On the importance of timing specifications in market microstructure research

Thomas Henker and Jian-Xin Wang*

This version: 3/30/2005

Abstract

This paper highlights the importance of timing specifications in empirical market microstructure studies. Small changes in the data matching process and the timing specification of economic variables can significantly alter the outcomes of empirical research. Using the methodology developed by Lee and Ready (1991), we find that their “5-second rule” is no longer appropriate for matching quotes with transactions for NYSE stocks in the TAQ data set. Quotes should be delayed one second when matched with transactions. We demonstrate the significance of the timing specifications of economic variables using the Huang and Stoll (1997) spread decomposition model. Seemingly minor variations from the theoretical model result in severe biases in the estimated parameters. Correcting the timing errors provide much more realistic spread component estimates than those achieved in the literature.

* Thomas Henker and Jian-Xin Wang are from the University of New South Wales, School of Banking and Finance, UNSW Sydney, NSW 2052, Australia (tel: +61 2 9385-5854 and +61 2 9385-5863, fax: +61 2 9385-6347, e-mail: t.henker@unsw.edu.au and jx.wang@unsw.edu.au). We thank Martin Martens, Roger Huang, Steven Brown, and participants at the 2004 Financial Management Association Conference for helpful comments. We thank Julia Henker for valuable research assistance. The usual disclaimer applies.
I Introduction

Details of timing specifications in the empirical adaptation of theoretical microstructure models often receive less scrutiny than they deserve. While the theoretical models are carefully crafted, with considerable attention given to the sequencing of events and the information sets of agents, often less consideration is given to the empirical adaptations necessary for testing. Using three examples, we show that timing specifications can dramatically affect the empirical outcomes in testing of a theoretical microstructure model and therefore determine the model’s validity. Although our examples are based on one particular data set (TAQ) and one particular model [Huang and Stoll (1997)], the timing issues discussed are relevant to many empirical microstructure studies.

Many studies require the matching of high frequency data from multiple sources. The most common is the matching of trades with prevailing quotes. Our first example shows that changes in data recording methods means that previously established data matching conventions are no longer applicable to current conditions. Specifically, the “5-second rule” of Lee and Ready (1991) is no longer appropriate for matching quotes with trades for NYSE stocks recorded in TAQ. Instead we show that a 1-second delay is more appropriate. The difference of four seconds can significantly affect the empirical results in microstructure studies. For example, when measuring the information content of trades or price impact, the “5-second rule” ignores the information content in quotes posted within 5 seconds of each trade and systematically overstates the price impact of trades. As we show later, this leads to the overestimation of the adverse selection cost component of the spread.

In testing a theoretical model, the empirically tested specification is often different from the outcome of the theoretical model, either by choice or by necessity.
Timing specifications in the empirical model implementation must be consistent with the theoretical model. Seemingly minor variations from the theoretical model that are inconsistent with the underlying assumptions and will lead to biased empirical results. A case in point is the empirical implementation of the Huang and Stoll (1997) spread decomposition model, where time-varying spreads replace the constant spread in the theoretical model. In our second example we show that matching the timing of the spread with the timing of order flow is inconsistent with the spread component model with time-varying spreads. The third example argues that using time-varying spreads is inconsistent with the model’s theoretical basis. Both inconsistencies lead to a large downward bias in the estimated adverse selection cost component. These examples demonstrate that subtle differences in timing specifications can have severe effects on empirical results. The theoretical model, not the availability of data, should dictate the empirical specification. As more intraday data become available, the temptation to use all available information may actually be detrimental to a model’s performance.

Our choice of Huang and Stoll (1997) as an example is motivated by two factors. First, the analysis of the spread components is regarded as “an example of market microstructure work, theoretical and empirical, at its best and most successful” [Goodhart and O’Hara (1997)]. The adverse selection component provides an empirical measure for the level of information asymmetry for a listed company. The measure has found many applications in asset pricing [Kumar et.al. (1998)], fund management [Neal and Wheatley (1998)], accounting [Affleck-Graves et.al. (2002)] and corporate finance [Clarke et.al. (2004)] studies. The ability to accurately measure the adverse selection component is critical to these and future applications.

Second, the Huang and Stoll (1997) model (HS) makes a significant theoretical contribution to the spread decomposition literature. It is among the few
models that allow for the separate estimation of the adverse selection and inventory holding costs. It nests many previous models that rely on trade indicators or serial covariance in the trade flow. Its theoretical framework has been adapted by others to examine such issues as tick size and the true equilibrium spread (for example, Ball and Chordia, 2001). However, the empirical performance of the HS model is less than satisfactory. The estimates of the adverse selection component, $\alpha$, are often negative. Clarke and Shastri (2000) find negative $\alpha$s for approximately 60 percent of the NYSE stocks in their sample. Van Ness, et.al. (2001) report negative $\alpha$ for over 50 percent of their sample. As a result, the HS model has not been as widely adopted as other spread component models [e.g. Lin, et.al. (1995), Madhavan, et.al. (1997)]. We show that timing specifications contribute to the poor performance of the HS model. Correcting the timing mis-specifications results in a lower cross sectional dispersion of spread cost components and almost no negative adverse selection cost estimates.

The paper proceeds as follows. The next section describes the characteristics of the data. Section three examines the appropriateness of Lee and Ready’s 5-second quote delay rule for matching trades to quotes in more recent data. The fourth section describes a timing inconsistency in the empirical implementation of the theoretical HS model, and examines how it affects empirical estimation results. Section five argues that the empirical implementation of the HS model should not be based on time-varying spreads. The theoretical model calls for the use of the average spread which, when implemented empirically, produces more consistent and higher estimates of $\alpha$. The final section concludes.
II. Data and Sample Selection

Our stock universe comprises all stocks in the S&P 500 index at the beginning of 1999 that had a primary listing on the NYSE and traded for at least 200 trading days in 1999 without a change to the ticker symbol or primary listing. Our resulting sample comprises 401 stocks, including all of the MMI stocks of the original HS study. Trade and quote data are taken from the Trade and Quote (TAQ) data from the NYSE for the year 1999. To assure data integrity, we subject the data to a series of error checks and further exclusions\(^1\). Following the original HS study, all consecutive trades at the same price are bunched if the bid and ask quotes did not change between trades.

Table 1 provides descriptive statistics of the data. Since the S&P 500 index rose 18 percent in 1999, it is not surprising that most stocks had more buyer-initiated trades than seller-initiated trades. The average buy/sell ratio is 1.08; the buy/sell ratio of 360 stocks (90 percent) are greater than 1. The average trade reversal probability, i.e. a buy(sell) followed by a sell(buy), is 0.589. There were 33 stocks with a trade reversal probability of less than 0.5. These stocks are later excluded from the sample for spread decomposition. Trade size and market depth are smaller than average during the first 15 minutes of the trading day, and are larger than average during the last 15 minutes of the trading day. In 1999 all stocks were traded in minimum price increments of one sixteenth (6.25 cents). The median quoted spread for our sample of stocks is 12.5 cents, while the median of the effective spread (2x|mid-quote – trade price|) is around 10 cents. Spreads are much higher at opening and are lower at close. The next section examines in more detail the procedure used to link trades and quotes.

---

\(^1\) All trades labeled as errors or out-of-order and all trades with non-standard settlement arrangements are excluded. We also exclude non-firm quotes and quotes with an associated depth of zero. A further price reversal filter captures rare egregious recording errors. A complete list of error checks for trades and quotes is available from the authors on request.
III. Quote delay to sequence trades and quotes

The NYSE trade and quote (TAQ) data are widely used in high-frequency microstructure studies. It requires researchers to match intraday quotes with trades to construct a causal event sequence. Using 1988 data, Lee and Ready (1991) suggest a 5-second delay to match quotes with trades. In this section, we show that the data reporting and recording process has changed since 1988, and a 1-second delay is more appropriate for recent data. The 4-second difference is highly significant for the classification of trades into buyer- and seller-initiated trades, and for studies of the information content of trades.

Lee and Ready (1991) examine the proper matching process using intraday quotes and transaction prices of 150 NYSE stocks in 1988. They calculate the frequency of quote arrival around isolated trades for which there are no other trades within a 2-minute window. Panel A of Figure 1 depicts the distribution of quote revisions and is an image replicate of Panel A of Figure 2 in Lee and Ready (1991). Because there are no other trades in the 2-minute window, these quote revisions are most likely triggered by the isolated trade. Therefore Lee and Ready suggest moving back 5 seconds to find the quote prevailing before the trade. Despite their warning that “a different delay may be appropriate for other time periods,” the “5-second rule” has gained nearly universal acceptance as the starting point for empirical research using TAQ data.

We apply the Lee and Ready procedure to 1999 TAQ data for our sample of S&P 500 stocks. Because of the increased number of trades, it is difficult to find isolated trades defined by a 2-minute window for frequently traded securities. Instead, we define isolated trades by a 40-second window, resulting in 27 percent of

---

2 Permission from the Journal of Finance is being requested for the use of the graph.
all trades being included in the analysis as isolated trades. Panel B of Figure 1 depicts
the distribution of quote revisions for 1999 data. Compared to Panel A, the
probability mass now clearly begins to accrue at the zero second mark, the timestamp
when the trade occurs. Panel B indicates that quote revisions attributable to the
isolated trade occur either simultaneously with the trade or immediately after the
trade. Therefore we only need to move back 1 second from the trade time stamp to
find the prevailing quote before the trade. Panel B also shows that using the 40-
second window is sufficient to exclude any contaminating effects on quote arrival
from trades outside the 40-second window. Our large number of isolated trades (38
percent of all trades) increases the robustness of our conclusion.

The quote arrivals in Panel B include repetitious quotes, i.e. both bid and ask
prices are the same as the prior quote. As a further robustness check, Panel C
describes the distribution of quote revisions where quoted prices actually change,
excluding repetitious quotes and quotes in which only the depth changes. It displays
the same probability density as Panel B, indicating that a 1-second delay is
appropriate for our sample. Because all quote arrivals involve actual price changes,
Panel C also shows that using a 5-second delay would lead to a number of incorrect
quote matches.

It is difficult to speculate on the origin for the change in the recorded quote
revision pattern, and thus the change from 5 to 1 second delay. According to Lee and
Ready (1991), when the specialist announces a trade and the revised quotes, the trade
is recorded by an employee of the exchange while the new quotes are recorded by the
specialist’s clerk. “If the specialist’s clerk is fast, the new quotes can be entered
before the trade.” (Lee and Ready 1991, p737). There have been changes in the data

5 In 1999 more than 90 percent of the quotes for frequently traded stocks do not contain a price change
from the prior quote. When the depth available changes, it is considered as a new quote arrival.
recording process since 1988. TAQ User’s Guide indicates that since June 1995, the
time stamps of all NYSE trades are set by the Consolidated Trade System (CTS), and
since March 1996, the time stamps of all quotes are set by the Consolidated Quote
System (CQS). It appears that the recording of trades and quotes is now
synchronized. This is consistent with our evidence in Figure 1.

The widespread use of the TAQ data, thus the need to match quotes and
trades, makes the 4-second difference a significant issue in trade classification and in
estimating the price impact of a trade. For example, assume that the market is
trending down and the quoted spread is 10¢ over time. The midquote (m_t) for a stock
at time t is $12.50, at t+4 seconds m_{t+4}=$12.40. If a buyer-initiated order is filled at
t+5 seconds at price p_{t+5}=$12.45, the 5-second rule will misclassify this as a seller-
initiated trade because p_{t+5}<m_t. If a seller-initiated order is filled at t+5 seconds at
price p_{t+5}=$12.35, the 5-second rule will correctly classify it as a seller-initiated trade,
but will overstate the effective spread as 2|p_{t+5}-m_t|=$0.30, while the true effective
spread is 2|p_{t+5}-m_{t+4}|=$0.10. Therefore in a down trend, the 5-second rule tends to
misclassify buyer-initiated orders and overstates the price impact and the information
content of seller-initiated trades. In an up trend, the 5-second rule tends to misclassify
seller-initiated orders and overstates the price impact and the information content of
buyer-initiated trades. The exaggerated price impact is systematic because down
trends have more sales and up trends have more buys. As an example, we implement
the HS spread decomposition using 5-second delay and 1-second delay. The average
adverse selection cost for the 5-second delay is 8.5 percent, while the average adverse
selection cost for the 1-second delay is 3.4 percent\(^6\). Correcting the timing mis-
specifications discussed in sections IV and V does not change the conclusion.

\(^6\) These results are for MMI stocks in 1998.
IV. Spread decomposition with the time-varying spread

This section demonstrates that a small change in timing specification can determine the empirical success or failure of a theoretical model. The analysis is based on the Huang and Stoll (1997) spread decomposition model. We begin by deriving the spread decomposition model using time varying spreads. We show that the original HS timing specification for the spread is incorrect and leads to biased parameter estimation. Empirical estimation is then carried out to compare parameter estimates from the original HS specification with those from the timing error corrected specification.

The HS model with the time-varying spread

The HS model provides a three-way decomposition of the bid-ask spread into adverse selection, inventory holding, and order processing cost components. The model was developed for a constant spread and then adapted to allow for time varying spreads. In the empirical estimation, HS replace the constant traded spread, $S_t$, in HS equation (23) with $S_{t-1}$ and $S_{t-2}$ as follows:

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

(1)

where $M_t$ is the midpoint of the bid-ask quote; $S_t$ is the quoted spread; $Q_t$ is a trade indicator, $Q_t = 1 (-1)$ for buyer (seller) initiated trades; $\alpha$ is the adverse selection component of the spread; $\beta$ is the inventory component; $\pi$ is the probability of a trade reversal, e.g. a buy followed by a sell. It is estimated from $Q_t = (1-2\pi)Q_{t-1} + u_t$. The model requires $\pi > 0.5^7$.

The choice of timing for the spread in the above equation, HS equation (26), is determined by the timing of $Q_{t-1}$ and $Q_{t-2}$ which indicate order flows at t-1 and t-2. We

---

7 In a pure inventory model, a market maker manages his inventory by changing quotes to induce trade reversals. This leads to $\pi > 0.5$. See Stoll (1989) and Huang and Stoll (1997).
show that this choice is inconsistent with the theoretical model. In the original model
with three-way spread decomposition and constant spread, the expected order flow is
\[ E(Q_{t-1}|Q_{t-2}) = (1-2\pi)Q_{t-2}. \quad (2) \]
The change in the asset value at time \( t \) is driven by private information embedded in
the unexpected order flow, \( u_{t-1} = Q_{t-1}-(1-2\pi)Q_{t-2}, \) at time \( t-1, \) and the unexpected
public information \( \varepsilon_t: \)
\[ \Delta V_t = \alpha \frac{S_{t-1}}{2} u_{t-1} + \varepsilon_t \quad (3) \]
When the time-varying spread is introduced into the model, the timing of the spread
should be determined by the timing of the unexpected order flow \( u_{t-1} \) at time \( t-1 \)
\[ \Delta V_t = \alpha \frac{S_{t-1}}{2} u_{t-1} + \varepsilon_t = \alpha \frac{S_{t-1}}{2} (Q_{t-1}-(1-2\pi)Q_{t-2}) + \varepsilon_t \quad (4) \]
The choice of timing for the spread should not be determined by the timing of the
total order flows \( Q_{t-1} \) and \( Q_{t-2}, \) as suggested by HS and adopted by subsequent studies.
The midpoint of the quote, with time-varying spreads, is determined by:
\[ M_t = V_t + (\beta/2) \sum_{i=t-1}^{t-1} S_i Q_i \quad (5) \]
Combining the first difference of equation (5) with equation (4) results in the change
in the quote midpoint specified as:
\[ \Delta M_t = (\alpha + \beta) \frac{S_{t-1}}{2} Q_{t-1} - \alpha (1-2\pi) \frac{S_{t-1}}{2} Q_{t-2} + \varepsilon_t \quad (6) \]
Equation (6) represents a model incorporating time-varying spreads. It differs from
HS equation (26) in that the spread in the second term is \( S_{t-1} \) instead of \( S_{t-2}. \)

Using \( S_{t-2} \) instead of \( S_{t-1} \) introduces a downward bias in the estimated \( \alpha \) and an
upward bias in the estimated \( \beta. \) Define \( \Delta S_t = S_t - S_{t-1}, \) then \( S_{t-1} = S_{t-2} + \Delta S_{t-1}. \) Let \( \hat{\pi} \) be
the estimated \( \pi \) from \( Q_t = (1-2\pi)Q_{t-1}+u_t. \) Equation (6) can be re-written as follows:

---

8 We thank the referee for recommending the preceding exposition.
\[ \Delta M_t = (\alpha + \beta) \frac{S_{t-1}Q_{t-1}}{2} - \alpha (1 - 2\hat{\pi}) \frac{S_{t-2}Q_{t-2}}{2} - \alpha (1 - 2\hat{\pi}) \frac{\Delta S_{t-1}Q_{t-2}}{2} + \varepsilon_t \]  \hspace{1cm} (7)

Equation (26) of HS omits the third term. Unless \( \Delta S_{t-1}Q_{t-2} \) in the third term is orthogonal to the first two terms, this omission leads to the omitted-variable bias in the estimation of \( \alpha \) and \( \beta \) [Greene (2003), p148]. The size and direction of the biases can be ascertained by rearranging equation (7) as \( \Delta M_t = \alpha X_1 + \beta X_2 + \alpha X_3 + \varepsilon_t \), where \( X_1 = S_{t-1}Q_{t-1}/2 - (1-2\hat{\pi})S_{t-2}Q_{t-2}/2 \), \( X_2 = S_{t-1}Q_{t-1}/2 \), and \( X_3 = -(1-2\hat{\pi})\Delta S_{t-1}Q_{t-2}/2 \). HS omits \( X_3 \), which causes biasness given by \( \text{E}(\hat{\alpha}) - \alpha = \alpha k_1 \) and \( \text{E}(\hat{\beta}) - \beta = \beta k_2 \). The coefficients \( k_1 \) and \( k_2 \) are determined by estimating \( X_3 = k_1 X_1 + k_2 X_2 + \varepsilon_t \). The Appendix shows that the \( \hat{k}_1 < 0 \) and \( \hat{k}_2 > 0 \), therefore \( \text{E}(\hat{\alpha}) - \alpha = \alpha k_1 < 0 \) and \( \text{E}(\hat{\beta}) - \beta = \beta k_2 > 0 \). The omission of \( X_3 \) leads to a downward bias in the estimated \( \alpha \) and an upward bias in the estimated \( \beta \).

The size of the bias depends on the true \( \alpha \).

Estimation from our data of S&P index stocks shows that \( \hat{k}_1 < 0 \) and \( \hat{k}_2 > 0 \) for all 401 stocks. The mean (median) of \( \hat{k}_1 \) and \( \hat{k}_2 \) are -0.14 (-0.139) and 0.14 (0.141) respectively, and the robust t-statistics are larger than 50. Furthermore, \( \hat{k}_1 + \hat{k}_2 = 0 \) implies that the estimation of \( (\alpha + \beta) \) in equation (7) is unaffected. These predictions are all confirmed in the estimation of \( \alpha \) and \( \beta \) in the following subsection.

**Empirical comparison of alternative timing specifications**

To assess the significance of the change in the model we compare parameter estimates from the original HS specification [equations (1) and (2)] to estimates obtained with the timing error corrected model [equations (2) and (6)]. Following HS, we use the generalized method of moments (GMM) procedure to estimate both models. The moment conditions are the normalizing equations from (1) or (6), plus the normalizing equation for \( Q_t = (1-2\pi)Q_{t-1} + u_t \) from (2). With three parameters
(α, β, and π) to be estimated for each stock, the models are exactly identified. The covariance matrix is estimated via Newey-West (1989), with the bandwidth selected by the automatic bandwidth estimator of Andrews (1991) using the Bartlett Kernel. For most stocks, the bandwidths are estimated to be between 6 to 9 lags.

The model and the estimation procedure requires further restriction of the sample data. Because the model requires π>0.5, we exclude the 33 stocks with π<0.5. If π is close to 0.5, (1-2π) is close to zero. If that is the case equation (6) shows that the parameters α and β cannot be separately identified. Therefore we remove 42 stocks with 0.5 < π < 0.525. This gives us the final sample of 326 stocks. We also exclude trades and quotes from the first 15 minutes after opening to remove the potential impact of the opening procedure. Following Madhavan et.al (1997), we remove trades at the quote midpoint because it is unlikely that the market makers undercut their own posted spreads to participate in these trades.

To facilitate comparison with the original HS results, we begin by comparing MMI constituent stocks for the two models. Table 2 presents the comparison. The average adverse selection cost estimated with the original HS specification is around 3 percent. Five out of the eighteen stocks have significantly negative α. The timing error corrected model has an average α of 15 percent with only one significantly negative α. The sharp increase in the estimated α by an average of 12 percent supports the claim that the original HS specification suffers a downward omitted-variable bias. The inventory holding cost component β decreases from an average of 31.5 percent for the original HS specification to an average of 19.4 percent for the timing error corrected model, a decrease of approximately 12 percent. Thus (α+β) remains relatively unchanged, as predicted by the previous analysis.

9 Alternative cutoff values give the same qualitative results.
10 In our sample AXP and DOW have π<0.5, therefore comparison is not made.
To further illustrate the importance of the timing specification, we expand our sample of stocks to include all S&P 500 stocks with a primary listing on the NYSE. This expanded sample includes smaller companies with less institutional holding and smaller transactions than those of the MMI index. Table 3 presents the cross-sectional statistics of the estimation results for the 326 S&P 500 index constituent stocks. When smaller stocks are included, the original HS specification produces negative average (-5.73 percent) and median (-4.5 percent) for the adverse selection cost estimates. Over half (67 percent) of the $\hat{\alpha}$s are negative. For the timing error corrected model, the average $\hat{\alpha}$ increases by 14 percent, from -5.73 percent to 8.37 percent, and the median $\hat{\alpha}$ increases from -4.5 percent to 5 percent. More than half of the negative $\hat{\alpha}$s from the original HS specification have become positive. Only 14 percent of the $\hat{\alpha}$s are significantly negative, compared to 52 percent before. The average $\hat{\beta}$ is now 30.6 percent, 14 percent lower than the average $\hat{\beta}$ for the original specification. Again ($\hat{\alpha} + \hat{\beta}$) remains relatively unchanged.

In summary our analysis illustrates that when using time-varying spread for spread decomposition, the timing of the spread should be based on the timing of the unexpected order flow, not the total order flow. The subtle change in the timing specification, i.e. replacing $S_{t-2}$ in equation (1) with $S_{t-1}$, significantly improves the estimation results.

V. Spread decomposition with the constant spread

Despite the significant improvement in the estimated parameters reported in the previous section, there is still a large portion of stocks (31 percent) with negative estimated $\alpha$ and nearly half of them (14 percent) are statistically significant (Table 3). Given that $\alpha$ should be none negative for all stocks, do these negative $\alpha$'s indicate
mis-specification in the modified model, or do they simply reflect estimation errors? In this section we argue that the decomposition of the spread into percentage components should be made using the average spread as the benchmark as specified in the theoretical HS model. Using intraday time-varying spreads for the empirical estimation is inconsistent with the model’s theoretical foundation and is actually detrimental to it’s empirical performance. As in section IV, we first show that the use of time-varying spread is inappropriate and leads to biased parameter estimates. We then estimate the spread decomposition model using the average spread and then compare them to the parameter estimates from section IV.

**Biasness from Using the Time-Varying Spread**

In the theoretical model of Huang and Stoll, the percentage spread components $\alpha$ and $\beta$ are measured relative to a constant spread $S$. The model starts by assuming $\Delta V_t = 0Q_{t-1} + \varepsilon_t$ where $\theta = \alpha S/2$ is the average dollar impact of a trade. Ignoring the inventory cost for the moment, the change in the midpoint of the quotes is $\Delta M_t = 0Q_{t-1} + \varepsilon_t$. The estimated $\theta$ equals $E(|\Delta V_t|)$ and $E(|\Delta M_t|)$ and reflects the average information content of a trade over the sample period. It is the dollar component of the spread attributed to the presence of adverse information, and the estimated $\alpha$ is $\hat{\alpha} = \frac{\hat{\theta}}{S/2} = \frac{E(|\Delta M_t|)}{S/2}$, where $S$ can be the average posted or effective spread.

By using the time-varying spread $S_{t-1}$ in the empirical estimation, the dollar price impact of a trade becomes $\alpha S_{t-1}/2$: $\Delta M_t = (\alpha S_{t-1}/2)Q_{t-1} + \varepsilon_t$. Clearly $\alpha S_{t-1}/2$ is no longer the average price impact of a trade. Nor is it the price impact specific to $Q_{t-1}$, which should be the time-specific adverse selection component $^{11} \alpha_{t-1}$ multiplied by the

---

$^{11}$ Studies have shown that $\alpha$ change with transaction size (Lin et.al. (1995)), and $\alpha$ is higher at opening and reduces over the trading day (Madhavan et.al. (1997)).
half-spread $S_{t-1}/2$ at the time. The estimated $\alpha$ in this setting is $\mathbb{E}\left(\frac{|\Delta M_t|}{S_{t-1}/2}\right)$, the average of the dollar price impact $|\Delta M_t|$ relative to the prevailing half spread $S_{t-1}/2$. It is different from $\mathbb{E}(|\Delta M_t|)/\mathbb{E}(S_{t-1}/2) = \mathbb{E}(|\Delta M_t|)/S/2$, the way $\alpha$ is defined in the theoretical model. We show below that when $\alpha$ is defined as in the theoretical model, using the time-varying spread underestimates $\alpha$ and overestimates $\beta$.

When both adverse selection and inventory effects are present, the theoretical model calls for the estimation of the following equation:

$$\Delta M_t = \frac{\alpha S}{2} u_{t-1} + \frac{\beta S}{2} Q_{t-1} + \varepsilon_t$$  \hspace{1cm} (8)$$

where $u_{t-1}$ is the unexpected order flow. Setting $S_{t-1} = S + e_{t-1}$, thus $S = S_{t-1} - e_{t-1}$, gives

$$\Delta M_t = \frac{\alpha S_{t-1}}{2} u_{t-1} + \frac{\beta S_{t-1}}{2} Q_{t-1} - \frac{\alpha}{2} e_{t-1} u_{t-1} - \frac{\beta}{2} e_{t-1} Q_{t-1} + \varepsilon_t$$  \hspace{1cm} (9)$$

Replacing $S$ in equation (8) with $S_{t-1}$ gives the modified HS model of equation (6). But equation (6) still omits the two terms involving $e_{t-1}$ in equation (9) and therefore produces biased estimates for $\alpha$ and $\beta$: $E(\hat{\alpha}) - \alpha = \alpha k_1 + \beta k_2$ and $E(\hat{\beta}) - \beta = \alpha k_3 + \beta k_4$. Let $X_1 = S_{t-1} u_{t-1}$, $X_2 = S_{t-1} Q_{t-1}$, $X_3 = -e_{t-1} u_{t-1}$, and $X_4 = -e_{t-1} Q_{t-1}$. The biasness parameters $k_1$, $k_2$, $k_3$, and $k_4$ can be estimated by regressing the omitted variables $X_3$ and $X_4$ on the included variables $X_1$ and $X_2$: $X_3 = k_1 X_1 + k_3 X_2 + u$ and $X_4 = k_2 X_1 + k_4 X_2 + v$. Table 4 reports the estimated coefficients $k_1$, $k_2$, $k_3$, and $k_4$. $k_1$ and $k_2$ are overwhelmingly negative, thus the biasness $E(\hat{\alpha}) - \alpha = \alpha k_1 + \beta k_2 < 0$, i.e. using time varying spread underestimates $\alpha$. The coefficient $k_3$ is overwhelmingly positive. But $k_4$ is on average not statistically different from zero: the mean and the median have opposite signs and the median t-statistics is not significant. Therefore on average $E(\hat{\beta}) - \beta \approx \alpha k_3 > 0$, i.e. $\beta$ is overestimated. For individual stocks, $E(\hat{\beta}) - \beta = \alpha k_3 + \beta k_4$ depends on the value of true
\( \alpha \) and \( \beta \). The numerical values of \( k_1, k_2, k_3, \) and \( k_4 \) indicate that \( E(\hat{\alpha} + \hat{\beta}) - (\alpha + \beta) = \alpha(k_1+k_3)+\beta(k_2+k_4) \) is likely to be negative.

**Empirical Results**

Table 5 presents the summary statistics of the spread decomposition using the average posted spread. For MMI stocks, the average and median \( \hat{\alpha} \) are 29 percent and 30 percent respectively, substantially higher than the 15 percent and 11.45 percent from the timing error corrected model in Table 2. For S&P stocks, the average and median \( \hat{\alpha} \) are 19.6 percent and 18 percent respectively, substantially higher than the 8.37 percent and 5 percent from the timing error corrected model in Table 3. The estimated \( \beta \)'s are slightly lower, with mean and median 28.5 percent and 31 percent for S&P stocks, compared to 31 percent and 33 percent before. The results confirm that when the true \( \alpha \) and \( \beta \) are measured relative to the constant spread as in the theoretical model, using time-varying spread underestimates \( \alpha \) and overestimates \( \beta \). Unlike the timing error corrected model, the estimated \( \alpha + \beta \) is higher thus the order processing cost \( (1 - \alpha - \beta) \) is lower.

Using the constant spread also greatly reduces the estimation errors and increases the robust t-statistics. The mean and median t-statistics of the estimated \( \alpha \) are 13.82 and 9.38, compared to 3.41 and 2.08 before. The median t-statistics of the estimated \( \beta \) is also higher, indicating lower estimation errors for \( \beta \). The cross-sectional range of the component estimates is slightly lower. The parameter ranges for \( \alpha \) and \( \beta \) (Max-Min) are 0.8 and 0.86, compared to 0.96 and 1.09 before. Combining the higher estimated \( \alpha \) with lower estimation errors, the number of negative \( \alpha \) drops sharply from 102 to 5, and the number of significantly negative \( \alpha \) drops from 46 to 2.
Even with lower estimated $\beta$’s, the smaller estimation errors reduce the number of negative $\beta$ by half.

VI. Conclusions

In this paper we draw attention to the timing issues in empirical microstructure research, particularly in data matching and in the model adaptation for empirical testing. We first show that a 1-second quote delay, rather than the 5-second delay adopted in previous studies, should be used to match quotes with trades for NYSE TAQ data. Using a 5-second quote delay misclassifies trades and overstates the information content of trades.

We then derive the Huang and Stoll (1997) spread decomposition model with time-varying spreads. The model indicates an error in the timing specification of the spread in the original empirical implementation. Correcting the timing error in the model removes over half of the negative adverse selection cost estimates compared to the original specification. If the adaptation of the model to available data with time varying spreads is reversed and the model is estimated using a constant average traded spread almost all spread cost component estimates for a large cross-section of stocks are positive.

The paper illustrates that apparently innocuous details of timing specifications can ultimately change the empirical outcomes. It also shows that the Huang and Stoll spread decomposition model, when properly implemented, produces much more reasonable spread cost component estimates than those documented in the literature.
Appendix

Equation (7) can be written as $\Delta M_t = \alpha X_1 + \beta X_2 + \alpha X_3 + \epsilon_t$, where $X_1 = S_{t-1}Q_{t-1}/2 - (1-2 \hat{\pi})S_{t-2}Q_{t-2}/2$, $X_2 = S_{t-1}Q_{t-1}/2$, and $X_3 = -(1-2 \hat{\pi})\Delta S_{t-1}Q_{t-2}/2$. The bias is given by $E(\hat{\alpha}) - \alpha = \alpha k_1$ and $E(\hat{\beta}) - \beta = \alpha k_2$. The coefficients $k_1$ and $k_2$ are determined by estimating $X_3 = k_1X_1 + k_2X_2 + \epsilon_t$. In this estimation $\hat{k}_1 = \left(\frac{X_1'X_3}{X_1'X_1} - \frac{X_1'X_2}{X_1'X_1}X_3\right)$.

We now show $\hat{k}_1 < 0$. The denominator is $\text{Var}(X_1)\text{Var}(X_2) - [\text{Cov}(X_1,X_2)]^2 > 0$, therefore the sign of $\hat{k}_1$ is determined by the numerator. Note that

$$X_1'X_3 = \text{cov}(X_1,X_3) = \text{cov}(S_{t-1}Q_{t-1}/2 - (1-2 \hat{\pi})S_{t-2}Q_{t-2}/2, S_{t-1}Q_{t-1}/2 - (1-2 \hat{\pi})\Delta S_{t-1}Q_{t-2}/2)$$

$$= -(1-2 \hat{\pi})(1/4)\text{cov}(S_{t-1}Q_{t-1}, \Delta S_{t-1}Q_{t-2}) + (1-2 \hat{\pi})^2(1/4)\text{cov}(S_{t-2}Q_{t-2}, \Delta S_{t-1}Q_{t-2})$$

Since $S_{t-1}$ (thus $\Delta S_{t-1}$) is set before observing $u_{t-1}$, $\text{cov}(\Delta S_{t-1}, u_{t-1}) = 0$. After substituting $Q_{t-1} = (1-2 \hat{\pi})Q_{t-2} + u_{t-1}$ the first term becomes $-(1-2 \hat{\pi})^2(1/4)\text{cov}(S_{t-1}Q_{t-2}, \Delta S_{t-1}Q_{t-2})$, which is negative because $\text{cov}(S_{t-1}, \Delta S_{t-1}) > 0$. The second term is also negative assuming mean-reversion in spreads. Mean reversion implies that a narrow spread at $t-2$ is likely to lead to trades that widens the spread ($\Delta S_{t-1} > 0$), and a wide spread at $t-2$ is likely to draw competing quotes or limit orders that narrow the spread ($\Delta S_{t-1} < 0$), therefore $\text{cov}(S_{t-2}, \Delta S_{t-1}) < 0$. Together we have $X_1'X_3 < 0$. Similarly

$$X_1'X_2 = \text{cov}(X_1,X_2) = \text{cov}(S_{t-1}Q_{t-1}/2 - (1-2 \hat{\pi})S_{t-2}Q_{t-2}/2, S_{t-1}Q_{t-1}/2)$$

$$= \text{var}(S_{t-1}Q_{t-1}/2) - (1-2 \hat{\pi})(1/4)\text{cov}(S_{t-2}Q_{t-2}, S_{t-1}Q_{t-1})$$

$$= \text{var}(S_{t-1}Q_{t-1}/2) - (1-2 \hat{\pi})^2(1/4)\text{cov}(S_{t-2}Q_{t-2}, S_{t-1}Q_{t-1})$$

Mean reversion in spreads implies $\text{cov}(S_{t-2}, S_{t-1}) < 0$ therefore $X_1'X_3 > 0$. Similarly

$$X_2'X_3 = \text{cov}(X_2,X_3) = \text{cov}(S_{t-1}Q_{t-1}/2 - (1-2 \hat{\pi})\Delta S_{t-1}Q_{t-2}/2)$$

$$= -(1-2 \hat{\pi})(1/4)\text{cov}(S_{t-1}Q_{t-1}, \Delta S_{t-1}Q_{t-2})$$

$$= -(1-2 \hat{\pi})^2(1/4)\text{cov}(S_{t-1}Q_{t-2}, \Delta S_{t-1}Q_{t-2}) < 0$$ because $\text{cov}(S_{t-1}, \Delta S_{t-1}) > 0$.

Combining the terms shows that the numerator is negative, thus $\hat{k}_1 < 0$. A similar analysis shows $\hat{k}_2 > 0$. 

17
References


Boehmer, B., and E. Boehmer, 2002, Trading your neighbor’s ETFs: Competition or fragmentation?, working paper NYSE.


Table 1: Descriptive statistics

Cross-sectional descriptive statistics for the 401 stocks with more than 200 trading days in 1999. The buy/sell ratio is the ratio of buyer-initiated trades to seller-initiated trades. The probability of reversal is the probability that a buy (sell) trade is followed by a sell (buy) trade. “At opening” and “at close” are the first and last 15 minutes of the trading day. The effective spread is defined as 2|mid-quote – trade price|.

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of trading days</td>
<td>251</td>
<td>252</td>
<td>205</td>
<td>252</td>
</tr>
<tr>
<td>Number of trades per day</td>
<td>210</td>
<td>155</td>
<td>17</td>
<td>1182</td>
</tr>
<tr>
<td>Average midquote price</td>
<td>45.98</td>
<td>42.8</td>
<td>2.84</td>
<td>144.28</td>
</tr>
<tr>
<td>Buy/sell ratio</td>
<td>1.084</td>
<td>1.081</td>
<td>0.877</td>
<td>1.322</td>
</tr>
<tr>
<td>Probability of reversal</td>
<td>0.589</td>
<td>0.571</td>
<td>0.45</td>
<td>0.917</td>
</tr>
</tbody>
</table>

**Trade Size** (shares)
- Buyer-initiated trades: 1166, 1000, 300, 4700
- Seller-initiated trades: 1096, 1000, 300, 4200
- Trade size at opening: 1015, 1000, 200, 5000
- Trade size at close: 1347, 1100, 300, 4600

**Depth** (shares)
- At bid: 2427, 2000, 500, 33700
- At ask: 2763, 2000, 400, 40000
- At opening (bid+ask): 4268, 3000, 700, 46700
- At close (bid+ask): 8960, 6800, 1600, 86650

**Spread** (dollar)
- Posted spread: 0.1217, 0.125, 0.0625, 0.3125
- Posted spread at opening: 0.1691, 0.1875, 0.0625, 0.4375
- Posted spread at close: 0.1172, 0.125, 0.0625, 0.25
- Effective spread for buys: 0.1055, 0.1038, 0.0665, 0.2316
- Effective spread for sales: 0.1032, 0.1012, 0.0655, 0.2187
- Effective spread at opening: 0.1085, 0.125, 0.0625, 0.25
- Effective spread at close: 0.0775, 0.0625, 0.0625, 0.1875
## Table 2: Spread decomposition for MMI stocks

Spread cost component estimates for the stocks considered in the original Huang and Stoll (1997) paper. The “Original model” refers to the original Huang and Stoll model (equations (1) and (2) in this paper). The “Corrected model” refers to the timing error corrected model consisting of equations (2) and (6). The Newey-West t-statistics are in parentheses. In our sample AXP and DOW have π<0.5 and are therefore excluded.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Original Model</th>
<th>Corrected Model</th>
<th>Change in α</th>
<th>Change in β</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Probability of reversal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>π</td>
<td>α</td>
<td>β</td>
<td>α</td>
<td>β</td>
</tr>
<tr>
<td>CHV</td>
<td>0.5621</td>
<td>0.007</td>
<td>0.296</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(12.55)</td>
<td>(6.73)</td>
<td>(5.76)</td>
</tr>
<tr>
<td>DD</td>
<td>0.6227</td>
<td>0.026</td>
<td>0.330</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(3.27)</td>
<td>(42.25)</td>
<td>(10.93)</td>
<td>(29.98)</td>
</tr>
<tr>
<td>EK</td>
<td>0.5993</td>
<td>-0.031</td>
<td>0.455</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>(-2.34)</td>
<td>(34.29)</td>
<td>(4.52)</td>
<td>(24.05)</td>
</tr>
<tr>
<td>GE</td>
<td>0.6320</td>
<td>-0.010</td>
<td>0.263</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(-0.17)</td>
<td>(6.18)</td>
<td>(0.74)</td>
<td>(5.14)</td>
</tr>
<tr>
<td>GM</td>
<td>0.6183</td>
<td>0.111</td>
<td>0.281</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>(8.59)</td>
<td>(22.34)</td>
<td>(13.92)</td>
<td>(14.24)</td>
</tr>
<tr>
<td>IBM</td>
<td>0.5319</td>
<td>0.231</td>
<td>0.098</td>
<td>0.609</td>
</tr>
<tr>
<td></td>
<td>(7.05)</td>
<td>(3.07)</td>
<td>(15.65)</td>
<td>(-7.51)</td>
</tr>
<tr>
<td>IP</td>
<td>0.5758</td>
<td>-0.036</td>
<td>0.404</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>(-2.03)</td>
<td>(24.21)</td>
<td>(5.94)</td>
<td>(13.47)</td>
</tr>
<tr>
<td>JNJ</td>
<td>0.6234</td>
<td>0.068</td>
<td>0.343</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>(9.15)</td>
<td>(45.35)</td>
<td>(14.93)</td>
<td>(35.18)</td>
</tr>
<tr>
<td>KO</td>
<td>0.7046</td>
<td>0.039</td>
<td>0.397</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>(8.28)</td>
<td>(88.09)</td>
<td>(14.2)</td>
<td>(78.15)</td>
</tr>
<tr>
<td>MMM</td>
<td>0.5329</td>
<td>0.096</td>
<td>0.321</td>
<td>0.430</td>
</tr>
<tr>
<td></td>
<td>(2.23)</td>
<td>(7.42)</td>
<td>(10.03)</td>
<td>(-0.37)</td>
</tr>
<tr>
<td>MO</td>
<td>0.8111</td>
<td>0.025</td>
<td>0.256</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>(4.81)</td>
<td>(64.05)</td>
<td>(6.03)</td>
<td>(70.26)</td>
</tr>
<tr>
<td>MOB</td>
<td>0.5344</td>
<td>0.062</td>
<td>0.360</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td>(1.86)</td>
<td>(10.85)</td>
<td>(9.91)</td>
<td>(2.28)</td>
</tr>
<tr>
<td>MRK</td>
<td>0.6352</td>
<td>0.096</td>
<td>0.161</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>(13.34)</td>
<td>(21.99)</td>
<td>(17.97)</td>
<td>(10.71)</td>
</tr>
<tr>
<td>PG</td>
<td>0.5442</td>
<td>-0.065</td>
<td>0.422</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>(-3.34)</td>
<td>(21.75)</td>
<td>(5.61)</td>
<td>(11.85)</td>
</tr>
<tr>
<td>S</td>
<td>0.6642</td>
<td>-0.010</td>
<td>0.294</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(-1.03)</td>
<td>(30.84)</td>
<td>(5.13)</td>
<td>(24.18)</td>
</tr>
<tr>
<td>T</td>
<td>0.7147</td>
<td>0.142</td>
<td>0.195</td>
<td>0.163</td>
</tr>
<tr>
<td></td>
<td>(31.06)</td>
<td>(45.19)</td>
<td>(34.56)</td>
<td>(38.64)</td>
</tr>
<tr>
<td>X</td>
<td>0.5944</td>
<td>-0.118</td>
<td>0.495</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(-5.87)</td>
<td>(24.12)</td>
<td>(-1.29)</td>
<td>(18.79)</td>
</tr>
<tr>
<td>XON</td>
<td>0.6164</td>
<td>-0.097</td>
<td>0.299</td>
<td>-0.047</td>
</tr>
<tr>
<td></td>
<td>(-9.85)</td>
<td>(30.51)</td>
<td>(-4.68)</td>
<td>(25.36)</td>
</tr>
<tr>
<td>Average</td>
<td>0.6176</td>
<td>0.0298</td>
<td>0.3150</td>
<td>0.1503</td>
</tr>
<tr>
<td></td>
<td>(3.63)</td>
<td>(29.73)</td>
<td>(9.49)</td>
<td>(22.23)</td>
</tr>
<tr>
<td>Median</td>
<td>0.6174</td>
<td>0.0255</td>
<td>0.3100</td>
<td>0.1145</td>
</tr>
</tbody>
</table>
Table 3: Spread decomposition for S&P 500 stocks using time-varying spreads

The “Original model” refers to the original Huang and Stoll model (equations (1) and (2) in this paper). The “Corrected model” refers to the timing error corrected model consisting of equations (2) and (6). The Newey-West t-statistics are in parentheses. “Sig. neg.” is the number of significantly negative coefficient estimates at a significance level of 95 percent.

<table>
<thead>
<tr>
<th></th>
<th>Original Model</th>
<th>Corrected Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Probability of reversal</td>
<td>Adverse Selection</td>
</tr>
<tr>
<td></td>
<td>π</td>
<td>α</td>
</tr>
<tr>
<td>Average</td>
<td>0.6128</td>
<td>-0.0573</td>
</tr>
<tr>
<td></td>
<td>(-1.13)</td>
<td>(24.19)</td>
</tr>
<tr>
<td>Median</td>
<td>0.5917</td>
<td>-0.0450</td>
</tr>
<tr>
<td></td>
<td>(-2.19)</td>
<td>(20.10)</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.0781</td>
<td>0.1417</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.5255</td>
<td>-0.6550</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.9239</td>
<td>0.3710</td>
</tr>
<tr>
<td>Negative</td>
<td>218</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>67 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Sig. Neg.</td>
<td>173</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>53 %</td>
<td>0 %</td>
</tr>
</tbody>
</table>

Table 4: The Biasness Coefficients

This table reports the cross-sectional summary of regression coefficients from $X_3=k_1X_1+k_3X_2+u$ and $X_4=k_2X_1+k_4X_2+v$ for 401 stocks, where $X_1=S_{t-1}u_{t-1}$, $X_2=S_{t-1}Q_{t-1}$, $X_3=-e_{t-1}u_{t-1}$, and $X_4=-e_{t-1}Q_{t-1}$. The hypothesized signs are $k_1<0$, $k_2<0$, $k_3>0$, and $k_4>0$.

|      | Mean  | Median | Mean  | Median | Stocks with hypothesized sign ($|t|>1.96$) |
|------|-------|--------|-------|--------|------------------------------------------|
| $k_1$ | -0.472| -0.490 | -67.1 | -48.5  | 99 % (99 %)                              |
| $k_2$ | -0.217| -0.240 | -32.1 | -24.0  | 92 % (90 %)                              |
| $k_3$ | 0.214 | 0.226  | 29.8  | 24.5   | 93 % (90 %)                              |
| $k_4$ | -0.022| 0.011  | 2.35  | 1.09   | 55 % (47 %)                              |
Table 5: Spread decomposition using constant spread

The t-statistics are in parentheses. Sig. neg. is the number of significantly negative coefficient estimates at a significance level of 95 percent.

<table>
<thead>
<tr>
<th></th>
<th>MMI Stocks</th>
<th>S&amp;P 500 Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of reversal</td>
<td>π</td>
<td>α</td>
</tr>
<tr>
<td>Average</td>
<td>0.6128</td>
<td>0.2899</td>
</tr>
<tr>
<td></td>
<td>(29.34)</td>
<td>(15.72)</td>
</tr>
<tr>
<td>Median</td>
<td>0.5917</td>
<td>0.3040</td>
</tr>
<tr>
<td></td>
<td>(26.12)</td>
<td>(10.95)</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.0781</td>
<td>0.1205</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.5255</td>
<td>-0.0210</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.9239</td>
<td>0.4780</td>
</tr>
<tr>
<td>Negative</td>
<td>6 %</td>
<td>6 %</td>
</tr>
<tr>
<td>Sig. Neg.</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

22
Figure 1: The probability density of quote revisions around isolated trades

The probability density of quotes per transaction in a 40-second window centered on isolated trades defined as the first trade after 9:30:20 but not after 15:59:40, with no other trade within a 40-second window centered on the trade.

Panel A: 1988 NYSE

(61.7% of the price changes recorded ahead of the trade)

Panel B: 1999 NYSE
Figure 1: The probability density of quote revisions around isolated trades (Continued)

Panel C: 1999 NYSE non-repetitious quote arrivals