# Three Essays on the Cost Components of the Bid-Ask Spread

Paul Leventhal

A Thesis

in

The Faculty of Commerce

and Administration

Presented in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

December 1999

0-612-48306-1

## ABSTRACT

## THREE ESSAYS ON THE COST COMPONENTS OF THE BID-ASK SPREAD

**Paul Leventhal, Ph. D.**
**Concordia University, 1999**

This dissertation consists of three interrelated essays. The first essay focuses on the adverse selection component of the bid-ask spread. A regime switching model applied to the trading process leads to a parsimonious model of the time-series evolution of the bid-ask spread in which market participants use trade data to answer the following question: Is there currently private information in the market for a given stock? If there is a high probability of private information in the market, this leads contemporaneously to a greater revision in beliefs about the true price. Together with compensation for transactions costs, this leads to a greater revision in transaction price. Using TSE 35 trade and quote data for March and May 1996, the pooled cross-section and time series results support this view.

The second essay examines the costs of adverse information and order processing in light of transaction size, type of trader and type of trading method. Specifically, it is found that adverse selection increases with the trade size (consistent with numerous empirical studies relating trade size and the cost components of the bid-ask spread). However, whether the trade was undertaken by the designated market maker, by a principal trader or by an individual belonging to neither of these two categories is also of

importance. In addition, we show that trades consummated within the investment dealer's firm have a lower adverse information cost component than trades executed externally. For order processing, it is found that the single most important determinant of cost is whether the transaction is internal or external to the investment dealer firm, with internal trades being more costly.

The third essay examines the robustness of the Huang and Stoll (1997) model estimates to the use of different clustering methods and to a minimum quotation increment reduction (MQIR) on the Toronto Stock Exchange. We find that adequate reversal of trade flow is a necessary but not sufficient condition for model performance. We also find that model estimates are quite sensitive to the data clustering method selected. In addition, we show that this model fails to provide adequate cost component estimates of the spread in the post-MQIR period due to a fundamental change in market-maker behavior.

## ACKNOWLEDGEMENT

The completion of this Doctoral Degree has taken a long time, and a great deal of effort. The effort has been aided by a number of people. First, I wish to thank my thesis committee members Dr. Lorne Switzer of Concordia Univerity and Dr. Minh Chau To of H.E.C. for their advice, encouragement and support. Dr. Gregory Lypny of Concordia University provided those things as well while serving on a prior committee. It has been a pleasure working with them.

I believe it accurate to say that this thesis would never have been completed without the dedication, encouragement and abilities of my supervisor, Dr. Lawrence Kryzanowski. In the difficult times, often accompanied by his wife Louise Carpentier, he provided the support and friendship I needed to continue. In the good times, he encouraged me to do more. In his regard, my gratitude is very great indeed.

## DEDICATION

This thesis is dedicated to two people: my grandfather, Sol Goldstein, who, although not here to witness its completion, would have been very proud; and to my sister, Jill, who has always provided encouragement.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

Information is imbedded in market prices as the result of the interplay of three elements of price: bid prices, ask prices and transaction prices. Market makers, as providers of liquidity, set the bid and ask prices while other market participants, informed and uninformed, decide how best to transact. Informed market participants, who are impatient to benefit from their information advantage, trade at the bid or ask. Uninformed investors set limit prices or, by disclosing their lack of information, may transact at prices within the bid-ask spread.

In this type of market, market makers not only have costs which they must cover, but also bear risk for which they require compensation. Specific costs include the cost of maintaining inventory and the fixed cost of processing orders. Since market makers are required to transact in the absence of public orders, they bear the risk that trades have to be undertaken with informed market participants. The compensation for this risk bearing, the compensation for the risk of information asymmetry, is an additional cost captured by the bid-ask spread. Thus, the primary focus of this dissertation is the cost of information asymmetry, how it is measured, and how it is affected by specific characteristics of transactions and markets.

Microstructure models generally make a number of assumptions. First, providers of market liquidity, generally market makers, set the bid-ask spread on the basis of immediately preceding trades. The direction of that preceding trade (i.e., buyer- or seller-initiated) may reveal information unknown to the market maker which may lead to an increase in bid and ask prices as a defensive measure. In addition, assuming that the market maker has provided liquidity for the transaction, the market maker's inventory subsequently must be rebalanced. Second, trades occur between anonymous market participants who are divisible into two categories, the public and the market maker. Third, the mechanism by which market makers set bid and ask prices are invariant to changing market structures. In the chapters to follow, we relax these assumptions in specific ways, and show that empirical results are dependent on these changes.

In Chapter 2, "Information Regimes, Transaction Prices and the Information Content of the Trading Process," we assume that market makers rely on more information than that contained in the immediately preceding trade. Specifically, it is assumed that from market open until market close, the market maker uses the entire sequence of transactions during each trading day in order to determine if informed traders are currently in the market. The trading day is divided into 26, 15-minute intervals. In each interval, the probability of informed trade is determined using a bivariate regime switching model. In order to analyze the microstructure effects, we begin with a microstructure model developed by Madhavan, Richardson and Roomans (1997), and then modify it to include the probability measure. The model is estimated using TSE35 transaction data for two months in 1996; namely, March and May. In April 1996, the TSE instituted a reduction in the minimum quotation increment (commonly called the tick size). We use the newly-developed model to examine the impact of this change in market structure. Three major conclusions are reached. First, modeling the price impact of information asymmetry using a bivariate regime switching model does a good job of explaining the observed features of the marketplace, such as the U-shaped pattern of bid-ask spreads over the trading day. Second, uncertainty about the presence of informed trade is lower throughout most of the trading day when compared to that at the market open. Third, there is evidence that information asymmetry increases when the tick size is reduced. However, this increase in information asymmetry does not result in an increase in transaction cost.

Lacking in virtually all microstructure models is the empirical reality of market fragmentation. Market participants are more diverse than simple models suggest. Brokers may create an internal market for transactions – known as the "upstairs" market – in which some trades are filled and then reported to the exchange, without ever reaching the exchange for possible or actual execution. In addition, market participants differ in terms of the degree to which they may be informed traders. For example, designated market makers differ from professional traders and the trading public in this regard. Chapter 3, "Cost Components of the Bid-Ask Spread: Trade Size, Trader and Internalization," continues the examination of the cost components of the bid-ask spread by incorporating market fragmentation and trader type into the microstructure model of Lin, Sanger and Booth (1995). Using TSE35 data prior to the reduction in tick size, and exploiting the trader identification provided with our source of transactions data, we find that trade size is an important determinant of both the costs of information asymmetry and order processing. This finding is consistent with the theoretical implications in the existing microstructure literature. We also conclude that trader type (market maker, professional trader, or member of the trading public) is an important determinant of the cost of information asymmetry. Trades executed with either the designated market maker or professional traders have the highest costs. We further conclude that internal trades have the highest order processing cost, as theory predicts. We test the robustness of these conclusions with data following the TSE tick size reduction. We conclude that these results are invariant to the change and, consistent with the results of Chapter 2, we do not observe an increase in the cost of information asymmetry.

In general, it is difficult to disaggregate the three elements of cost embedded within the bid-ask spread: information asymmetry, inventory management, and order processing. However, it is important to do so since the quality of a market depends, in part, on the relationship between these costs. In an effort to measure the costs separately, Huang and Stoll (1997) develop a general microstructure model. Huang and Stoll test their model using NYSE-Amex data, and find that the model performs counter to expectations using all transactions available over a period of time. Moreover, their estimates of the cost of information asymmetry are negative – an outcome that no theoretical model can explain. Huang and Stoll find that by aggregating specific types of transactions, model performance and measurement improve dramatically. In Chapter 4, "Robustness of Spread Component Estimates to Bunching Method and Decimalization," we find that the method of data aggregation has a significant impact on the performance of their model, and that the model fails to perform well following the reduction in tick size on the TSE using either unbunched or bunched data. We exploit the trader identification feature of the TSE data set to provide a justifiable type of aggregation or bunching. However, we do confirm the Huang and Stoll result using TSE35 data for the months prior to the reduction in tick size. As in Huang and Stoll, the model is estimated herein using the Generalized Method of Moments.

The three chapters of the dissertation are linked through a common data set, TSE35 transaction data immediately prior to and subsequent to the change to a smaller tick size, and through a common focus, the measurement of the cost components of the bid-ask spread. Individual chapters either develop different models (Chapters 2 and 3), or employ existing models in new ways (Chapters 2 and 4). In all cases, the work builds on the existing literature. In sum, the dissertation extends the literature in at least four ways. First, it is likely that market makers use more than the immediately preceding transaction in setting bid and ask prices. Second, although information asymmetry increases when the tick size is smaller, that increase in asymmetry may not be priced within the spread. Another mechanism may be at play protecting the market maker from the risk of trading with informed traders. That mechanism may be a reduction in quoted depth. Third, aggregating transactions is likely to change inferences. Since these inferences may be used by policy makers, the choice of bunching method must be done with considerable care. Fourth, trade size is important in determining the costs of both information asymmetry as well as the cost of order processing, but other elements such as trader type and market structure also significantly affect the magnitudes of these costs.

# CHAPTER 2

# INFORMATION REGIMES, TRANSACTION PRICES AND THE INFORMATION CONTENT OF THE TRADING PROCESS

## 2.1. INTRODUCTION

In perfect capital markets, transaction prices vary with the random arrival of information and the bid-ask spread is zero. However, in real markets, information is not freely and simultaneously available to all market participants. Moreover, the bid-ask spread must compensate market makers for transaction costs, such as the cost of maintaining inventory, providing liquidity to the market and order processing. In addition, if market makers trade with better (i.e., privately) informed traders, losses will result. Thus, compensation also is required for this risk. Theoretical microstructure models (e.g. Ho and Stoll (1981), Glosten (1987), Glosten and Harris (1988)) model transaction costs and information asymmetry. The relative economic importance of the components of the bid-ask spread varies from study to study. Stoll (1989) finds that 43% of the quoted spread is due to information asymmetry (referred to as adverse selection), while George, Kaul and Nimalendran (1991) find that adverse selection accounts for 8% to 13% of the quoted spread with order processing costs accounting for the rest. All studies, however, attribute some importance to information asymmetry. Modeling this asymmetry is the principal focus of this chapter.

The source of information asymmetry is exogenous to the trading process. The information possessed by informed traders is obtained outside of the market. However, the existence of private information must be reflected in the sequence of transaction prices. Those prices are affected by the revisions in beliefs about true price. Informed traders earn excess profits because of their information advantage, and uninformed traders scan the transaction flow for evidence of their presence. Buy (sell) transactions initiated by informed traders lead the uninformed to revise their estimate of true prices upward (downward). Simple models of this process have been proposed in the literature.[1] They assume that uninformed market participants revise their beliefs about true price by observing the most recent trade. These models provide insight into the significance of market frictions, such as the determinants of the volatility in transaction prices (see for example, Huang and Stoll (1997)). They have been less successful at explaining empirical features of the trading process. For example, theoretical models predict that information asymmetry should decline monotonically throughout the trading day as market participants learn through the trading process. This resolution of uncertainty should be reflected in a narrowing of spreads. However, spreads exhibit a U-shaped pattern throughout the day (for example, see Harris 1986).

This chapter extends the existing literature in the following way: We assume that uninformed market participants, including market makers, observe more that just the most recent trade in revising their beliefs about true price. Rather, they observe the past sequence of transactions throughout the trading day. Using that information, they determine whether or not there is private information in the market at the "current" time. The probability of private information in the market leads to a revision in beliefs about true price. We model the probability of information in the market as a change in regime.

[1] See, for example, Ho and Stoll (1981), Glosten (1987), and Stoll (1989).

This modeling and resulting estimations lead to three conclusions. First, statistical evidence exists to suggest that market participants revise beliefs about true price by adding information gleaned from new trading to that already obtained from the previous sequence of trades observed since the beginning of the day. Second, consistent with theoretical models, the probability of the existence of information in the market declines from market open. However, following mid-day, the probability of informed trade rises. As a result, the intra-day evolution of the probability of informed trade is U-shaped. We return to this issue in the conclusion to the chapter. Our third conclusion, based on data prior to and subsequent to a reduction in minimum price variation on the Toronto Stock exchange, is that information asymmetry increases subsequent to a reduction in tick size. This result is consistent with the models of Anshuman and Kalay (1993) and Harris (1994) and differs from the empirical results of Bacidore (1997) who finds no increase in information asymmetry subsequent to tick reduction.

The chapter is organized as follows: In Section 2.2, we set out the model to be used to examine the relationship between private information and transaction price changes. In Section 2.3, we show how the probability of the existence of a private information regime is determined. The nature of the data set is discussed in Section 2.4. Section 2.5 presents the empirical results. In section 2.5.1 some of the basic statistical characteristics of the data are described. In Section 2.5.2 we focus specifically on the regime probabilities. Sections 2.5.3 and 2.5.4 discuss the parametric estimates of the model developed in Section 2.2 and the robustness of those results, respectively. Section 2.6 concludes this chapter.

## 2.2 THE MODEL

In Madhavan, Richardson and Roomans (MRR) (1996), as in microstructure model specifications since Glosten and Milgrom (1985), the impact of private information on transaction prices is modeled as an innovation in belief about true price. In MRR, this innovation in belief is some function of a trade initiation variable. In general, buyer-initiated trades, seller-initiated trades and neither buyer- nor seller-initiated trades have an impact on transaction prices through an impact on the unobserved true price of the stock. Letting:

$V_t \equiv$ the unobservable true price of the stock;

$x_t \in \{-1,0,+1\} \equiv$ the trade initiation variable (seller-initiated, neither buyer- nor seller-initiated, and buyer-initiated);

$\gamma \equiv$ the coefficient of private information; and

$\varepsilon_t \equiv$ the impact of public information, assumed to arrive randomly;

then:

$$\Delta V_t \equiv V_t - V_{t-1} = \gamma g(x_t \mid \Omega_{t-1}) + \varepsilon_t \, .$$

where g(•) is some function of trade initiation, and $\Omega_{t-1}$ is the information set consisting of $x_{t-1},\ldots,x_{t-n}$. It is assumed that $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$.

MRR model g(•) as:

$$x_t - E(x_t | x_{t-1}) = z_t$$

which they show is equivalent to:

$$x_t - \rho x_{t-1} = z_t \qquad (1)$$

where ρ is the correlation coefficient of the trade initiation variable, and the process on $z_t$ is AR(1). In their formulation, the true price of the stock depends on the innovation in trade flow, $z_t$, and on the arrival of public information such that:

$$V_t = V_{t-1} + \gamma(x_t - \rho x_{t-1}) + \varepsilon_t. \tag{2}$$

Trade flow is autocorrelated since continuations are more likely than reversals given that informed traders camouflage their information advantage by breaking large orders into smaller transaction lots.

In this chapter we propose an alternate modeling of the impact of private information on true (and transaction) price. Specifically, it is hypothesized that trades take place during periods characterized by a belief in the presence or absence of privately informed traders. We refer to these periods as information regimes, $s_t$, such that $s_t \in \{1,2\}$. In regime $s_t = 1$, it is believed that privately informed traders with either positive or negative information are in the market. In $s_t = 2$, it is believed that there are no privately informed traders. True price in this model is a function of the probability that we are in a private information regime. In

$$V_t = V_{t-1} + \gamma I_t + \varepsilon_t \tag{3}$$

$I_t$ is the probability that the current regime is the one in which private information exists. Formally that probability is:

$$P(s_t = 1 \mid \mathbf{y}_t, \dots, \mathbf{y}_{t-n})$$

where $\mathbf{y}_t$ is a bivariate vector process made up of the sum of buyer-initiated and seller-initiated trades (within a given time interval) and the number of non-buyer- or seller-initiated trades (during that same interval). Those two variables are assumed to be independently and normally distributed.

All past realizations of $\mathbf{y}_t$, up to and including the current realization, are used to infer whether the current regime is $s_t = 1$. For example, if $p(s_t = 1 \mid \mathbf{y}_t,...,\mathbf{y}_{t-n})$ is high (i.e., it is likely that traders with private information are in the market for that stock), then $V_t$ should change more in value than when $p(s_t = 1 \mid \mathbf{y}_t,...,\mathbf{y}_{t-n})$ is low, ceterus peribus. By contrast, if $p(s_t = 1 \mid \mathbf{y}_t,...,\mathbf{y}_{t-n})$ is close to zero (i.e., it is believed unlikely that informed traders are in the market), then $V_t$ should not be affected.

Equation (3) cannot form the basis of estimation since $V_t$ is unobservable. Transaction price, $P_t$, depends on beliefs about the true price and also on other factors such as the dealer's cost of providing liquidity, the cost of holding inventory, and compensation for risk bearing. Following MRR, we assume that:

$$P_t = V_t + \phi x_t + \xi_t \qquad (4)$$

where $\xi_t$ is a mean-zero, normally distributed random variable capturing the effect of price discreteness, and $\phi$ is the dealer cost of providing liquidity and maintaining inventory. Note that (3) and (4) together imply the following maximum ask and minimum bid prices conditional on the trade initiation variable:

$$P_t^{Ask} = V_{t-1} + \gamma I_t + \phi, \qquad (5)$$

$$P_t^{Bid} = V_{t-1} - \gamma I_t - \phi \qquad (6)$$

When the current trade is neither buyer- nor seller- initiated, trades occur at the mid-point of the difference between the bid and ask prices.

Lagging (3) by one period and combining it with (4) gives:

$$P_t - P_{t-1} = \gamma I_t + \phi(x_t - x_{t-1}) + \varepsilon_t + \xi_t - \xi_{t-1} \qquad (7)$$

Equation (7) serves as the basis for estimation.

## 2.3. ESTIMATION OF REGIME PROBABILITIES

Estimation of Equation (7) is performed using a two-step procedure. First, the probability of being in the private information regime is estimated from the data. Second, (7) is estimated using pooled time series and cross sectional regressions. In this section, we explain how the regime probabilities are determined.

In general, consider a single variable, $y_t$. Assume that $y_t \sim N(\mu, \sigma^2)$. Now consider the possibility that $y_t \sim N(\mu_{s_t}, \sigma_{s_t}{}^2)$. That is, $y_t$ may be drawn from a different normal distribution depending on the unobserved state variable, $s_t$, which can take on one of two values, 1 or 2. The evolution of this state, or regime, is governed by the following Markov transition probabilities:

$$\begin{aligned} p(s_t = 1 \mid s_{t-1} = 1) &= p_{11} \\ p(s_t = 2 \mid s_{t-1} = 1) &= 1 - p_{11} \\ p(s_t = 1 \mid s_{t-1} = 2) &= 1 - p_{22} \\ p(s_t = 2 \mid s_{t-1} = 2) &= p_{22} \end{aligned} \qquad (8)$$

Engel and Hamilton (1990) use this specification to examine exchange rate behavior. While simple, it permits time series patterns with features observed in exchange rate data. For example, with large $\mu_1$ and small $p_{11}$, the time series for regime 1 exhibits brief, abrupt movements. With a small $\mu_2$ and large $p_{22}$, changes in regime 2 are small and the trend persistent (Engel and Hamilton (1990), p. 692).

Following Engel and Hamilton (1990), the probability law governing $\{y_t\}$ is given by:

$$\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, p_{11}, p_{22})', \qquad (9)$$

or the vector of population parameters for the process. With this information, we can determine the probability of either state conditional on the past realizations of the process up to and including the current realization. This is known as the filter inference about the regime in effect at time t:

$$p(s_t \mid y_t, ..., y_1; \theta). \qquad (10)$$

Note that, while this model resembles the mixture of (normal) distributions, it is different in that the realizations of $\{y_t\}$, which are governed by the Markov process in (8), are not independent. The probability that a particular draw comes from $s_1$ or $s_2$ depends on the realization of the process at other points in time.

The specification in (9) is sufficient to describe the conditional distributions of $y_t$ given $s_t$, and $s_t$ given $s_{t-1}$, as well as the unconditional distribution of the state of the first observation. Doing such yields:

$$p(s_1 = 1; \theta) = \frac{(1 - p22)}{(1 - p11) + (1 - p22)} \qquad (11)$$

Next, form the joint probability distribution for all sample data $y_T$ with $s_T$:

$$p(y_1,\ldots,y_T,s_1,\ldots s_T;\theta)=p(y_T\mid s_T;\theta)\cdot p(s_T\mid s_{T-1};\theta)$$
$$\cdot\, p(y_{T-1}\mid s_{T-1};\theta)\cdot p(s_{T-1}\mid s_{T-2};\theta)\cdot\ldots\cdot p(s_2\mid s_1;\theta)\cdot p(y_1\mid s_1;\theta)\cdot p(s_1;\theta) \quad (12)$$

Summing (12) over all possible values of $(s_1,\ldots s_T)$ gives:

$$p(y_1,\ldots,y_T;\theta)=\sum_{s_1=1}^{2}\cdots\sum_{s_T=1}^{2}p(y_1,\ldots,y_T,s_1,\ldots,s_T;\theta) \qquad (13)$$

Hamilton (1990, 1993) provides a technology for obtaining estimates of $\theta$ by maximizing (13) using the EM algorithm of Dempster, Laird and Rubin (1977). In general, the solution to (10) assumes knowledge of the population parameters, $\theta$. Substituting MLE $\hat{\theta}$ based on (13) provides estimates of the filter probabilities. Hamilton (1993) shows how to arrive at the filter inferences in (10) by a recursive process (Hamilton (1993), pp. 237-240).

We note, in passing, that the filter inferences in (10) are not generally useful to Hamilton (1989, 1990, 1993) or Engel and Hamilton (1990). They are more interested in isolating changes in regime given the full sample of observations. An ex post determination of which regime is the current regime at a particular point in the series is given by:

$$p(s_t\mid y_T,\ldots,y_1;\theta) \qquad (14)$$

These are known as the full sample or "smoothed" inferences.

Hamilton (1989) and Engel and Hamilton (1990) show how to generalize this methodology to vector processes. In our case of a bivariate process, the probability law, $\theta$, consists of 14 elements and becomes:

$$\theta = (\mathbf{u}_1, \mathbf{u}_2, \Omega_1, \Omega_2, p_{11}, p_{22})', \qquad (15)$$

where $(\mathbf{u}_1, \mathbf{u}_2)$ are each two elements, and $(\Omega_1, \Omega_2)$ are 2X2 matrices.

## 2.4. THE SAMPLE AND DATA

Data for the study consist of the 35 stocks of the Toronto Stock Exchange 35 Index (TSE 35) for the 21 trading days of March 1996 and the 22 trading days of May 1996. Prior to April 15, 1996, the minimum tick size on the TSE was $0.125. Subsequent to April 15, 1996, the minimum tick size was reduced to $0.05 for shares trading above $5.00. The selection of March and May is designed to avoid any short-term pre- and post-decimalization effects. Trade and quote data are taken from the TSE Equity History (EH) data base. First, each trading day is divided into 26, 15-minute intervals from the opening bell at 9:30 a.m. until market close at 4:00 p.m. Since the dependent variable in our model (Equation (7)) is price change, the first 15-minute interval for all except price is eliminated from consideration. The elimination of opening trades is consistent with evidence that such trades belong to a distribution different from trades occurring later in the day (Amihud and Mendelson (1987)).

Within each 15-minute interval, the numbers of buyer-initiated trades, seller-initiated trades and neither buyer- nor seller-initiated trades are counted. Buyer- (seller-) initiated trades are those trades executed at the most recent quoted ask (bid). We do not use a tick test since quote data are available. Quotes are aligned with transactions, following the Lee and Ready (1991) procedure. Thus, quotes must occur at least five seconds prior to a transaction in order for that quote to be linked to that transaction.

Since the TSE tape provides quotes and transactions in six-second intervals, we require simply that quotes and trades not be contemporaneous.

Trades occurring within the bid-ask spread are counted in the "neither" category, as are all crosses (trades in which both the buying and selling broker are the same).[2] To estimate the filter probabilities of a private information regime, we add together trades that are buyer- or seller-initiated since those are the trades most likely to come from informed traders. Trades that are neither buyer- nor seller-initiated make up the source of non-information-based trading data. Thus, in the bivariate estimation model described in the preceding section, $\mathbf{u}_1$ is the mean vector for the private information regime where the first element is the mean number of information-based trades, and the second is the mean number of non-information-based trades. Similarly $\mathbf{u}_2$ is the mean vector for the no-private-information regime.[3]

Price change is based on the volume-weighted average price during each 15-minute interval. Our pooled regressions use 25 trading periods for each trading day.

The trade initiation variable, like that used by MRR, is based on the difference between the number of buyer-initiated trades and seller-initiated trades. However, unlike MRR, we focus on trade initiation over each 15-minute interval. Specifically, if buyer-

[2] There is anecdotal evidence that crosses are not always reported to the Exchange in a timely manner. Thus the impact of counting such trades on the time series properties of the data cannot be determined. However, internal trades account for more than 50% of transaction volume. Our methodology of using 15-minute intervals may mitigate such timing problems.

(seller-) initiated trades are greater, $x_t$ is +1 (-1). If they are equal, $x_t$ is zero. The choice of a 15-minute interval is motivated by a desire to capture a reasonably short period over which market participants might retain a belief about the existence of a particular regime. This needs to be balanced against the technical problems associated with an insufficient number of trades over a shorter interval.[4]

Although we begin the estimation process with 35 stocks for each trading day, the full sample is never available. This is because the filter probabilities are sometimes unavailable for a given stock due to the lack of model convergence or as a result of the inability to converge at a global maximum. Over the 43 trading days of March and May 1996, as few as 14 firms and as many as 34 firms could be used on any given day. The average number of firms over the sample period per day is 29. The robustness of the results given these deletions is discussed below.

## 2.5. EMPIRICAL FINDINGS

### 2.5.1. Some Descriptive Statistics

Table 2.1[5] provides descriptive statistics on the firms used in this study. Average per firm daily trade volume for March 1996 is about 438,000 shares, while for May it is about 448,000 shares. Medians are somewhat lower suggesting that the sample is biased

[3] It would have been preferable to present a trivariate model based on buyer-initiated, seller-initiated and neither buyer- nor seller-initiated trades, and three regimes (positive information, negative information, and no information). That model, with 27 parameters, was constructed but failed to converge given the data.
[4] Model estimation was attempted with shorter time periods (as short as five minutes) and longer time periods (as long as thirty minutes). The use of shorter time periods results in too few trades of the three types to permit convergence of the EM algorithm. The use of longer time periods reduces the length of the time series. This makes estimation of (7) problematic when pooling cross section (firm) and time series (periods) data.
[5] Please note that all tables and figures appear at the end of chapters.

towards higher volume firms. Average share price for both months is in the range of $30 for any given transaction. The average per transaction spread for March 1996 is $0.147 reflecting the $0.125 tick size. For May 1996 (that is, following the April 15, 1996 tick size reduction to $0.05), the average spread narrows to $0.083.

Table 2.2 provides average data on variables used in the study on a firm-by-firm basis. Note that the probability measure is almost always less than 0.50. On average, the probability of informed trading is less than 50%, and this is true for both the pre- and post-decimalization months. Easley, Kiefer and O'Hara (1997b) reject the hypothesis that there is more than one information event during a given trading day. A lower than 50% average probability of informed traders in the market is consistent with this result.[6] We also note from Table 2.2 that the average probability of informed trade is higher post- than pre-decimalization. A higher probability of informed trade is consistent with the theoretical predictions of Anshuman and Kalay (1993) and Harris (1994) that information asymmetry should increase following a reduction in tick size.

The average number of buyer- and seller- initiated trades in both the pre- and post-decimalization periods is approximately the same at about three. On average, more trades are considered to be information motivated than not (given that "Neither" is less than two on average).

---

[6] Recall, however, that our model does not hypothesize that privately informed trading makes prices change, rather prices change with the strength of the *belief* that privately informed traders are in the market.

Figure 2.1 shows the change in the probability measure on a period-by-period basis. It is comparable to the MRR periodic estimates of $\rho$, the autocorrelation of order flow. While that value is not generally statistically significant in MRR, its intra-day variation exhibits the same U-shaped pattern indicated in Figure 2.1. This suggests that our modeling of the regime transition probabilities as a Markov process inherently models the (casually observed) autocorrelation of trade flow.

### 2.5.2. Regime Probability

In this section we discuss the construction of the probability measure. This is the probability that the current regime is the one with information. It is computed based on the relationship between information-based trades and non-information-based trades.

The filter inferences used as our probability measure are a result of finding that value of the parameter vector, $\theta$, that maximizes (13). The procedure uses a GAUSS program developed by Hamilton and modified for our purposes. As in applications of the EM algorithm employed by Hamilton (1989, 1990, 1993) and Engel and Hamilton (1990), the use of diffuse priors often leads to a lack of convergence. As explained in Hamilton (1991), a global maximum may fail to exist when, for one of the distributions, the mean of the regime is equal to one of the observations and has zero variance. Such an occurrence is likely in the first period used in our estimates, given the nature of our data. As a result, we use the quasi-Bayesian estimates suggested by Hamilton (1991). While this procedure improves the ability of obtaining model convergence, it does not guarantee that the resulting parameter estimates result from a global maximum. In fact, it is likely that in some cases we have only managed to achieve local maxima.[7]

Figure 2.1 depicts the relationship between the probability measure and different types of trade. These graphs are of the unconditional firm means computed intra-day per period. The graphs for all of the variables are very similar, exhibiting the same U-shaped pattern throughout the trading day. In addition, there is little difference between the pre- and post decimalization results. The correlation between the sum of buyer- and seller-initiated trades versus neither buyer- nor seller-initiated trades is very high at 0.95 (0.91 post-decimalization). However, the regime probability measure is determined on a firm-by-firm basis and for any given firm the correlation may be lower.[8] Ideally, of course, they should be uncorrelated. The fact that they are correlated suggests that our computed regime probability measure is less than perfect. Part of the problem must rest with the definitions of what constitutes informed and uninformed trading. Without knowing the specific intentions of traders and using only trade and quote data, we are unlikely to develop a highly precise discriminating measure.

Statistical tests are conducted to determine if the estimates of mean vectors for each regime are different. At the 5% level, we fail to reject the hypothesis of no difference between regimes for 18.4% of the time in March (24.3% in May) for information-based trades, and 28.6% in March (25.6% in May) for non-information-

[7] Given the large number of estimates (35 firms X 43 days = 1,505), we did not verify the stability of parameter estimates (or filter inferences) for all cases. However, tests for stability show that the estimates are robust to modest changes in the initial parameters.
[8] The lowest correlation for firms in the study, either pre- or post-decimalization, is 0.60.

based trades. Thus, in spite of a high degree of correlation between informed and non-informed trade, the model is able to discriminate between them.

On average, the probability of informed trade declines until mid-day and then rises until market close. Much of the increase occurs in the two 15-minute periods prior to close. From the open to thirty minutes prior to market close, there is a reduction in the probability of informed trade. This result is consistent with the measure of information asymmetry in MRR as well as with the theoretical predictions of Easley and O'Hara (1992).

### 2.5.3. Parameter Estimates

The specification in (7) is estimated using pooled cross-section and time series data. Cross sections consist of between 14 and 34 firms, and the time series is the 25 trading periods of a given day beginning 15 minutes after the opening bell. It is reasonable to expect a degree of cross-sectional dependence since the TSE 35 consists of the largest firms in the Canadian market in terms of market capitalization. These firms tend to respond somewhat similarly to market-wide shocks. In addition, given the specification in (7), we expect a degree of autocorrelation in the time series. As a result, our cross section-time series regressions assume cross-sectionally correlated and autoregressive errors. The use of the cross sectional dependence procedure is not without cost. Specifically, the variance/covariance matrix is singular when the number of cross sections (potentially 35 firms) is equal to the number of time series (25 periods). As a result, the following procedure is adopted: for each day, the procedure is run with the first twenty firms, the last twenty firms, the middle twenty firms, and finally the first and last ten firms combined. This procedure is followed for all trading days except for the single day, during the pre-decimalization period, with data for only 14 firms.

Tables 2.3 and 2.4 provide estimates of the parameters of the model for information asymmetry and transaction cost for the pre-decimalization (March) and post-decimalization (May) months, respectively. In the left panel of Table 2.3, the coefficient of the probability of an information regime is significant and positive in sign for 8 of the 21 trading days in March 1996. In these cases, we reject the null of a zero coefficient at the 1% level. In three cases, the sign is negative and significant at the 1% level. In two other cases, the sign is negative and the coefficient is significant at the 10% level. A negative (positive) coefficient of the regime probability implies that the balance of information that day was negative (positive). This interpretation is straightforward given the uniformly positive values of the transaction cost estimates (right panels of Tables 2.3 and 2.4). The value of the probability coefficient ranges from +0.0156 to –0.0148. Thus, assuming that private information is in the market (PROB=1.0), this belief translates into a maximum share price impact of $0.0152 during a given 15 minute period. The interpretation of this coefficient is not identical to the "information asymmetry" parameter in MRR, although the (absolute) values are similar. In our model, information asymmetry is accounted for over time by the variation in the probability of information in the market on a period-by-period basis. MRR examine asymmetry for individual periods. In a sense, the coefficient in this study represents the direct impact on share price of the belief that there is information in the market.

In the right panels of Tables 2.3 and 2.4, we present parameter estimates for the combined cost of maintaining inventory and providing liquidity (the coefficient of the change in trade initiation). For each day, the period cost is uniformly significant (at the 1% level) and positive. It ranges between 0.0186 and 0.0302 with a mean of 0.0239. These values are a little more than half those of MRR but are generally consistent. For a major period-to-period shift from (for example) seller-initiated trades to buyer-initiated trades (i.e. $x_t - x_{t-1} = 2$), the mean cost translates into a \$0.0478 impact on the periodic change in share price.

Table 2.4 provides the same information as Table 2.3 for the post-decimalization month, May 1996. In general, the statistical results are superior to those obtained for the pre-decimalization month. Specifically, of the 22 trading days in May 1996, 14 of the information asymmetry coefficients are significant at the 1% level, and one is significant at the 10% level. The mean of 0.0059 over all days is statistically significant at the 1% level. We might interpret this difference between Tables 2.3 and 2.4 as suggesting that there is a greater probability of positive information post-decimalization. However, a t-test for differences between means is not significant. The information asymmetry coefficient estimates range from a high of 0.0691 to a low of –0.0037. This suggests a maximum price impact of \$0.0364 per 15-minute period. Transaction costs post-decimalization are lower than in the pre-decimalization period. For each day, period cost is uniformly significant (at the 1% level) and positive. Period cost ranges between 0.0024 and 0.0170 with a mean of 0.0116. These values are less than half those of MRR. For a major period-to-period shift from (for example) seller-initiated trades to buyer-initiated trades (i.e. $x_t - x_{t-1} = 2$), the mean cost translates into a $0.0232 impact on the periodic change in share price.

### 2.5.4. Robustness of the Results

As was mentioned in a previous section, while the regime estimation procedure is used on all 35 firms in the TSE 35, lack of model convergence results in fewer than 35 firms in the daily regressions. It is reasonable to inquire as to the impact such deletions may have on the overall results.

In some cases filter inferences are not produced at all when the EM algorithm is applied. We have little to say as to what impact these types of deletions may have on the results. Fortunately, for any trading day, this type of problem occurs in fewer than four of 35 firms on average during March and six on average during May. More common is the failure of the algorithm to produce global maxima. In a number of cases, at least one gradient is clearly non-zero, suggesting a corner solution. Ironically, for some of these cases, inclusion in the regressions actually improves the overall results. This occurs because the source of the problem, lack of trading (which in some cases leads to a low probability of an information regime) correlates well with the lack of change in the share price. While one might argue that such a relationship is anticipated, it is not consistent for all firms with less frequent trading. On balance, if we include all firms for which filter inferences are obtained, the overall conclusions remain essentially unaffected, although the statistical results deteriorate.[9]

## 2.6. CONCLUSION

We have presented a bivariate model of information regimes in which market participants use trade data in order to answer the following question: Is there currently private information in the market for a given stock? If there is a high probability of private information in the market, this leads contemporaneously to a greater revision in beliefs about the true price. Together with compensation for transactions costs, this leads to a greater revision in transaction price. Using TSE 35 trade and quote data for March 1996 (prior to the reduction in minimum tick size on the TSE) and for May 1996 (subsequent to a tick size reduction), our pooled cross-section and time series results support this view.

The time series changes in the probability of information in the market averaged over firms in our sample is interesting in its own right. The probability of information in the market is lower thirty minutes prior to market close than at market open. This lends support to models incorporating the temporal resolution of uncertainty into the trading process. However, including the last two 15-minute periods results in a clear U-shaped pattern.[10] This may account for the often reported U-shaped pattern in bid-ask spreads.

[9] Alternate specifications of the data were used for months in addition to March and May 1996. Specifically, all months from January through June 1996 were run but without the requirement that quotes not be contemporaneous with the transactions. In spite of this, statistical inferences are largely unchanged.
[10] Further study is required in order to determine why the probability of asymmetric information increases towards the end of the day. One possibility is that investors expect news releases immediately following market close. A second reason could be that informed traders, given greater market activity at the end of the day, could more easily camouflage their information advantage.

We have examined the time series characteristics of our measure of information for both pre-decimalization and post-decimalization months. We measure the price impact of information asymmetry using pooled time series and cross-sectional regressions. Based on the pooling, we find some evidence of an increase in the cost of information asymmetry. However, the time series patterns of the probability of informed traders pre- and post-decimalization (Figure 2.1), clearly indicate an increase in the probability of informed trade during the post-decimalization month (May). This leads to the question that, if the probability of informed trade is not priced in the bid-ask spread, what protective mechanism is employed by the designated market marker? One possibility is that the market maker reduces quote depth in order to attenuate the impact of information asymmetry.[11]

This study may have a number of limitations. First, our regime switching algorithm did not allow for convergence for all firms of the TSE 35 for all days. Therefore, we are left with a possible survivorship bias in the results, caused by the use of only those firms for which results are obtained, in spite of the fact that no firms are consistently eliminated from the sample on a day-to-day basis. The bias may be more subtle. For example, perhaps convergence is dependent on the volatility of the bivariate process such that highly volatile trading periods may lower the probability of convergence. In such a case we would be eliminating potentially important time series features from the results. Second, the variables making up the bivariate process on which our regime probabilities depend are correlated. We might argue that the impact of this correlation can be determined empirically. If the correlation is problematic, the results will be poor. However, tests for the equality of the estimates of the regime mean vectors suggest that this correlation does not hinder model performance.

A number of extensions can be made to our model. First, we have provided a very parsimonious model of the time series pattern for the means of our bivariate process. Although the Markov transition probabilities may sufficiently account for an underlying correlation in *regimes*, we have not accounted for the possible autocorrelation in the underlying variables. Conceptually, only the "informed" component of the process should be correlated, based on the empirical fact that informed traders tend to camouflage their information advantage by breaking up (desired) large order into smaller ones (Chan and Lakonishock (1993, 1995)). The "uninformed" component should be uncorrelated. However, evidence provided by Easley, Kiefer and O'Hara (1997b) finds that uninformed trade is correlated as well. The second extension that might be undertaken is to use the regime probabilities as a basis for understanding reported regularities in market data. For example, a regime model may help explain the Monday effect, reported by Lakonishok and Maberly (1990). Finally, following the lead of Easley and O'Hara (1992) and Easley, Kiefer and O'Hara (1997a,b), the data set could be extended to include non-trading intervals in determining the probability of information regimes.

[11] Such evidence is provided in Chapter 4, and elsewhere in the literature.

**Table 2.1** Descriptive Statistics

| | Volume | | Share Price | | Spread | |
|---|---|---|---|---|---|---|
| | March | May | March | May | March | May |
| Mean | 437,892 | 448,500 | $30.22 | $31.36 | $ 0.147 | $ 0.083 |
| Median | 373,787 | 448,534 | $28.11 | $29.86 | $ 0.140 | $ 0.069 |
| High | 1,069,842 | 949,229 | $66.54 | $71.67 | $ 0.233 | $ 0.214 |
| Low | 75,908 | 78,278 | $11.41 | $11.73 | $ 0.115 | $ 0.048 |

The above table presents statistics for the TSE 35 for the 21 trading days of March and 22 trading days of May 1996. Volume is the number of shares traded per firm per day. Share Price is the average per firm share price per transaction, which is not volume weighted. Spread is the average spread per transaction.

## Table 2.2 Statistical Characteristics of Key Variables

| | PROBABILITY | | BUYER | | SELLER | | NEITHER | |
|---|---|---|---|---|---|---|---|---|
| Firm | March | May | March | May | March | May | March | May |
| 1 | 0.44 | 0.41 | 2.7 | 1.7 | 2.2 | 1.3 | 1.4 | 1.0 |
| 2 | 0.42 | 0.37 | 4.9 | 2.4 | 4.5 | 2.7 | 2.8 | 2.0 |
| 3 | 0.38 | 0.40 | 3.5 | 2.5 | 3.6 | 2.4 | 2.4 | 1.7 |
| 4 | 0.35 | 0.30 | 3.4 | 3.1 | 1.9 | 2.0 | 2.0 | 1.6 |
| 5 | 0.42 | 0.46 | 7.1 | 5.1 | 6.5 | 16.3 | 4.3 | 5.3 |
| 6 | 0.28 | 0.33 | 7.0 | 6.6 | 4.0 | 4.3 | 3.5 | 2.7 |
| 7 | 0.32 | 0.37 | 3.7 | 4.8 | 3.2 | 4.6 | 1.9 | 2.2 |
| 8 | 0.41 | 0.45 | 4.9 | 5.3 | 3.3 | 4.0 | 2.0 | 2.6 |
| 9 | 0.39 | 0.37 | 2.8 | 2.2 | 4.8 | 3.4 | 2.6 | 1.9 |
| 10 | 0.36 | 0.48 | 1.0 | 1.1 | 1.9 | 1.5 | 0.6 | 0.6 |
| 11 | 0.52 | 0.43 | 1.0 | 1.2 | 0.9 | 0.7 | 0.6 | 0.6 |
| 12 | 0.39 | 0.47 | 1.7 | 0.8 | 2.1 | 1.7 | 1.2 | 1.0 |
| 13 | 0.52 | 0.60 | 3.2 | 2.4 | 1.5 | 1.5 | 1.8 | 1.8 |
| 14 | 0.44 | 0.48 | 1.9 | 1.8 | 2.0 | 1.5 | 0.9 | 1.0 |
| 15 | 0.46 | 0.51 | 2.2 | 1.2 | 2.9 | 1.9 | 1.4 | 1.2 |
| 16 | 0.32 | 0.33 | 2.1 | 1.3 | 4.4 | 4.8 | 1.7 | 1.1 |
| 17 | 0.46 | 0.50 | 2.9 | 3.6 | 1.6 | 2.1 | 1.1 | 1.6 |
| 18 | 0.51 | 0.59 | 1.0 | 1.0 | 0.8 | 0.6 | 0.6 | 0.4 |
| 19 | 0.40 | 0.42 | 2.0 | 1.8 | 2.1 | 1.9 | 1.8 | 1.7 |
| 20 | 0.31 | 0.48 | 1.3 | 1.3 | 1.6 | 1.6 | 1.1 | 1.0 |
| 21 | 0.38 | 0.47 | 4.0 | 1.9 | 2.7 | 2.7 | 1.7 | 1.2 |
| 22 | 0.45 | 0.59 | 1.7 | 1.4 | 2.3 | 2.0 | 1.5 | 1.8 |
| 23 | 0.31 | 0.34 | 4.7 | 3.1 | 4.2 | 4.7 | 3.5 | 2.4 |
| 24 | 0.39 | 0.35 | 4.3 | 2.0 | 3.6 | 2.8 | 2.5 | 1.9 |
| 25 | 0.38 | 0.41 | 1.3 | 1.8 | 1.2 | 1.7 | 0.9 | 1.1 |
| 26 | 0.43 | 0.54 | 1.3 | 1.2 | 1.7 | 0.9 | 0.7 | 0.7 |
| 27 | 0.34 | 0.36 | 4.6 | 4.6 | 4.0 | 4.0 | 2.5 | 2.1 |
| 28 | 0.39 | 0.45 | 3.5 | 5.5 | 2.9 | 4.5 | 1.9 | 2.7 |
| 29 | 0.37 | 0.36 | 5.2 | 5.4 | 4.1 | 4.8 | 2.9 | 2.6 |
| 30 | 0.54 | 0.51 | 0.9 | 0.7 | 1.0 | 0.9 | 0.6 | 0.6 |
| 31 | 0.54 | 0.52 | 1.3 | 1.2 | 1.9 | 2.0 | 1.0 | 0.8 |
| 32 | 0.43 | 0.46 | 1.9 | 1.7 | 1.2 | 1.6 | 1.2 | 1.2 |
| 33 | 0.39 | 0.47 | 5.6 | 7.2 | 2.7 | 2.9 | 3.6 | 3.2 |
| 34 | 0.42 | 0.42 | 5.2 | 4.6 | 3.4 | 2.9 | 2.3 | 2.2 |
| 35 | 0.50 | 0.50 | 3.3 | 2.1 | 1.9 | 1.5 | 2.0 | 1.7 |
| | | | | | | | | |
| Mean | 0.41 | 0.44*** | 3.12 | 2.75** | 2.70 | 2.88 | 1.84 | 1.69* |
| Median | 0.40 | 0.45 | 2.85 | 2.04 | 2.27 | 2.00 | 1.80 | 1.66 |
| High | 0.54 | 0.60 | 7.14 | 7.18 | 6.54 | 16.28 | 4.31 | 5.32 |
| Low | 0.28 | 0.30 | 0.92 | 0.69 | 0.84 | 0.58 | 0.57 | 0.40 |

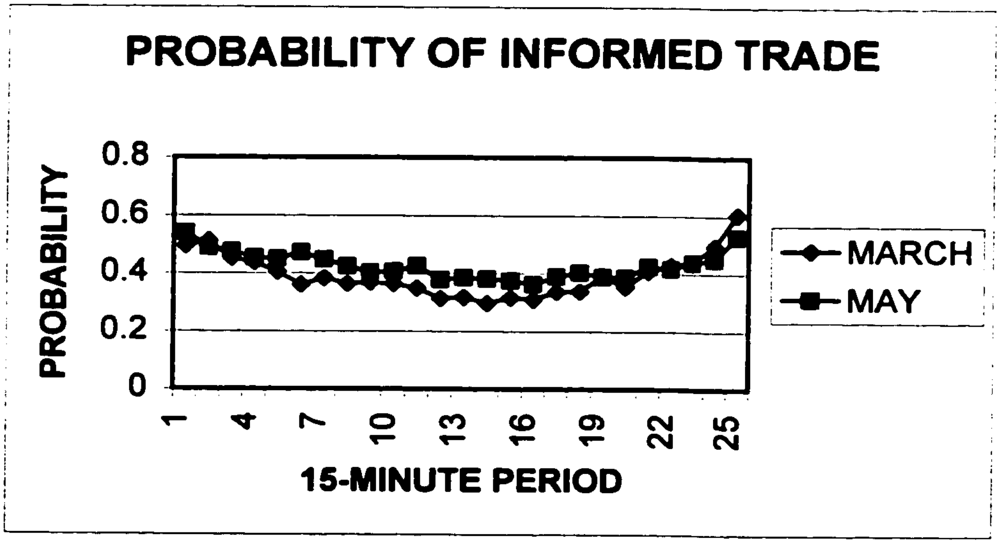
All firm statistics are calculated over 25 trading periods for the firms retained for the pooled cross section-time series regressions. In March, data for each firm was available for 17.2 of the 21 days on average, and in May, 19 of the 22 days on average. The means at the bottom of the table are the overall means. Thus, in the second column (FIRM 1, PROBABILITY, MARCH), the value 0.44 is the average probability of private information in the market for Firm 1 during March 1996. The mean for that column, 0.41, is calculated over the 35 firms. BUYER is the number of buyer-initiated trades per 15-minute period on average for a given firm, SELLER is the number of seller-initiated trades and NEITHER is the number of neither buyer- nor seller-initiated trades. The means for March are interpreted as follows. The average probability of the "information" regime during any 15-minute period over the 35 firms is 41%. There are 3.12 buyer-initiated trades on average per 15-minute period, 2.70 seller-initiated trades, the 1.84 trades that are neither buyer- nor seller-initiated. "***", "**" and "*" indicate that means are significantly different at the 1%, 5% and 10% levels, respectively. PROBABILITY is estimated using a modified version of the Hamilton GAUSS algorithm. Buyer-(seller-) initiated trades are those trades executed at the most recent quoted ask (bid). "Neither" trades are those trades not at the bid nor ask, or are trades internal to the broker.

**Figure 2.1** Per Period Characteristics of Key Variables

The following graphs provide visual representations of key variables. These variables are the probability that the current regime is the "information" regime, the number of buyer-initiated trades, the number of seller initiated trades, and trades that are neither buyer- nor seller-initiated. All graphs are on a per period basis aggregated over all firms used in the pooled regressions. The horizontal axis is "15-Minute Trading Period within the Day". Pre-decimalization (March 1996) and post-decimalization (May 1996) are shown on each graph.

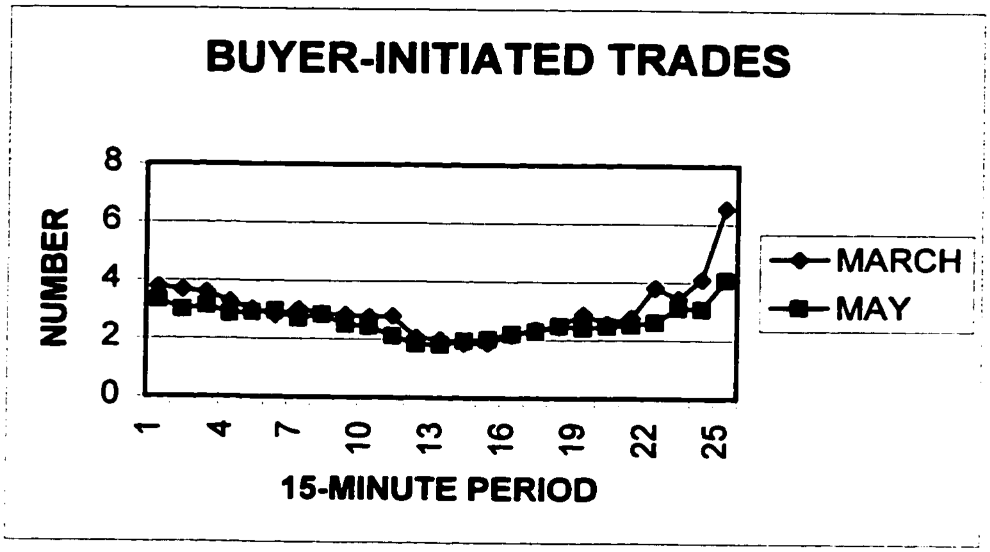
**Graph 2.1** Probability that the current regime is the "information" regime

This graph shows the average probability that the current regime is the information regime on a period-by-period basis over the day. Data are averaged for all firms used in the pooled time series and cross section regressions. Probability is estimated using a modified version of the Hamilton GAUSS algorithm based on the number of buyer- and seller-initiated trades versus trades that are neither buyer- nor seller-initiated. The March and May mean probabilities are significantly different at the 1% level.



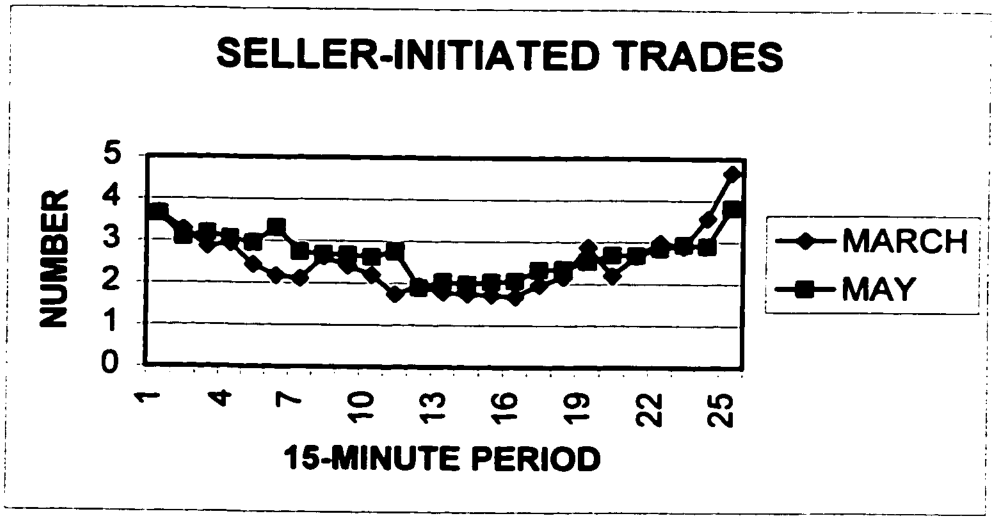
**Graph 2.2.** Number of buyer-initiated trades averaged over all firms and days period-by-period

A trade is categorized as buyer-initiated if it occurs at the most recent quoted ask price.




**Graph 2.3.** Number of seller-initiated trades averaged over all firms and days period-by-period

A trade is categorized as seller-initiated if it occurs at the most recent quoted bid price.



**Graph 2.4.** Average number of trades per period that are neither buyer- nor seller-initiated

A trade is categorized as neither buyer- nor seller initiated if it does not occur at either the most recent quoted ask price or bid price, or if the transaction is internal to the broker.

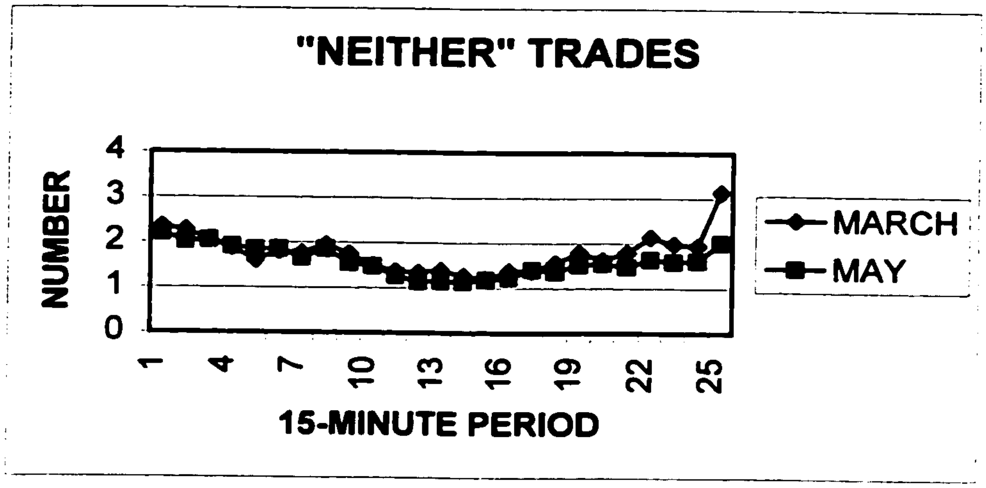


**Table 2.3.** Regression Results for "Information" Regime Probabilities and Transaction Costs for the Pre-decimalization Period, March 1996

$\gamma$ (coefficient of information asymmetry)

| Day | $\gamma$ | Std Error | Sign. | Firms |
|---|---|---|---|---|
| | | | | |
| 1 | 0.0031 | 0.0029 | | 30 |
| 2 | 0.0069 | 0.0023 | *** | 14 |
| 3 | -0.0016 | 0.0026 | | 25 |
| 4 | -0.0094 | 0.0028 | *** | 28 |
| 5 | 0.0063 | 0.0021 | *** | 30 |
| 6 | -0.0148 | 0.0032 | *** | 34 |
| 7 | 0.0156 | 0.0020 | *** | 32 |
| 8 | -0.0039 | 0.0022 | * | 29 |
| 9 | 0.0131 | 0.0024 | *** | 30 |
| 10 | 0.0081 | 0.0021 | *** | 32 |
| 11 | -0.0056 | 0.0029 | * | 29 |
| 12 | 0.0083 | 0.0024 | *** | 31 |
| 13 | 0.0018 | 0.0023 | | 26 |
| 14 | 0.0031 | 0.0025 | | 31 |
| 15 | 0.0034 | 0.0030 | | 31 |
| 16 | 0.0004 | 0.0023 | | 30 |
| 17 | 0.0005 | 0.0026 | | 31 |
| 18 | 0.0005 | 0.0025 | | 30 |
| 19 | 0.0051 | 0.0021 | *** | 31 |
| 20 | 0.0063 | 0.0024 | *** | 31 |
| 21 | -0.0106 | 0.0024 | *** | 30 |
| | | | | |
| Mean | 0.0017 | 0.0025 | | 29 |
| Median | 0.0031 | 0.0024 | | 30 |
| High | 0.0156 | 0.0032 | | 34 |
| Low | -0.0148 | 0.0020 | | 14 |

$\phi$ (coefficient of transaction cost)

| Day | $\phi$ | Std Error | Sign. | Firms |
|---|---|---|---|---|
| | | | | |
| 1 | 0.0208 | 0.0013 | *** | 30 |
| 2 | 0.0212 | 0.0011 | *** | 14 |
| 3 | 0.0302 | 0.0012 | *** | 25 |
| 4 | 0.0270 | 0.0010 | *** | 28 |
| 5 | 0.0206 | 0.0010 | *** | 30 |
| 6 | 0.0220 | 0.0011 | *** | 34 |
| 7 | 0.0222 | 0.0010 | *** | 32 |
| 8 | 0.0250 | 0.0008 | *** | 29 |
| 9 | 0.0226 | 0.0009 | *** | 30 |
| 10 | 0.0265 | 0.0009 | *** | 32 |
| 11 | 0.0262 | 0.0008 | *** | 29 |
| 12 | 0.0222 | 0.0011 | *** | 31 |
| 13 | 0.0211 | 0.0010 | *** | 26 |
| 14 | 0.0265 | 0.0011 | *** | 31 |
| 15 | 0.0242 | 0.0013 | *** | 31 |
| 16 | 0.0228 | 0.0010 | *** | 30 |
| 17 | 0.0249 | 0.0012 | *** | 31 |
| 18 | 0.0253 | 0.0012 | *** | 30 |
| 19 | 0.0227 | 0.0010 | *** | 31 |
| 20 | 0.0290 | 0.0012 | *** | 31 |
| 21 | 0.0186 | 0.0010 | *** | 30 |
| | | | | |
| Mean | 0.0239 | 0.0010 | *** | 29 |
| Median | 0.0228 | 0.0010 | | 30 |
| High | 0.0302 | 0.0013 | | 34 |
| Low | 0.0186 | 0.0008 | | 14 |

Table 2.3 reports coefficient estimates, $\gamma$, of the probability of an information regime, $I_t$, and transactions cost, $\phi$, of ($x \; - x_{t-1}$), based on pooled cross section and time series regressions for the period prior to a reduction in tick size on the TSE. The pooling method used assumes cross sectional dependence and autocorrelated errors.

$P_t - P_{t-1} = \gamma I_t + \phi(x_t - x_{t-1}) + \varepsilon_t + \xi_t - \xi_{t-1}$ is the model, where $P_t - P_{t-1}$ is the change in price and the last three terms are the error component. For the column Sign. (next to the standard errors), asterisks indicate the level of significance. Three asterisks mean that the coefficient estimate is significant at the 1% level, two at the 5% level, and one at the 10% level. The last column of each panel shows the number of TSE 35 firms included in the pooled regressions. Firms do not appear in the sample when the regime algorithm fails to converge. For all days but one, the number of firms exceeds the number (25) of time periods. In those cases, the following procedure is used: For each day, the procedure is run with the first twenty firms, the last twenty firms, the middle twenty firms, and finally the first and last ten firms. Statistics reported reflect the average results of these four regressions.

**Table 2.4.** **Regression Results for "Information" Regime Probabilities and Transaction Costs for the Post-decimalization Period, May 1996**

$\gamma$ (coefficient of information asymmetry)

| Day | $\gamma$ | Std Error | Sign. | Firms |
|---|---|---|---|---|
| | | | | |
| 1 | 0.0085 | 0.0012 | *** | 28 |
| 2 | 0.0015 | 0.0016 | | 31 |
| 3 | 0.0011 | 0.0012 | | 30 |
| 4 | 0.0691 | 0.0033 | *** | 28 |
| 5 | 0.0037 | 0.0013 | *** | 31 |
| 6 | 0.0001 | 0.0014 | | 30 |
| 7 | -0.0016 | 0.0012 | | 30 |
| 8 | -0.0019 | 0.0015 | | 30 |
| 9 | 0.0048 | 0.0010 | *** | 32 |
| 10 | 0.0011 | 0.0014 | | 30 |
| 11 | 0.0038 | 0.0014 | *** | 34 |
| 12 | 0.0022 | 0.0012 | * | 32 |
| 13 | 0.0055 | 0.0016 | *** | 30 |
| 14 | 0.0023 | 0.0015 | | 27 |
| 15 | 0.0041 | 0.0015 | *** | 27 |
| 16 | 0.0067 | 0.0012 | *** | 31 |
| 17 | 0.0030 | 0.0011 | *** | 33 |
| 18 | 0.0053 | 0.0012 | *** | 33 |
| 19 | 0.0099 | 0.0016 | *** | 28 |
| 20 | 0.0024 | 0.0011 | *** | 32 |
| 21 | 0.0029 | 0.0011 | *** | 32 |
| 22 | -0.0037 | 0.0014 | *** | 32 |
| | | | | |
| Mean | 0.0059 | 0.0014 | *** | 31 |
| Median | 0.0030 | 0.0013 | | 31 |
| High | 0.0691 | 0.0033 | | 34 |
| Low | -0.0037 | 0.0010 | | 27 |

$\phi$ (coefficient of transaction cost)

| Day | $\phi$ | Std Error | Sign. | Firms |
|---|---|---|---|---|
| | | | | |
| 1 | 0.0106 | 0.0005 | *** | 28 |
| 2 | 0.0089 | 0.0006 | *** | 31 |
| 3 | 0.0092 | 0.0005 | *** | 30 |
| 4 | 0.0024 | 0.0009 | *** | 28 |
| 5 | 0.0057 | 0.0006 | *** | 31 |
| 6 | 0.0085 | 0.0006 | *** | 30 |
| 7 | 0.0140 | 0.0006 | *** | 30 |
| 8 | 0.0099 | 0.0006 | *** | 30 |
| 9 | 0.0133 | 0.0004 | *** | 32 |
| 10 | 0.0121 | 0.0006 | *** | 30 |
| 11 | 0.0120 | 0.0005 | *** | 34 |
| 12 | 0.0170 | 0.0006 | *** | 32 |
| 13 | 0.0139 | 0.0006 | *** | 30 |
| 14 | 0.0130 | 0.0006 | *** | 27 |
| 15 | 0.0159 | 0.0006 | *** | 27 |
| 16 | 0.0158 | 0.0005 | *** | 31 |
| 17 | 0.0134 | 0.0005 | *** | 33 |
| 18 | 0.0155 | 0.0005 | *** | 33 |
| 19 | 0.0092 | 0.0006 | *** | 28 |
| 20 | 0.0134 | 0.0005 | *** | 32 |
| 21 | 0.0105 | 0.0005 | *** | 32 |
| 22 | 0.0115 | 0.0005 | *** | 32 |
| | | | | |
| Mean | 0.0116 | 0.0006 | *** | 31 |
| Median | 0.0120 | 0.0006 | | 31 |
| High | 0.0170 | 0.0009 | | 34 |
| Low | 0.0024 | 0.0004 | | 27 |

Table 2.4 reports coefficient estimates, $\gamma$, of the probability of an information regime, $I_t$, and transactions cost, $\phi$, of $(x - x_{t-1})$, based on pooled cross section and time series regressions for the period subsequent to a reduction in tick size on the TSE. The pooling method used assumes cross sectional dependence and autocorrelated errors. $P_t - P_{t-1} = \gamma I_t + \phi(x_t - x_{t-1}) + \varepsilon_t + \xi_t - \xi_{t-1}$ is the model, where $P_t - P_{t-1}$ is the change in price and the last three terms are the error component. For the column Sign. (next to the standard errors), asterisks indicate the level of significance. Three asterisks mean that the coefficient estimate is significant at the 1% level, two at the 5% level, and one at the 10% level. The last column of each panel shows the number of TSE 35 firms included in the pooled regressions. Firms do not appear in the sample when the regime algorithm fails to converge. For all, the number of firms exceeds the number (25) of time periods. As a result, the following procedure is used: for each day, the procedure was run with the first twenty firms, the last twenty firms, the middle twenty firms, and finally the first and last ten firms. Statistics reported reflect the average results of these four regressions.

## CHAPTER 3

## COST COMPONENTS OF THE BID-ASK SPREAD: TRADE SIZE, TRADER AND INTERNALIZATION

### 3.1. INTRODUCTION

In perfect capital markets, the bid-ask spread is zero and all market participants are price takers. Differences in trade size, identity of traders or execution method have no impact on price. In real markets, however, the bid-ask spread compensates market makers for order processing costs, inventory costs and the risk of trading against the better-informed. In such markets, trade volume, trader identity and execution method may be important in understanding the factors affecting the size and component composition of the bid-ask spread.

Trade size has long been hypothesized to affect both the costs of maintaining and managing inventory as well as the imbedded cost of asymmetric information. Filled orders may require subsequent inventory rebalancing, and larger orders may signal the existence of private information. Thus, we would expect a priori that larger orders have an impact on price movement generally and, more specifically, on the components of the bid-ask spread. A number of studies examine the role of volume in models of the spread. A recent model by Huang and Stoll (1997) finds that greater volume results in a smaller information asymmetry component of the spread. Specifically, they examine the components of the bid-ask spread given the sequence of trades at the bid or ask and whether those trades are from small to large, small to medium, small to small and so on. They conclude that information asymmetry is greatest for medium trades not exhibiting trade reversal. Larger trades, they argue are negotiated upstairs. Thus the potential for a surprise is mitigated. Trades that reverse in direction (for example a trade at the bid followed by a trade at the ask) are expected as being caused by the normal bid-ask bounce. Therefore, a trade that is not pre-negotiated and follows another in the same direction may suggest the presence of private information.

In a model similar to that of Huang and Stoll (1997), Madhavan, Richardson and Romans (1997), while not testing for the impact of volume, suggest that volume may not have an important effect on the large, frequently traded NYSE stocks they examine. They suggest, however, that for less frequently traded stocks, volume effects may be large.

As noted by Bacidore (1997) and Harris (1994), market makers have implicit schedules consisting of a set of prices for different trading quantities, despite the fact that they may only quote two price/quantity pairs, one for the bid and one for the ask. Since larger trades may reflect the presence of privately informed traders, the implicit schedule is such that spreads are an increasing function of trade volume. Easley and O'Hara (1987) explicitly model the relationship between trade volume and spread under this assumption. Glosten and Harris (1988) provide empirical support for that view.

Trader identity is imbedded in models of the components of the bid-ask spread at least since Glosten (1987). His decomposition of bid-ask components, and those which follow such as Glosten and Harris (1988) and Stoll (1989), relies on the existence of two fundamentally different types of market participant: informed traders and uninformed traders. Informed traders have private information and seek to use their information to advantage, potentially at the expense of uninformed market participants including, by assumption, designated market makers. Informed trading is important in the determination of asset prices. Market makers set bid and ask prices and adjust those prices in keeping with their belief about "true" value. The beliefs of market makers, as modeled by Easley and O'Hara (1992) for example, are updated as the market maker learns from the sequence of trades. Equilibrium is reached when prices have risen (or fallen) to reflect the information of informed traders.

Different types of market participants are likely to be better or worse informed than others and this information may be available to other market participants. For example, Battalio and Holden (1997) in modeling payment for order flow and internalization, make a distinction between externally verifiable characteristics and internally verifiable characteristics. They argue that brokers alone may be able to use information that they have about trading clients to identify those likely to be informed. Externally verifiable characteristics are widely known. Professional traders, for example, are more likely to be informed. Other work looks directly at the impact that changes in market features may have on different types of market participants. For example, Griffiths, Smith, Turnbull and White (1998) examine the impact on trading activity of upstairs traders and market makers resulting from a minimum tick size reduction on the TSE. They find that the reduction in tick size led to a migration of smaller orders away from the upstairs market and to the designated market maker, as well as to a reduction in profits earned upstairs.

Upstairs market activity actually contains two features of importance as far as trader type is concerned. First, the upstairs market is an internal market of a brokerage firm, distinct from the trading floor. Second, two types of trades may be undertaken. The first type of trade is principal trades, i.e. trades between the brokerage firm and clients from or to the broker's own inventory. The second type of trade is non-principal trades; i.e. trades between clients of the firms facilitated in-house. In addition, the upstairs market acts as a filter for trades eventually reaching the trading floor and the designated market maker. Upstairs traders have 15 minutes to decide to handle trades internally or to gate them to the exchange. Internalization has been cited by Huang and Stoll (1996) as one explanation for the persistent higher spreads on NASDAQ as compared to the NYSE. However, little formal modeling of internalization has been done (Battalio and Holden (1997) being an exception), and no linking of internalization and the cost components of the bid-ask spread has appeared in the literature.

The purpose of this chapter is to examine the cost components of the bid-ask spread in relation to those features of real markets believed to be significant; namely, trade size, trader type, and internalization. We do this for the 35 firms included in the TSE35 index first for the 73 trading days from January 2, 1996 through April 12, 1996 that just precedes the switch to decimalization and a lower tick size by the TSE. We identify trades as being large or small. Trades are classified as being one of three trader types -- between principals, trades involving the designated market maker (which we refer to as the registered trader (RT)), and trades between principals and traders who are neither principals nor RTs. Finally, trading may be either within the firm (internal) or outside of the firm (external). We measure the relative importance of different types of trader activity for both external and internal trades. We conduct three major tests of robustness. First, we account for cross-sectional dependencies across firms by using a market proxy much as is done using a traditional market model in the asset pricing literature. Second, we use a Scholes-Williams (1997) type of adjustment to capture any non-synchronous trading biases in the market proxy used to capture the cross-sectional dependencies across firms. Finally, we estimate the cost components for each of the TSE35 firms for a different time period; namely, the period immediately after the TSE switch to decimalization and a lower tick size on April 15, 1996.

This chapter makes three contributions to the existing literature. First, we find that trade size is the single most important determinant of the cost of adverse information or information asymmetry. Second, we find that whether trades are executed internally or externally is the single most important determinant of order processing cost, but is an important determinant of the cost of adverse information as well . This conclusion is consistent with payment for order flow. Third, we find that trader type (e.g. principal-principal) is a determinant of both the cost of adverse information and the cost of order processing. In addition, we find that our initial results are very robust for the three major tests of robustness.

The chapter is organized as follows: Section 3.2 describes the data set used and provides some preliminary analysis. Section 3.3 presents a model for the decomposition of the bid-ask spread first, as an example, on the basis of trade size, and then on the basis of the interrelationship of trade size, trader type and internalization. Section 3.4 presents and discusses the main empirical findings of the chapter. Section 3.5 concludes.

## 3.2. DESCRIPTION AND ANALYSIS OF THE DATA AND SAMPLE

The data first used in this study are for the 35 firms making up the TSE 35 index as of December 1995, for the period from January 2, 1996 through April 12, 1996 (a period of approximately 3.5 months). On April 15, the TSE reduced the minimum tick size from 1/8 to $0.05 for stocks trading above $5. In Section 3.4, we present results for the post-tick size reduction period from April 15, 1996 through June 30, 1996 (a period of about 2.5 months), as a test of robustness of our model estimates.

The TSE data, obtained from the TSE Equity History tape, include prices and quotes in six second intervals from market open at 9:30 am to market close at 4:00 pm.[1] Volume data include trade size and quote depth on both sides of the market. Buying and selling brokerage firms are identified by TSE number as well as by the type of buyer and seller. Specifically, buyers and sellers are one of three types: Registered trader (RT), principal, and neither RT nor principal.

[1] In fact, the data includes quotes well prior to market open and trades with time stamps past market close. Trades past the market close are rare – fewer than 1 per day on average.

The Lee and Ready (1991) procedure is applied to the data. Trades are matched against prevailing quotes with the quotes time-stamped at least 5 seconds prior to the trade.[2] For each firm, transactions are placed into two size categories: large and small. The large category consists of trades in the top size quartile of all trades for that firm. Thus, "large" trades are defined relative to all of a firm's transaction volumes over the period. This categorization method is consistent with the Lin et al. (1995) sorting procedure. In contrast to their method in which there are 7 percentile categories, we collapse their first three (lower three quartiles) and last 4 categories (upper quartile) into 2 categories, small (lower 3 quartiles) and large (upper quartile).

We provide discussion and analysis of both trades external to brokerage houses and internal to brokerage houses. External trades are defined as those trades in which the buying and selling firms are different. Buying and selling firms are the same for internal trades. In reality, the situation is likely to be more complex since a number of brokerage firms having seats on the TSE and identified by different numbers are, in fact, affiliated. Our definitions of internal and external may thus understate the extent to which internal trading exists on the TSE.

### 3.2.1. All Trades

We begin our discussion without concern as to whether trades are internal or external. Tables 3.1, 3.2a, 3.2b and 3.3 summarize the data in terms of type of trader

[2] The Lee and Ready (1991) procedure also characterizes trades as buyer-initiated or seller-initiated. The Lin et al. model, which we follow, does not use this information.

(Tables 3.1, 3.2a, 3.2b), transactions and share volumes (Table 3.3), and quote and price data (last 2 rows of Table 3.3).

In Table 3.1, we present transaction data sorted by type of trader and whether that trader was buying or selling. Thus, the first panel (BUYS) reveals 112,231 RT buy transactions out of a total of 558,534 buy transactions. We also present firm mean and median data by type of trader. The average number of transactions per firm is 15,958, and RTs are buyers for 3,207 of these transactions. Table 3.2a gives this data in percentage terms. In the first panel of Table 3.2a, (ALL), these RT buys represent 20.1% of all buy transactions, (and, since buyers sell, 20.1% of all transactions). Medians suggest that there are more small average-transaction firms in the sample since medians are below the means. This feature is present to the same degree in all trader categories and for both buys and sells.

From Table 3.2a, we see that, in general, RTs are buyers or sellers in approximately equal numbers of transactions; they are net buyers for only 0.3% of transactions. However, principals sell in more transactions than they buy, and traders in the neither category buy in more transactions than they sell. In fact, in the last 2 panels of Table 3.2a, we note that principals are net sellers to "the field" while the field is a net buyer from principals.[3]

The picture changes in significant ways when we look at transaction *volume* instead of the number of transactions. That data is presented in Table 3.2b. Whereas RTs account for 20% of all buy transactions, they account for only 8.8% of all trade volume. On the other hand, while principals are involved in 30% of buy transactions, they account for 40% of buy volume. A comparison of Tables 3.2a and 3.2b suggests that principal transactions are larger than those of either RTs or the field. We note that the data is virtually the same as that presented in Table 3.2b (volume) when we examine percentage shares of transaction *value* (not presented).

We conclude on the basis of the data, that since RTs are involved in about 9% of trading volume and about 20% of transactions that their service of providing liquidity and immediacy is focused more on small volume transactions. This conclusion is consistent with the results of Lin et al. (1995). RTs trade with both principals and the field in similar proportions. Principals account for 30% of transactions and 40% of volumes. They trade with other principals and with the field. However, given the role of the RT, principal trades with the field are apparently larger in volume than RT trades with the field.

### 3.2.2. Internal and External Trades

Table 3.3 provides data on the relative importance of internal and external trading. Internal transactions account for 17.4% of all transactions. However, Table 3.4 shows that 58.6% of total trading volume is internal.

Tables 3.5 through 3.8 examine internal and external trading activity percentages by traders for transactions (Tables 3.5 and 3.6), and share volumes (Tables 3.7 and 3.8).

---

[3] These imbalances are probably temporary and should move to netted positions over longer horizons.

Both views of these data provide similar inferences although the volume perspective is clearer. First, RTs are more active externally than internally. Second, Principal trades are more predominant externally than internally. Finally, trades by others in neither category are higher internally than externally. In addition, internal trading is more likely to be between principal and the neither category, and between-principal transactions are more likely to be external. RTs trade more often externally with principals than internally.

All of these conclusions from the data are consistent with models of payment for order flow (POF) and trade internalization, and specifically with that of Battalio and Holden (1997).[4] Under the assumption that principals are professional and sophisticated (i.e. informed) traders, orders from them are less likely to be candidates for POF since potential purchasers of such orders tend to reject trades from informed traders. The broker (another potential principal in a transaction) who could have sold that now-rejected flow must decide to route the order through to the exchange (to the RT) or to handle it internally. But the broker is faced with the same likelihood of dealing with an informed individual (in fact according to Battalio and Holden (1997) the broker may have internal knowledge of the trader confirming sophistication). Thus trades not candidates for POF are also poor candidates for internalization. Trades from the less-likely-to-be-informed neither category are candidates for such internalization. Therefore, when the informed (principals) trade together, it is likely to be external. Most internal trades are between traders in the neither category or between principal and neither.

## 3.3. ESTIMATING COST COMPONENTS OF THE BID-ASK SPREAD

### 3.3.1. The Model

In order to investigate the interrelationships discussed in the previous section, we begin with a quote revision model developed by Lin, Sanger and Booth (1995). This model is itself a development of models of Huang and Stoll (1994), Lin (1992) and Stoll (1989).

From Stoll (1989), the specialist's trade-by-trade expected gross profit at time t+1 conditional on a trade at time t is:

$$\delta(B_{t+1} - B_t) + (1-\delta)(A_{t+1} - B_t) = E_t(P_{t+1}) - P_t \qquad (1)$$

where:

$E_t(P_{t+1}) = \delta(B_{t+1}) + (1-\delta)(A_{t+1})$;

$\delta \equiv$ probability of a buy order following a buy or sell following a sell (order persistence);

$A, B \equiv$ are ask and bid prices, respectively;

$P \equiv$ transaction price.

Following Huang and Stoll (1994), we assume that the effective half spread is:

$z_t = | P_t - Q_t |$

and $Q_t = (A_t + B_t)/2$ is the quote midpoint.

Under the assumption that specialist quotes are adjusted upward (downward) following a buy (sell), the adverse information component of the spread is now modeled.

---

[4] Payment for order flow occurs when an originating broker sells orders to a market maker or another broker for execution. The buying party makes a side payment to the originating broker. Subsequent

Assume $B_{t+1} = B_t + \lambda z_t$ and $A_{t+1} = A_t + \lambda z_t$, where $\lambda$ is the proportion of the bid-ask spread due to adverse information. Combining these assumptions with (1) yields the specialist's expected gross profit following a sell or buy, namely:

$E_t(P_{t+1}) - P_t = -(1-\lambda-\theta)z_t$ with $\theta = 2\delta - 1$. Since $\theta$ is directly related to sequences of trades through $\delta$, it can be thought of as an order persistence parameter.

Since z is signed, the gross profit as a percentage of the effective spread is:

$\gamma = 1 - \lambda - \theta$.

Following Huang and Stoll (1994) and Lin (1992), the parameters $\lambda$ and $\theta$ can be estimated with the following regression equations:

$$Q_{t+1} - Q_t = \lambda z_t + e_{t+1} \qquad (2)$$

$$z_{t=1} = \theta z_t + n_{t+1} \qquad (3)$$

where the disturbance terms are assumed to be uncorrelated.

Combining equations (2) and (3) with the definition of the effective half spread yields:

$$P_{t+1} - P_t = -\gamma z_t + u_{t+1} \qquad (4)$$

Equations (2) and (4) are used to estimate the adverse information component of the bid-ask spread and the order processing costs of the bid-ask spread respectively.

### 3.3.2. Estimation: Trade Size, Adverse Selection and Order Processing Cost

Equations (2) and (4) form the basis of the analysis of the relationship between trade size and the components of the bid-ask spread in Lin et al. (1995). They also serve

---

transactions seldom occur within the bid-ask spread.

as the basis of our investigation of the interrelationships that exist between trade size, type of trader (or supplier of immediacy), and trade execution path (external or internal to the broker).

In general, our estimation approach differs from that of Lin et al. (1995). Lin et al. (1995) are interested essentially in the relationship between trade size and the cost components of the bid-ask spread. They sort transactions on a firm-by-firm basis by trade size and then assign trades to percentile ranks. The regression equations (2) and (4) are applied to all transactions within those ranks. The same result can be obtained by assigning transactions to percentile ranks and then setting those ranks (less one) as dummy variables. We begin our estimations by confirming the Lin et al. (1995) result of the positive relationship between trade size and bid-ask spread components using the following regression equations:

$$Q_{t+1} - Q_t = \lambda z_t + \alpha_1 (size * z_t) + e_{t+1} \qquad (5)$$

$$P_{t+1} - P_t = -\gamma z_t + \beta_1 (size * z_t) + u_{t+1} \qquad (6)$$

In this model, we define a size indicator variable, $size$, equal to 1 if the trade is in the top quartile of trades for that firm. Thus, for our purposes, a large trade is top quartile, and small trades are those in the lower three quartiles. Lin et al. (1995) use seven percentile categories. Their last four categories together compose the upper quartile, and their first three categories are the lower three quartiles.

Based on Table 3.9, small transactions (i.e. transactions in the lower three quartiles) represent about 10% of total volume, and about three times the number of large transactions. The average size of a small trade in our sample is 312 shares compared to 8,287 shares for large trades. Large transactions are those above 1,157 shares on average. However, this figure ranges from a low of 400 shares to a high of 2,500. Of course, these differences are created in major part by variations in share price, which also are reported in Table 3.9.

The results of the regressions are presented in Table 3.10. The average trade adverse information cost parameter is 0.064 and 0.136 (the sum of the two coefficients) for small and large trades, respectively. While these values are low compared to the Lin et al. (1995) estimates (0.29 on average for the lower 3 quartiles, and 0.44 on average for the upper quartile in Lin et al. (1995)), they are proportionate.

For order execution cost, the estimate is 0.300 (similar to the average Lin et al. (1995) estimate of 0.37) and 0.268 for large trades (also similar to average Lin et al. (1995) estimates).

Differences in our results with those of Lin et al. (1995) may be due to a number of factors. Specifically, Lin et al. (1995) eliminate trades at the same price and volume as immediately preceding trades. We do not. Moreover, Lin et al. (1995) sort trades by percentile rank and then perform the regressions. This could give undue weight to a few firms with higher or lower numbers of trades within those ranks. Our procedure, since it is based on individual firms does not have this problem. If such an influence did exist, it would be apparent from the differences between means and medians for the sample. Our sample is smaller than that of Lin et al. (1995), 35 firms versus 150. Data reporting differences (i.e. TSE data versus data used in the Lin et al. (1995) study) also may account for the differences between our results and those reported for U.S. markets. Finally, while the portfolio-like approach used by Lin et al. (1995) accounts for any contemporaneous correlations for quote and price changes across firms, it does not account for firm-specific differences in cost components across firms. In contrast, our non-portfolio procedure does not (initially) account for contemporaneous correlations for quote and price changes across firms but does account for firm-specific differences in cost components across firms.

As is shown in the next section, our approach is easily extended to capture the effect of trader and internalization differences across firms on the cost components of trade costs. Furthermore, in a subsequent section where we test for model robustness, we find that our estimates are not materially affected by further modifying our model to account for contemporaneous cross-equation correlations, with and without a correction for non-synchronous trading, across the TSE 35 firms.

### 3.3.3. Estimation of the Cost Components of the Bid-Ask Spread Accounting for Trade Size, Trader, and Internalization Differences

In order to examine the relationships between trader type, trade size and trading mechanism, we estimate two regression equations. First, for adverse information, we estimate:

$$
\begin{aligned}
Q_{t+1} - Q_t = \lambda z_t &+ \alpha_1 * LARGE * z_t + \alpha_2 * EXTERNAL * z_t \\
&+ \alpha_3 * PRINPRIN * z_t + \alpha_4 * PRINNEITH * z_t \\
&+ \alpha_5 * NEITHNEITH * z_t + l_{t+1}
\end{aligned} \tag{7}
$$

Second, for order processing, we estimate:

$$
\begin{aligned}
P_{t+1} - P_t = -\gamma z_t &- \beta_1 * LARGE * z_t - \beta_2 * EXTERNAL * z_t \\
&- \beta_3 * PRINPRIN * z_t - \beta_4 * PRINNEITH * z_t \\
&- \beta_5 * NEITHNEITH * z_t + m_{t+1}
\end{aligned} \tag{8}
$$

The capitalized elements in equations (7) and (8) are defined as follows:

LARGE ≡ 1 if the trade is in the upper quartile of trades for that firm, and is 0 otherwise;

EXTERNAL ≡ 1 if both sides of the trade are different brokerage firms, and is 0 otherwise;

PRINPRIN ≡ 1 if a Principal is trading with another Principal, and is 0 otherwise;

PRINNEITH ≡ 1 if a Principal is trading with Neither a principal nor a RT, and is 0 otherwise;

NEITHNEITH ≡ 1 if Neither a principal nor RT is trading with Neither, and is 0 otherwise.

The five indicator variables allow us to examine the cost components of the bid-ask spread in a parsimonious manner. Relationships are examined by examining the coefficient sums conditional on values of different indicator variables. For example, adverse information conditional on trades being large, external and between principals is $(\lambda + \alpha_1 + \alpha_2 + \alpha_3)$ based on equation (7). All indicator variables at 0 (the base case) represent small, internal RT trades.

The underlying specifications in (7) and (8) implicitly assume that interactions are not relevant to the question of relative transaction costs. For example, "large external trades" are not examined for their differential impact on cost as a separate category. To the extent that interactions are important, regression equations (7) and (8) may be misspecified. However, since interaction terms have components included in other regression terms, they create a potential for severe multicollinearity that may increase the standard errors of the coefficient estimates.

## 3.4. EMPIRICAL RESULTS FOR COST COMPONENTS REFLECTING VARIOUS TRADE FEATURES SUCH AS TRADE SIZE, TRADER TYPE, AND INTERNALIZATION

### 3.4.1. Basic Regression Results

Regression results are reported in Table 3.11 (adverse information) and Table 3.12 (order processing). Table 3.13 provides the values of the adverse selection component (Panel A) and order processing (Panel B) conditional on indicator variables. Table 3.14 looks at the sum for those two cost components.

First, based on Tables 3.11 and 3.12 (Equations (7) and (8), respectively), all of the coefficient estimates in both equations are significantly different from zero at the 5% confidence level or above. The lowest average t-statistic is 2.807 (for the coefficient of $z_t$ in Equation (7)). Median values for estimates of the coefficients, standard errors and t-statistics are provided in order to evaluate the influence of outliers on t-statistic averages. The lowest median t-statistic is –2.648 (the coefficient of PRINNEITH in Equation (7)).

Thus, it is reasonable to conclude that inferences based on the significance of the means are robust.

Table 3.13, Panel A converts the coefficient estimates from Table 3.11 into estimates of the cost of adverse information for the conditional features of interest. Those features are trade size, trader type and internalization. The values of adverse information are then sorted from lowest to highest. We note from the outset that estimates of the adverse information component are lower than in U.S. studies such as that of Lin et al. (1995) or Huang and Stoll (1997). This is true even for the highest estimate in our results. In addition, only the first estimate, small internal transactions between transactors who are neither RTs nor principals is negative (and significant). We have no interpretation for this negative estimate of adverse information. However, given that this estimate is close to zero, it appears to be of little economic significance.

In general, the observed order for increases in the cost of adverse information is consistent with expectations. That is, we would expect, on the basis of the market microstructure literature, to observe a particular order in the results with large, external, between-principal trades having the greatest degree of adverse information. Dividing Panel A of Table 3.13 in half, we note that six of the eight lowest estimates are small size trades, and six are large in the top eight. We conclude that even considering trader and internalization, size is the fundamental determinant of the cost of adverse information. This conclusion is consistent with Lin et al. (1995) and Easley and O'Hara (1987).

Evidence exists to support the notion that internal trades are lower in adverse information cost than external trades, since six of the first eight categories are internal. Type of trader appears to be of significance as well. Of the four possible Principal-Principal combinations, three of them appear in the highest eight adverse information cost estimates. Similarly of the four possible Neither-Neither combinations, three of them are in the lowest eight. The fourth Neither-Neither combination is the second-lowest of the top eight. We conclude from this that trades between principals have a greater cost of adverse information than trades between transactors who are neither principals nor RTs. This conclusion is consistent with the model of POF and internalization of Battalio and Holden (1997). Thus large, between principal trades have the highest adverse information cost, while small trades between principals and neither RTs nor principals have the lowest cost. These features may be amplified to some degree by whether the trade is external.

Panel B of Table 3.13 presents a very different story for the costs of order processing. Large trades may have a somewhat lower cost than small trades, and this is consistent with the economies of scale in trade size. However, the fundamental determinant of order processing cost is whether the trade is internal or external with internal having the highest cost. The eight highest cost combinations are all internal while the eight lowest are all external. This result is consistent with POF in which firms incur additional cost for order flow that is executed internally. In terms of trader type, there does not appear to be any category exhibiting lower or higher order processing cost than another.

Table 3.14 adds the costs of adverse information and order processing together to provide an overall view. The highest cost type of transaction is a large internal trade with RT participation. The lowest cost transactions are small external ones between transactors who are neither principals nor RTs. On balance, the total cost of trading is clearly more dependent upon whether the trade is internal or external, and whether or not it involves a RT than whether the trade is large or small. Since size is important in determining the cost of adverse information, it appears that POF or some other cost consistent with internalization, is very expensive.

### 3.4.2 Tests of Robustness

Three tests of robustness are conducted on the data. First, the model indicated by equations (7) and (8) is modified to include and additional regressor as follows:

$$\begin{aligned} Q_{t+1} - Q_t = \lambda z_t &+ \alpha_1 * LARGE * z_t + \alpha_2 * EXTERNAL * z_t \\ &+ \alpha_3 * PRINPRIN * z_t + \alpha_4 * PRINNEITH * z_t \\ &+ \alpha_5 * NEITHNEITH * z_t + \alpha_6 * TIPSQ + l_{t+1} \end{aligned} \tag{9}$$

where $TIPSQ$ is the log change in the quote mid-point of TIPs, a single security similar to a Standard and Poor Depository Receipt but for the TSE 35 Index.

Equation (8) was similarly modified:

$$\begin{aligned} P_{t+1} - P_t = -\gamma z_t &- \beta_1 * LARGE * z_t - \beta_2 * EXTERNAL * z_t \\ &- \beta_3 * PRINPRIN * z_t - \beta_4 * PRINNEITH * z_t \\ &- \beta_5 * NEITHNEITH * z_t + \beta_6 * TIPSP + m_{t+1} \end{aligned} \tag{10}$$

where *TIPSP* is the log of the TIPs transaction price change.

As noted earlier, these regressions are performed in order to account for possible cross sectional dependency among the 35 firms used in the study. Since the coefficient estimates resulting from these equations are very similar to those obtained from equations (7) and (8), with only minor changes in the t-statistics, we do not formally present these results.[5] Inferences are unchanged.

The second test of robustness also is designed to account for cross sectional dependency but applies the Scholes and Williams (1977) adjustment for possible non-synchronous trading in TIPs to the additional terms described in equations (9) and (10). Specifically, each equation was run with a contemporaneous TIPs term (as in (9) and (10)) and with both a lagged and leading TIPs term.[6] Inferences remain unchanged as a result of the application of this procedure. Since the coefficient estimates and levels of significance again are essentially unchanged, these results are not reported herein.

The third test of robustness is an application of the model in equations (7) and (8) to the post-tick-size-reduction period (post) immediately following the sample period used in the earlier part of this study. Thus, the model is rerun for the 54 trading days from April 15, 1996 through June 28, 1996. Unreported coefficient estimates are similar to those obtained in the pre-tick-size-reduction period (pre). While the unreported t-statistics are

---

[5] These unreported results are available upon request from the author.

[6] TIPs quotes and transactions are not, in general, available for each quote or transaction for individual firms. Therefore the following procedure is adopted both for equations (9) and (10) as well as for the Scholes and Williams (1977) procedure: A file was constructed in which TIPs quotes and transaction

somewhat lower, all average t-statistics are significantly different from zero at the 95% level or above. Adverse information coefficients in the post period range from a low of -0.018 to a high of 0.208 (as compared to the pre period of –0.011 to 0.214), while order processing costs are uniformly higher in the post period ranging from a low of 0.166 to a high of 0.689 (as compared to the pre period of 0.102 to 0.664). The order of the unreported conditional estimates for the post period is very similar to that indicated in both panels of Table 3.13 as well as for the total cost estimates reported in Table 3.14. Thus, all previous inferences basically are unchanged. We conclude, therefore, that the model is robust to changes in minimum tick size.

## 3.5. CONCLUSION

This chapter examined the differential impact of trade size, trader type and internalization on the cost components of the bid ask spread. It first used data for the TSE 35 stocks for the period from January through mid-April 1996 (pre-switch to decimalization). We document the relative importance of internal trading for TSE member firms (i.e 58.6% of all total trading volume is internal), as well as the differential importance of trader types for both internal and external trades. The three types of traders we identify are designated market markers (RT), principal to principal trades, and trades involving principals and traders who are neither principals nor RTs. We find that principal trades by volume are more important externally than internally. This is consistent with the conclusions of Battalio and Holden (1997) that internal trading with

prices were present for each six-second trading period. If a TIPs transaction did not take place during a given six-second period, the most recent transaction price and related quote was used.

the more-likely-to-be-informed is undesirable. Trades involving neither RTs nor principals are more likely to be internal than external. This is also consistent with the inferences of Battalio and Holden (1997). Both of these results would be observed if firms pay for order flow.

Indicator variable regressions are performed to isolate the relationships between trade size, trader type and internalization on the cost components of the bid-ask spread, specifically adverse information cost and order processing cost. We draw three major conclusions. First, trade size is the single most important determinant of the cost of adverse information. Second, whether trades are executed internally or externally is the single most important determinant of order processing cost, but is an important determinant of the cost of adverse information as well. This conclusion is consistent with payment for order flow. Third, identification of the trader type (e.g. principal-principal) is a determinant of both the cost of adverse information as well as the cost of order processing.

This chapter has some potential limitations. First, the data set includes 35 firms for about 6 months. However, the TSE 35 firms used herein are the most actively traded stocks in the market and their use may therefore attenuate problems created by non-synchronous trading. A recent study by Griffiths et al. (1998) uses a shorter time period. Finally, our definition of internal trading assumes that the internal trading activity is evidenced uniquely by identically numbered firms on the TSE Equity History tape. We believe that affiliations between brokers may mask additional internal trading. However, correcting for any misclassifications by internal and external trading should improve our results by further accentuating the differences identified herein.

**Table 3.1.** Number of Transactions by Type of Trader

| | | | | | RT From: | | | Principal From: | | Neither From: |
|---|---|---|---|---|---|---|---|---|---|---|
| | Total | RT | Principal | Neither | RT | Principal | Neither | Principal | Neither | Neither |
| Panel A: Buys | | | | | | | | | | |
| Total | 558,534 | 112,231 | 168,601 | 277,702 | 7,187 | 27,587 | 77,457 | 63,080 | 78,928 | 105,637 |
| Mean | 15,958 | 3,207 | 4,817 | 7,934 | 205 | 788 | 2,213 | 1,802 | 2,255 | 3,018 |
| Median | 13,923 | 2,594 | 3,495 | 6,283 | 59 | 607 | 1,754 | 1,310 | 1,618 | 2,499 |
| Max | 47,222 | 17,078 | 11,761 | 20,022 | 1,299 | 2,836 | 14,632 | 6,024 | 5,928 | 8,924 |
| Min | 3,754 | 525 | 1,313 | 1,525 | 5 | 142 | 347 | 524 | 574 | 444 |
| Panel B: Sells | | | | | | | | | | |
| Total | 558,534 | 110,777 | 185,735 | 262,022 | | 26,593 | 76,997 | | 95,068 | |
| Mean | 15,958 | 3,165 | 5,307 | 7,486 | | 760 | 2,200 | | 2,716 | |
| Median | 13,923 | 2,973 | 4,560 | 6,384 | | 557 | 2,136 | | 2,139 | |
| Max | 47,222 | 9,209 | 12,727 | 29,484 | | 2,507 | 6,611 | | 8,236 | |
| Min | 3,754 | 550 | 1,526 | 1,450 | | 165 | 343 | | 706 | |
| Panel C: Net Buys | | | | | | | | | | |
| Total | | 1,454 | (17,134) | 15,680 | | 994 | 460 | | (16,140) | |
| Mean | | 42 | (490) | 448 | | 28 | 13 | | (461) | |
| Median | | (157) | (301) | 495 | | 28 | (275) | | (318) | |
| Max | | 7,869 | 2,199 | 9,727 | | 329 | 8,021 | | 2,379 | |
| Min | | (4,614) | (5,113) | (9,462) | | (213) | (4,683) | | (5,044) | |

This table in three panels shows the number of buy trades, sell trades and net buy trades undertaken by one of three types of traders: Registered Traders (RT), Principals, and Neither principals nor RTs. The upper panel shows total buys by trader. For example, the whole sample consists of 558,534 trades over the 35 TSE 35 firms, with an average of 15,958 trades per firm, a median of 13,923, a maximum of 47,222 and minimum of 3,754. The next columns of the top panel show the distribution of total trades by type of trader. Thus RT buys number 112,231. The sixth column of the table shows buys by the RT from RT as 7,187. Similarly, columns nine and eleven show the same breakdown for Principal and Neither trades. The second panel shows the same data for sells. The third panel shows net buys.

**TABLE 3.2a. Transactions and Types of Traders in Percentages**

| Panel A: All with itself and other two trade categories | | | |
|---|---|---|---|
| Transaction type | All trades with: | | |
| | RT | Principal | Neither |
| Buys | 20.1% | 30.2% | 49.7% |
| Sells | 19.8% | 33.3% | 46.9% |
| Net buys | 0.3% | -3.1% | 2.8% |
| **Panel B: RT with itself and other two trade categories** | | | |
| Transaction type | RT trades with: | | |
| | RT | Principal | Neither |
| Buys | 1.3% | 4.9% | 13.9% |
| Sells | 1.3% | 4.8% | 13.8% |
| Net buys | 0.0% | 0.2% | 0.1% |
| **Panel C: Principal with itself and other two trade categories** | | | |
| Transaction type | Principal trades with: | | |
| | RT | Principal | Neither |
| Buys | 4.8% | 11.3% | 14.1% |
| Sells | 4.9% | 11.3% | 17.0% |
| Net buys | -0.2% | 0.0% | -2.9% |
| **Panel D: Neither with itself and other two trade categories** | | | |
| Transaction type | Neither trades with: | | |
| | RT | Principal | Neither |
| Buys | 13.8% | 17.0% | 18.9% |
| Sells | 13.9% | 14.1% | 18.9% |
| Net buys | -0.1% | 2.9% | 0.0% |

This table shows data presented in Table 3.1 in percentage terms. The top panel, divided into buys, sells and net buys, is for all trader types without regard to the trade counterparty. Thus, registered traders (RT) account for 20.1% of all buys and 19.8% of all sells. The next three panels show how total trades break down with respect to counterparty. Thus, RT buys from RTs, principals and neither account for 1.3%, 4.9% and 13.9% of all trades, respectively.

**Table 3.2b. Trader Shares of Trade Volumes in Percentages**

| Panel A: All with itself and other two trade categories | | | |
|---|---|---|---|
| | All trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 8.8% | 40.0% | 51.3% |
| Sells | 8.4% | 42.9% | 48.7% |
| Net buys | 0.4% | -3.0% | 2.6% |
| **Panel B: RT with itself and other two trade categories** | | | |
| | RT trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 0.8% | 3.7% | 4.3% |
| Sells | 0.8% | 3.5% | 4.1% |
| Net buys | 0.0% | 0.2% | 0.2% |
| **Panel C: Principal with itself and other two trade categories** | | | |
| | Principal trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 3.5% | 15.0% | 21.5% |
| Sells | 3.7% | 15.0% | 24.2% |
| Net buys | -0.2% | 0.0% | -2.8% |
| **Panel D: Neither with itself and other two trade categories** | | | |
| | Neither trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 4.1% | 24.2% | 23.0% |
| Sells | 4.3% | 21.5% | 23.0% |
| Net buys | -0.2% | 2.8% | 0.0% |

This table shows percentage data similar to Table 3.2a but based on share volumes rather than transactions. The top panel, divided into buys, sells and net buys, is for all trader types without regard to the trade counterparty. Thus, registered traders (RT) account for 8.8% of all buys by volume and 8.4% of all sells by volume. The next three panels show how total trades break down with respect to counterparty. Thus, RT buys from RTs, principals and neither account for 0.8%, 3.7% and 4.3% of all trades by volume, respectively.

**Table 3.3.** Summary Statistics on External and Internal Transactions by Type of Trader

| Statistic | | | | | RT trades with: | | | Principal trades with: | | | Neither trades with: | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | RT | Principal | Neither | RT | Principal | Neither | RT | Principal | Neither | RT | Principal | Neither |
| Panel A: External buy transactions | | | | | | | | | | | | | |
| Total | 461,322 | 103,269 | 141,925 | 216,128 | 6,770 | 25,772 | 70,727 | 24,827 | 56,241 | 60,857 | 68,959 | 72,234 | 74,935 |
| Mean | 13,181 | 2,951 | 4,055 | 6,175 | 193 | 736 | 2,021 | 709 | 1,607 | 1,739 | 1,970 | 2,064 | 2,141 |
| Median | 11,456 | 2,536 | 2,922 | 5,207 | 58 | 586 | 1,675 | 547 | 1,183 | 1,282 | 1,839 | 1,719 | 1,856 |
| Max | 38,912 | 16,007 | 10,288 | 14,667 | 1,206 | 2,590 | 13,702 | 2,326 | 5,470 | 4,468 | 5,910 | 5,686 | 5,670 |
| Min | 3,081 | 418 | 1,101 | 1,244 | 4 | 109 | 278 | 127 | 439 | 454 | 335 | 573 | 336 |
| Panel B: External sell transactions | | | | | | | | | | | | | |
| Total | 461,322 | 100,556 | 154,247 | 206,519 | 6,770 | 24,827 | 68,959 | 25,772 | 56,241 | 72,234 | 70,727 | 60,857 | 74,935 |
| Mean | 13,181 | 2,873 | 4,407 | 5,901 | 193 | 709 | 1,970 | 736 | 1,607 | 2,064 | 2,021 | 1,739 | 2,141 |
| Median | 11,456 | 2,755 | 3,800 | 5,067 | 58 | 547 | 1,839 | 586 | 1,183 | 1,719 | 1,675 | 1,282 | 1,856 |
| Max | 38,912 | 8,365 | 11,238 | 23,840 | 1,206 | 2,326 | 5,910 | 2,590 | 5,470 | 5,686 | 13,702 | 4,468 | 5,670 |
| Min | 3,081 | 493 | 1,296 | 1,150 | 4 | 127 | 335 | 109 | 439 | 573 | 278 | 454 | 336 |
| Panel C: Internal sell transactions | | | | | | | | | | | | | |
| Total | 97,212 | 8,962 | 26,676 | 61,574 | 417 | 1,815 | 6,730 | 1,766 | 6,839 | 18,071 | 8,038 | 22,834 | 30,702 |
| Mean | 2,777 | 256 | 762 | 1,759 | 12 | 52 | 192 | 50 | 195 | 516 | 230 | 652 | 877 |
| Median | 2,195 | 154 | 599 | 1,255 | 3 | 33 | 120 | 38 | 163 | 373 | 169 | 449 | 574 |
| Max | 8,310 | 1,071 | 1,884 | 5,827 | 93 | 246 | 930 | 245 | 554 | 1,460 | 1,226 | 2,550 | 3,254 |
| Min | 628 | 4 | 212 | 281 | - | - | 2 | - | 76 | 120 | - | 119 | 108 |
| Panel D: Internal sell transactions | | | | | | | | | | | | | |
| Total | 97,212 | 10,221 | 31,488 | 55,503 | 417 | 1,766 | 8,038 | 1,815 | 6,839 | 22,834 | 6,730 | 18,071 | 30,702 |
| Mean | 2,777 | 292 | 900 | 1,586 | 12 | 50 | 230 | 52 | 195 | 652 | 192 | 516 | 877 |
| Median | 2,195 | 198 | 630 | 1,268 | 3 | 38 | 169 | 33 | 163 | 449 | 120 | 373 | 574 |
| Max | 8,310 | 1,293 | 2,967 | 5,644 | 93 | 245 | 1,226 | 246 | 554 | 2,550 | 930 | 1,460 | 3,254 |
| Min | 628 | 2 | 201 | 300 | - | - | - | - | 76 | 119 | 2 | 120 | 108 |

This table in four panels shows the number of external buy trades and sell trades (Panels 1 and 2) and internal buy trades and sell trades (Panels 3 and 4) undertaken by one of three types of traders: Registered Traders (RT), Principals, and Neither principals nor RTs. In the first panel, there are 461,322 external trades over the 35 TSE 35 firms, with an average of 13,181 external trades per firm, a median of 11,456, a maximum of 38,912 and minimum of 3,081. The next columns of the top panel show the distribution of external trades by type of trader. Thus RT buys number 103,269. The fifth column of the table shows external buys by the RT from RT as 6,770. Similarly, columns 9 and 12 show the same breakdown for Principal and Neither trades. The second panel shows the same data for external sells. The third and fourth panels show the same data as panels 1 and 2 except for internal transactions.

**Table 3.4.** Summary Statistics for External and Internal Volumes by Type of Trader

| Statistic | | | | | RT trades with: | | | Principal trades with: | | | Neither trades with: | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | RT | Principal | Neither | RT | Principal | Neither | RT | Principal | Neither | RT | Principal | Neither |
| Panel A: External buy transactions | | | | | | | | | | | | | |
| Total | 541,727,169 | 88,285,297 | 278,959,557 | 174,482,315 | 9,785,561 | 44,413,791 | 34,085,945 | 42,488,253 | 149,693,740 | 86,777,564 | 33,204,316 | 89,668,896 | 51,609,103 |
| Mean | 15,477,919 | 2,522,437 | 7,970,273 | 4,985,209 | 279,587 | 1,268,965 | 973,884 | 1,213,950 | 4,276,964 | 2,479,359 | 948,695 | 2,561,968 | 1,474,546 |
| Median | 11,330,970 | 1,198,737 | 4,688,389 | 4,513,089 | 41,200 | 520,253 | 583,884 | 525,466 | 2,388,200 | 1,855,350 | 694,133 | 2,084,750 | 1,441,700 |
| Max | 45,404,087 | 11,545,011 | 26,092,580 | 11,598,095 | 2,054,970 | 7,035,172 | 3,483,743 | 5,790,442 | 14,098,800 | 7,835,100 | 2,756,945 | 6,857,900 | 2,928,200 |
| Min | 2,497,478 | 170,475 | 1,268,103 | 902,312 | 2,000 | 65,009 | 88,694 | 97,391 | 526,000 | 546,300 | 101,862 | 530,200 | 270,250 |
| Panel B: External sell transactions | | | | | | | | | | | | | |
| Total | 541,727,169 | 85,478,130 | 283,776,427 | 172,472,612 | 9,785,561 | 42,488,253 | 33,204,316 | 44,413,791 | 149,693,740 | 89,668,896 | 34,085,945 | 86,777,564 | 51,609,103 |
| Mean | 15,477,919 | 2,442,232 | 8,107,898 | 4,927,789 | 279,587 | 1,213,950 | 948,695 | 1,268,965 | 4,276,964 | 2,561,968 | 973,884 | 2,479,359 | 1,474,546 |
| Median | 11,330,970 | 1,209,971 | 5,412,764 | 4,405,916 | 41,200 | 525,466 | 694,133 | 520,253 | 2,388,200 | 2,084,750 | 583,884 | 1,855,350 | 1,441,700 |
| Max | 45,404,087 | 10,285,816 | 27,184,478 | 14,247,043 | 2,054,970 | 5,790,442 | 2,756,945 | 7,035,172 | 14,098,800 | 6,857,900 | 3,483,743 | 7,835,100 | 2,928,200 |
| Min | 2,497,478 | 216,025 | 1,363,909 | 917,544 | 2,000 | 97,391 | 101,862 | 65,009 | 526,000 | 530,200 | 88,694 | 546,300 | 270,250 |
| Panel C: Internal buy transactions | | | | | | | | | | | | | |
| Total | 767,811,165 | 26,579,073 | 244,232,224 | 496,999,868 | 903,653 | 3,648,216 | 22,027,204 | 2,934,093 | 47,083,857 | 194,214,274 | 20,199,153 | 227,517,374 | 249,283,341 |
| Mean | 21,937,462 | 759,402 | 6,978,064 | 14,199,996 | 25,819 | 104,235 | 629,349 | 83,831 | 1,345,253 | 5,548,979 | 577,119 | 6,500,496 | 7,122,381 |
| Median | 20,920,152 | 226,900 | 5,349,959 | 13,331,392 | 3,100 | 54,400 | 144,294 | 42,550 | 861,011 | 3,948,449 | 132,520 | 5,483,937 | 5,550,847 |
| Max | 62,144,502 | 4,536,400 | 20,887,798 | 39,777,813 | 265,700 | 612,255 | 4,246,600 | 517,680 | 5,934,506 | 15,860,450 | 5,862,600 | 21,262,550 | 16,008,035 |
| Min | 2,248,378 | 375 | 400,825 | 1,749,673 | - | - | 175 | - | 133,234 | 228,804 | - | 1,010,300 | 403,245 |
| Panel D: Internal sell transactions | | | | | | | | | | | | | |
| Total | 767,811,165 | 24,036,899 | 278,249,447 | 465,524,819 | 903,653 | 2,934,093 | 20,199,153 | 3,648,216 | 47,083,857 | 227,517,374 | 22,027,204 | 194,214,274 | 249,283,341 |
| Mean | 21,937,462 | 686,769 | 7,949,984 | 13,300,709 | 25,819 | 83,831 | 577,119 | 104,235 | 1,345,253 | 6,500,496 | 629,349 | 5,548,979 | 7,122,381 |
| Median | 20,920,152 | 163,400 | 6,979,751 | 11,990,630 | 3,100 | 42,550 | 132,520 | 54,400 | 861,011 | 5,483,937 | 144,294 | 3,948,449 | 5,550,847 |
| Max | 62,144,502 | 6,094,200 | 24,573,150 | 31,477,152 | 265,700 | 517,680 | 5,862,600 | 612,255 | 5,934,506 | 21,262,550 | 4,246,600 | 15,860,450 | 16,008,035 |
| Min | 2,248,378 | 200 | 1,179,476 | 686,621 | - | - | - | - | 133,234 | 1,010,300 | 175 | 228,804 | 403,245 |

This table in four panels shows the number of external buy trades and sell trades by volume (Panels 1 and 2) and internal buy trades and sell trades by volume (Panels 3 and 4) undertaken by one of three types of traders: Registered Traders (RT), Principals, and Neither principals nor RTs. In the first panel, there were 541,727,169 shares traded externally over the 35 TSE 35 firms, with an average of 15,477,919 shares traded externally per firm, a median of 11,330,970, a maximum of 45,404,087 and minimum of 2,497,478. The next columns of the top panel show the distribution of external trades by volume by type of trader. Thus RT buy volume is 88,285,297. The sixth column of the table shows external buy volume by the RT from RT as 9,785,561. Similarly, columns 9 and 12 show the same breakdown for Principal and Neither trade volumes. The second panel shows the same data for external sell volume. The third and fourth panels show the same data as panels 1 and 2 except for internal transactions by volume.

Table 3.5. **External Transactions By Type Of Trader**

| Panel A: All with itself and other two trade categories | | | |
|---|---|---|---|
| | ALL trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 22.4% | 30.8% | 46.8% |
| Sells | 21.8% | 33.4% | 44.8% |
| Net buys | 0.6% | -2.7% | 2.1% |

| Panel B: RT with itself and other two trade categories | | | |
|---|---|---|---|
| | RT trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 1.5% | 5.6% | 15.3% |
| Sells | 1.5% | 5.4% | 14.9% |
| Net buys | 0.0% | 0.2% | 0.4% |

| Panel C: Principal with itself and other two trade categories | | | |
|---|---|---|---|
| | Principal trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 5.4% | 12.2% | 13.2% |
| Sells | 5.6% | 12.2% | 15.7% |
| Net buys | -0.2% | 0.0% | -2.5% |

| Panel D: Neither with itself and other two trade categories | | | |
|---|---|---|---|
| | Neither trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 14.9% | 15.7% | 16.2% |
| Sells | 15.3% | 13.2% | 16.2% |
| Net buys | -0.4% | 2.5% | 0.0% |

This table shows the number of external transactions by type of trader in percentage terms. The top panel, divided into buys, sells and net buys, is for all trader types without regard to the trade counterparty. Thus, registered traders (RT) account for 22.4% of all external buys and 21.8% of all external sells. The next three panels show how total trades break down with respect to counterparty. Thus, external RT buys from RTs, principals and neither account for 1.5%, 5.6% and 15.3% of all trades, respectively.

Table 3.6. Descriptive Statistics on Internal Transactions by Type of Trader

| Panel A: All with itself and other two trade categories | | | |
|---|---|---|---|
| Transaction type | All trades with: | | |
| | RT | Principal | Neither |
| Buys | 9.2% | 27.4% | 63.3% |
| Sells | 10.5% | 32.4% | 57.1% |
| Net buys | -1.3% | -5.0% | 6.2% |
| **Panel B: RT with itself and other two trade categories** | | | |
| Transaction type | RT trades with: | | |
| | RT | Principal | Neither |
| Buys | 0.4% | 1.9% | 6.9% |
| Sells | 0.4% | 1.8% | 8.3% |
| Net buys | 0.0% | 0.1% | -1.3% |
| **Panel C: Principal with itself and other two trade categories** | | | |
| Transaction type | Principal trades with: | | |
| | RT | Principal | Neither |
| Buys | 1.8% | 7.0% | 18.6% |
| SELLS | 1.9% | 7.0% | 23.5% |
| Net buys | -0.1% | 0.0% | -4.9% |
| **Panel D: Neither with itself and other two trade categories** | | | |
| Transaction type | Neither trades with: | | |
| | RT | Principal | Neither |
| Buys | 8.3% | 23.5% | 31.6% |
| Sells | 6.9% | 18.6% | 31.6% |
| Net buys | 1.3% | 4.9% | 0.0% |

This table shows the number of internal transactions by type of trader in percentage terms. The top panel, divided into buys, sells and net buys, is for all trader types without regard to the trade counterparty. Thus, registered traders (RT) account for 9.2% of all internal buys and 10.5% of all internal sells. The next three panels show how total trades break down with respect to counterparty. Thus, internal RT buys from RTs, principals and neither account for 0.4%, 1.9% and 6.9% of all trades, respectively.

Table 3.7. Descriptive Statistics on External Volumes by Type of Trader

| Panel A: All with itself and other two trade categories | | | |
|---|---|---|---|
| | All trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 16.3% | 51.5% | 32.2% |
| Sells | 15.8% | 52.4% | 31.8% |
| Net buys | 0.5% | -0.9% | 0.4% |
| **Panel B:** RT with itself and other two trade categories | | | |
| | RT trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 1.8% | 8.2% | 6.3% |
| Sells | 1.8% | 7.8% | 6.1% |
| Net buys | 0.0% | 0.4% | 0.2% |
| **Panel C:** Principal with itself and other two trade categories | | | |
| | Principal trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 7.8% | 27.6% | 16.0% |
| Sells | 8.2% | 27.6% | 16.6% |
| Net buys | -0.4% | 0.0% | -0.5% |
| **Panel D:** Neither with itself and other two trade categories | | | |
| | Neither trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 6.1% | 16.6% | 9.5% |
| Sells | 6.3% | 16.0% | 9.5% |
| Net buys | -0.2% | 0.5% | 0.0% |

This table shows the share volume of external transactions by type of trader in percentage terms. The top panel, divided into buys, sells and net buys, is for all trader types without regard to the trade counterparty. Thus, registered traders (RT) account for 16.3% of all external buys by volume and 15.8% of all external sells by volume. The next three panels show how the volume of total trades break down with respect to counterparty. Thus, external RT buys from RTs, principals and neither account for 1.8%, 8.2% and 6.3% of all trades by volume, respectively.

**Table 3.8.** Summary Statistics on Internal Volumes by Type of Trader

| Panel A: All with itself and other two trade categories | | | |
|---|---|---|---|
| | All trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 3.5% | 31.8% | 64.7% |
| Sells | 3.1% | 36.2% | 60.6% |
| Net buys | 0.3% | -4.4% | 4.1% |
| **Panel B: RT with itself and other two trade categories** | | | |
| | RT trades with: | | |
| Transaction type | RT | Principal | Neither |
| Buys | 0.1% | 0.5% | 2.9% |
| Sells | 0.1% | 0.4% | 2.6% |
| Net buys | 0.0% | 0.1% | 0.2% |
| **Panel C: Principal with itself and other two trade categories** | | | |
| | Principal trades with: | | |
| Transaction type | RT | Principal | Neither |
| BUYS | 0.4% | 6.1% | 25.3% |
| SELLS | 0.5% | 6.1% | 29.6% |
| NET BUYS | -0.1% | 0.0% | -4.3% |
| **Panel D: Neither with itself and other two trade categories** | | | |
| | Neither trades with: | | |
| Transaction type | RT | Principal | Neither |
| BUYS | 2.6% | 29.6% | 32.5% |
| SELLS | 2.9% | 25.3% | 32.5% |
| NET BUYS | -0.2% | 4.3% | 0.0% |

This table shows the share volume of internal transactions by type of trader in percentage terms. The top panel, divided into buys, sells and net buys, is for all trader types without regard to the trade counterparty. Thus, registered traders (RT) account for 3.5% of all internal buys by volume and 3.1% of all internal sells by volume. The next three panels show how the volume of total trades break down with respect to counterparty. Thus, internal RT buys from RTs, principals and neither account for 0.1%, 0.5% and 2.9% of all trades by volume, respectively.

**Table 3.9.** Summary Descriptive Statistics on Large and Small Transactions

| | | Cross-sectional | | | |
|---|---|---|---|---|---|
| Type of Transaction | TOTAL | Mean | Median | High | Low |
| # Transactions | 558,534 | 15,958 | 13,923 | 47,222 | 3,754 |
| Volume-shares | 1,309,538,350 | 37,415,381 | 29,846,000 | 107,548,600 | 6,159,945 |
| Small volume-shares | 125,450,717 | 3,584,306 | 2,678,219 | 13,407,160 | 710,030 |
| Large volume-shares | 1,184,087,620 | 33,831,075 | 27,650,390 | 97,787,700 | 5,117,695 |
| # Small transactions | 406,799 | 11,623 | 10,316 | 34,258 | 2,519 |
| # Large transactions | 151,735 | 4,335 | 3,938 | 12,964 | 1,212 |
| Small trans size-shares | | 312 | 277 | 733 | 116 |
| Large trans size-shares | | 8,287 | 8,237 | 14,409 | 1,891 |
| Cutoff for upper 25% | | 1,157 | 1,000 | 2,500 | 400 |
| Quoted spread | | $ 0.146 | $ 0.139 | $ 0.228 | $ 0.116 |
| Price/share | | $ 30.16 | $ 27.87 | $ 63.82 | $ 11.35 |

This table shows transaction size and other data for the TSE 35 firms from January 2, 1996 to April 12, 1996. Where appropriate, data is provided in total for all firms and cross-sectionally across firms (namely, the mean, median, highest value and lowest value among the firms). Thus, there are a total of 558,534 transactions, 15,958 per firm on average, with a per firm median of 13,923. The highest number of transactions for any firm was 47,222 and the lowest was 3,754. A transaction is large if it is in the top quartile of all trades within the period. A transaction is small if it is in the lower three quartiles of all trades within the period.

**Table 3.10.** OLS Estimates of Cost Components of the Bid-ask Spread Including a Trade Size Indicator

Panel A: Parameter estimates for the adverse information component of the bid-ask spread

| | $\lambda$ | Stan. Err. | t-value | $\alpha_1$ | Stan. Err | t-value |
|---|---|---|---|---|---|---|
| Average | 0.064 | 0.005 | 11.348 | 0.072 | 0.010 | 7.808 |
| Median | 0.048 | 0.005 | 10.125 | 0.091 | 0.009 | 9.352 |
| High | 0.224 | 0.015 | 22.040 | 0.160 | 0.027 | 16.180 |
| Low | 0.011 | 0.002 | 4.914 | -0.224 | 0.003 | -17.430 |

Panel A shows parameter estimates of the adverse information component of the bid-ask spread, $\lambda$, and for the coefficient, $\alpha_1$, of the top quartile indicator variable, $size$. The regression equation is: $Q_{t+1} - Q_t = \lambda z_t + \alpha_1(size * z_t) + e_{t+1}$, where the left-hand side is the change in the log quoted midpoint, $z_t = P_t - Q_t$, and $P_t$ is the log transaction price at time t. $size$ is 1 if the trade is in the top quartile of all trades for the firm over the entire period. Averages, medians, highs and lows are computed over the 35 firms in the study (TSE 35). Both coefficient estimates are significant above a 5% confidence level.

Panel B: Parameter estimates of the order-processing component of the bid-ask spread

| | $\gamma$ | Stan. Err. | t-value | $\beta_1$ | Stan. Err. | t-value |
|---|---|---|---|---|---|---|
| Average | 0.300 | 0.008 | 40.736 | -0.032 | 0.016 | -2.762 |
| Median | 0.292 | 0.007 | 39.295 | -0.051 | 0.014 | -3.401 |
| High | 0.423 | 0.004 | 91.580 | 0.627 | 0.008 | 30.310 |
| Low | 0.193 | 0.019 | 12.270 | -0.151 | 0.029 | -15.270 |

Panel B shows parameter estimates of the order processing component of the bid-ask spread, $\gamma$ and for the coefficient, $\beta_1$, of the top quartile indicator variable, $size$. The regression equation is: $P_{t+1} - P_t = -\gamma z_t + \beta_1(size * z_t) + u_{t+1}$, where the left-hand side is the change in the log transaction price, $z_t = P_t - Q_t$, and $P_t$ is the log transaction price at time t. $size$ is 1 if the trade is in the top quartile of all trades for the firm over the entire period. Averages, medians, highs and lows are computed over the 35 firms in the study (TSE 35). Both coefficient estimates are significant above a 5% confidence level.

**Table 3.11.** Estimates of the Adverse Information Component of the Bid-Ask Spread Accounting for Trade size, Trader Type, and Internalization

| Statistic | Coef. of Z | s.e. | t-stat | Coef. of Principal-Principal | s.e. | t-stat |
|---|---|---|---|---|---|---|
| Average | 0.050 | 0.015 | 2.807 | 0.035 | 0.017 | 2.823 |
| Median | 0.029 | 0.012 | 2.707 | 0.038 | 0.016 | 2.886 |
| High | 0.502 | 0.048 | 10.550 | 0.191 | 0.037 | 9.967 |
| Low | -0.004 | 0.004 | -0.941 | -0.187 | 0.007 | -5.200 |
| # signif. | 24 | | | 24 | | |
| | | | | | | |
| | Large | | | Principal-Neither | | |
| Average | 0.081 | 0.012 | 7.914 | -0.037 | 0.012 | -2.893 |
| Median | 0.091 | 0.010 | 8.520 | -0.027 | 0.010 | -2.648 |
| High | 0.164 | 0.032 | 14.140 | 0.047 | 0.034 | 1.578 |
| Low | -0.034 | 0.004 | -1.964 | -0.345 | 0.002 | -10.290 |
| # signif. | 33 | | | 22 | | |
| | | | | | | |
| | External | | | Neither-Neither | | |
| Average | 0.048 | 0.013 | 4.104 | -0.061 | 0.014 | -4.051 |
| Median | 0.030 | 0.010 | 3.623 | -0.043 | 0.012 | -3.988 |
| High | 0.219 | 0.042 | 13.900 | 0.000 | 0.041 | -0.509 |
| Low | 0.017 | 0.004 | 0.595 | -0.413 | 0.004 | -10.150 |
| # signif. | 29 | | | 29 | | |

This table presents average, median, high and low coefficient estimates for the TSE 35 firms used in the study. Estimates are based on the following model of adverse information:

$$\begin{aligned} Q_{t+1} - Q_t = {} & \lambda z_t + \alpha_1 * LARGE * z_t + \alpha_2 * EXTERNAL * z_t \\ & + \alpha_3 * PRINPRIN * z_t + \alpha_4 * PRINNEITH * z_t \\ & + \alpha_5 * NEITHNEITH * z_t + l_{t+1} \end{aligned}$$

where $Q_t$ is the log quote midpoint and $z_t = P_t - Q_t$, where $P_t$ is the log transaction price. All coefficient estimates are significantly different from zero at the 5% confidence level or greater. "# Signif." indicates the number of individual firms among the 35 having coefficient estimates significantly different from zero at the 5% confidence level or greater, and of the expected sign. 's.e.' refers to the standard error, and 't-stat' refers to the t-value.

**Table 3.12.** Estimate of the Order Processing Component of the Bid-Ask Spread Accounting for Trade size, Trader Type, and Internalization

| | Coef. of: | s.e. | t-stat | Coef. of: | s.e. | t-stat |
|---|---|---|---|---|---|---|
| | Z | | | Principal-Principal | | |
| Average | -0.664 | 0.022 | -35.749 | 0.186 | 0.026 | 8.211 |
| Median | -0.639 | 0.019 | -32.290 | 0.175 | 0.024 | 6.391 |
| High | -0.363 | 0.041 | -8.795 | 0.343 | 0.049 | 20.850 |
| Low | -0.879 | 0.010 | -67.620 | 0.037 | 0.012 | 0.969 |
| # signif. | 35 | | | 34 | | |
| | | | | | | |
| | Large | | | Principal-Neither | | |
| Average | 0.052 | 0.017 | 3.837 | 0.189 | 0.018 | 12.708 |
| Median | 0.060 | 0.015 | 3.894 | 0.193 | 0.015 | 10.140 |
| High | 0.161 | 0.031 | 13.790 | 0.356 | 0.036 | 29.380 |
| Low | -0.172 | 0.009 | -5.909 | -0.104 | 0.009 | -4.428 |
| # signif. | 27 | | | 33 | | |
| | | | | | | |
| | External | | | Neither-Neither | | |
| Average | 0.288 | 0.019 | 17.745 | 0.222 | 0.021 | 13.417 |
| Median | 0.285 | 0.016 | 15.720 | 0.230 | 0.018 | 10.700 |
| High | 0.545 | 0.035 | 39.440 | 0.399 | 0.048 | 32.990 |
| Low | 0.063 | 0.009 | 1.791 | -0.171 | 0.010 | -6.625 |
| # signif. | 34 | | | 32 | | |

This table presents average, median, high and low coefficient estimates for the TSE 35 firms used in the study. Estimates are based on the following model of order processing costs:

$$P_{t+1} - P_t = -\gamma z_t - \beta_1 * LARGE * z_t - \beta_2 * EXTERNAL * z_t$$
$$- \beta_3 * PRINPRIN * z_t - \beta_4 * PRINNEITH * z_t$$
$$- \beta_5 * NEITHNEITH * z_t + m_{t+1}$$

where $P_t$ is the log transaction price and $z_t = P_t - Q_t$, where $Q_t$ is the log quote midpoint. All average coefficient estimates are significantly different from zero at the 5% confidence level or greater. "# Signif." indicates the number of individual firms among the 35 having coefficient estimates significantly different from zero at the 5% confidence level or greater, and of the expected sign. 's.e.' refers to the standard error, and 't-stat' refers to the t-value.

**Table 3.13.** **Conditional Estimates of the Costs of Adverse Information and Order Processing**

Panel A: Conditional estimates of adverse information

| Indicator Variable | | | | | Coefficient | | |
|---|---|---|---|---|---|---|---|
| Large | External | Principal-principal | Principal-neither | Neither-neither | Estimate | # Signif. | For: |
| 0 | 0 | 0 | 0 | 1 | -0.011 | 9 | Small internal n/n trades |
| 0 | 0 | 0 | 1 | 0 | 0.013 | 8 | Small internal p/n trades |
| 0 | 1 | 0 | 0 | 1 | 0.037** | 27 | Small external n/n trades |
| 0 | 0 | 0 | 0 | 0 | 0.050 | 24 | Small internal RT trades |
| 0 | 1 | 0 | 1 | 0 | 0.061** | 35 | Small external p/n trades |
| 1 | 0 | 0 | 0 | 1 | 0.070* | 31 | Large internal n/n trades |
| 0 | 0 | 1 | 0 | 0 | 0.085** | 32 | Small internal p/p trades |
| 1 | 0 | 0 | 1 | 0 | 0.094** | 34 | Large internal p/n trades |
| 0 | 1 | 0 | 0 | 0 | 0.098** | 35 | Small external RT trades |
| 1 | 1 | 0 | 0 | 1 | 0.118** | 35 | Large external n/n trades |
| 1 | 0 | 0 | 0 | 0 | 0.131** | 34 | Large internal RT trades |
| 0 | 1 | 1 | 0 | 0 | 0.133** | 34 | Small external p/p trades |
| 1 | 1 | 0 | 1 | 0 | 0.142** | 35 | Large external p/n trades |
| 1 | 0 | 1 | 0 | 0 | 0.166** | 35 | Large internal p/p trades |
| 1 | 1 | 0 | 0 | 0 | 0.179** | 35 | Large external RT trades |
| 1 | 1 | 1 | 0 | 0 | 0.214** | 35 | Large external p/p trades |

Panel B: Conditional estimates for order processing

| Indicator variable | | | | | Coefficient | | |
|---|---|---|---|---|---|---|---|
| Large | External | Principal-principal | Principal-neither | Neither-neither | Estimate | # Signif. | For: |
| 1 | 1 | 0 | 0 | 1 | 0.102* | 31 | Large external n/n trades |
| 1 | 1 | 0 | 1 | 0 | 0.135** | 34 | Large external p/n trades |
| 1 | 1 | 1 | 0 | 0 | 0.138** | 33 | Large external p/p trades |
| 0 | 1 | 0 | 0 | 1 | 0.154** | 35 | Small external n/n trades |
| 0 | 1 | 0 | 1 | 0 | 0.187** | 34 | Small external p/n trades |
| 0 | 1 | 1 | 0 | 0 | 0.190** | 33 | Small external p/p trades |
| 1 | 1 | 0 | 0 | 0 | 0.324** | 35 | Large external RT trades |
| 0 | 1 | 0 | 0 | 0 | 0.376** | 35 | Small external RT trades |
| 1 | 0 | 0 | 0 | 1 | 0.391** | 35 | Large internal n/n trades |
| 1 | 0 | 0 | 1 | 0 | 0.423** | 35 | Large internal p/n trades |
| 1 | 0 | 1 | 0 | 0 | 0.426** | 34 | Large internal p/p trades |
| 0 | 0 | 0 | 0 | 1 | 0.443** | 35 | Small internal n/n trades |
| 0 | 0 | 0 | 1 | 0 | 0.475** | 35 | Small internal p/n trades |
| 0 | 0 | 1 | 0 | 0 | 0.478** | 35 | Small internal p/p trades |
| 1 | 0 | 0 | 0 | 0 | 0.612** | 35 | Large internal RT trades |
| 0 | 0 | 0 | 0 | 0 | 0.664** | 35 | Small internal RT trades |

The above table presents conditional coefficient estimates, sorted from low to high, for adverse information costs (Panel A) and order processing costs (Panel B) for the models estimated in Tables 3.11 and 3.12 respectively. Column headings are indicator variables in the regression equations. Each indicator is equal to 1 or zero. The conditional estimate for adverse information costs for a large, external trade between parties neither RTs nor principals would be computed as: $(\lambda + \alpha_1 + \alpha_2 + \alpha_5)$, and similarly for other estimates. For order processing costs, the same conditional estimate would be: $-(\gamma + \beta_1 + \beta_2 + \beta_5)$. *# Signif* gives the number of TSE35 stocks for which the coefficient sum is significant at the 0.05 level. * and ** indicate significance at the 0.10 and 0.05 levels, respectively, based on average values of F-Statistics.

**Table 3.14.** Summed Conditional Estimates of the Costs of Adverse Information and Order Processing

| Estimate | Conditional on : |
|---|---|
| 0.191 | Small external n/n trades |
| 0.220 | Large external n/n trades |
| 0.247 | Small external p/n trades |
| 0.277 | Large external p/n trades |
| 0.322 | Small external p/p trades |
| 0.351 | Large external p/p trades |
| 0.432 | Small internal n/n trades |
| 0.461 | Large internal n/n trades |
| 0.473 | Small external RT trades |
| 0.488 | Small internal p/n trades |
| 0.503 | Large external RT trades |
| 0.517 | Large internal p/n trades |
| 0.563 | Small internal p/p trades |
| 0.592 | Large internal p/p trades |
| 0.714 | Small internal RT trades |
| 0.743 | Large internal RT trades |

This table presents the sum of adverse information costs and order processing costs presented in Table 3.13. The combined (or total) cost is sorted from low to high value. In the table, 'RT' refers to Registered trade, 'p' refers to a Principal trade, and 'n' refers to neither a RT or p trade.

# CHAPTER 4

# ROBUSTNESS OF SPREAD COMPONENT ESTIMATES TO BUNCHING METHOD AND DECIMALIZATION

## 4.1. INTRODUCTION

In microstructure models, transaction costs are reflected by the effective bid-ask spread, measured as the difference between the price paid (generally the market maker's ask price) and the price at which the transactor sells (generally the market maker's bid price). This effective bid-ask spread is compensation for the provision of liquidity by the market maker. It accounts for order processing costs, for the cost of maintaining inventory, and for a premium for the risk that the market maker may be transacting with an informed trader. These three costs, order processing, inventory, and adverse selection, are modeled frequently in the microstructure literature. Huang and Stoll (1997) and Madhavan, Richardson and Roomans (1997) use Generalized Method of Moments (GMM) and NYSE-AMEX data to estimate the components of their respective bid-ask spread models. These models are similar and come to similar conclusions. The effective spread mostly reflects order processing costs, followed by the cost of maintaining inventory, and the (small) cost of adverse selection.

In general, models of the decomposition of the spread are of two types; namely, trade indicator models in which trades are categorized as either buyer- or seller-initiated, and covariance-based models in which the covariance of either prices or quotes is used to infer cost components. An example of the former is the Madhavan, Richardson and Roomans (1997) model. Examples of the latter are the Roll (1984) and Stoll (1989) models. Both approaches share a common problem. Since trade prices tend to reverse due to both the presence of order processing costs as well as inventory holding costs, it is difficult to separate these two cost components. Huang and Stoll (1997) offer a method for separating all three cost components in an efficient manner, and for reconciling previous attempts at determining those costs.

Huang and Stoll (1997) estimate their model using different data sets. The first data set includes all transactions while the second clusters or groups transactions in a particular manner. This chapter examines the performance of their model for different clustered data sets, and examines the robustness of their model and clustering method for the period subsequent to a reduction in minimum price variation (tick size) on the Toronto Stock Exchange (TSE).

We come to three conclusions. First, clustering (or bunching or grouping) transactions has a significant impact on the measurement of the cost components of the spread in predicable ways. This basic result is not entirely new and is discussed briefly by Huang and Stoll (1997). We provide additional evidence that any clustering method has an impact on cost component measurement. Second, we provide an alternative data clustering method that uses additional information provided by our data relating to trader type. We show that model performance may improve even though a necessary condition of model performance – adequate reversal of trade flow – may be less apparent. Third, we show that the Huang and Stoll (1997) model (and by inference, any simple trade indicator model) fails to provide adequate measurement of the cost components of the spread in the period following a lowering of the minimum tick size because of a fundamental change in market maker behavior. This change in behavior, which involves the market maker providing less depth at the bid and ask to offset a tighter quoted bid-ask spread, is not captured in these trade indicator models.

The remainder of the chapter is organized as follows: Section 4.2 provides an overview of the Huang and Stoll (1997) model. Section 4.3 describes the data. Section 4.4 provides the estimates of the model for different clustering methods. Section 4.5 applies the model to (non)clustered data in the post-decimalization period as a further test of robustness. Section 4.6 concludes.

## 4.2. THE MODEL FOR DECOMPOSING THE COMPONENTS OF THE BID-ASK SPREAD

Modern attempts to decompose the bid-ask spread start with Stoll (1989). More recently, Huang and Stoll (1997) estimate a general, structural model using Generalized Method of Moments (GMM) on NYSE-Amex trade data. Their model permits the direct measurement of two components of the bid-ask spread: the cost of information asymmetry, and the cost of inventory adjustment. The third cost component, order processing, is found by subtracting the other two components from total cost, which is assumed to be 100% of the effective spread.

The Huang and Stoll (1997) model for a three component decomposition consists of a two-equation system. The first equation captures the autocorrelation hypothesized to be present in trade flow and allows for the identification of the inventory cost component. Specifically:

$$E(Q_{t-1} \mid Q_{t-2}) = (1-2\pi)Q_{t-2} \tag{1}$$

where $\pi$ is the probability of a reversal in trade flow and Q is a trade initiation variable taking on the value of $+1$ when the transaction is buyer-initiated and $-1$ when it is seller-initiated. A trade is buyer-initiated (seller-initiated) if the transaction occurs above (below) the mid-point of the bid-ask spread. Trades at the mid-point are coded as zero and thus are ignored in the estimation.[1]

Equation (1) assumes an autocorrelation structure in trade flow that attempts to capture the fact that the liquidity provider (generally the market maker) adjusts inventory relative to previous trades. For example, a buy at the market maker's ask leads the market maker to raise both the bid and ask quotes in order to reestablish inventory. The higher bid increases the probability of a subsequent market sell at the bid (i.e. seller-initiated). The higher ask price lowers the probability of another market buy. Therefore, seller-initiated trades are assumed, in terms of inventory adjustment, to follow buyer-initiated trades. This generates a negative serial covariance in the sequence of either the mid-point of the bid-ask spread or in transaction prices. In Equation (1), the probability of a reversal in trade flow must be greater than 0.50 for the autocorrelation coefficient to be negative (as required).

---

[1] The implicit assumption, discussed by Madhavan et al. (1997), is that trades at the mid-point are both buyer- and seller-initiated. A pre-negotiated cross is an example of this type of trade.

The second equation models the process of the change in the mid-point of the bid-ask spread, $m$, conditional upon the past realizations of the trade initiation variable. Specifically:

$$\Delta m_t = (\alpha + \beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1-2\pi)\frac{S_{t-2}}{2}Q_{t-2} + \varepsilon_t, \qquad (2)$$

where $\alpha$ is the cost of adverse selection, $\beta$ is the cost of inventory management and $(1-\alpha-\beta)$ is the cost of order processing. $S$ is the observed quoted spread. All costs are estimated as percentages of the bid-ask spread. Conceptually, the quoted spread is constant. In estimation, using (variable) quote data increases the degrees of freedom since the spread need not be estimated.

The induced negative serial covariance in the inventory adjustment process modeled in Equation (1) represents the major contribution of the Huang and Stoll (1997) model. Specifically it allows for the separation of the adverse selection component, $\alpha$, and the inventory cost component, $\beta$. A non-negative autocorrelation coefficient, $(1-2\pi)$, creates serious estimation problems for this model. To illustrate, for the case where $\pi = 0.5$, the second term of Equation (2) collapses to zero. If the mid-point of the bid-ask spread remains constant for a sequence of either buyer-initiated or seller-initiated trades, $\alpha$ has to be negative if $\beta$ is positive. It is straightforward to show that the same result is obtained if $\pi$ is less than 0.5. This is a likely result if little change occurs in the mid-point of the bid-ask spread. Therefore, a necessary condition for the successful estimation of this model is that buyer- and seller-initiations (and, therefore, trades) must, on average, reverse.

## 4.3. THE DATA

We estimate the Huang and Stoll (1997) two-equation model given by Equations (1) and (2) using data from the TSE equity history (EH) data set for the 127 trading days from 2 January 1996 through 28 June 1996. For shares trading above $5.00, the minimum tick decreased from $0.125 (1/8) to $0.05 on the TSE on 15 April 1996. Our data consists of 73 trading days prior to the change to decimalization, and 54 days subsequent to the change. In addition, we examine only large TSE-traded firms – specifically, the firms comprising the TSE 35 index.

In general, all trades are used in estimation. Quotes are available and are aligned with trades following Lee and Ready (1993). Thus, quotes must occur at least five seconds prior to the transaction in order to be linked to that transaction. Since the TSE EH data base records transactions and quotes in six-second intervals, alignment of quotes with trades generally requires quotes and trades not to be contemporaneous.

In addition to price and quote data, the TSE tape also provides an indication of buyer and seller trader type. Specifically, the tape identifies trades from the designated market maker, principals and trades by neither market makers nor principals (the so-called public investors).

Table 4.1 provides an indication of the nature of trading for firms examined in this study prior to the reduction in tick size. Average daily trading volumes differ substantially from firm to firm as does the mean number of trades. Prices during the six months under consideration range from about $11 to $64 per share on average. Interestingly, the number of average daily trades for our sample firms (219) is more than twice as large as that (95) for the sample used by Madhavan, Richardson and Roomans (1997) and similar to that used by Huang and Stoll (1997). The volume per transaction is almost identical to the 22.7 board lot transactions per hour reported by Madhavan, Richardson and Roomans (1997) at 23.4. Average spreads are much smaller than those reported by Madhavan, Richardson and Roomans (1997) (0.146 versus 0.228) but similar to those reported by Huang and Stoll (1997) (0.146 versus 0.174). On balance, the TSE data are similar in many respects to the NYSE data in general, and more similar to the largest of the NYSE stocks that are examined in the Huang and Stoll (1997) study.

Table 4.2 presents the same data as in Table 4.1, except that Table 4.2 computes the data using the Huang and Stoll (1997) methodology for "bunching" transactions. Their bunching method entails considering all consecutive transactions on the same side of the market, at the same price, based on the same quote, as parts of a broken-up order or program. Thus, the Huang and Stoll bunching method converts transactions into reconstituted orders. Table 4.2 shows that bunching has an impact on the reported number of transactions and on quoted spreads. Quoted spreads are 90 basis points higher for bunched data on average, and the average number of transactions per day falls to 50 from 208. Thus, the bunching methodology assumes *de facto* that each program consists of approximately 4 separate transactions on average.

## 4.4. EMPIRICAL RESULTS

### 4.4.1 Basic Results for Non-bunched and Huang and Stoll-Bunched Transactions Data

Table 4.3 presents GMM results for the estimations of Equations (1) and (2) using non-bunched data for the pre-decimalization period. Estimates of $\pi$ are low in general and significant. Even the highest estimate of $\pi$, 0.17, is below 0.5. As a result, given the nature of the model, we expect the coefficient estimates of $\alpha$ to be poor. In fact, they are generally opposite in sign to expectations and statistically significant. These results are entirely consistent with those obtained by Huang and Stoll (1997) for NYSE-Amex data.

Recall that $\pi$ measures the probability of a reversal in trade flow. That is, $\pi$ is the estimated probability that buyer- (seller-) initiated trades will be followed by seller- (buyer-) initiated trades. The last column of Table 4.3 presents a similar measure determined independently of the model, the transaction price ex post reversal rate. This independent measure is based only upon the sequence of tick changes in transaction prices. The transaction price only reverses 13.3% of the time on average. Moreover, we find that most transactions (82.9%) are at the same price as the preceding transaction.

Table 4.4 provides the same types of results as Table 4.3 but for data bunched according to the Huang and Stoll (1997) method. Specifically, all transactions in which neither the price nor quote has changed are considered to be a single transaction. Huang and Stoll (1997) recognize that this bunching method is unlikely to be a perfect proxy for the correct groupings of same-program transactions by traders.

There are at least two reasons why "true" transactions may be separated into smaller transactions. First, informed transactors, who wish to trade in larger volume, may camouflage their intentions by breaking up larger orders into smaller ones. However, it is unlikely that individual order components can be executed with such rapidity that they appear as sequential and contiguous transactions at the same price in the data set. In addition, such rapid order submission probably would lead to the same reaction on the part of the market maker to order flow as a single large transaction. The second reason why "true" transactions may be broken up is that a large order may be executed against several smaller orders at the same price. For example, a 1,000 share buy may require ten separate 100 share sells to be filled. All ten transactions may appear in a contiguous sequence and possibly at the same price, without any change in quotes.

If traders are not identified, it is not possible to separate sequences of trades by individual parties. A reasonable alternative is to assume, a priori, that all transactions at the same price and quote are from the same trader. Since this is unlikely to be the case, this bunching method will overstate true bunching, and its associated parameter estimates will be different than those under the true bunching method.

In general, bunched data coefficient estimates are far "superior" in comparison to those obtained with non-bunched data (Table 4.3). For the 35 firms studied herein, 20 of the $\alpha$ estimates are of the expected sign and significant at the 5% level. Seven firms have negative $\alpha$, but these coefficients are significant in only one case. Estimates of $\beta$ are significant, with one exception, and all are of the expected sign. Estimates of $\pi$ are much higher than in Table 4.3 at 0.88 on average. The ex post reversal rate for the bunched data of 55.6%, is also much higher than that for the non-bunched data (13.3%).

These results are similar to those obtained by Huang and Stoll (1997) for their bunched data. Specifically in Huang and Stoll (1997), two of twenty firms are of the wrong sign but with large standard errors. All of their other results are entirely consistent with ours. Interestingly, the coefficient estimates themselves also are very similar. In Huang and Stoll (1997), $\alpha$ is 0.096 on average, while our data yields 0.080. Similarly, in Huang and Stoll (1997), $\beta$ is 0.287, while in this study it is 0.326. Huang and Stoll (1997) estimate $\pi$ at 0.868 while our estimate is 0.885. Given the different data sets, these differences (which probably are not statistically significant) are remarkably small and provide a basic validation of the Huang and Stoll (1997) model.

As we note in the previous section, the Huang and Stoll (1997) model requires $\pi$ to be greater than 0.50. Therefore, it is important to ask if the bunching technique employed by Huang and Stoll (1997) is simply a mechanism by which $\pi$ is increased. In other words, does bunching make the model work by increasing the rate of trade reversal, or does the model work because the data are bunched properly? We attempt to provide some initial answers to this question in the remaining subsections of this section of the chapter.

### 4.4.2. Price-only Bunched Data

In order to test the relationship between trade reversal and model validity, we rebunch the data in a way that clearly increases the rate of trade reversal while, at the same time, ignoring what happens to quotes. Specifically, we bunch the data only on the basis of price change. Thus, all sequential transactions occurring at the same price are considered as being part of the same transaction. This bunching method should yield poorer results for estimations of the Huang and Stoll (1997) model since the link between market maker activity and trade sequencing is broken. Table 4.5 provides coefficient estimates based on this "price only" bunching method.

In Table 4.5, we note that coefficient estimates of $\alpha$ are inferior to those of the Huang and Stoll (1997)-bunching method. Standard errors are generally larger with $\alpha$ estimates for only 13 of the 35 firms significant at the 5% level, although fewer firms are of the wrong sign. Coefficient estimates are similar for $\alpha$ but lower for $\beta$ compared to those obtained using the Huang and Stoll-bunching method. Interestingly, the estimate of the trade reversal parameter, $\pi$, at 0.836 is lower than in Table 4.4 (0.885), and the difference is statistically significant. However, the ex post reversal rate is higher for price-only bunching at 74.2% compared to Huang and Stoll bunching at 55.6%.

To further examine the relationship between trade reversal and model performance, we regress the firm ex post reversal rates against the t-statistics obtained from an application of the model. For Huang and Stoll (1997) bunched data, the coefficient estimate has a p-value of zero and an R-square of 0.75, suggesting that the reversal rate alone does a good job of forecasting model performance. However, using price only bunching, the coefficient p-value is 0.095 and the R-square falls to 0.05. We conclude from these results that while trade reversal is a necessary condition for the satisfactory performance of the Huang and Stoll (1997) model, it is not a sufficient condition. However, the regressions results also suggest that the Huang and Stoll (1997) bunching method may "fit" the model better simply because of its reversal rate enhancement – a fortuitous bunching method.

### 4.4.3 Trader-Identified Bunching

As indicated above, the TSE data provide an indication of trader type. Specifically, traders are identified as market maker, principal or neither market maker nor principal. In the previous section, we show that the Huang and Stoll (1997) bunching method may work because the transaction sequence created by the method may correspond better to the assumptions of the model while not necessarily bunching transactions appropriately. In this section, we use the trader type data to improve upon the Huang and Stoll (1997) bunching method.

The Huang and Stoll (1997) bunching method implicitly assumes that price change is an indication of a change of trader. That is, traders submit orders for execution (either large or small), and those orders are broken up (either by the trader or by the market maker) into smaller transaction lots. This model may be valid for larger limit orders that are to be filled at a stipulated price, but a larger market order is subject to changes in execution price. This is an action taken by the market maker to protect himself from the possibility that larger orders come from informed traders. Programs are unlikely to be captured by looking for sequences of unchanged prices since a larger program, according to Chan and Lakonishok (1993, 1995), is often broken up and executed in smaller lots at different times, and, often, at different prices. Even in the case of limit orders, Huang and Stoll (1997) cannot know if larger transactions have been broken up, or small trades have simply been grouped together due to, for example, brokerage firm internalization. Their method simply assumes that all trades at the same price and quote have been submitted by the same trader.

We can use the TSE data to solve a part of the problem. We are still unable to identify program trades, and we do not know the identity of traders. However, we can reduce some of the error by bunching trades on the basis of trader type and price. That is, we can bunch all consecutive transactions together if they are from the same trader type on the same side of the market and at the same price without an intervening change in quotes. In other words, we build on the Huang and Stoll (1997) bunching method by adding the requirement that the trader type not change in order for transactions to be bunched.

The results of this trader-identified bunching are presented in Table 4.6. First, we note that all $\alpha$ estimates are of the expected sign. For the 35 firms studied herein, 22 of the $\alpha$ estimates are significant at the 5% level. However, only 15 of the $\beta$ estimates are both significant and of the right sign. In addition, five of the $\beta$ estimates are negative, although only one of them is significant. The coefficient estimates are different on average from the Huang and Stoll (1997)-bunched results. Specifically, the mean $\alpha$ is much higher (0.188) in comparison to our Huang and Stoll (1997)-bunched results (0.080). In addition, the mean $\beta$ is much lower (0.111 versus 0.326).[2] The estimate of $\pi$ also is lower at 0.693 versus 0.884, as is the ex post reversal rate (0.472 versus 0.556). On balance, there appears to be a tradeoff between bunching method and the relative values of $\alpha$ and $\beta$. The trader-identified bunching method allows more trades to remain in the data set since trader types change even when prices and quotes do not. Thus, both trade reversal measures fall. However, the additional bunching on the basis of trader type, by doing a better job of gathering together those trades that may be information-based, increases the values and significance of $\alpha$. However, the reduction in reversals leads to a reduction in the $\beta$ estimates. As with the Huang and Stoll (1997) bunching method, reversal rates do a good job of forecasting model performance but not as good. While the p-value of the coefficient of a regression of the ex post reversal rate against the t-statistic is zero, the R-square is lower than for the Huang and Stoll bunching method at 0.61.

### 4.4.4. Other Bunching Methods

The use of other bunching methods results in a different tradeoff between relative estimates of $\alpha$ and $\beta$ obtained using the various bunching methods. For example, bunching on the basis of trader type alone (that is, without concern as to whether there is a change in price or quotes) should mask changes in price and quotes that would be part of information-based trading. Thus, we would expect $\alpha$ estimates to fall, and $\beta$

[2] These differences are statistically significant.

estimates to rise (if the reversal rate increases). In fact, although the results are not statistically significant, some evidence exists that this type of bunching produces this expected result.

#### 4.4.5. Conclusions on the Impact of Bunching Method on Model Performance

The Huang and Stoll (1997) bunching method leads to "successful" parameter estimation of the Huang and Stoll (1997) model. While high reversal rates are necessary for model performance, they are not sufficient. In addition, alternate bunching techniques provide equal (or superior) performance, but lead to statistically significant differences in the parameter estimates. An important question is whether there is a single correct bunching method. Unfortunately, the answer appears to be no.

Consider the problem of estimation with a perfect database, one that provides complete trader identification, trader intentions, trade order flow and execution, and execution location. We know that some bunching is required in order to obtain a sufficient reversal rate. The obviously correct bunching method is to bunch on the basis of transactor identification (and also trader trade execution strategy). This would lead to an isolation of information-based trading and provide correct estimates of $\alpha$. However, by bunching any transactions, we are reducing the (unitary) increase in transaction volumes. As a result, we will not be able to obtain correct estimates of the cost of managing inventory, $\beta$. Of course, we could find that, since there are many different individual transactors, the reversal rate will be insufficient to allow the model to work at all.[3]

## 4.5. EMPIRICAL RESULTS FOR THE POST-TICK SIZE REDUCTION PERIOD

On 15 April 1996, the TSE converted to decimal trading, such that the minimum tick size for shares above $5.00 fell to $0.05. We use the 54 trading days subsequent to the conversion to gain additional insight into the robustness of the performance of the Huang and Stoll (1997) model.

Other studies examine the impact of tick size reduction on spreads and the cost components of the spread. Chung, Kryzanowski and Zhang (1996), Porter and Weaver (1997) and Bacidore (1997) all find significant reductions in quoted and effective spreads in the post-decimalization period. Bacidore (1997) tests for a change in adverse selection and finds no significant change from the pre- to the post-decimalization period.

Table 4.7 provides the results of an application of the Huang and Stoll (1997) model to data bunched according to the Huang and Stoll (1997) method. First, we note that estimates of $\alpha$ are of the wrong sign, but not statistically significant, in thirteen cases. Furthermore, of the 9 significant $\alpha$ estimates, two are greater than one, and one of these is above three. The average $\alpha$ estimate is greater than that in the pre-decimalization period, but not if we eliminate the estimate above three. Thus, we do not find a significant change in $\alpha$ pre- versus post-decimalization.

[3] The use of simulations could potentially shed some light on this issue.

Estimates of $\beta$ in Table 4.7 are better with all but two estimates of the expected sign and significant. In one case, the $\beta$ estimate is an insignificant –2.89 (the same firm with the insignificant $\alpha$ estimate above three). On average, it appears that $\beta$ has declined with respect to the pre-decimalization period, but eliminating the unrealistic negative value, shows an increase in $\beta$.

We further note from Table 4.7 that the estimate of $\pi$ as well as the ex post trade reversal rate, have fallen in the post-decimalization period, but not to the extent that would lead one to expect, a priori, poor model performance.

Table 4.8 provides the same data as Table 4.7 but for trader-identified bunching. The results are poorer than those obtained with the Huang and Stoll (1997) bunching method, but are generally consistent. In addition, estimates of $\pi$ are lower although the average value of the ex post reversal rate is unchanged. From Tables 4.7 and 4.8, we conclude that poor model performance in the post-decimalization period is not related to the bunching method.

What then is the cause of poor model performance in the post-decimalization period? Table 4.9 provides an indication that the mechanism by which market makers adjust to the presence of informed trading in the market may be different subsequent to a reduction in minimum price variation. Specifically, two assertions made by Anshuman and Kalay (1993) are relevant. First, they suggest that adverse selection costs should increase with a smaller tick size since informed traders are able to camouflage their trades more easily. Second, they argue that the market maker's reaction to the increase in adverse selection is a decrease in quoted market depth. The first assertion is based on the notion that an informed trader is impatient to make use of his information and therefore transacts at the market at either the quoted bid or ask. Market makers may provide uninformed traders with prices better than the quoted bid or ask since the risk to the market maker is lower. However, with a lower minimum price variation, it becomes easier for an informed trader to hide since more uninformed trades take place at either the bid or ask given the smaller spread. Since the market maker has difficulty identifying informed trade by the sequence of transactions, he may decide to mitigate the problem by offering to transact fewer shares at any quoted inside bid and ask.

Table 4.9 provides a clear indication that this may be the case. We note that the average quoted bid depth declines by 50%, and average quoted ask depth declines by 52% (as in Chung, Kryzanowski and Zhang, (1996)). Although medians suggest that a few firms with large quotes are responsible for high average bid and ask depths, the changes in median quote depths are similar. Therefore, it appears that market maker quoting behavior is very different in the post-decimalization period and may be due to the difficulty of identifying the possible existence of informed trade. Huang and Stoll (1997) do not model changes in depth behavior or significant change in depth provision behavior. This omission accounts for the poor performance of their model, regardless of bunching technique, in the post-decimalization period.

## 4.6. CONCLUSION

This chapter applies the Huang and Stoll (1997) model and different variations of their bunching technique to a data set consisting of the 35 stocks making up the TSE 35 index for the period from January through June 1996. It includes both the period prior to and subsequent to a reduction in tick size on the TSE on 15 April 1996. It draws three major conclusions. First, we show, consistent with Huang and Stoll (1997) that clustering (or bunching) transactions has a significant impact on the measurement of the cost components of the spread. Specifically, any bunching method tends to emphasize the adverse selection cost component and to reduce the inventory holding cost component. Second, we provide an alternative data clustering method that uses additional information provided by our data relating to trader identification. We show that model performance may improve even though a necessary condition of model performance – reversal of trade flow – may be less apparent. Third, we show that the Huang and Stoll (1997) model, and by inference any simple trade indicator model, fails to provide adequate measurement of the cost components of the spread in the period following a change to minimum tick size because of a fundamental change in market maker depth provision behavior. This change in behavior is not captured in the Huang and Stoll model.

These conclusions suggest that the use of this type of trade indicator variable model will not provide consistent measurement of cost components unless data sets provide complete identification of transactors and their trade strategies – an unlikely development. Furthermore, although consistency may be obtained in this case, it is not clear that the resultant measurements of the cost components will be accurate. Thus, assessing market quality using models of this type should be done with caution. Additionally, these types of models are not likely to be robust to changes in minimum price variations that cause a systematic lowering of depth provision by market makers.

The results of this chapter provide a suggestion for further study. Trade indicator variable models, in order to provide consistent and accurate decompositions of the bid-ask spread, should be developed incorporating the empirical lack of trade reversal in transactions data without the use of clustering rules. In addition, models must incorporate alternative market maker strategies for dealing with the existence of adverse selection, such as changing the number of shares offered at the inside bid and ask quotes.

**Table 4.1.** Average Daily Volumes, Trades, Prices and Spreads for the Pre-Decimalization Period For Unbunched Transactions

| Firm | Volume | Trades | Price | Spread |
|---|---|---|---|---|
| 1 | 404,470 | 177 | 20.60 | 0.139 |
| 2 | 1,180,444 | 391 | 40.98 | 0.144 |
| 3 | 717,057 | 273 | 43.09 | 0.151 |
| 4 | 579,449 | 199 | 20.10 | 0.116 |
| 5 | 606,351 | 647 | 48.80 | 0.137 |
| 6 | 641,158 | 407 | 31.92 | 0.131 |
| 7 | 625,991 | 281 | 30.94 | 0.132 |
| 8 | 615,557 | 275 | 41.43 | 0.137 |
| 9 | 763,277 | 277 | 26.63 | 0.138 |
| 10 | 172,723 | 79 | 15.33 | 0.141 |
| 11 | 156,720 | 62 | 44.78 | 0.185 |
| 12 | 234,278 | 141 | 20.11 | 0.144 |
| 13 | 84,383 | 131 | 51.39 | 0.173 |
| 14 | 242,242 | 111 | 26.11 | 0.150 |
| 15 | 676,717 | 180 | 14.06 | 0.136 |
| 16 | 298,109 | 191 | 17.69 | 0.135 |
| 17 | 164,289 | 162 | 26.14 | 0.143 |
| 18 | 116,963 | 51 | 57.82 | 0.228 |
| 19 | 556,947 | 178 | 45.30 | 0.165 |
| 20 | 272,591 | 116 | 11.35 | 0.131 |
| 21 | 479,262 | 214 | 27.87 | 0.139 |
| 22 | 338,353 | 214 | 63.82 | 0.176 |
| 23 | 1,151,685 | 320 | 12.28 | 0.128 |
| 24 | 955,781 | 308 | 38.36 | 0.149 |
| 25 | 408,849 | 105 | 14.34 | 0.139 |
| 26 | 280,864 | 95 | 35.09 | 0.151 |
| 27 | 934,992 | 323 | 31.98 | 0.132 |
| 28 | 208,891 | 228 | 14.77 | 0.130 |
| 29 | 722,329 | 351 | 24.19 | 0.130 |
| 30 | 266,218 | 68 | 28.35 | 0.166 |
| 31 | 486,735 | 124 | 30.23 | 0.158 |
| 32 | 390,396 | 112 | 19.98 | 0.119 |
| 33 | 365,869 | 343 | 19.26 | 0.129 |
| 34 | 1,473,268 | 358 | 12.95 | 0.135 |
| 35 | 365,672 | 159 | 47.45 | 0.171 |
| | | | | |
| Mean | 512,539 | 219 | 30.16 | 0.146 |
| Median | 408,849 | 191 | 27.87 | 0.139 |
| High | 1,473,268 | 647 | 63.82 | 0.228 |
| Low | 84,383 | 51 | 11.35 | 0.116 |

Data is for TSE 35 firms for 2 January 1996 through 12 April 1996, and includes all transactions during that period. Share prices are not volume-weighted. All data are computed based on averages of the daily transactions on a firm by firm basis, and then averaged over the 73 trading days during that period.

**Table 4.2.** Average Daily Volumes, Trades, Prices and Spreads for the Pre-decimalization Period Using Huang and Stoll (1997)-Bunched Transactions

| Firm | Volume | Trades | Price | Spread |
|---|---|---|---|---|
| 1 | 404,470 | 37 | 20.56 | 0.150 |
| 2 | 1,180,444 | 93 | 41.00 | 0.151 |
| 3 | 717,057 | 84 | 43.15 | 0.161 |
| 4 | 579,449 | 40 | 20.05 | 0.123 |
| 5 | 606,351 | 139 | 48.71 | 0.142 |
| 6 | 641,158 | 78 | 31.95 | 0.135 |
| 7 | 625,991 | 61 | 30.93 | 0.138 |
| 8 | 615,557 | 68 | 41.41 | 0.142 |
| 9 | 763,277 | 53 | 26.62 | 0.148 |
| 10 | 172,723 | 17 | 15.31 | 0.154 |
| 11 | 156,720 | 23 | 44.77 | 0.195 |
| 12 | 234,278 | 32 | 20.09 | 0.157 |
| 13 | 84,383 | 41 | 51.36 | 0.177 |
| 14 | 242,242 | 30 | 26.16 | 0.160 |
| 15 | 676,717 | 33 | 14.06 | 0.144 |
| 16 | 298,109 | 34 | 17.65 | 0.144 |
| 17 | 164,289 | 37 | 26.20 | 0.153 |
| 18 | 116,963 | 23 | 58.31 | 0.243 |
| 19 | 556,947 | 63 | 45.44 | 0.177 |
| 20 | 272,591 | 17 | 11.35 | 0.138 |
| 21 | 479,262 | 49 | 27.83 | 0.148 |
| 22 | 338,353 | 79 | 63.91 | 0.186 |
| 23 | 1,151,685 | 50 | 12.28 | 0.132 |
| 24 | 955,781 | 90 | 38.35 | 0.158 |
| 25 | 408,849 | 23 | 14.34 | 0.151 |
| 26 | 280,864 | 31 | 35.01 | 0.165 |
| 27 | 934,992 | 63 | 31.98 | 0.137 |
| 28 | 208,891 | 37 | 14.77 | 0.136 |
| 29 | 722,329 | 61 | 24.20 | 0.135 |
| 30 | 266,218 | 23 | 28.31 | 0.179 |
| 31 | 486,735 | 34 | 30.20 | 0.169 |
| 32 | 390,396 | 28 | 20.02 | 0.125 |
| 33 | 365,869 | 51 | 19.26 | 0.134 |
| 34 | 1,473,268 | 57 | 12.97 | 0.144 |
| 35 | 365,672 | 60 | 47.54 | 0.181 |
| | | | | |
| Mean | 512,539 | 50 | 30.17 | 0.155 |
| Median | 408,849 | 41 | 27.83 | 0.150 |
| High | 1,473,268 | 139 | 63.91 | 0.243 |
| Low | 84,383 | 17 | 11.35 | 0.123 |

Data is for TSE 35 firms for 2 January 1996 through 12 April 1996, and bunches together all transactions for which consecutive prices are the same on the same side of the market with no change in quotes. Share prices are not volume-weighted. All data are computed based on averages of the daily transactions on a firm by firm basis, and then averaged over all 73 trading days during the period.

**Table 4.3.** GMM Parameter Estimates of Spread Components for the Pre-decimalization Period for Non-bunched trades

| Firm | Estimates of: $\alpha$ | $\beta$ | $\pi$ | Ex Post Reversal Rate |
|---|---|---|---|---|
| 1 | -0.01882*** | 0.06622*** | 0.07762*** | 0.103 |
| 2 | -0.02257*** | 0.08779*** | 0.08857*** | 0.102 |
| 3 | -0.03117*** | 0.10478*** | 0.12478*** | 0.127 |
| 4 | -0.04079*** | 0.09400*** | 0.09727*** | 0.144 |
| 5 | -0.01750*** | 0.03305*** | 0.09431*** | 0.166 |
| 6 | -0.01845*** | 0.04049*** | 0.08993*** | 0.141 |
| 7 | -0.02297*** | 0.04853*** | 0.09552*** | 0.151 |
| 8 | -0.03022*** | 0.05944*** | 0.11428*** | 0.161 |
| 9 | -0.02010*** | 0.06093*** | 0.07090*** | 0.106 |
| 10 | -0.03886*** | 0.06730*** | 0.09864*** | 0.111 |
| 11 | -0.05835*** | 0.13413*** | 0.17406*** | 0.170 |
| 12 | -0.01023 | 0.05557*** | 0.08656*** | 0.128 |
| 13 | -0.03462*** | 0.07734*** | 0.15110*** | 0.169 |
| 14 | -0.02875*** | 0.08383*** | 0.11262*** | 0.133 |
| 15 | -0.00871* | 0.04053*** | 0.06677*** | 0.111 |
| 16 | -0.02497*** | 0.04452*** | 0.07927*** | 0.114 |
| 17 | -0.01017 | 0.05195*** | 0.10367*** | 0.135 |
| 18 | -0.07312*** | 0.33039*** | 0.17376*** | 0.183 |
| 19 | -0.02633** | 0.13948*** | 0.14314*** | 0.120 |
| 20 | 0.00371 | 0.02868*** | 0.05907*** | 0.097 |
| 21 | -0.01621*** | 0.05743*** | 0.10247*** | 0.137 |
| 22 | -0.04281*** | 0.13018*** | 0.15807*** | 0.153 |
| 23 | -0.00987*** | 0.02572*** | 0.05465*** | 0.122 |
| 24 | -0.05449*** | 0.13270*** | 0.11870*** | 0.115 |
| 25 | -0.00922 | 0.05331*** | 0.07687*** | 0.114 |
| 26 | -0.04015*** | 0.12143*** | 0.14423*** | 0.146 |
| 27 | -0.01645*** | 0.04763*** | 0.07987*** | 0.127 |
| 28 | -0.00551** | 0.01564*** | 0.06385*** | 0.131 |
| 29 | -0.01296*** | 0.02994*** | 0.06146*** | 0.123 |
| 30 | -0.03078 | 0.12286*** | 0.14781*** | 0.127 |
| 31 | -0.02716*** | 0.07758*** | 0.11712*** | 0.140 |
| 32 | -0.04213*** | 0.08527*** | 0.14369*** | 0.189 |
| 33 | -0.01164*** | 0.02038*** | 0.07138*** | 0.121 |
| 34 | -0.00975** | 0.05172*** | 0.04901*** | 0.078 |
| 35 | -0.05545*** | 0.12621*** | 0.16873*** | 0.163 |
| | | | | |
| Mean | -0.02622*** | 0.07848** | 0.10456** | 0.133*** |
| Median | -0.02297 | 0.06093 | 0.09727 | 0.128 |
| High | 0.00371 | 0.33039 | 0.17406 | 0.189 |
| Low | -0.07312 | 0.01564 | 0.04901 | 0.078 |

GMM parameter estimates are based on Equations (1) and (2) estimated jointly for 73 trading days from 2 January 1996 through 12 April 1996 (the pre-decimalization period). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively Ex post reversal rate represents the percentage of reversals in transaction prices based on tick changes. All trades are used in the estimates.

**Table 4.4.** GMM Parameter Estimates of Spread Components for the Pre-decimalization Period for Huang and Stoll (1997)-bunched Transactions

| Firm | Estimates of: $\alpha$ | $\beta$ | $\pi$ | Ex Post Reversal Rate |
|---|---|---|---|---|
| 1 | -0.01293 | 0.48276*** | 0.89367*** | 0.469 |
| 2 | -0.00824 | 0.56777*** | 0.87092*** | 0.424 |
| 3 | -0.06756*** | 0.54886*** | 0.79962*** | 0.408 |
| 4 | 0.12629*** | 0.30674*** | 0.95903*** | 0.696 |
| 5 | 0.06987*** | 0.12274*** | 0.96807*** | 0.771 |
| 6 | 0.13094*** | 0.20221*** | 0.97518*** | 0.721 |
| 7 | 0.08881*** | 0.23105*** | 0.96520*** | 0.687 |
| 8 | 0.08696*** | 0.24669*** | 0.94406*** | 0.648 |
| 9 | 0.09141*** | 0.37936*** | 0.90611*** | 0.543 |
| 10 | 0.10962*** | 0.30863*** | 0.88543*** | 0.481 |
| 11 | -0.07161 | 0.40564*** | 0.77470*** | 0.433 |
| 12 | 0.11946*** | 0.26490*** | 0.89354*** | 0.542 |
| 13 | 0.03546 | 0.25086*** | 0.81048*** | 0.528 |
| 14 | 0.04917 | 0.40164*** | 0.87468*** | 0.494 |
| 15 | 0.13238*** | 0.27601*** | 0.94225*** | 0.593 |
| 16 | 0.06532** | 0.27155*** | 0.93936*** | 0.620 |
| 17 | 0.03142 | 0.28930*** | 0.89874*** | 0.580 |
| 18 | 0.56996*** | 0.13578 | 0.61298*** | 0.405 |
| 19 | -0.05838 | 0.61366*** | 0.73803*** | 0.335 |
| 20 | 0.13245*** | 0.26088*** | 0.97726*** | 0.593 |
| 21 | 0.08591*** | 0.29799*** | 0.92365*** | 0.595 |
| 22 | 0.05583 | 0.36000*** | 0.71778*** | 0.412 |
| 23 | 0.16836*** | 0.17441*** | 0.98863*** | 0.759 |
| 24 | -0.00371 | 0.56450*** | 0.80898*** | 0.389 |
| 25 | 0.08395** | 0.35796*** | 0.88114*** | 0.492 |
| 26 | 0.05689 | 0.42778*** | 0.82145*** | 0.437 |
| 27 | 0.09059*** | 0.30983*** | 0.96375*** | 0.640 |
| 28 | 0.15067*** | 0.08406*** | 0.98749*** | 0.780 |
| 29 | 0.11155*** | 0.24376*** | 0.97335*** | 0.697 |
| 30 | -0.05599 | 0.59820*** | 0.80071*** | 0.366 |
| 31 | 0.06701** | 0.30077*** | 0.83998*** | 0.501 |
| 32 | 0.13587*** | 0.16977*** | 0.96248*** | 0.730 |
| 33 | 0.14620*** | 0.10249*** | 0.98672*** | 0.789 |
| 34 | 0.04601* | 0.53656*** | 0.92951*** | 0.485 |
| 35 | 0.04651 | 0.32116*** | 0.74823*** | 0.431 |
| | | | | |
| Mean | 0.08018** | 0.32618** | 0.88466** | 0.556** |
| Median | 0.08395 | 0.30077 | 0.89874 | 0.542 |
| High | 0.56996 | 0.61366 | 0.98863 | 0.789 |
| Low | -0.07161 | 0.08406 | 0.61298 | 0.335 |

GMM parameter estimates are based on Equations (1) and (2) estimated jointly for 73 trading days from 2 January 1996 through 12 April 1996 (the pre-decimalization period). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively Ex post reversal rate represents the percentage of reversals in transaction prices based on tick changes. Trades are bunched using the Huang and Stoll (1997) method where all transactions at the same price without an intervening change in quotes are considered to be a single transaction.

**Table 4.5.** GMM Parameter Estimates of Spread Components for the Pre-decimalization Period Using Price-only Bunched Transactions

| Firm | Estimates of: $\alpha$ | $\beta$ | $\pi$ | Ex Post Reversal Rate |
|---|---|---|---|---|
| 1 | -0.04109 | 0.21400*** | 0.82555*** | 0.717 |
| 2 | 0.05987 | 0.25384*** | 0.74205*** | 0.695 |
| 3 | 0.12891** | 0.16477*** | 0.67942*** | 0.661 |
| 4 | 0.12492*** | 0.16419*** | 0.92668*** | 0.811 |
| 5 | 0.04145*** | 0.04959*** | 0.96053*** | 0.892 |
| 6 | 0.05008*** | 0.06242*** | 0.95055*** | 0.866 |
| 7 | 0.03781 | 0.07324*** | 0.93227*** | 0.840 |
| 8 | 0.02574 | 0.11545*** | 0.89449*** | 0.808 |
| 9 | 0.09096*** | 0.09135*** | 0.86955*** | 0.784 |
| 10 | 0.15790*** | 0.04899 | 0.82415*** | 0.660 |
| 11 | -0.05702 | 0.26698*** | 0.73104*** | 0.617 |
| 12 | 0.06045* | 0.14371*** | 0.87219*** | 0.748 |
| 13 | 0.07075* | 0.12741*** | 0.78739*** | 0.705 |
| 14 | 0.01360 | 0.22047*** | 0.82246*** | 0.706 |
| 15 | 0.05067 | 0.08396*** | 0.91699*** | 0.794 |
| 16 | 0.10539*** | 0.04137* | 0.90432*** | 0.793 |
| 17 | 0.01626 | 0.12946*** | 0.86150*** | 0.765 |
| 18 | 1.23260** | -0.58050 | 0.55310*** | 0.538 |
| 19 | 0.17270 | 0.30081*** | 0.62439*** | 0.561 |
| 20 | 0.03598 | 0.06068*** | 0.96112*** | 0.758 |
| 21 | 0.06200** | 0.12347*** | 0.88621*** | 0.792 |
| 22 | 0.07289 | 0.27697*** | 0.64503*** | 0.616 |
| 23 | 0.02773 | 0.04476*** | 0.97705*** | 0.880 |
| 24 | -0.05925 | 0.38583*** | 0.69638*** | 0.651 |
| 25 | 0.17292*** | 0.07465* | 0.84174*** | 0.702 |
| 26 | 0.10886 | 0.21968*** | 0.74421*** | 0.653 |
| 27 | 0.00383 | 0.09054*** | 0.92613*** | 0.834 |
| 28 | 0.02558 | 0.02196** | 0.98424*** | 0.876 |
| 29 | 0.03499 | 0.06912*** | 0.94150*** | 0.850 |
| 30 | 0.04981 | 0.40075*** | 0.70831*** | 0.564 |
| 31 | 0.05512 | 0.18071*** | 0.79264*** | 0.707 |
| 32 | 0.13591*** | 0.09494*** | 0.94912*** | 0.811 |
| 33 | 0.05696*** | 0.02320*** | 0.98209*** | 0.890 |
| 34 | 0.06063* | 0.13250*** | 0.86379*** | 0.776 |
| 35 | 0.18273*** | 0.12858* | 0.67675*** | 0.642 |
| | | | | |
| Mean | 0.09625* | 0.12285* | 0.83586** | 0.742** |
| Median | 0.05696 | 0.12347 | 0.86379 | 0.758 |
| High | 1.23260 | 0.40075 | 0.98424 | 0.892 |
| Low | -0.05925 | -0.58050 | 0.55310 | 0.538 |

GMM parameter estimates are based on Equations (1) and (2) estimated jointly for 73 trading days from 2 January 1996 through 12 April 1996 (the pre-decimalization period). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively Ex post reversal rate represents the percentage of reversals in transaction prices based on tick changes. Trades are bunched using a price-only method where all transactions at the same price, regardless of quote revisions, are considered to be a single transaction.

**Table 4.6.** GMM Parameter Estimates of Spread Components for the Pre-decimalization Period Using Trader-identified Bunched Transactions

| Firm | Estimates of: $\alpha$ | $\beta$ | $\pi$ | Ex Post Reversal Rate |
|---|---|---|---|---|
| 1 | 0.08649 | 0.29620*** | 0.68839*** | 0.387 |
| 2 | 0.12608** | 0.28198*** | 0.64897*** | 0.339 |
| 3 | 0.08348 | 0.33332*** | 0.66313*** | 0.357 |
| 4 | 0.17131*** | 0.11646*** | 0.74998*** | 0.568 |
| 5 | 0.16794*** | 0.02705** | 0.82382*** | 0.658 |
| 6 | 0.18237*** | -0.00104 | 0.73523*** | 0.600 |
| 7 | 0.13728*** | 0.07066* | 0.72798*** | 0.575 |
| 8 | 0.12115*** | 0.02719 | 0.73415*** | 0.548 |
| 9 | 0.19913*** | 0.15282*** | 0.69805*** | 0.443 |
| 10 | 0.63613** | -0.36086 | 0.58450*** | 0.411 |
| 11 | 0.12389 | 0.00335 | 0.63637*** | 0.401 |
| 12 | 0.27165*** | 0.08434 | 0.71899*** | 0.477 |
| 13 | 0.12841* | 0.10193 | 0.68017*** | 0.494 |
| 14 | 0.15415** | 0.19401*** | 0.73930*** | 0.442 |
| 15 | 0.24515*** | 0.07380 | 0.70411*** | 0.473 |
| 16 | 0.24132*** | 0.06964 | 0.66801*** | 0.501 |
| 17 | 0.20601*** | 0.13039* | 0.73360*** | 0.524 |
| 18 | 0.70408** | -0.13751 | 0.58304*** | 0.387 |
| 19 | 0.08190 | 0.39050*** | 0.65123*** | 0.301 |
| 20 | 0.09427 | 0.11436 | 0.62858*** | 0.431 |
| 21 | 0.08359* | 0.18764*** | 0.71614*** | 0.507 |
| 22 | 0.10230 | 0.22610** | 0.60566*** | 0.374 |
| 23 | 0.14578*** | 0.10947*** | 0.76263*** | 0.610 |
| 24 | 0.22001*** | 0.23767*** | 0.64602*** | 0.331 |
| 25 | 0.25272* | 0.10937 | 0.65212*** | 0.411 |
| 26 | 0.13380 | 0.21052 | 0.64262*** | 0.386 |
| 27 | 0.15448*** | 0.06081 | 0.71148*** | 0.527 |
| 28 | 0.25607*** | -0.09042*** | 0.75698*** | 0.655 |
| 29 | 0.14890*** | 0.00280 | 0.69578*** | 0.565 |
| 30 | 0.13925 | 0.20654 | 0.66064*** | 0.337 |
| 31 | 0.10799 | 0.10245 | 0.67817*** | 0.434 |
| 32 | 0.18421*** | 0.08104*** | 0.84137*** | 0.623 |
| 33 | 0.24056*** | -0.03478 | 0.78692*** | 0.664 |
| 34 | 0.08426 | 0.33761*** | 0.64413*** | 0.366 |
| 35 | 0.17150** | 0.17246** | 0.64627*** | 0.398 |
| | | | | |
| Mean | 0.18822** | 0.11108** | 0.69270** | 0.472** |
| Median | 0.15415 | 0.10937 | 0.68839 | 0.443 |
| High | 0.70408 | 0.39050 | 0.84137 | 0.664 |
| Low | 0.08190 | -0.36086 | 0.58304 | 0.301 |

GMM parameter estimates are based on Equations (1) and (2) estimated jointly for 73 trading days from 2 January 1996 through 12 April 1996 (the pre-decimalization period). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively Ex post reversal rate represents the percentage of reversals in transaction prices based on tick changes. Trades are bunched using the trader-identified bunching method where all sequential transactions by the same trader type, which are on the same side of the market at the same price and without an intervening quote revision, are considered to be a single transaction.

**Table 4.7.** GMM Parameter Estimates of the Spread Components for the Post-decimalization Period Using the Huang and Stoll (1997)-bunched Transactions

| Firm | Estimates of: $\alpha$ | $\beta$ | $\pi$ | Ex Post Reversal Rate |
|---|---|---|---|---|
| 1 | 0.00841 | 0.43670*** | 0.77619*** | 0.413 |
| 2 | -0.07213 | 0.50966*** | 0.67569*** | 0.379 |
| 3 | 0.02877 | 0.36498*** | 0.69907*** | 0.387 |
| 4 | 0.23608*** | 0.28592*** | 0.90183*** | 0.645 |
| 5 | 0.06635*** | 0.14479*** | 0.87751*** | 0.687 |
| 6 | 0.01662 | 0.26780*** | 0.91875*** | 0.644 |
| 7 | 0.02288 | 0.33489*** | 0.91607*** | 0.608 |
| 8 | 0.05479*** | 0.23416*** | 0.89114*** | 0.612 |
| 9 | 0.08356** | 0.29802*** | 0.76463*** | 0.453 |
| 10 | 0.06863 | 0.45024*** | 0.79999*** | 0.453 |
| 11 | -0.14101 | 0.25295*** | 0.73604*** | 0.381 |
| 12 | -0.25156 | 0.59024*** | 0.78150*** | 0.485 |
| 13 | 0.03039 | 0.20954*** | 0.70862*** | 0.510 |
| 14 | 0.08760 | 0.34772*** | 0.74564*** | 0.429 |
| 15 | -0.02185 | 0.39232*** | 0.79910*** | 0.452 |
| 16 | 0.04399 | 0.25338*** | 0.82495*** | 0.546 |
| 17 | -0.01742 | 0.23261*** | 0.79767*** | 0.556 |
| 18 | 3.50030 | -2.89360 | 0.51708*** | 0.401 |
| 19 | 1.09320*** | -0.20895 | 0.68991*** | 0.344 |
| 20 | -0.01647 | 0.28728*** | 0.89588*** | 0.573 |
| 21 | 0.01193 | 0.33727*** | 0.82682*** | 0.503 |
| 22 | -0.07565 | 0.37585*** | 0.60235*** | 0.406 |
| 23 | 0.05373** | 0.29327*** | 0.93924*** | 0.628 |
| 24 | -0.06649 | 0.54908*** | 0.68281*** | 0.383 |
| 25 | -0.00390 | 0.41074*** | 0.83426*** | 0.471 |
| 26 | -0.03980 | 0.36846*** | 0.70897*** | 0.374 |
| 27 | 0.04183* | 0.32174*** | 0.91348*** | 0.611 |
| 28 | 0.06544*** | 0.08910*** | 0.97981*** | 0.793 |
| 29 | 0.09409*** | 0.27623*** | 0.94644*** | 0.650 |
| 30 | -0.42719 | 0.75016*** | 0.68515*** | 0.347 |
| 31 | -0.02655 | 0.34778*** | 0.78183*** | 0.467 |
| 32 | 0.28490*** | 0.15314*** | 0.83477*** | 0.625 |
| 33 | 0.10316*** | 0.17964*** | 0.94748*** | 0.714 |
| 34 | -0.06800 | 0.54984*** | 0.82682*** | 0.432 |
| 35 | 0.07014 | 0.19165*** | 0.67317*** | 0.441 |
| | | | | |
| Mean | 0.13825 | 0.22813 | 0.79716** | 0.509** |
| Median | 0.02877 | 0.29802 | 0.79910 | 0.471 |
| High | 3.50030 | 0.75016 | 0.97981 | 0.793 |
| Low | -0.42719 | -2.89360 | 0.51708 | 0.344 |

GMM parameter estimates are based on Equations (1) and (2) estimated jointly for 73 trading days from 2 January 1996 through 12 April 1996 (the pre-decimalization period). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Ex post reversal rate represents the percentage of reversals in transaction prices based on tick changes. Trades are bunched using the Huang and Stoll (1997) method where all transactions at the same price without an intervening change in quotes are considered to be a single transaction.

**Table 4.8.** GMM Parameter Estimates of the Spread Components for the Post-decimalization Period Using Trader-identified Bunched Transactions

| Firm | Estimates of: | | | Ex Post Reversal Rate |
|---|---|---|---|---|
| | $\alpha$ | $\beta$ | $\pi$ | |
| 1 | 0.08951 | 0.24669*** | 0.69906*** | 0.394 |
| 2 | 0.16093 | 0.18982 | 0.59945*** | 0.356 |
| 3 | 0.16784* | 0.15403 | 0.65062*** | 0.372 |
| 4 | 0.39234*** | 0.06482 | 0.76169*** | 0.579 |
| 5 | 0.17143*** | 0.02420 | 0.75422*** | 0.609 |
| 6 | 0.05702 | 0.12103*** | 0.73157*** | 0.579 |
| 7 | 0.03849 | 0.09662* | 0.71515*** | 0.541 |
| 8 | 0.13185*** | 0.02082 | 0.74526*** | 0.557 |
| 9 | 0.03941 | 0.27636*** | 0.67078*** | 0.429 |
| 10 | 0.11830 | 0.39921*** | 0.69803*** | 0.420 |
| 11 | -0.89725*** | 1.03590*** | 0.62764*** | 0.371 |
| 12 | -0.08467 | 0.49855*** | 0.65337*** | 0.469 |
| 13 | 0.03737 | 0.13609 | 0.64330*** | 0.514 |
| 14 | 0.20025 | 0.19434 | 0.66298*** | 0.423 |
| 15 | -0.01401 | 0.32752*** | 0.66701*** | 0.402 |
| 16 | 0.14636 | 0.13818 | 0.66291*** | 0.495 |
| 17 | 0.10101 | 0.16384** | 0.71784*** | 0.560 |
| 18 | 3.70000 | -3.12610 | 0.51974*** | 0.388 |
| 19 | 0.25228 | 0.12905 | 0.61180*** | 0.327 |
| 20 | 0.09233 | 0.12476 | 0.68627*** | 0.495 |
| 21 | 0.10828 | 0.20656*** | 0.69089*** | 0.470 |
| 22 | -0.24817 | 0.50864* | 0.55280*** | 0.406 |
| 23 | 0.07965 | 0.14992*** | 0.74002*** | 0.536 |
| 24 | 0.01598 | 0.38990*** | 0.63147*** | 0.365 |
| 25 | -0.03415 | 0.22957 | 0.63482*** | 0.425 |
| 26 | -0.22045 | 0.50596*** | 0.61772*** | 0.363 |
| 27 | 0.10838* | 0.05737 | 0.71420*** | 0.544 |
| 28 | 0.14293*** | 0.05181* | 0.85800*** | 0.722 |
| 29 | 0.13174*** | 0.03425 | 0.72861*** | 0.569 |
| 30 | -0.48805 | 0.63427* | 0.62923*** | 0.355 |
| 31 | -0.01890 | 0.22024*** | 0.68535*** | 0.439 |
| 32 | 0.33208*** | 0.04623 | 0.76231*** | 0.571 |
| 33 | 0.17943*** | 0.05542* | 0.81773*** | 0.652 |
| 34 | -0.03678 | 0.35738*** | 0.66150*** | 0.375 |
| 35 | 0.15849 | 0.06752 | 0.62889*** | 0.440 |
| | | | | |
| Mean | 0.14604 | 0.13516 | 0.68092** | 0.472** |
| Median | 0.10101 | 0.15403 | 0.67078 | 0.440 |
| High | 3.70000 | 1.03590 | 0.85800 | 0.722 |
| Low | -0.89725 | -3.12610 | 0.51974 | 0.327 |

GMM parameter estimates are based on Equations (1) and (2) estimated jointly for 73 trading days from 2 January 1996 through 12 April 1996 (the pre-decimalization period). *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Ex post reversal rate represents the percentage of reversals in transaction prices based on tick changes. Trades are bunched using the trader-identified bunching method where all sequential transactions by the same trader type, which are on the same side of the market at the same price and without an intervening quote revision, are considered to be a single transaction.

**Table 4.9.** Pre- Versus Post-decimalization Bid and Ask Quote Depth

| | Bid Quote Depth | | Ask Quote Depth | |
|---|---|---|---|---|
| Firm | Pre-decimal | Post-decimal | Pre-decimal | Post-decimal |
| 1 | 146 | 77 | 141 | 69 |
| 2 | 189 | 68 | 178 | 67 |
| 3 | 93 | 51 | 99 | 57 |
| 4 | 206 | 97 | 217 | 113 |
| 5 | 234 | 140 | 214 | 92 |
| 6 | 267 | 128 | 270 | 131 |
| 7 | 307 | 137 | 265 | 133 |
| 8 | 173 | 95 | 181 | 95 |
| 9 | 311 | 118 | 278 | 116 |
| 10 | 137 | 88 | 127 | 60 |
| 11 | 51 | 49 | 56 | 54 |
| 12 | 142 | 84 | 128 | 77 |
| 13 | 45 | 31 | 53 | 33 |
| 14 | 99 | 70 | 104 | 57 |
| 15 | 428 | 215 | 408 | 172 |
| 16 | 166 | 97 | 167 | 80 |
| 17 | 78 | 36 | 96 | 53 |
| 18 | 28 | 27 | 27 | 25 |
| 19 | 80 | 69 | 82 | 66 |
| 20 | 394 | 143 | 469 | 178 |
| 21 | 131 | 69 | 136 | 69 |
| 22 | 66 | 33 | 57 | 29 |
| 23 | 639 | 300 | 629 | 270 |
| 24 | 149 | 74 | 150 | 70 |
| 25 | 182 | 115 | 175 | 122 |
| 26 | 72 | 51 | 79 | 53 |
| 27 | 298 | 132 | 285 | 138 |
| 28 | 451 | 194 | 400 | 235 |
| 29 | 332 | 164 | 326 | 161 |
| 30 | 79 | 73 | 75 | 60 |
| 31 | 128 | 94 | 105 | 71 |
| 32 | 217 | 100 | 273 | 111 |
| 33 | 421 | 163 | 469 | 184 |
| 34 | 425 | 132 | 418 | 136 |
| 35 | 62 | 43 | 64 | 43 |
| | | | | |
| Mean | 206 | 102*** | 206 | 99*** |
| Median | 166 | 94 | 167 | 77 |
| High | 639 | 300 | 629 | 270 |
| Low | 28 | 27 | 27 | 25 |

Bid and ask quoted depth is presented for the 73 days prior to a reduction in minimum price variation on the TSE (pre-decimalization period) from 2 January 1996 through 12 April 1996, and for the 54 days subsequent to the reduction (post-decimalization period) from 15 April 1996 through 28 June 1996. Bid and ask depth is the per-firm average depth in board lots per transaction for the respective periods. "***" indicates that the mean bid (and ask) quote depths are significantly different in the pre- and post-decimalization periods at the 1% level.

# CHAPTER 5

# CONCLUSION

The preceding three chapters have focused on the cost elements of the bid-ask spread, on types of market participants, on types of transactions, and on how transactions might be aggregated. They have used the same data set, the thirty-five firms making up the TSE35 index, and the same time frame, the periods prior to and following a move to decimalization and a lower tick size on the TSE. Based on the reported findings in Chapters 2, 3 and 4, we arrive at four major conclusions. First, it is likely that market makers use more than the immediately preceding transaction in setting bid and ask prices as the trading session advances. Second, although information asymmetry increases when the tick size is smaller, the increase in asymmetry may not be priced within the spread. Third, the action of aggregating transactions is likely to change inferences, and affect economic and regulatory decisions that are predicated on the differing empirical estimates that result from different choices of aggregation criteria. Fourth, trade size is important in determining the costs of both information asymmetry as well as the cost of order processing. However, other elements, such as trader type and market structure, also significantly affect these costs. Furthermore, the differences in these costs across firms suggest that market maker differences in risk bearing and capitalization, among other factors, are likely to be important in determining the component costs of bid-ask spreads.

These conclusions are subject to limitations. First, the data used was the thirty-five largest capitalization stocks of the TSE over a relatively short period of time, the first six months of 1996 in aggregate. Second, by their nature, models abstract certain features of reality, and must necessarily exclude others in order to be parsimonious. The ability of a model to reveal systematic features of the data is directly related to these choices. The models we have used, the adaptations we have made to them, and the way they have been estimated with the data have an impact on the results reported herein. While we have used tests to determine whether our models and methods lead to robust conclusions, other choices may lead to different conclusions.

The results open avenues for further research as well as providing direction for policy makers who have the responsibility of choosing alternate market mechanisms. In terms of further research, Chapter 2 presented a model for determining the probability of informed trade period-by-period on the basis of transaction data. This model could be used to examine how beliefs about information may change for different periods throughout the year, or for different days of the week. For example, is there a systematic difference in the belief about the presence of informed trade on Mondays as opposed to the other days of the week? This model also could be used to examine the information impact of corporate announcements of different types. That is, does the release of public information condition beliefs about the presence of private information? In addition, this model could be used to examine investor anticipation of the existence of asymmetric information around events. One might expect the probability of private information to decline subsequent to an information release.

In Chapter 3, the measurement of the costs of adverse information and order processing accounted not only for trade size but for the type of transactor and the trading mechanism. Relatively recent innovations to trading mechanisms have entered the scene such as payment for order flow and "upstairs" trading. We have shown that "upstairs" trading can increase the cost of trading in general. Overall, our results suggest that there may be a tendency to filter out low information asymmetry trades from exchange trading, thus leaving problematic transactions to the "floor". This de factor segregation of trade type may be at odds with the desires of both the exchanges and regulators if this segregation results in higher transaction costs being shifted to the trading public whose orders are traded in the so-called "downstairs" market. This finding should be of interest to regulators in dealing with proposals that may result in the further fragmentation of secondary markets.

In all three chapters, we have provided evidence that a move to a lower minimum price variation or decimal pricing creates a problem both for theory and measurement. Specifically, more sophisticated trade indicator-based microstructure models, such as that of Huang and Stoll (1997), are likely to perform less well since trade reversal rates fall following decimalization. As noted earlier, while information asymmetry may be higher following decimalization, that increase in risk does not appear to be priced. A different mechanism, quote depth reduction, appears to be used by market makers to attenuate the impact of the increase in risk. Thus, we conclude that in an era of decimal pricing, microstructure models must be extended to directly model the relationship between spreads and quote depth. In other words, the model must be extended to add a second dimension of market making (i.e. quote depth management) to the first dimension (i.e., spread management) that is commonly incorporated in most existing models of bid-ask cost components.

## REFERENCES


Amihud, Y., and H. Mendelson, 1987, "Trading Mechanisms and Stock Returns: An Empirical Investigation," *Journal of Finance*, 42, 533-553.

Anshuman, V, and A. Kalay, 1993, "Market Making Rents under Discrete Prices", working paper, University of Utah.

Bacidore, J., 1997, "The Impact of Decimalization on Market Quality: An Empirical Investigation of the Toronto Stock Exchange," *Journal of Financial Intermediation*, 6, 92-120.

Battalio, R., and C. Holden, 1997, "Why Doesn't Decimal Trading Eliminate Payment for Order Flow and Internalization?", working paper

Chan, L., and J. Lakonishok, 1995, "The Behavior of Stock Prices Around Institutional Trades," *Journal of Finance*, 50, 1147-1174.

Chan, L., and J. Lakonishok, 1993, "Institutional Trades and Intra-day Stock Price Behavior," *Journal of Financial Economics*, 33, 173-199.

Chung, R., L. Kryzanowski and H. Zhang, 1996, "Decimalization's Winners and Losers," *Canadian Investment Review*, winter, 35-39

Dempster, A., N. Laird, and D. Rubin, 1977, "Maximum Likelihood Estimation from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B. 39, 1-38.

Easley, D., N. Kiefer, and M. O'Hara, 1997a, "One Day in the Life of a Very Common Stock," *Review of Financial Studies*, 10, 805-835.

Easley, D., N. Kiefer, and M. O'Hara, 1997b, "The Information Content of the Trading Process," *Journal of Empirical Finance*, 4, 159-186.

Easley, D., and M. O'Hara, 1992, "Time and the Process of Security Price Adjustment," *Journal of Finance*, 47, 577-606.

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

Engel, C. and J. Hamilton, 1990, "Long Swings in the Dollar: Are They in the Data and Do the Markets Know It?," *American Economic Review*, 80:4, 689-713.

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.

Glosten, L., 1987, "Components of the Bid-Ask Spread and the Statistical Properties of Transaction Prices," *Journal of Finance*, 42, 1293-1307.

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.

Griffiths, M., B. Smith, D. Turnbull, and R. White, 1998, "The Role of Tick Size in Upstairs Trading and Downstairs Trading," *Journal of Financial Intermediation*, 7, 393-417

Hamilton, J., 1993, "Estimation, Inference and Forecasting of Time Series Subject to Changes in Regime," in *Handbook of Statistics*, Vol. 11, G. Maddala, C. Rao, and H. Vinod, eds., Elsevier Science Publishers, 231-260.

Hamilton, J., 1991, "A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions," *Journal of Business & Economic Statistics*, 9:1, 27-39.

Hamilton, J., 1990, "Analysis of Time Series Subject to Changes in Regime," *Journal of Econometrics*, 45, 39-70.

Hamilton, J., 1989, "A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Harris, L., 1986, "A Transaction Data Study of Weekly and Intradaily Patterns in Stock Returns," *Journal of Financial Economics*, 16, 99-118.

Harris, L., 1994, "Minimum Price Variations, Discrete Bid-Ask Spreads, and Quotation Sizes," *Review of Financial Studies*, 7:1, 149-178.

Ho, T., and H. Stoll, 1981, "Optimal Dealer Pricing under Transaction and Return Uncertainty," *Journal of Financial Economics*, 9, 47-73

Huang, R. and H. Stoll, 1997, "The Components of the Bid-Ask Spread: A General Approach," *Review of Financial Studies*, 10, 995-1034.

Huang, R. and H. Stoll, 1996, "Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE," *Journal of Financial Economics*, 41, 313-357.

Huang, R. and H. Stoll, 1994, "Market Microstructure and Stock Return Predictions," *Review of Financial Studies*, 7, 179-213.

Lee, C., and M. Ready, 1991, "Inferring Trade Direction from Intraday Data," *Journal of Finance*, 46, 733-746.

Lin, J-C., 1992, "Order Persistence, Adverse Selection, and Gross Profits Earned by NYSE Specialists," working paper, Louisiana State University.

Lin, J-C, G. Sanger, G. Booth, 1995, "Trade Size and Components of the Bid-Ask Spread," *Review of Financial Studies*, Volume 8, No. 4, 1153-1183.

Madhavan, A., M. Richardson, and M. Roomans, 1997, "Why Do Security Prices Change? A Transaction-Level Analysis of NYSE Stocks," *Review of Financial Studies*, 10, 1035-1064.

Lakonishok, J. and E. Maberly, 1990, "The Weekend Effect: Trading Patterns of Individual and Institutional Investors," *Journal of Finance*, 45, 231-243.

Porter, D. and D. Weaver, 1997, "Tick Size and Market Quality", *Financial Management*, 26:4 (winter), 5-26.

Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance*, 39, 1127-1139

Scholes, M. and J. Williams, 1977, "Estimating Betas from Nonsynchronous Data," *Journal of Financial Economics*, 5, 309-327.

Stoll, H., 1989, "Inferring the Components of the Bid-Ask Spread: Theory and Empirical Tests," *Journal of Finance*, 44, 115-134.