study. The sample period is from January 5 to March 30, 2000.

Panel A: Mean parameter estimates of VAR coefficients on Q_t and $q_t Q_t$

	a_{01}	a_{02}
Group 1	0.821	0.006
(N=171)	(0.026)	(0.001)
Group 2	2.831	0.055
(N=11)	(0.093)	(0.011)
Group 3	9.691	0.208
(N=22)	(1.163)	(0.049)

Panel B: Bid-ask half-spread and trade size

	De Jong, Nijman, and Roell (1997) and Glosten and Harris (1988) model				Hasbrouck (1988, 1991a, b) VAR model				MRR model
	q = 1	q = median	q = 10	q=99% MTU	q = 1	q = median	q = 10	q=99% MTU	
Group 1 (N=171)									
Mean	0.835	0.842	0.885	1.026	0.827	0.833	0.880	1.021	0.855
Mean s.e.	(0.027)	(0.027)	(0.027)	(0.027)	(0.026)	(0.026)	(0.027)	(0.029)	(0.026)
Group 2 (N=11)									
Mean	2.942	2.984	3.430	4.089	2.886	2.931	3.381	4.063	3.073
Mean s.e.	(0.100)	(0.106)	(0.196)	(0.222)	(0.103)	(0.109)	(0.196)	(0.228)	(0.112)
Group 3 (N=22)									
Mean	10.006	10.077	12.303	12.550	9.899	9.971	11.774	12.216	10.269
Mean s.e.	(1.204)	(1.197)	(1.750)	(1.360)	(1.192)	(1.185)	(1.498)	(1.255)	(1.208)