Journal of Financial Markets 9 (2006) 162–179

Journal of FINANCIAL MARKETS

www.elsevier.com/locate/finmar

On the importance of timing specifications in market microstructure research ☆

Thomas Henker, Jian-Xin Wang*

School of Banking and Finance, the University of New South Wales, Sydney, NSW 2052, Australia

Available online 28 February 2006

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. Journal of Finance 46(2) 733–746], we show that their "5-second rule" is not appropriate for matching quotes with transactions for NYSE stocks in the TAQ data set, and the prevailing quotes are the ones immediately before the trades. We demonstrate the significance of the timing specifications of economic variables using the Huang and Stoll [1997. Review of Financial Studies 10, 995–1034] 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.

JEL classification: G10

Keywords: Timing specification; Market microstructure; Quote delay; Spread decomposition; Adverse selection

☆Comments from an anonymous referee have significantly improved the quality of the paper and are greatly appreciated. We also thank Steven Brown, David Feldman, Doug Foster, Bruce Lehmann (the Editor), Martin Martens, David Michayluk, Peter Swan, Terry Walter, Xianming Zhou, and seminar participants at the University of New South Wales, the University of Hong Kong, and the 2004 Financial Management Association Conference for helpful comments. We thank Julia Henker and Vincent Tang for valuable research assistance. The usual disclaimer applies.

*Corresponding author. Tel.: 0061 2 9385 5863; fax: 0061 2 9385 6347.

E-mail addresses: t.henker@unsw.edu.au (T. Henker), jx.wang@unsw.edu.au (J.-X. Wang).

1386-4181/$ - see front matter
doi:10.1016/j.finmar.2006.01.001

1. 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 two 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), i.e. matching trades with quotes at least 5 seconds before the trade timestamp, is not appropriate for NYSE stocks in TAQ data. Instead we show that the prevailing quotes are the ones immediately before each trade, i.e. with timestamps at least 1 second before the trade timestamp. We label our result as "1-second delay". It is different from the zero-delay recommended by Bessembinder (2003) and Peterson and Sirri (2003). The timing choice for matching trades with quotes affects a wide range of microstructure analysis, including trade classification, calculating effective spread, and estimating the information content of trades. We show that using the 5 seconds delay 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 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. We show that the timing of the spread should be matched with the timing of the unexpected order flow, not the total order flow. Furthermore using time-varying spreads in the empirical implementation is inconsistent with the model's theoretical basis. These inconsistencies in the timing specifications lead to a large downward bias in the estimated adverse selection cost component. The 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.

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, α, are often negative. Clarke and Shastri (2000) find negative αs for approximately 60 percent of the NYSE stocks in their sample. Van Ness, et al. (2001) report negative α 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 inconsistencies in the timing 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 3 examines the appropriateness of Lee and Ready's 5 seconds 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 5 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 α. The final section concludes.

2. 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. The resulting sample comprises 401 stocks, including all of the MMI stocks in the original HS study. Intraday trades and quotes 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,

[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.

Table 1
Descriptive statistics

Cross-sectional descriptive statistics for 401 stocks from the S&P 500 index with primary listing on the NYSE and 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|.

	Average	Median	Min	Max
Number of trading days	251	252	205	252
Number of trades per day	210	155	17	1182
Average midquote price	45.98	42.8	2.84	144.28
Buy/sell ratio	1.084	1.081	0.877	1.322
Probability of reversal	0.589	0.571	0.45	0.917
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	33,700
At ask	2763	2000	400	40,000
At opening (bid + ask)	4268	3000	700	46,700
At close (bid + ask)	8960	6800	1600	86,650
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

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.

3. Quote delay to sequence trades and quotes

The NYSE TAQ data cover all trades and quotes on NYSE, AMEX, and NASDAQ since 1993. Since the data are supplied in separate files for trades and quotes, researchers have to match intraday quotes with trades to construct a causal event sequence. Lee and Ready (1991) recommend a 5 seconds delay of quote time for matching quotes with trades on NYSE and AMEX. In this section, we show that a 1-second delay is more appropriate for TAQ data. The 4 seconds difference is highly significant for the classification of trades into buyer- and seller-initiated trades, and for studies of the information content of trades.

The analysis of Lee and Ready (1991) is based on the event sequence that underlies a typical specialist market in the microstructure literature: bid and ask quotes are set by the specialist; floor-brokers trade either with each other or with the specialist; the specialist then revises the quotes to reflect the information content of the trade and to achieve the desired inventory position (see Madhavan and Smidt, 1991). Using 1988 ISSM[2] data on 150 NYSE stocks, Lee and Ready examine quote revisions around isolated trades for which there are no other trades within a 2-minute window. These quote revisions are most likely triggered by the isolated trades. Panel A of Fig. 1 depicts the distribution of quote revisions taken directly from Lee and Ready (1991). It is clear that quote revisions are recorded up to 5 seconds before the isolated trade. Therefore the prevailing quote should be the last quote 5 seconds before the trade timestamp. Despite their warning that "a different delay may be appropriate for other time periods,"[3] the "5 seconds rule" has gained nearly universal acceptance for empirical research using TAQ data.

Relatively few studies have directly examined reporting delays on NYSE since Lee and Ready (1991). Using TORQ[4] data, Hasbrouck et al. (1993) examine reporting delays for NYSE trades, and find a medium delay of 14 seconds between the SuperDot execution report time and the Consolidated Trade System (CTS) print time (the TAQ time stamp). Their analysis does not extend to quotes therefore one cannot make any inference on matching quotes with trades from this paper.[5] Peterson and Sirri (2003) and Bessembinder (2003) examine the impact of reporting delays on the success rate of trade classification algorithms and on the percentage of trades executed inside the prevailing quotes respectively. Neither study investigates the actual misalignment of quote and trade timestamps. They both experiment with quote delays of 0–30 seconds with 5 seconds increments, and recommend a zero delay in matching trades with quotes.[6]

We follow the Lee and Ready methodology as it is based on the event sequence on the NYSE trading floor. It does not rely on proprietary or audit trail data so the analysis can be applied to a long sample period. We start with 1999 TAQ data for our sample of S&P 500 stocks. Lee and Ready (1991) define an isolated trade as the first trade between 11:00 a.m. to 2:30 p.m. with no other trades within a 2-minute window. So they select at most one isolated trade per stock per day. With increased trading activity since 1988, this definition is too restrictive and would lead to a very small number of isolated trades. Instead, we define isolated trades by a 40-second window at any time during the trading

[2]The Institute for the Study of Securities Markets (ISSM) at the University of Memphis supplied intraday data on NYSE and AMEX up to 1992. Similar to TAQ data the ISSM data also come in two separate files for quotes and trades.

[3]Footnote 10 of Lee and Ready (1991).

[4]The Trade, Order, Report, and Quote (TORQ) dataset contains audit trail data for 144 NYSE stocks over a three-month period starting November 1990.

[5]Blume and Goldstein (1997) use the trade reporting delays found in TORQ data for the matching of trades and quotes in TAQ. This is valid only if there is no reporting delay in quotes.

[6]Several studies have examined reporting delays on NASDAQ which has a different trading and reporting process than NYSE. Ellis et al. (2000) find reporting delays for trades at 15 or 16 seconds but not delays for quotes, but still recommend zero delay in matching trades with quotes based on the success rate of trade classification. Schultz (2000) reports that delays on NASDAQ vary over time and were up to several minutes in 1995 and 1996. He concludes that "it appears to be impossible to find the actual quotes in effect when trades took place."

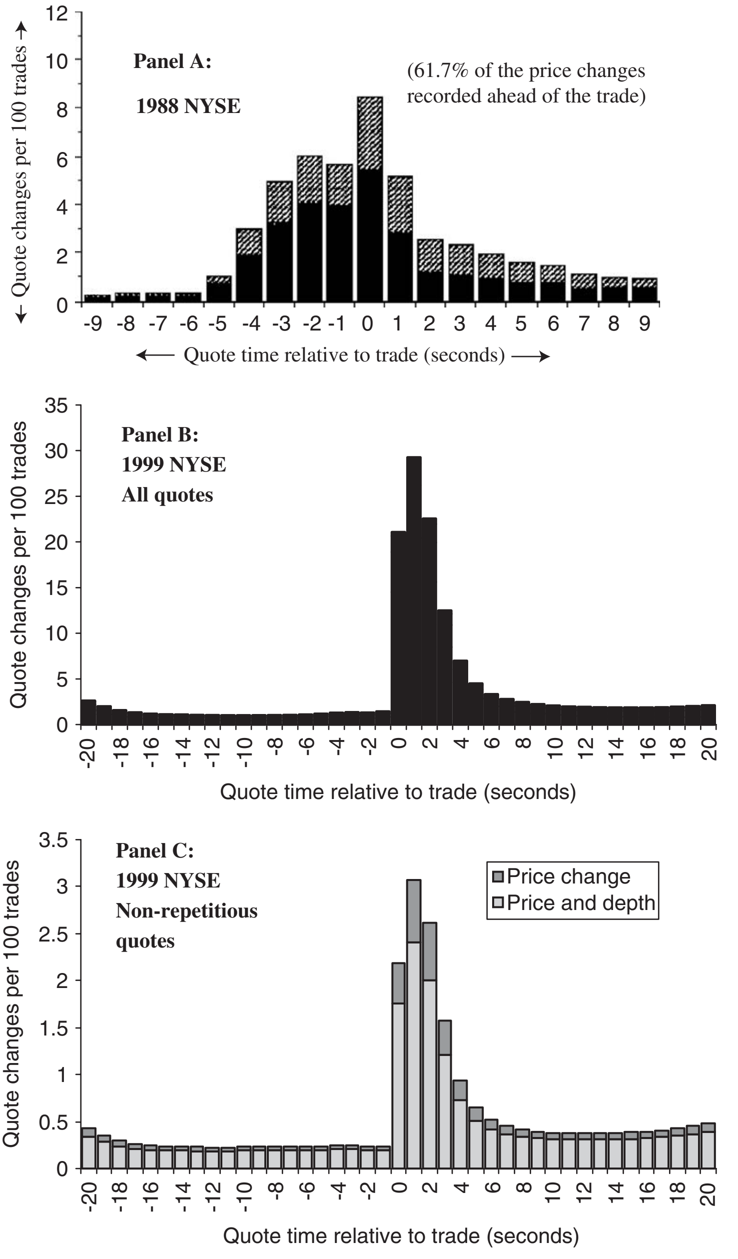

Fig. 1. The number of quote revisions around isolated trades. Panel A is the number of quote revisions around isolated trades for 150 NYSE stocks in 1988 and is taken from Fig. 2 of Lee and Ready (1991). Panels B and C are the number of quote revisions around isolated trades for 401 stocks in the S&P 500 index with primary listing on the NYSE and more than 200 trading days in 1999.

day, resulting in 38 percent of all trades being included in the analysis as isolated trades. Panel B of Fig. 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 the prevailing quote should be the last quote before the trade. Since the timestamp in TAQ is accurate to the second, the prevailing quote has to be at least 1-second before the trade, with the actual time difference usually longer than 1-second. Panel B also shows that using the 40-second window is sufficient to exclude 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.[7] Panel C describes the distribution of quote revisions where quoted prices actually change, excluding repetitious quotes and quotes in which only the depth changes. The quote distribution remains the same as in Panel B, indicating that the last quote, at least 1-second before the trade timestamp, should be the prevailing quote. Because all quote arrivals involve actual price changes, Panel C also shows that using a 5-seconds delay would lead to incorrect quote matches.

As a robustness check and to see how this result holds over time, we repeat the above analysis for the 20 stocks in the Major Market Index (MMI) from 1994 to 2002. The results are reported in Table 2. The MMI stocks are large and frequently traded. The fraction of trades considered isolated is much lower than in the sample of S&P 500 stocks, and has declined over the years from 1994 to 2002. Panel A of Table 2 considers all quote revisions while panel B only considers quote revisions where at least one side of the quote price changes. The results are consistent with our findings in Fig. 1 for all years since the creation of the TAQ database. The number of quote revisions is small before the trade, and has a sharp jump at the trade time and after. The pattern indicates the causal effect from trades to quote revisions. From 1994 onward, the prevailing quote is the one immediately before the trade.

Table 2 also shows some interesting patterns of quote activity over time. From 1994 to 2002, trading intensified and the portion of trades considered isolated dropped from 39% to 2% for MMI stocks. Quoting also intensified and the number of quotes per 100 isolated trades increased by a factor of six from 67 to 413. Before the trade, the number of quote revisions is flat at 1–2 quotes per second and did not increase significantly. The increased quoting activity concentrated at the trade time and immediately after. Quote revisions increased 16 times at the time of the trade (from 5.5 to 90.6) and by a factor of 20 at one second after the trade (from 4.6 to 93). Panel B shows that the results for quote revisions with price changes (excluding depth changes and repetitious quotes) are even more extreme; the increases in quotes at trade time and one second after are 38 and 46 times, respectively. The pattern confirms that these quote revisions are indeed triggered by the trades. If quotes were revised independent of trades, one would see an equal rise in quote arrival before and after the trade.

The analyses for S&P 500 stocks in 1999 and for MMI stocks from 1994 to 2002 indicate that the prevailing quotes are the one immediately *before* the trades. Bessembinder (2003) and Peterson and Sirri (2003) recommend "trade lag 0", which implies that zero second is deducted from the trade report time before matching to quotes. Therefore they recommend using the *contemporaneous* quote, when available, as the prevailing quote. Our evidence

[7]In 1999 more than 90 percent of the quotes for frequently traded stocks do not contain a price change from the prior quote. The TAQ data set considers each depth change or quote repetition as a new quote arrival.

Table 2
The number of quote revisions around isolated trades for MMI stocks

This table reports the number quote revisions per 100 isolated trades. Isolated trades are those with no other trades within a 40 seconds window. "Quote time" is measured relative to trade time in seconds. Per second average number of revisions is used for intervals longer than one second. Panel A considers all quote revisions. Panel B includes only quotes where at least one side of the quote has changed from the previous quote.

Quote time	1994	1995	1996	1997	1998	1999	2000	2001	2002
Panel A: all quote revisions per 100 trades									
−20 to −16	1.0	1.0	1.3	2.3	2.5	2.9	3.5	4.2	4.8
−15 to −11	1.0	0.8	0.8	1.4	1.5	1.6	1.7	1.7	1.6
−10 to −6	1.3	0.9	0.7	1.3	1.4	1.5	1.7	1.6	1.6
−5	1.2	0.9	0.7	1.4	1.5	1.7	1.9	1.8	1.8
−4	1.2	1.0	0.8	1.5	1.7	1.9	2.1	2.0	2.0
−3	1.1	1.0	0.9	1.6	1.8	2.1	2.2	2.0	1.8
−2	1.0	1.0	1.0	1.7	1.9	2.1	2.1	2.1	2.0
−1	1.2	1.3	1.3	1.9	2.2	2.1	2.1	2.2	2.3
0	5.5	7.4	10.9	23.9	28.1	27.6	37.0	61.8	90.6
1	4.6	6.8	11.6	23.4	28.1	33.0	47.0	70.5	93.0
2	4.0	6.0	10.0	16.9	18.7	22.5	30.4	35.6	41.7
3	3.6	5.1	7.6	11.1	11.1	12.1	14.9	15.9	17.3
4	2.9	3.9	5.2	7.6	7.3	7.4	8.4	9.9	10.4
5	2.5	3.1	3.8	5.8	5.4	5.4	5.9	7.1	7.3
6 to 10	1.8	2.0	2.2	3.7	3.7	3.8	4.2	4.9	5.4
11 to 15	1.4	1.4	1.5	2.7	2.8	3.1	3.8	4.8	6.0
16 to 20	1.3	1.2	1.3	2.6	3.0	3.4	4.4	6.6	9.2
Quotes per 100 trades	67	74	92	167	182	199	251	330	413
Panel B: price change per 100 trades									
−20 to −16	0.1	0.1	0.1	0.3	0.4	0.5	0.5	0.9	0.9
−15 to −11	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.4	0.3
−10 to −6	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.4	0.4
−5	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.5	0.4
−4	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.5	0.5
−3	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.5	0.4
−2	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.5	0.5
−1	0.1	0.1	0.1	0.2	0.3	0.3	0.3	0.5	0.5
0	0.4	0.5	0.6	1.9	3.0	3.1	3.1	10.5	14.7
1	0.3	0.4	0.7	2.0	3.2	4.0	4.2	11.7	15.3
2	0.3	0.4	0.7	1.7	2.4	3.0	3.3	6.4	7.6
3	0.3	0.4	0.6	1.1	1.5	1.8	1.8	3.1	3.5
4	0.2	0.3	0.4	0.8	1.1	1.1	1.1	2.1	2.1
5	0.2	0.2	0.3	0.6	0.8	0.9	0.8	1.5	1.5
6 to 10	0.2	0.2	0.2	0.4	0.6	0.6	0.6	1.1	1.2
11 to 15	0.2	0.1	0.2	0.4	0.6	0.6	0.6	1.3	1.4
16 to 20	0.2	0.1	0.2	0.4	0.6	0.7	0.7	1.8	2.3
Quotes per 100 trades	6	6	8	19	27	30	30	67	79
Isolated trades/all NYSE trades	0.39	0.36	0.29	0.19	0.15	0.11	0.08	0.04	0.02

suggests that the contemporaneous quote is most likely driven by the trade. For example in 2002, the number of non-repetitious quotes before the trade is around 0.5 per second or one every 2 s (Panel B of Table 2). This is the arrival rate for quotes that are not triggered by the trade. Therefore 14.2 of the 14.7 non-repetitious quote revisions at trade time (time 0) are triggered by the trade, representing a probability of 97% that contemporaneous quotes are *not* the prevailing quotes before the trades. Contemporaneous quotes represent 6.7% (0.4 out of 6) of non-repetitious quotes in 1994 and 18.6% (14.7 out of 79) in 2002. As more quotes arrive simultaneously with trades, using a zero delay will result in more contemporaneous quotes being wrongly used as the prevailing quotes.

The change from 5 to 1-second delay is likely driven by the changes in the data recording process at the NYSE. In 1988 when the specialist announced a trade and the revised quotes, the trade was recorded by a floor reporter and the new quotes were typed into the Display Book 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). In addition, the floor reporter had to record the trade information onto a mark-sense card first and then feed the card into an optical reader. The optical reader takes a few seconds to scan the card and put a timestamp to the trade. In 1994, the optical reader was abolished and trades were typed into a hand-held device. This is likely to be the critical change that reduced or even eliminated the time difference between recording trades and quotes. Overtime more trades are reported by the specialist's clerk via Display Book. Since July 2001, the floor reporter position is eliminated and all trades are reported via the Display Book by the specialist's clerk [NYSE 2001 Factbook]. These changes result in more synchronized timestamps of trades and quotes and more simultaneous quote revisions with trades as reported in Fig. 1 and Table 2.

The timing choice for matching trades with prevailing quotes has a direct impact on the classification of trades and on the estimation of the price impact of trades. 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 $t+4$ s $m_{t+4} = \$12.40$. If a buyer-initiated order is filled at $t+5$ s at price $p_{t+5} = \$12.45$, the 5 seconds 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 seconds 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 seconds rule tends to misclassify buyer-initiated orders and overstates the price impact of seller-initiated trades. In an up trend, the 5 seconds rule tends to misclassify seller-initiated orders and overstates the price impact of buyer-initiated trades. The exaggerated price impact is systematic because down trends have more sales and up trends have more buys. This analysis is confirmed by the empirical findings of Bessembinder (2003). He reports that the size of the effective spread and the average price impact of trades, all increase monotonically with longer time adjustments. We use the HS spread component model to examine the impact of using 5 seconds versus 1-second delays on the estimated information content of trades. The average adverse selection component is 8.5 percent for the 5 seconds delay, but only 3.4 percent for 1-second delay.[8] Correcting the timing misspecifications discussed in Sections 4 and 5 does not change the conclusion.

4. 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

[8] These results are for MMI stocks in 1998.

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 estimates. Empirical estimation is then carried out to compare parameter estimates from the original HS specification with those from the timing error corrected specification.

4.1. 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, in HS equation (23) with S_{t-1} and S_{t-2} as follows:

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

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; α is the adverse selection component of the spread; β is the inventory component; π 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 $\pi > 0.5$.[9]

The choice of timing for the spread in the above equation, HS Eq. (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 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

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

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 ε_t:

$$\Delta V_t = \alpha\frac{S}{2}u_{t-1} + \varepsilon_t. \tag{3}$$

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 $t-1$[10]

$$\Delta V_t = \alpha\frac{S_{t-1}}{2}u_{t-1} + \varepsilon_t = \alpha\frac{S_{t-1}}{2}\left(Q_{t-1} - (1-2\pi)Q_{t-2}\right) + \varepsilon_t. \tag{4}$$

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

[9]All covariance spread models, starting from Roll (1984), require negative covariance between consecutive price changes thus negative covariance in trade flows. This implies $\pi > 0.5$. Positive serial covariance in trade flows leads to the long-run accumulation or deprivation of inventory which will lead to the "eventual failure" of the market maker [of Garman (1976), p. 263]. Although $\pi < 0.5$ is empirically possible, it is theoretically impermissible. Therefore we remove all stocks with $\pi < 0.5$ in our empirical analysis. See further discussion by Huang and Stoll (1997).

[10]We thank the referee for recommending the preceding exposition.

midpoint of the quote, with time-varying spreads, is determined by:

$$M_t = V_t + (\beta/2)\sum_{i=1}^{t-1} S_i Q_i. \tag{5}$$

Combining the first difference of Eq. (5) with Eq. (4) results in the change in the quote midpoint specified as

$$\Delta M_t = (\alpha+\beta)\frac{S_{t-1}}{2}Q_{t-1} - \alpha(1-2\pi)\frac{S_{t-1}}{2}Q_{t-2} + \varepsilon_t. \tag{6}$$

Eq. (6) represents a model incorporating time-varying spreads. It differs from HS Eq. (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 α and an upward bias in the estimated β. 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 π from $Q_t = (1-2\pi)Q_{t-1} + u_t$. Eq. (6) can be re-written as follows:

$$\Delta M_t = (\alpha+\beta)\frac{S_{t-1}Q_{t-1}}{2} - \alpha(1-2\hat{\pi})\frac{S_{t-2}Q_{t-2}}{2} - \alpha(1-2\hat{\pi})\frac{\Delta S_{t-1}Q_{t-2}}{2} + \varepsilon_t. \tag{7}$$

Eq. (26) of HS omits the third term. Defining the regressors on the right hand side as $X_1 = S_{t-1}Q_{t-1}/2 - (1-2\hat{\pi})S_{t-2}Q_{t-2}/2$, $X_2 = S_{t-1}Q_{t-1}/2$, and $X_3 = -(1-2\hat{\pi})\Delta S_{t-1}Q_{t-2}/2$, Eq. (7) becomes $\Delta M_t = \alpha X_1 + \beta X_2 + \alpha X_3 + \varepsilon_t$. The biasness from omitting X_3 is given by $E(\hat{\alpha}) - \alpha = \alpha k_1$ and $E(\hat{\beta}) - \beta = \beta k_2$, with coefficients k_1 and k_2 determined by estimating $X_3 = k_1 X_1 + k_2 X_2 + e_t$ (Greene, 2003, p. 148). Thus the size and the direction of the bias depends on the impact of X_3 on ΔM_t, i.e. the true α, as well as the partial correlations k_1 and k_2 between X_3 and the included variables X_1 and X_2. The Appendix shows that $\hat{k}_1 < 0$ and $\hat{k}_2 > 0$. Therefore the omission of X_3 leads to a downward bias in the estimated α and an upward bias in the estimated β.

4.2. 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 [Eqs. (1) and (2)] to estimates obtained with the timing error corrected model [Eqs. (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 and 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 and 9 lags.

The model and the estimation procedure require further restriction of the sample data. Because the model requires $\pi > 0.5$, we exclude the 33 stocks with $\pi < 0.5$. If π is close to 0.5, $(1-2\pi)$ 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 $0.5 < \pi < 0.525$.[11] 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.

[11]Alternative cutoff values give the same qualitative results.

Table 3
Spread decomposition for MMI stocks

Spread cost component estimates for the MMI stocks in 1999, using time-vary spread decomposition of Eq. (6). Stocks AXP and DOW have $\pi < 0.5$ and are therefore excluded. The "Original model" refers to the original Huang and Stoll model (Eqs. (1) and (2) in this paper). The "Corrected model" refers to the timing error corrected model consisting of Eqs. (2) and (6). The Newey–West t-statistics are in parentheses.

	Probability of reversal π	Original model		Corrected model		Change in α	Change in β
		Adverse selection α	Inventory holding β	Adverse selection α	Inventory holding β		
CHV	0.5621	0.007	0.296	0.163	0.139	0.156	−0.157
		(0.29)	(12.55)	(6.73)	(5.76)		
DD	0.6227	0.026	0.330	0.096	0.258	0.070	−0.072
		(3.27)	(42.25)	(10.93)	(29.98)		
EK	0.5993	−0.031	0.455	0.066	0.357	0.097	−0.098
		(−2.34)	(34.29)	(4.52)	(24.05)		
GE	0.6320	−0.010	0.263	0.043	0.209	0.053	−0.054
		(−0.17)	(6.18)	(0.74)	(5.14)		
GM	0.6183	0.111	0.281	0.199	0.193	0.088	−0.088
		(8.59)	(22.34)	(13.92)	(14.24)		
IBM	0.5319	0.231	0.098	0.609	−0.285	0.378	−0.383
		(7.05)	(3.07)	(15.65)	(−7.51)		
IP	0.5758	−0.036	0.404	0.114	0.252	0.150	−0.152
		(−2.03)	(24.21)	(5.94)	(13.47)		
JNJ	0.6234	0.068	0.343	0.122	0.287	0.054	−0.056
		(9.15)	(45.35)	(14.93)	(35.18)		
KO	0.7046	0.039	0.397	0.072	0.366	0.033	−0.031
		(8.28)	(88.09)	(14.2)	(78.15)		
MMM	0.5329	0.096	0.321	0.430	−0.016	0.334	−0.337
		(2.23)	(7.42)	(10.03)	(−0.37)		
MO	0.8111	0.025	0.256	0.039	0.244	0.014	−0.012
		(4.81)	(64.05)	(6.03)	(70.26)		
MOB	0.5344	0.062	0.360	0.341	0.078	0.279	−0.282
		(1.86)	(10.85)	(9.91)	(2.28)		
MRK	0.6352	0.096	0.161	0.158	0.099	0.062	−0.062
		(13.34)	(21.99)	(17.97)	(10.71)		
PG	0.5442	−0.065	0.422	0.115	0.241	0.180	−0.181
		(−3.34)	(21.75)	(5.61)	(11.85)		
S	0.6642	−0.010	0.294	0.050	0.236	0.060	−0.058
		(−1.03)	(30.84)	(5.13)	(24.18)		
T	0.7147	0.142	0.195	0.163	0.173	0.021	−0.022
		(31.06)	(45.19)	(34.56)	(38.64)		
X	0.5944	−0.118	0.495	−0.027	0.403	0.091	−0.092
		(−5.87)	(24.12)	(−1.29)	(18.79)		
XON	0.6164	−0.097	0.299	−0.047	0.249	0.050	−0.050
		(−9.85)	(30.51)	(−4.68)	(25.36)		
Average	0.6176	0.0298	0.3150	0.1503	0.1935	0.1206	−0.1215
		(3.63)	(29.73)	(9.49)	(22.23)		
Median	0.6174	0.0255	0.3100	0.1145	0.2385	0.0790	−0.0800
		(2.05)	24.17)	(8.32)	(16.52)		

Table 4
Spread decomposition for S&P 500 stocks using time-varying spreads

This table reports the estimated spread components for 326 stocks in 1999, using time-vary spread decomposition of Eq. (6). Stocks are selected from the S&P index with primary listing on the NYSE, over 200 trading days in 1999, and the probability of reversal $\pi > 0.525$. The "Original model" refers to the original Huang and Stoll model (Eqs. (1) and (2) in this paper). The "Corrected model" refers to the timing error corrected model consisting of Eqs. (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.

	Probability of reversal π	Original model		Corrected model	
		Adverse selection α	Inventory holding β	Adverse selection α	Inventory holding β
Average	0.6128	−0.0573	0.4477	0.0837	0.3062
		(−1.13)	(24.19)	(3.41)	(18.00)
Median	0.5917	−0.0450	0.4460	0.0500	0.3290
		(−2.19)	(20.10)	(2.08)	(13.20)
Std. dev.	0.0781	0.1417	0.1630	0.1640	0.1703
Minimum	0.5255	−0.6550	−0.0670	−0.3040	−0.3570
Maximum	0.9239	0.3710	1.0090	0.6600	0.7430
Negative		218	1	102	20
		67%	0%	31%	6%
Sig. neg.		173	0	46	10
		53%	0%	14%	3%

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 3 presents the comparison. The average adverse selection cost estimated with the original HS specification is around 3 percent. Five out of the eighteen stocks[12] 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. The findings confirm that the original HS model suffers the omitted-variable bias: $E(\hat{\alpha}) - \alpha = \alpha k_1 < 0$ and $E(\hat{\beta}) - \beta = \alpha k_2 > 0$. When we construct the regressors in Eq. (7) X_1, X_2, and X_3 as defined previously and run the auxiliary regression $X_3 = k_1 X_1 + k_2 X_2 + e_t$, we find that $\hat{k}_1 < 0$ and $\hat{k}_2 > 0$ for all 326 stocks as predicted by the theoretical analysis in the Appendix. The mean (median) of $\hat{k}_1$ and $\hat{k}_2$ are −0.137 (−0.136) and 0.137 (0.138) respectively, and the robust t-statistics are larger than 50. Furthermore, $\hat{k}_1 + \hat{k}_2 \approx 0$ implies that the estimation of $(\alpha + \beta)$ in equation (7) is unaffected. Indeed Table 3 shows that $\hat{\alpha} + \hat{\beta}$ remains relatively unchanged.

[12]In our sample AXP and DOW have $\pi < 0.5$, 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 4 presents the cross-sectional statistics of the estimation results for the 326 S&P 500 index constituent stocks in our sample. When smaller stocks are included, the original HS specification produces negative average (-5.73 percent) and median (-4.5 percent) 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 Eq. (1) with S_{t-1}, significantly improves the estimation results.

5. 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 α and nearly half of them (14 percent) are statistically significant. Since α should be positive for all stocks, the question remains whether these negative α's indicate misspecification in the modified model, or simply reflect estimation errors? In this section we argue that the decomposition of the spread into percentage components should be made using the average quoted spread as specified in the HS theoretical model and not time-varying spreads as suggested in the later empirical adaptation of the model. We first show that using intraday time-varying spreads for the empirical estimation is inconsistent with the model's theoretical foundation and leads to biased parameter estimates. We then estimate the spread decomposition model using the average quoted spread and show that inappropriately using time-varying spreads is quite detrimental to the empirical performance of the model.

5.1. Biasness from using the time-varying spread

In the theoretical model of Huang and Stoll, the percentage spread components α and β are measured relative to a constant spread S. The model derivation starts by assuming $\Delta V_t = \theta Q_{t-1} + \varepsilon_t$, 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 $\Delta M_t = \theta Q_{t-1} + \varepsilon_t$. The estimated θ 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 α is $\frac{\hat{\theta}}{S/2} = \frac{E(|\Delta M_t|)}{S/2}$, 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

Table 5
The biasness coefficients

This table reports the cross-sectional regression coefficients from $X_3 = k_1X_1 + k_3X_2 + u$ and $X_4 = k_2X_1 + k_4X_2 + v$, 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 regressions are estimated for 326 stocks in the S&P 500 index with primary listing on the NYSE, over 200 trading days in 1999, and the probability of reversal $\pi > 0.525$. The hypothesized signs are $k_1 < 0$, $k_2 < 0$, $k_3 > 0$, and $k_4 > 0$.

			t-stats		Stocks with hypothesized sign ($\lvert t\rvert > 1.96$)
	Mean	Median	Mean	Median	
k_1	−0.516	−0.511	−78.4	−56.8	100% (100%)
k_2	−0.265	−0.260	−39.3	−26.9	100% (100%)
k_3	0.255	0.243	36.1	29.6	100% (100%)
k_4	0.028	0.028	6.18	3.45	67% (57%)

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[13] α_{t-1} multiplied by the half-spread $S_{t-1}/2$ at the time. The estimated α in this setting is $E(|\Delta M_t|/S_{t-1}/2)$, the average of the dollar price impact $|\Delta M_t|$ relative to the prevailing half spread $S_{t-1}/2$. It is different from $E(|\Delta M_t|)/S/2$, the way α is defined in the theoretical model. We show below that when α is defined as in the theoretical model, using the time-varying spread underestimates α and overestimates β.

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

$$\Delta M_t = \frac{\alpha S}{2}u_{t-1} + \frac{\beta S}{2}Q_{t-1} + \varepsilon_t, \tag{8}$$

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

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

Replacing S in Eq. (8) with S_{t-1} gives the modified HS model of Eq. (6). But Eq. (6) still omits the two terms involving e_{t-1} in Eq. (9) and therefore produces biased estimates for α and β: $E(\hat{\alpha}) - \alpha = \alpha k_1 + \beta k_2$ 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: $X_3 = k_1X_1 + k_3X_2 + u$ and $X_4 = k_2X_1 + k_4X_2 + v$. Table 5 reports the estimated coefficients k_1, k_2, k_3, and k_4 for our sample of 326 stocks in 1999. Again k_1 and k_2 are negative for all stocks, thus the biasness $E(\hat{\alpha}) - \alpha = \alpha k_1 + \beta k_2 < 0$. This means using time varying spreads instead of the average quoted spread underestimates α. The coefficient k_3 is positive for all stocks while k_4 is positive for 217 out of 326 stocks (67%) and significantly positive for 185 out of 326 stocks (57%). Therefore on average $E(\hat{\beta}) - \beta = \alpha k_3 + \beta k_4 > 0$, i.e. β is overestimated. 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.

[13]Studies have shown that α changes with transaction size (Lin et.al., 1995), and α is higher at the opening and reduces over the trading day (Madhavan et.al., 1997).

5.2. Empirical results

Table 6 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 3. 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 for the timing error corrected model in Table 4. The estimated βs 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 α and β are estimated with the constant spread as in the theoretical model, using time-varying spreads underestimates α and overestimates β. 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 α are 13.82 and 9.38, compared to 3.41 and 2.08 before. The median t-statistics of the estimated β is also higher, indicating lower estimation errors for β. The cross-sectional range of the component estimates is slightly lower. The parameter ranges for α and β (Max–Min) are 0.8 and 0.86, compared to 0.96 and 1.09 before. Combining the higher estimated α with lower estimation errors, the number of negative α drops sharply from 102 to 5, and the number of significantly negative α drops from 46 to 2. Even with lower estimated β's, the smaller estimation errors reduce the number of negative β by half.

Table 6
Spread decomposition using constant spread

This table reports the estimated spread components for 326 stocks in 1999, using constant spread decomposition of Eq. (8). Stocks are selected from the S&P index with primary listing on the NYSE, over 200 trading days in 1999, and the probability of reversal $\pi > 0.525$. The "Original model" refers to the original Huang and Stoll model (Eqs. (1) and (2) in this paper). The "Corrected model" refers to the timing error corrected model consisting of Eqs. (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.

	Probability of reversal π	MMI stocks		S&P 500 stocks	
		Adverse selection α	Inventory holding β	Adverse selection α	Inventory holding β
Average	0.6128	0.2899	0.1337	0.1958	0.2855
		(29.34)	(15.72)	(13.82)	(17.99)
Median	0.5917	0.3040	0.1545	0.1800	0.3140
		(26.12)	(10.95)	(9.38)	(15.32)
Std. dev.	0.0781	0.1205	0.1025	0.1229	0.1458
Minimum	0.5255	−0.0210	−0.1200	−0.1140	−0.2360
Maximum	0.9239	0.4780	0.3120	0.6840	0.6340
Negative		1	1	5	10
		6%	6%	1.5%	3%
Sig. neg.		1	1	2	8
		6%	6%	0.6%	2.5%

6. 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 seconds delay adopted in previous studies, should be used to match quotes with trades for NYSE TAQ data. Using a 5 seconds 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 that the timing of the spread should be based on the timing of the unexpected order flow, not the total order flow. Correcting the timing error in the model, removes over half of the negative adverse selection cost estimates compared to the original model specification. If 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. This 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 A

Eq. (7) can be written 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 $X_3 = -(1-2\hat{\pi})\Delta S_{t-1}Q_{t-2}/2$. 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 $X_3 = k_1 X_1 + k_2 X_2 + e_t$. In this estimation,

$$\hat{k}_1 = \frac{(X_2'X_2)X_1'X_3' - (X_1'X_2)X_2'X_3}{(X_1'X_1)(X_2'X_2) - (X_1'X_2)(X_2'X_1)}.$$

To show $\hat{k}_1 < 0$, we need two conditions: $\mathrm{cov}(S_{t-1}, S_t) > 0$ and $\mathrm{cov}(S_{t-1}, \Delta S_t) < 0$. The first implies persistence in spread. The second requires spread to be mean reverting: a narrow spread at $t-1$ is likely to lead to trades that widen the spread ($\Delta S_t > 0$), and a wide spread at $t-1$ is likely to draw competing quotes or limit orders that narrow the spread ($\Delta S_t < 0$). These properties hold in the actual spread data and are similar to those of volatility. The conditions are not inconsistent with each other: $\mathrm{cov}(S_{t-1}, \Delta S_t) = \mathrm{cov}(S_{t-1}, S_t - S_{t-1}) = \mathrm{cov}(S_{t-1}, S_t) - \mathrm{var}(S_{t-1}) < 0$ requires $0 < \mathrm{cov}(S_{t-1}, S_t) < \mathrm{var}(S_{t-1})$.

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

$$\begin{aligned} X_1'X_3 &= \mathrm{cov}(X_1, X_3) = \mathrm{cov}(S_{t-1}Q_{t-1}/2 - (1-2\hat{\pi})S_{t-2}Q_{t-2}/2,\ -(1-2\hat{\pi})\Delta S_{t-1}Q_{t-2}/2) \\ &= -(1-2\hat{\pi})(1/4)\mathrm{cov}(S_{t-1}Q_{t-1}, \Delta S_{t-1}Q_{t-2} + (1-2\hat{\pi})^2(1/4)\mathrm{cov}(S_{t-2}Q_{t-2}, \Delta S_{t-1}Q_{t-2}). \end{aligned}$$

Since S_{t-1} (thus ΔS_{t-1}) is set before observing u_{t-1}, $\mathrm{cov}(\Delta S_{t-1}, u_{t-1}) = 0$. After substituting $Q_{t-1} = (1-2\hat{\pi})Q_{t-2} + u_{t-1}$ the first term becomes $-(1-2\hat{\pi})^2(1/4)\mathrm{cov}(S_{t-1}Q_{t-2}, \Delta S_{t-1} Q_{t-2}) \equiv A < 0$ because $\mathrm{cov}(S_{t-1}, \Delta S_{t-1}) > 0$. The second term, $(1-2\hat{\pi})^2(1/4)$ $\mathrm{cov}(S_{t-2}Q_{t-2}, \Delta S_{t-1}Q_{t-2}) \equiv B$, is also negative because $\mathrm{cov}(S_{t-2}, \Delta S_{t-1}) < 0$. Similarly

$$\begin{aligned} X_1'X_2 &= \mathrm{cov}(X_1, X_2) = \mathrm{cov}(S_{t-1}Q_{t-1}/2 - (1-2\hat{\pi})S_{t-2}Q_{t-2}/2,\ S_{t-1}Q_{t-1}/2) \\ &= \mathrm{var}(S_{t-1}Q_{t-1}/2) - (1-2\hat{\pi})(1/4)\mathrm{cov}(S_{t-2}Q_{t-2}, S_{t-1}Q_{t-1}) \\ &= X_2'X_2 - (1-2\hat{\pi})^2(1/4)\mathrm{cov}(S_{t-2}Q_{t-2}, S_{t-1}Q_{t-2} = X_2'X_2 + C. \end{aligned}$$

Persistence in spreads implies $\text{cov}(S_{t-2}, S_{t-1}) > 0$ therefore $C < 0$. Finally

$$\begin{aligned} X'_2X_3 &= \text{cov}(X_2, X_3) = \text{cov}(S_{t-1}Q_{t-1}/2, -(1-2\hat{\pi})\Delta S_{t-1}Q_{t-2}/2) \\ &= -(1-2\hat{\pi})(1/4)\text{cov}(S_{t-1}Q_{t-1}, \Delta S_{t-1}Q_{t-2}) \\ &= -(1-2\hat{\pi})^2(1/4)\text{cov}(S_{t-1}Q_{t-2}, \Delta S_{t-1}Q_{t-2}) = A < 0. \end{aligned}$$

The numerator equals $(X'_2X_2)(A+B)-(X'_2X_2+C)A = (X'_2X_2)B-CA < 0$, thus $\hat{k}_1 < 0$. A similar analysis shows $\hat{k}_2 > 0$.

References

Affleck-Graves, J., Callahan, C.M., Chipalkatti, N., 2002. Earnings predictability, information asymmetry, and market liquidity. Journal of Accounting Research 40, 561–583.

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Ball, C., Chordia, T., 2001. True spreads and equilibrium prices. Journal of Finance 56, 1801–1835.

Bessembinder, H., 2003. Issues in assessing trade execution costs. Journal of Financial Markets, 233–257.

Blume, M., Goldstein, M., 1997. Quotes, order flow, and price discovery. Journal of Finance 52 (1), 221–244.

Clarke, J., Fee, C., Thomas, S., 2004. Corporate diversification and asymmetric information: Evidence from stock market trading. Journal of Corporate Finance 10, 105–129.

Clarke, J., Shastri, K., 2000. On information asymmetry metrics. University of Pittsburgh working paper.

Ellis, K., Roni, M., O'Hara, M., 2000. The Accuracy of trade classification rules: Evidence from NASDAQ. Journal of Financial and Quantitative Analysis 35 (4), 529–561.

Garman, M., 1976. Market microstructure. Journal of Financial Economics 3, 257–275.

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

Greene, W., 2003. Econometric Analysis, 5th international edition, Prentice Hall.

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

Kumar, R., Sarin, A., Shastri, K., 1998. The impact of options trading on market quality of the underlying security: An empirical analysis. Journal of Finance 53, 717–732.

Lee, C., Ready, M., 1991. Inferring trade direction from intraday data. Journal of Finance 46 (2), 733–746.

Lin, J., Sanger, G., Booth, G., 1995. Trade Size and Components of the Bid-Ask Spread. Review of Financial Studies 8, 1153–1183.

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

Madhavan, A., Smidt, S., 1991. A Bayesian Model of Intraday Specialist Pricing. Journal of Financial Economics 30 (1), 99–134.

Neal, R., Wheatley, S., 1998. Adverse selection and bid–ask spreads: Evidence from closed-end funds. Journal of Financial Markets 1, 121–149.

Newey, W., West, K., 1989. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Peterson, M., Sirri, E., 2003. Evaluation of the biases in execution cost estimation using trade and quote data. Journal of Financial Markets, 259–280.

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

Schultz, P., 2000. Regulatory and legal pressures and the costs of NASDAQ trading. Review of Financial Studies 13, 917–957.

Van Ness, B., Van Ness, R., Warr, R., 2001. How Well Do Adverse Selection Components Measure Adverse Selection? Financial Management, 5–30.