Overestimated effective spreads: Implications for investors*

Björn Hagströmer

March 23, 2017

*Björn Hagströmer, Stockholm Business School, Stockholm University, and the Swedish House of Finance. E-mail: bjh@sbs.su.se. I thank Jonathan Brogaard, Petter Dahlström, Albert Menkveld, Lars Nordén, Andreas Park, and Patrik Sandås for helpful comments. Research funding from the Jan Wallander Foundation and the Tom Hedelius Foundation is gratefully acknowledged.

Overestimated effective spreads: Implications for investors

Abstract

I show that the effective spread measured relative to the spread midpoint overstates the theoretical effective spread in markets with discrete prices and elastic liquidity demand. The average overestimation is 16% for S&P 500 stocks in general, and 60% for stocks with high relative tick size. The overestimation increases with price discreteness and maker fees. I propose an alternative measure that overcomes the overestimation problem: the microprice effective spread. Investors using the microprice effective spread are better positioned to minimize illiquidity costs through liquidity timing and order routing, evaluate adverse selection, and optimize their portfolio weights and rebalancing frequency.

The effective spread is one of the most prevalent measures of illiquidity, applied in diverse applications ranging from the evaluation of market structure changes (e.g., Hendershott et al., 2011) to asset pricing (e.g., Korajczyk and Sadka, 2008) and corporate finance (e.g., Fang et al., 2009). In addition, the effective spread has regulatory status in Rule 605 of the US Regulation National Market Systems (Reg NMS), which mandates that all exchanges publish their effective spread for each stock on a monthly basis.

Conceptually, the effective spread measures the cost of immediate execution, defined as the difference between the effective quote (where transactions actually take place) and the fundamental value. Empirically, it is defined by Blume and Goldstein (1992) and Lee (1993), as well as in Rule 605, as the difference between the transaction price and the bid–ask spread midpoint. The spread midpoint is an approximation of the fundamental value, following the convention introduced by Demsetz (1968). I refer to the conceptual definition as the *effective spread* and to its empirical implementation as the *midpoint effective spread*.

I challenge the use of the spread midpoint when measuring the effective spread. Anshuman and Kalay (1998) show that, when trade prices are discrete, market makers are unable to quote prices symmetrically around the fundamental value, implying that the cost of immediacy may differ for buy and sell market orders. In line with the finding of Epps (1976) that liquidity demand is elastic, I argue that investors respond to the quote asymmetry by submitting more market orders on the side of the market where the spread is tighter. I show in a simple model that, when the liquidity demand is elastic, use of the spread midpoint leads to overestimation of the effective spread.

The overestimation of the effective spread can be illustrated by a simple example. Consider a stock with an fundamental value of USD 10.0025 that has liquidity supplied at the nearest prices where trading is allowed, USD 10.00 and USD 10.01. The effective spread is then asymmetric: 0.25 cents for trades on the bid side and 0.75 cents (three times higher) on the ask side. If the liquidity demand is elastic, market orders in this example are more likely to arrive on the bid side than on the ask side. The effective spread is then, on average, smaller than the midpoint effective spread (which is 0.5 cents).

The purpose of this paper is to document the overestimation implied by the midpoint effective

spread and propose an alternative measure. The motivation is twofold: First, investors who manage their illiquidity costs based on the midpoint effective spread overpay for liquidity. All else being equal, such investors overlook fundamental value variation that does not trigger a midpoint price change, which undermines their liquidity timing ability. In addition, if the effective spread overestimation varies across trading venues, it may bias investors' order routing decisions.

Second, for academics and other researchers interested in market microstructure, asset pricing, and corporate finance, the effective spread is an important measure of liquidity (see previous references). In particular, if the effective spread overestimation varies systematically across securities, such applications may be misled by the use of the midpoint effective spread.

The model that I propose is a version of Glosten's (1994) limit order book model with adverse selection. My model features a time priority rule, such as that of Seppi (1997), and discrete prices, as in the reduced-form approach of Sandås (2001), with the added feature that liquidity demand may be elastic. The approach to allow liquidity demand to be elastic is consistent with the work of Hendershott and Menkveld (2014), who use the elasticity to model price pressures (in a different model framework).

The main model outcome is that the expected effective spread is a function of the expected midpoint effective spread and the imbalance of depth posted at the price quotes (the *depth asymmetry*). I show that this function equals the expected midpoint effective spread only when the liquidity demand elasticity is zero. In all other cases, the expected effective spread is lower than the expected midpoint effective spread. Importantly, the overestimation is not mitigated by averaging across a large set of trades.

The first point of my empirical analysis is to assess the elasticity of liquidity demand. The null hypothesis of inelastic liquidity demand is strongly rejected by the data. In a one-week sample (November 30 to December 4, 2015) of trades in the S&P 500 index constituent stocks, I find a significant relation between the depth asymmetry and the direction of trade. For example, when the depth asymmetry indicates that the ask-side spread is three times larger than the bid-side spread (as in the example above), only 24% of all trades are buyer initiated (paying the wide side of the spread) and 76% are seller initiated (paying the tight side). The evidence indicates that the effective

spread asymmetry is an important determinant in the traders' decision to submit a market order, the depth asymmetry is informative about the current fundamental value, and the average midpoint effective spread indeed overstates the effective spread.

The model also suggests a lower bound for the expected effective spread. It is given by the difference between the transaction price and the microprice, and I refer to it as the *microprice effective spread*. The microprice, which is the midpoint adjusted for the depth asymmetry, is a quote-based proxy of the fundamental value that is commonly applied in the financial industry (for a discussion, see Harris, 2013). I use the lower bound to quantify the overestimation of the effective spread, defined as the difference between the average midpoint and microprice effective spreads.

I find strong empirical support for overestimation of the effective spread. On average, across all S&P 500 stocks, I find that the effective spread interval is 1.30–1.47 basis points (bps). The upper bound, the midpoint effective spread, thus implies a potential overestimation of 16%. With the 2015 trading volume in S&P 500 stocks being USD 8.7 trillion, use of the midpoint effective spread overstates the annual illiquidity costs of these stocks by USD 148 million.

In addition to the level effect, I report significant variation in the overestimation across stocks, trading venues, and time. The model indicates that the overestimation should be increasing with price discreteness and decreasing with order processing costs. In the cross section of stocks, I find strong support for the price discreteness prediction. Because the minimum tick size is fixed at USD 0.01 for most US stocks, firms with low share prices have high price discreteness and the effective spread overestimation of such stocks is indeed higher than for other stocks. For example, S&P 500 stocks priced below USD 15 exhibit an average overestimation of 60%.

I also report evidence consistent with the order processing cost prediction. Overestimation of the effective spread is higher in trading venues that apply lower fees (or higher rebates) to liquidity suppliers. Finally, in a 20-year sample of stocks listed on the New York Stock Exchange (NYSE) and the American Stock Exchange (AMEX / NYSE MKT), I find that the overestimation falls sharply after the tick size reforms in 1997 and 2001. The overestimation is close to zero in 2003 and then gradually increases to around 10% in 2015.

The economic magnitude of the effective spread overestimation can be illustrated by relating it to findings on the midpoint effective spread in response to market structure changes. For example, Bessembinder (2003, Table 3, Panel D) reports that the decimalization of minimum tick sizes in the United States (US) leads to reductions in value-weighted effective spreads for large-cap stocks of 32.6% on NYSE and 5.1% on NASDAQ. Hendershott et al. (2011) find that a one standard deviation increase in algorithmic trading is associated with a 22.5% reduction in the large-cap effective spread.[1] O'Hara and Ye (2011) investigate the effects of market fragmentation and report that fragmented stocks have 8% lower spreads than consolidated stocks.[2] Given that the studies above investigate major US market structure events, the overestimation of up to 60% reported here is economically sizable.

To overcome the effective spread bias, I propose the microprice should replace the midpoint as a benchmark for the effective spread. The microprice effective spread has the advantage that it potentially captures fluctuations in the fundamental value that do not influence the midpoint. The microprice is already widely used in the financial industry as a proxy for the fundamental value and its additional computational burden relative to the spread midpoint is negligible. The data required are readily available to practitioners through Security Information Processor (SIP) consolidated data feeds and to academics in the most common databases for intraday liquidity analysis, such as the Daily Trade and Quote (DTAQ) and Thomson Reuters Tick History (TRTH) databases.

A potential concern is that the expected microprice effective spread in my model constitutes the lower bound of the expected effective spread, implying that it may underestimate the illiquidity cost. I show empirically that, in environments where the microprice is a noisy predictor of the fundamental value, traders are less inclined to trade on the tight side of the spread. That is, any bias present in the microprice is counteracted by diverse market orders. This "self-correcting property" of the microprice effective spread mitigates the risk of underestimation.

The main contribution of this paper is to document that, if liquidity demand is elastic, the

[1] Using the estimate reported by Hendershott et al. (2011) for Q1 in their Table III (-0.18) multiplied by the standard deviation of their algorithmic trading measure (4.54) and relating it to the level effective spread reported in Table I (3.63 bps), I calculate the reduction as $\frac{0.18 \times 4.54}{3.63} = 0.225$. See the corresponding calculation of Hendershott et al. (2011, p.22) for the quoted spread.

[2] Based on the results reported in Table 7 of O'Hara and Ye (2011), $\frac{0.29}{3.61} = 0.080$.

midpoint effective spread, on average, overstates the effective spread. Though a perfect proxy of the fundamental value is infeasible, the bias can be mitigated by measuring the effective spread relative to the microprice instead of the spread midpoint.

My findings add to the literature on the measurement of effective spreads, including the works of Blume and Goldstein (1992), Lee (1993), and Petersen and Fialkowski (1994). The results are also relevant to the liquidity measurement literature more generally. Roll (1984), Hasbrouck (2009), Holden (2009), and Corwin and Schultz (2012) develop effective spread proxies based on daily equity data. The evaluation of such proxies is typically based on average intraday midpoint effective spreads (e.g., Goyenko et al., 2009). Holden and Jacobsen (2014) show that the use of intraday data from the Monthly Trade and Quote (MTAQ) database results in distorted metrics of the effective spread. My findings indicate that the benchmark used for all these studies, the midpoint effective spread, is itself a biased measure of the effective spread.

The paper also contributes to the literature on the motives for initiating trades by submitting market orders. Sarkar and Schwartz (2009) report that market orders are more frequent on one side of the book during times of asymmetric information (e.g., ahead of merger news). In times of belief heterogeneity (e.g., ahead of macroeconomic news and earnings announcements), in contrast, the authors find that the distribution of market orders on the bid and ask sides is more balanced. My evidence shows that the arrival rates of market orders at the best bid and ask prices are strongly related to the asymmetry of the effective spreads, driven by the elasticity of liquidity demand.

The overestimation of effective spreads has important implications for investors and regulators. I show that investors who use the spread midpoint to gauge the fundamental value overlook about one-third of the liquidity variation, which undermines their liquidity timing ability. Furthermore, I find that the midpoint effective spread reports mandated by Rule 605 of RegNMS are misleading. In particular, for low-priced S&P 500 stocks (priced at or below USD 25), the venue ranked lowest in terms of the midpoint effective spread (NYSE MKT) is ranked highest in terms of the microprice effective spread! Investors relying on the midpoint effective spread thus overpay for liquidity through both suboptimal liquidity timing and misguided order routing. In addition, I report that the overestimation carries over to the evaluation of adverse selection costs (price impact) in liquidity

supply and introduces bias in portfolio selection and rebalancing decisions.

Finally, overestimation of the effective spread is relevant to applications where liquidity is analyzed as an outcome variable of market structure features. For example, my results relate to the recent literature on price discreteness (O'Hara et al., 2015; Werner et al., 2015; Yao and Ye, 2015; Chao et al., 2017). The finding that the overestimation problem is increasing with price discreteness is directly relevant to the evaluation of the tick size pilot in the US market, where the price discreteness of randomly selected low-priced stocks is increased from 1 cent to 5 cents.[3] Another market structure aspect under debate is the use of differential fees for suppliers and demanders (Colliard and Foucault, 2012; Harris, 2013; Malinova and Park, 2015; Panayides et al., 2017). I find that the overestimation of the effective spread is more severe in venues with lower fees charged to liquidity suppliers.

1 The Model

This section presents a static model of a limit order book market with discrete prices and a price–time priority rule. The model is based on the reduced-form model of Sandås (2001), which in turn builds on the models of Glosten (1994) and Seppi (1997). The new feature of the model is that the liquidity demand may be elastic. The outcome is an expression for the expected effective spread where the lower and upper bounds are empirically observable.

1.1 Model Setup

The model market is populated by a large number of market makers and traders. The market makers are competitive, risk-neutral, profit-maximizing liquidity suppliers. The traders are impatient and potentially privately informed liquidity demanders.

The agents trade a risky security that, in period t, conditional on all publicly available information, has a fundamental value X_t. Since new information arrives in the next period, the fundamental

[3]For details, see the US Securities and Exchange Commission (SEC) press release from May 6, 2015, https://www.sec.gov/news/pressrelease/2015-82.html.

value is updated in accordance with

$$X_{t+1} = X_t + d_{t+1}, \tag{1}$$

where d_{t+1} is a random innovation with positive variance. Henceforth, I omit the time index t in equations without intertemporal dimensions.

Trading occurs sequentially over periods indexed by t and each period has three stages. First, market makers arrive and post limit orders until an equilibrium is reached where no limit order with a positive expected profit can be added. Then a trader arrives and, if the trader has a strong enough trading signal, submits a market order. Finally, the new fundamental value is announced and the process starts over.

The traders. The trader arriving is randomly drawn from a population of traders. Buyers and sellers are equally likely to arrive and they observe a trading signal that, in the absence of transaction costs, motivates a trade of size $\tilde{m}$. I assume that the market order quantity in the absence of transaction costs is exponentially distributed, with an expected quantity ϕ. The density function $f(\tilde{m})$ is equal for buyers and sellers and is given by

$$f(\tilde{m}) = \frac{1}{2\phi} e^{-\frac{\tilde{m}}{\phi}}. \tag{2}$$

The key innovation of the model relative to that of Sandås (2001) is that the traders may account for transaction costs in their trading decision. Traders who submit a market order pay the effective spread

$$s = D(p - X), \tag{3}$$

where D is a direction of trade indicator that equals +1 for buy market orders and -1 for sell market orders and p is the transaction price. When the trading signal is not strong enough to cover the transaction cost, the trader does not submit a market order. Specifically, the size of the market order submitted in t is given by

$$m = \begin{cases} 0 & \text{if } \tilde{m} \leq \delta s, \\ \tilde{m} & \text{if } \tilde{m} > \delta s, \end{cases} \tag{4}$$

where δ denotes the liquidity demand elasticity ($\delta \geq 0$).

The liquidity demand elasticity assumption is consistent with the empirical findings of, for example, Epps (1976), and the modeling approach of Hendershott and Menkveld (2014). It implies that the expected trade size, conditional on a market order being posted, is $\phi+\delta s$. When the effective spread is symmetric around the fundamental value or when $\delta = 0$, market orders to buy and sell have equal probabilities and equal expected trade sizes. When the effective spread is asymmetric and $\delta > 0$, the side of the limit order book with a smaller spread has a higher probability of incoming market orders and a smaller expected trade size.

The market makers. The market makers know that the traders are potentially informed about the fundamental value innovation d_{t+1} and account for that when forming their expectation of the profit of a limit order. I assume that the liquidity suppliers' expectation of the fundamental value conditional on the arrival of a market order takes the form

$$E[X_{t+1}|X_t, D_t m_t] = X_t + \lambda D_t m_t, \tag{5}$$

where λ is the per unit price impact of market orders, which I assume to be positive.

Given the price impact function in (5), the market makers are able to calculate the effective profit of a one-unit limit order placed at price level P_i in the limit order book, where $i = \{B, A\}$ denotes prices at the best bid and ask prices. Assuming a quantity-invariant and non-negative order processing cost denoted γ,[4] the expected profit of a limit order on the ask side of the book is given by

$$P_A - E[X_{t+1}|X_t, m_t \geq q] - \gamma, \tag{6}$$

where q is the minimum market order size required to execute the limit order in question. The conditional expectation term is the "upper-tail expectation" of Glosten (1994).

[4]A negative order processing cost may, in reality, be possible, for example, due to liquidity rebates earned by market makers. Sandåss (2001) estimates indicate that the order processing costs are negative and he argues that this may be due to agents with heterogeneous beliefs. If order processing costs are negative, the results indicating that the midpoint effective spread overstates the expected effective spread would be strengthened, but the expected effective spread lower bound would not be observable.

1.2 Equilibrium

Because market makers are profit maximizing, the order depth at the best bid and ask prices must be such that the last unit in the queue breaks even in expectation. Following the same steps as Sandås (2001, pp. 715–716), I derive break-even conditions for the depth Q_i posted at the best bid and ask prices. In equilibrium, the depth at the best ask price is

$$Q_A = \frac{P_A - X - \gamma}{\lambda} - \phi - \delta(P_A - X) \tag{7}$$

and the depth at the best bid price is

$$Q_B = \frac{X - P_B - \gamma}{\lambda} - \phi - \delta(X - P_B). \tag{8}$$

The fundamental value. I define the *spread midpoint* as $M = \frac{P_A + P_B}{2}$ and the *midpoint effective spread* as $s^M = \frac{P_A - P_B}{2}$. By solving (7) and (8) for λ and setting the resulting expressions equal to each other, I retrieve the following expression for the fundamental value (see Appendix A for the derivation):

$$X = M + (s^M - \gamma) \underbrace{\frac{Q_B - Q_A}{Q_B + Q_A + 2(\gamma\delta + \phi)}}_{\text{Depth asymmetry } (DA)}. \tag{9}$$

The term marked as depth asymmetry in (9) is, by definition, bounded between -1 and +1, because all the parameters are non-negative and the depth variables are positive.

An important insight from (9) is that the fundamental value is well represented by the spread midpoint when either the order processing cost is high or the depth asymmetry is small (or both). The smaller the order processing cost and the higher the (absolute) depth asymmetry, the less accurate is the midpoint as a proxy for the fundamental value.

The effective spread. By inserting (9) in (3), I obtain

$$s = s^M - D(s^M - \gamma)DA. \tag{10}$$

Upper and lower bounds for the expected effective spread. In equilibrium, the expected effective spread is

$$E(s) = \pi_A E(s|D = 1) + \pi_B E(s|D = -1), \tag{11}$$

where π_A and π_B represent the probability of a market order arriving at the best ask and bid prices, respectively ($\pi_A + \pi_B = 1$).

The upper bound of $E(s)$ is given by the expected midpoint effective spread $E(s^M)$. To see this, note that (10) implies $E(s|D = 1) + E(s|D = -1) = 2E(s^M)$ and that, as discussed above, the probability of market order arrivals is greater on the tighter side of the spread. The upper bound holds when the liquidity demand is inelastic ($\delta = 0$; see Appendix B for details).

The lower bound of $E(s)$ is reached when the absolute value of the second term of (10) is at its maximum, which is when $\gamma = 0$ and ϕ is approaching zero. In this special case, the expected effective spread is approximately

$$E\left[s^M\left(1 - D\frac{Q_B - Q_A}{Q_B + Q_A}\right)\right], \tag{12}$$

which I refer to as the expected *microprice effective spread* and denote $E(s^\mu)$. This name is due to the fact that the fundamental value, in this special case, is approximately $M + s^M \frac{Q_B - Q_A}{Q_B + Q_A}$, which is commonly referred to as the *microprice* (Harris, 2013). In this case, the depth asymmetry is bounded between -1 and +1, implying that the fundamental value can lie anywhere between the best bid and ask prices.

In conclusion, in equilibrium, the expected effective spread is bounded upward by the expected midpoint effective spread and downward by the expected microprice effective spread:

$$E(s^\mu) < E(s) \leq E(s^M). \tag{13}$$

1.3 Discussion

Relative tick size and order processing costs. In the US and European Union (EU) equity markets, the relative tick size varies only due to variation in the trade price. In such tick size

regimes, two otherwise identical securities with different trading prices also have different relative tick sizes. In the US equity market, where all stocks (with a price exceeding USD 1) have a tick size equal to one cent, the relative tick size is largest for low-priced stocks. In EU equity markets, where the tick size increases stepwise with the trade price, the relative tick size is largest for stocks trading just above the threshold levels that trigger tick size increases.

If order processing costs are equal across stocks, which is typically true for stocks traded in the same venues, then variation in the relative tick size should influence the degree of asymmetry in effective spreads. The model predicts that securities with a large relative tick size have larger differences between the midpoint and the microprice effective spreads.

In the fragmented equity markets seen in the US and in the EU nowadays, the fee schedule is an important competitive tool for exchanges (see the empirical evidence of Panayides et al., 2017). Exchanges often charge different fees for the makers and takers of liquidity (liquidity suppliers and demanders). In fact, it is common with negative fees for liquidity suppliers, referred to as maker rebates. The variation in maker fees and rebates introduces variation in order processing costs across venues. The model predicts that lower order processing costs lead to larger differences between the midpoint and the microprice effective spreads.

The empirical investigation in Section 3.3 shows results consistent with the model predictions.

The microprice and liquidity demand elasticity. There are several reasons to believe that the microprice is *not* an accurate proxy for the fundamental value. First, the accuracy of the microprice is decreasing in the order processing costs and the expected size of market order arrivals. Second, the derivation of the microprice builds on equilibrium conditions. If the limit order book is not in equilibrium when the market order arrives, then the depth asymmetry does not reflect the latest information about the fundamental value.

Third, outside the model, market makers may have private values that affect the break-even conditions of their quotes and undermine the accuracy of the microprice. For example, Hendershott and Menkveld (2014) show that market makers post more aggressive quotes on the side of the book where they have undesired inventory. Private values also emerge when liquidity demanders choose to post limit orders instead of market orders (see, e.g., Foucault et al., 2005; Roşu, 2009, for

theoretical models of the order submission decision). In conclusion, the relevance of the expected effective spread lower bound relies on the fundamental value being the dominant determinant of the depth asymmetry used to calculate the microprice.

If the variation in depth asymmetry is not due to the fundamental value, then the distribution of market orders arriving on the bid and ask sides of the book should be independent of the depth asymmetry. I use this observation to assess the liquidity demand elasticity and the accuracy of the microprice jointly in Section 3.1.

Limitations. It is important to recognize the limitations of this class of models. First, the model is static and does not make predictions about the timing of market orders. Furthermore, investors in the model are either traders or market makers; that is, they do not choose their type depending on market conditions, their degree of patience, or other variables. The assumption that traders are sensitive to transaction costs does, however, bring the model one step closer to reality in this regard. Empirical support for that liquidity demand is elastic is provided in Section 3.1.

Second, the model does not say anything about how the equilibrium is reached. The setup may be interpreted as market makers having time to revise their orders optimally between the trader arrivals, implying that the market makers are fast relative to the traders. In current markets, it is common for market makers to be high-frequency traders, whereas traders are frequently represented by human brokers. Even though the brokers are likely to use execution algorithms, the notion that market makers are relatively fast is not unrealistic (see, e.g., evidence presented by Brogaard et al., 2015).

Finally, the model does not account for market fragmentation. For a model incorporating both market fragmentation and differences in trading speed, see van Kervel (2015). In the empirical investigation below, I use limit order book information based on the National Best Bid and Offer (NBBO) of US markets. This ensures that liquidity at all the relevant exchanges is considered in the analysis.

2 Data and Sample

I use the TRTH to access trades and quotes for US equities. The data set is supplied by the Securities Industry Research Centre of Asia-Pacific. For the purpose of sample selection and stock characteristics, I also access data from the Centre for Research in Security Prices (CRSP). For most applications in the paper, the sample consists of the constituent stocks of the S&P 500 index for five trading days, from November 30 to December 4, 2015. To study the time series variation, I also consider a 20-year sample (1996–2015) of common stocks. I refer below to the five-day sample as the *cross-sectional sample* and to the 20-year sample as the *time series sample*.

2.1 Sample Selection

For the cross-sectional sample period, the S&P 500 index consists of 505 stocks, all available in the TRTH. I include trades from all relevant US national securities exchanges.[5] Trades in dark pools and over-the-counter markets are not included. Quotes are taken from the official NBBO feed from the SIPs.

For the time series sample, for each month, I include stocks that meet the following criteria: (i) It is a common stock; (ii) it has its primary listing on the NYSE, NYSE MKT (or AMEX in earlier parts of the sample), or NYSE Arca; (iii) it is listed in the CRSP database; and (iv) it is listed in the TRTH database. Out of the stocks satisfying criteria (i) to (iii), on average, 99.8% also satisfy criterion (iv) and the minimum monthly match rate between the CRSP and TRTH databases is 99.0% (see Appendix C for details on the matching methodology). Due to computational limitations, I restrict the time series of trading days to Wednesdays, following Yueshen (2016). In the 20-year sample, this yields 1044 trading days. As above, I use trades from all relevant exchanges (the set of relevant exchanges varies over time) and quotes from the official NBBO feed.

[5]The national exchanges are the following: Bats BZX Exchange (with TRTH exchange code *BAT*), Bats BYX Exchange (*BTY*), Bats EDGA Exchange (*DEA*, formerly Direct Edge EDGA), Bats EDGX Exchange (*DEX*, formerly Direct Edge EDGX), Chicago Stock Exchange (*MID*), NASDAQ BX (*BOS*, formerly Boston Stock Exchange), NASDAQ PHLX (*XPH*, formerly Philadelphia Stock Exchange), The Nasdaq Stock Market (*NAS*, *THM*), NYSE (*NYS*), NYSE Arca (*PSE*), and NYSE MKT (*ASE*, formerly AMEX).

2.2 Data Quality

The TRTH database is not commonly used for US equity research but it is comparable to the database analyzed by Holden and Jacobsen (2014). For US equities, the TRTH contains consolidated instruments that merge trades taken from the consolidated tape and quotes taken from the official NBBO feed.[6] This is the same data source as for the DTAQ database ((Holden and Jacobsen, 2014, p. 1735)). Holden and Jacobsen (2014) show that the DTAQ is strongly preferable to the monthly version of the database (MTAQ), due to the latter having problems with withdrawn quotes, low time stamp granularity, and canceled quotes. As the TRTH and DTAQ have the same data source, the TRTH does not have those problems.[7,8]

2.3 Data Screening

The following screening is applied to both samples. I include trades that are time stamped between 9:35 AM and 3:55 PM. To avoid opening and closing effects in the measurement of liquidity, the first five minutes and the last five minutes of the trading day are excluded. I exclude block trades, defined as trades of at least 10,000 shares. Additional screens, excluding less than 0.01% of all trades, are described in Appendix C. Each trade observation contains information on the date, stock, time, price, volume, and trading venue.

Retained trades are matched to the last quote observation in force in the preceding millisecond (as recommended by Holden and Jacobsen, 2014). Importantly, the quotes are screened after the quotes and trades are matched. This ensures that the trades are not matched to obsolete quotes. The details of the quote screening, setting around 5% of all matched quotes to missing, are also

[6]The Thomson Reuters support staff has confirmed in a personal communication that the consolidated instrument data sources are the SIPs (for NYSE- and AMEX-listed stocks, the SIP is the Consolidated Tape Association and, for NASDAQ-listed stocks, it is UTP). More information about the TRTH consolidated instruments is available at http://www.sirca.org.au/2011/08/consolidated-instruments-tick-history/.

[7]In addition, I confirm that the TRTH data for the example provided by Holden and Jacobsen (2014, Figure 2) conform to those of DTAQ. See Appendix D.

[8]The time stamps reported to the DTAQ and TRTH databases are given in milliseconds. For TRTH entries before October 23, 2006, however, the time stamps are given in seconds (I am unable to confirm whether this is the case for DTAQ as well). The TRTH also assigns its own time stamps with microsecond granularity when the data are received by Thomson Reuters. Though the TRTH time stamps have higher granularity than the official time stamps, they are subject to a reporting delay. I use the official time stamps when available at millisecond granularity and the internally assigned TRTH time stamps otherwise.

given in Appendix C.

Retained quote observations contain information on the date, stock, time, bid price, ask price, bid volume, ask volume, and the contributing trading venue for the bid and ask volumes. The NBBO presents the volume available at the venue that currently has the largest volume available at the best price. That is, if there are several venues with the same price, it is not the aggregate volume across venues that is reported. This is important because the same liquidity is often cross-posted at several trading venues (e.g., van Kervel, 2015). In addition, the use of NBBO quotes when measuring the effective spread is consistent with Rule 605 in Reg NMS.[9]

3 Liquidity Demand Elasticity and Effective Spread Overestimation

In this section, I first show that liquidity demand is elastic and that variation in the microprice reflects variation in the fundamental value. According to the model, elastic liquidity demand is a sufficient condition for the expected midpoint effective spread to overestimate the true expected effective spread.

Next, I quantify the overestimation and find it to be 16%, on average, across the S&P 500 stocks. I document significant variations across stocks and trading venues. Consistent with the model presented above, I show that the cross-sectional variation is driven by differences in price discreteness (the relative tick size) and order processing costs. Finally, I document that the overestimation of the effective spread falls sharply in response to the tick size reforms of June 1997 and January 2001 and that it increases gradually from 2003 to 2015.

3.1 Liquidity Demand Elasticity and the Relevance of the Microprice

Two key ingredients in the theoretical model of the effective spread are the fundamental value and the liquidity demand elasticity. Empirically, neither of the two is observable. The model shows that, in equilibrium, under certain parameter settings, the fundamental value can be proxied by

[9] https://www.sec.gov/rules/final/34-43590.htm

the microprice. If the microprice is an accurate proxy for the fundamental value and if liquidity demand is elastic, then more market orders should be arriving on the tight side of the spread than on the wide side.

The microprice for stock i in the moment before trade j is defined as

$$\mu_{ij} = M_{ij} + 0.5\Big(\underbrace{P^A_{ij} - P^B_{ij}}_{\text{Quoted spread } (QS_{ij})}\Big) \times \underbrace{\frac{Q^B_{ij} - Q^A_{ij}}{Q^B_{ij} + Q^A_{ij}}}_{\text{Observed depth asymmetry } (DA_{ij})}, \tag{14}$$

where P^A_{ij} and P^B_{ij} are the best bid and ask prices and Q^A_{ij} and Q^B_{ij} represent, respectively, the depths available at those prices. The variable M_{ij} is the bid–ask spread midpoint, defined as $0.5(P^A_{ij} + P^B_{ij})$. Below I refer to $P^A_{ij} - P^B_{ij}$ as the quoted spread, denoted QS_{ij}, and the ratio $\frac{Q^B_{ij} - Q^A_{ij}}{Q^B_{ij} + Q^A_{ij}}$ as the observed depth asymmetry, denoted DA_{ij}.

I assess the liquidity demand elasticity and the accuracy of the microprice jointly by investigating the relation between the observed depth asymmetry and market order arrivals. Under the null hypothesis, market orders arrive independently of the observed depth asymmetry, implying either that the liquidity demand is inelastic or that the microprice is not an accurate proxy for the fundamental value.

I categorize trades by the observed depth asymmetry in the moment before the trade. The variable DA_{ij} is bounded between -1 and +1, and I create 19 categories using the following breakpoints: -1, -0.85, -0.75, ..., 0.05, 0.05, ..., 0.75, 0.85, 1. The first and last categories are labeled *DA–0.9* and *DA0.9*, respectively, and all the other categories are labeled by the midpoint of their interval. In Figure 1, I report the frequency of buyer-initiated trades (i.e., market orders arriving on the ask side) for each trade category, as well as the category share of the total dollar volume.[10] Trades priced higher than the spread midpoint are assumed to be buyer-initiated, and trades priced lower than the midpoint are seen as seller-initiated.

The results show a positive relation between the observed depth asymmetry and the frequency of buyer-initiated trades. For example, consider the case when the observed depth asymmetry is in the *DA –0.9* category, where the microprice effective spread is three times higher for buy

[10] I exclude midpoint trades, which are trades priced exactly at the prevailing spread midpoint.

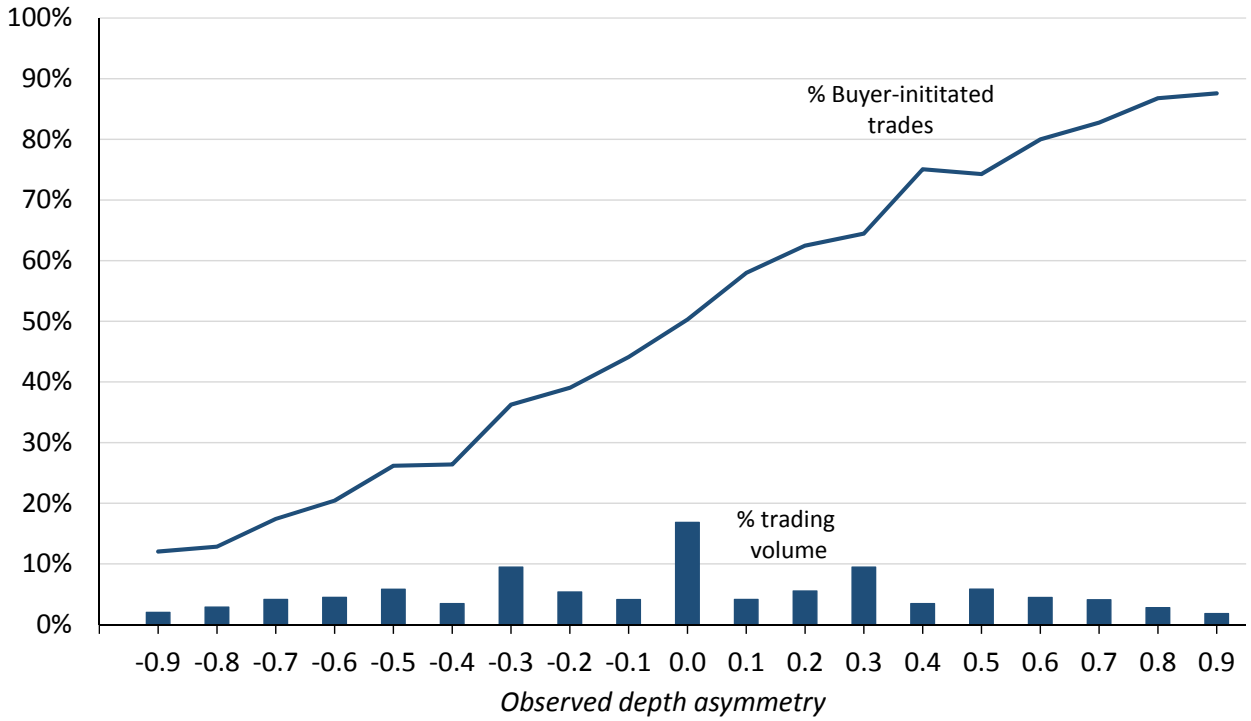


Figure 1: Liquidity demand elasticity in the S&P 500 stocks. This figure shows the frequency of buyer-initiated trades and the dollar volume market shares for different categories of the observed depth asymmetry (*DA*). Midpoint trades are excluded. The *DA* categories are determined by the breakpoints $-1, -0.85, -0.75, ..., 0.05, 0.05, ..., 0.75, 0.85, 1$. The first and last categories are labeled *DA –0.9* and *DA 0.9*, respectively, and all the other categories are labeled by the midpoint of their interval. The sample includes all constituents of the S&P 500 index for the five trading days in the period from November 30 to December 4, 2015

market orders than for sell market orders. For this case, I find that only 26% of all trades are buyer initiated. For trades in the *DA 0.5* depth asymmetry category, in contrast, 74% of the trades are buyer initiated. When the bid- and ask-side depths are in balance (in category *DA 0*), the split between buyer- and seller-initiated trades is even (50% of the trades are buyer initiated).

The depth asymmetry group with the highest trading volume is clearly category *DA 0* (with 17% of the total trading volume), but all the categories record non-negligible volumes (with around 2% in each of the extreme categories).

Figure 1 strongly indicates that liquidity demand is elastic and that variation in the microprice reflects variation in the fundamental value. To investigate the null hypothesis statistically, I model the probability of a trade being buyer initiated as a function of the observed depth asymmetry:

$$Buy_{ij} = \alpha + \beta DA_{ij} + \varepsilon_{ij}, \tag{15}$$

where Buy_{ij} equals one for buyer-initiated trades and zero for seller-initiated trades. Midpoint trades are excluded. Variation that is unexplained by the model is captured by the residual term ε_{ij}.

I estimate the model in (15) using a probit regression. The result shows a strongly significant relation between the direction of trade and the observed depth asymmetry. The β is estimated at 1.35 and the pseudo-R^2 value of the model is 0.152. There are 47.5 million trade observations in the sample, leading to narrow confidence bounds around the coefficient estimate. With a z-statistic of 2985.3, the null hypothesis is strongly rejected. This result indicates that liquidity demand is elastic and that the microprice is useful as a proxy for the fundamental value.[11]

According to the theoretical model, positive liquidity demand elasticity implies that the average midpoint effective spread overstates the expected effective spread. Next, I attempt to quantify the overestimation.

[11]The estimates are tabulated in Table 3), where I also consider extensions of the model.

I also consider potential alternative explanations for the results in Appendix E. The empirical results presented here lend further support to the use of the microprice as a proxy for the fundamental value.

3.2 Overestimation of the Effective Spread

The *upper bound* of the effective spread corresponds to the midpoint effective spread and the *lower bound* is the microprice effective spread. I define the effective spread *overestimation* as the difference between the upper and lower bounds, divided by the lower bound. The definition captures the maximum bias possible when using the midpoint effective spread. I refer to the difference between the average upper and lower bounds as the effective spread *Diameter*.

I measure the midpoint effective spread for all trades j ($j = 1, ..., J$) in each stock i as

$$s_{ij}^{M} = \frac{D_{ij}\left(p_{ij} - M_{ij}\right)}{M_{ij}}, \tag{16}$$

where D_{ij} is a direction of trade indicator that equals +1 when the trade is buyer initiated, -1 when it is seller initiated, and zero for midpoint trades; p_{ij} is the trade price; and M_{ij} is the spread midpoint. The definition is consistent with, for example, the works of Blume and Goldstein (1992) and Lee (1993).

The microprice effective spread follows the same logic, simply replacing the midpoint as the benchmark price with the microprice:

$$s_{ij}^{\mu} = \frac{D_{ij}\left(p_{ij} - \mu_{ij}\right)}{\mu_{ij}}. \tag{17}$$

To retrieve empirical counterparts for the expected effective spread upper and lower bounds, I calculate the dollar volume-weighted average for s_{ij}^{M} and s_{ij}^{μ}, respectively, for each stock.[12] In the time series sample, I calculate monthly dollar volume-weighted averages across all trades in each stock.

Table 1 reports the effective spread properties (upper and lower bounds, diameter, and overestimation) in the cross-sectional sample. The averages across stocks are dollar volume weighted. In addition to the effective spread properties, I report the *relative quoted spread*, which is the quoted spread divided by twice the spread midpoint (for comparability to the effective spread), and two

[12] Before averaging, I winsorize the data by setting observations below the 1% quantile equal to the 1% quantile and observations above the 99% quantile equal to the 99% quantile.

measures of the aggregate trading volume.

The results show that the average effective spread lies in the interval 1.30–1.47 bps. Consistent with the results of Petersen and Fialkowski (1994), the effective spread is lower than the relative quoted spread at 1.62 bps, which may be due to hidden liquidity or price improvements. The average diameter is 0.17 bps and the average overestimation is 15.75%.

The distributional properties reflected by the percentiles reported in Table 1 show that the upper and lower bounds have similar dispersion. The differences for each reported percentile between the two are in the interval 0.09–0.53 bps. This does not tell the whole story, however, because there is also considerable dispersion in the diameter and overestimation. The overestimation ranges from -1.2% for the fifth percentile to 54.36% for the 95th percentile.

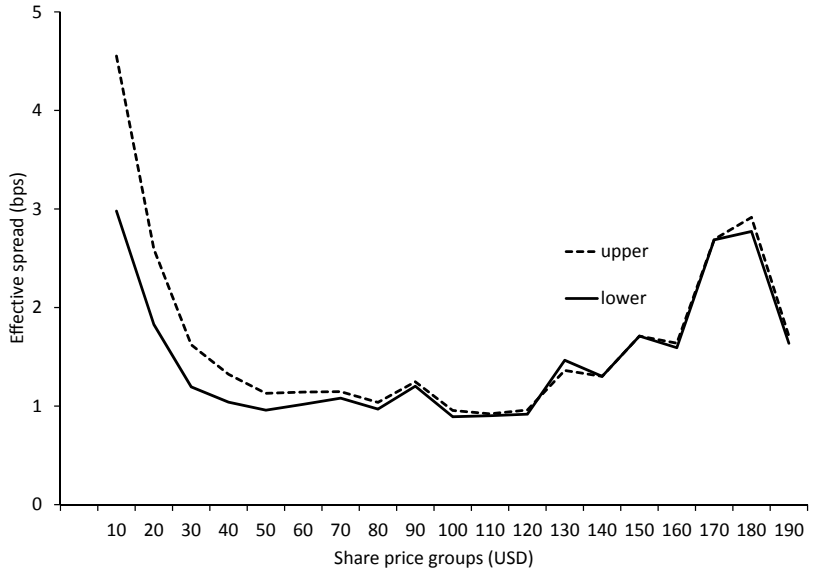
The cross-sectional overestimation variation is important because it implies that the measurement error potentially influences the relative liquidity across stocks. The next section seeks to understand the determinants of overestimation variation across stocks and trading venues.

3.3 Cross-Sectional Determinants of the Effective Spread Overestimation

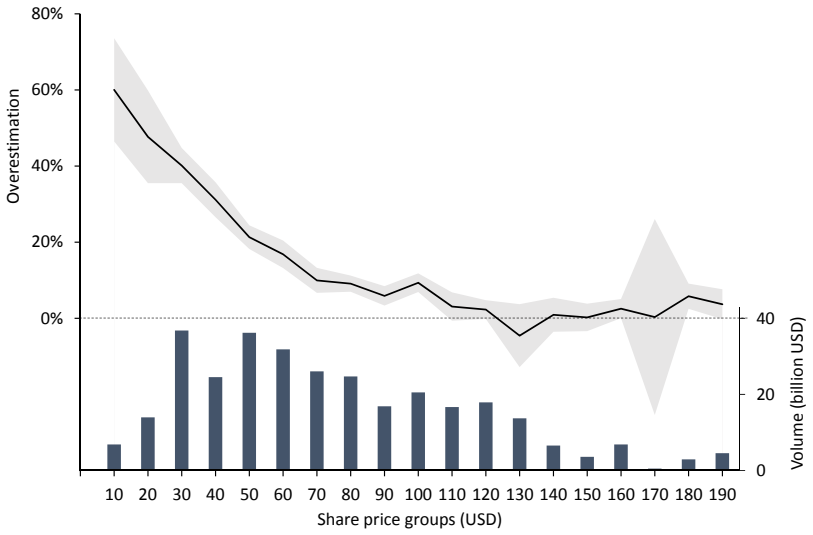
The results in Table 1 indicate that the overestimation of the effective spread varies in the cross section of S&P 500 stocks. The model predicts that stocks with a high relative tick size and venues with high maker rebates have a larger bias. To assess the relation between the overestimation and the relative tick size, I split the sample into share price groups. The *USD 10* group includes all stocks with an average trade price in the USD 5.01–15 interval, the *USD 20* group includes all trades in the USD 15.01–25 interval, and so on with 10-dollar intervals for each price group. The highest price group considered is for *USD 190*, corresponding to trades in the USD 185.01–195 interval. There is no stock in the S&P 500 that has an average price lower than USD 5 and only 4% of the stocks have an average price exceeding USD 195. Figure 2 shows the effective spread upper and lower bounds for each share price group in Panel (a) and the effective spread overestimation in Panel (b). As a point of reference, Panel (b) also shows the fraction of the total trading volume corresponding to each share price group. The effective spread properties are plotted as line charts and the trading volume as a bar chart.

Table 1: Effective spread properties in the S&P 500 stocks. This table shows effective spread properties (upper and lower bounds, diameter, and overestimation) as well the relative quoted spread and the trading volume for the constituents of the S&P 500 index, measured from November 30 to December 4, 2015. The reported statistics are based on stock-level measures of each variable and include the mean, the standard deviation, and the fifth, 25$^{\text{th}}$, 50$^{\text{th}}$, 75$^{\text{th}}$, and 95$^{\text{th}}$ percentiles. The effective spread upper bound is the relative spread between the trade price and the midpoint just before each trade and the lower bound is the relative spread between the trade price and the microprice. The upper and lower bounds are measured for each stock as the dollar-weighted averages across all trades. The effective spread diameter is the difference between the upper and lower bounds for each stock and the overestimation is the diameter divided by the lower bound. The variable *Relative quoted spread* is the quoted spread just before each trade, divided by twice the midpoint and measured for each stock as the dollar-weighted average across trades. The variable *Trade price* is the dollar-weighted average price across all trades for each stock. The mean reported for all the measures above is also dollar weighted across stocks. The volume measures, *Number of trades* (measured in thousands) and *Dollar volume* (measured in millions of US dollars), are reported as equal-weighted averages across stocks.

			Percentiles				
	Mean	Std. Dev.	5$^{\text{th}}$	25$^{\text{th}}$	50$^{\text{th}}$	75$^{\text{th}}$	95$^{\text{th}}$
Effective spread							
Upper bound, s_{ij}^{M} (bps)	1.47	0.99	0.72	1.08	1.41	1.99	3.58
Lower bound, s_{ij}^{μ} (bps)	1.30	0.83	0.63	0.92	1.20	1.75	3.05
Diameter, $s_{ij}^{M} - s_{ij}^{\mu}$ (bps)	0.17	0.35	-0.02	0.05	0.12	0.29	0.84
Overestimation $(s_{ij}^{M} - s_{ij}^{\mu})/s_{ij}^{\mu}$ (%)	15.75	18.46	-1.02	3.90	11.33	26.46	54.36
Rel. quoted spread (bps)	1.62	1.15	0.78	1.15	1.53	2.20	4.08
Trade price (USD)	124.73	101.39	17.02	38.29	60.76	95.61	186.93
Trade volume (thousands)	97.80	84.04	25.11	46.61	74.79	120.33	242.72
Dollar volume (millions)	681.04	877.59	143.53	274.47	437.86	738.39	1921.09



(a) Effective spread upper and lower bounds (s_{ij}^{μ} and s_{ij}^{M})



(b) Effective spread overestimation

Figure 2: Effective spread properties across share price groups in the S&P 500 stocks. Panel (a) shows the upper and lower bounds of the effective spread averaged across stocks in the same share price group. The average effective spread *Overestimation* across stocks in the same share price group and its confidence interval are presented in Panel (b). For variable definitions, see Table 1. The share price groups are formed based on the volume-weighted average price across all trades in the sample for each stock. Each share price group corresponds to a share price interval of USD 10. For example, the USD 10 share price group includes all stocks priced higher than USD 5 and USD 15. Panel (b) also includes the aggregate dollar trading volumes for each trade price category, plotted as a bar chart and measured on the right axis. The sample includes all constituents of the S&P 500 index from November 30 to December 4, 2015.

The share price groups from *USD 30* to *USD 130* span the lion's share of the trading activity (72% of the dollar trading volume and 74% of the stocks in the S&P 500 index). In that interval, the effective spreads are around 1.2 bps, on average. Stocks in the *USD 10* and *USD 20* categories have higher spreads, which may be due to the fact that the minimum tick size is more constraining than for higher-priced stocks. It is also clear from Panel (a) in Figure 2 that the diameter is higher in the low-priced stocks.

Panel (b) of Figure 2 shows the overestimation of the effective spread (solid line) and its 95% confidence interval (shaded area) based on the cross-sectional average across stocks in each trade price group. The results support the notion that stocks with lower prices (higher relative tick size) have a more severe overestimation problem. The price groups *USD 10* and *USD 20* display average overestimations of 60% and 48%. As indicated by the confidence interval, the overestimation is statistically significant for all price groups up to and including *USD 100*, corresponding to two-thirds of the trading volume and 78% of the S&P 500 index stocks. Stocks priced higher than USD 130 have relatively small overestimation problems and their relatively thin volumes make their confidence intervals large.

Next, I investigate the effective spread overestimation in the trading venue dimension. Figure 3, Panel (a), displays the average overestimation in all trades executed in the price interval USD 5.01–75 for each trading venue in the sample. I exclude MID (Chicago Stock Exchange), since it represents only 0.01% of the total trading volume. The results uncover substantial differences across exchanges, with overestimation ranging from 11% for *BOS* (NASDAQ BX) to 86%, on average, for trades executed for *ASE* (NYSE MKT).

According to the model, cross-venue differences may be due to trading fee variation. An important difference in pricing schedules between trading venues is whether they subsidize liquidity suppliers by giving rebates to passively executed trades and charge fees to actively executed trades (referred to as *maker/taker fees*) or vice versa (*inverted fees*). The model predicts that venues with lower fees for liquidity suppliers have larger problems with effective spread overestimation. In Figure 3, venues with inverted fees are indicated by an asterisk (*). Consistent with the model prediction, the three venues with an inverted fee structure tend to have lower overestimation, though

the relation is not monotonic.

In Panel (b) of Figure 3, I report typical fees for makers and takers of liquidity at the 11 venues in the sample. The exchanges offer rich variation in fees, depending on the order type and the status and volume traded of the member in question. The fees reported here are for trades executed using visible orders by members without preferential fee status.[13] Within the Bats exchange group, I observe a monotonic relation between the maker fees and the overestimation seen in Panel (a), which is in line with the model's prediction. The tendency is not consistent across all venues, but I consider the evidence interesting enough to merit a more formal investigation.

I assess the determinants of the effective spread overestimation across stocks (indexed by i) and trading venues (indexed by v) using the following linear regression model:

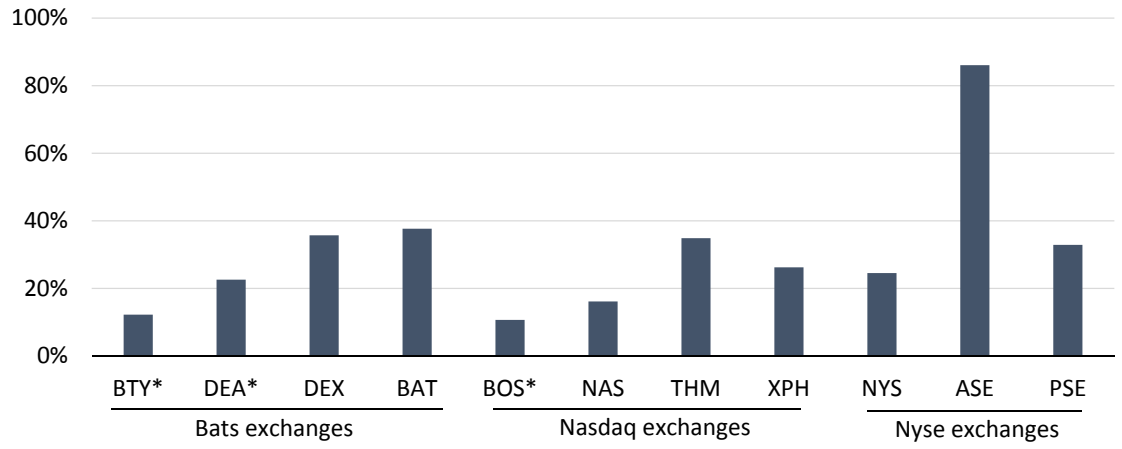
$$Overestimation_{iv} = \alpha + \sum \psi_k S_{k,i} + \sum \nu_l V_{l,v} + \varepsilon_{iv}, \tag{18}$$

where $Overestimation_{iv}$ is the effective spread overestimation for a one stock–venue combination, $S_{k,i}$ are variables indexed by k that vary in the stock dimension only, and $V_{l,v}$ are variables indexed by l that vary in the venue dimension only.

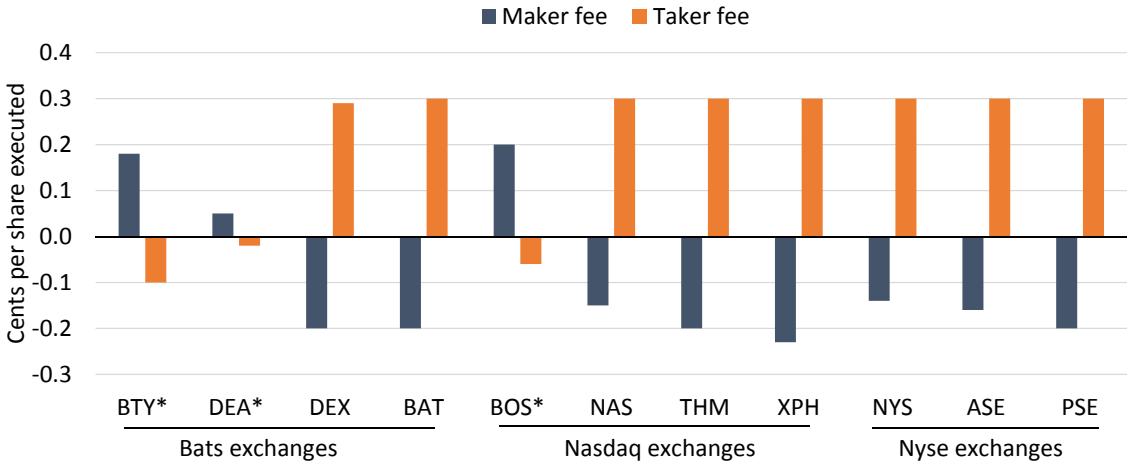
I present ordinary least squares estimates of the model in (18) in Table 2. I consider seven model specifications and include venue (stock) fixed effects when none of the venue (stock) dimension variables are included. The standard errors are clustered on stocks and venues, following the methodology of Petersen (2009).

To investigate the relation between price discreteness and the effective spread overestimation, I consider the variables *RelativeTickSize* (defined as the minimum tick size divided by the value-weighted average price across all trades in each stock) and *QuotedSpread*. The latter variable is motivated by the fact that the minimum tick size is more binding in more liquid stocks. As seen in Table 2, both variables have a significant effect on the overestimation when considered separately (see specifications [1] and [2]). As expected, higher price discreteness and higher liquidity are associated with higher overestimation. When considered in combination, however, the liquidity

[13]The fees are collected from the exchanges' websites. Exact documentation is available from the author upon request.



(a) Overestimation of the effective spread



(b) Trading fees

Figure 3: Overestimation of the effective spread across venues. Panel (a) shows the effective spread *Overestimation* (defined as in Table 1) for each trading venue in the cross-sectional sample, calculated as the dollar-weighted average across all trades executed at a price above USD 5 and below USD 75. The sample includes all constituents of the S&P 500 index for the five trading days from November 30 to December 4, 2015. Panel (b) shows the fees charged to the liquidity suppliers (*Maker fees*) and the liquidity demanders (*Taker fees*) for each exchange. The fees are obtained from the exchange websites and represent the amounts paid by exchange members who are not eligible for special discounts and for trades executed using non-hidden orders. In both panels, exchanges that apply an inverted fee schedule are indicated by an asterisk (*). The exchanges are categorized by their corporate ownership. For the full exchange names corresponding to the three-letter abbreviations, see footnote 5.

Table 2: Determinants of the effective spread overestimation. This table shows ordinary least squares estimates of the cross-sectional determinants of the effective spread *Overestimation*. The variable *Overestimation* is defined as in Table 1 and is measured for all constituents of the S&P 500 index for the five trading days from November 30 to December 4, 2015. The explanatory variables include *QuotedSpread*, which is the bid–ask spread prevailing just before each trade, measured in cents, *RelativeTickSize*, which is the minimum tick size (1 cent) divided by the average trade price; the *Maker/Taker*, which is a dummy indicating exchanges that charge liquidity takers and subsidize liquidity makers; and *MakerRebate*, which measures the rebate given to the liquidity supplier of each trade and is negative for exchanges that charge maker fees. Each estimate is reported along with *t*-statistics (within parentheses) and the superscripts *, **, and *** indicate significance at the 10%, 5%, and 1% confidence levels, respectively. The standard errors are clustered on stocks and venues, following Petersen (2009).

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
[intercept]						−0.021	0.032
						(-0.7)	(1.2)
QuotedSpread	−0.475***		−0.176*			−0.183**	−0.184**
(cents)	(-2.8)		(-1.9)			(-2.0)	(-2.0)
RelativeTickSize		5.499***	5.338***			5.418***	5.421***
×100		(4.2)	(4.2)			(4.2)	(4.2)
Maker/Taker				0.123***		0.124***	
(dummy)				(4.6)		(5.0)	
MakerRebate					0.364***		0.367***
(cents)					(5.8)		(6.2)
Venue FE	Yes	Yes	Yes	No	No	No	No
Stock FE	No	No	No	Yes	Yes	No	No
R^2 incl. FE	0.148	0.366	0.370	0.689	0.694	0.327	0.332
R^2 excl. FE	0.034	0.282	0.286	0.166	0.179	0.327	0.332
Number of obs.	4480	4480	4480	4480	4480	4480	4480

effect is only marginally significant (see specification [3]).

I assess the venue dimension using either a dummy variable taking the value one for venues with a maker/taker fee schedule and zero otherwise (*Maker/Taker*) or a continuous variable reflecting the typical maker rebate at each venue (*MakerRebate*). Both variables are significant at the 5% confidence level when considered separately (see specifications [4] and [5]). In line with the model's prediction, the estimated coefficients indicate that venues with higher maker rebates have higher overestimation. Judging from the R^2 value, I find the added variation of *MakerRebate* relative to the binary *Maker/Taker* variable yields a modest increase inexplanatory power (from 0.166 to 0.179). As can be inferred from Panel (b) of Figure 3, the two are highly correlated, leading me to not consider them in combination.

Finally, I consider the stock-level and venue-level variables in combination (see specifications [6] and [7] in Table 2). In these models, where no fixed effects are included, all the explanatory variables have the expected sign and are statistically significant at the 5% confidence level. Overall, my findings on the determinants of the effective spread overestimation support the predictions of the theoretical model.

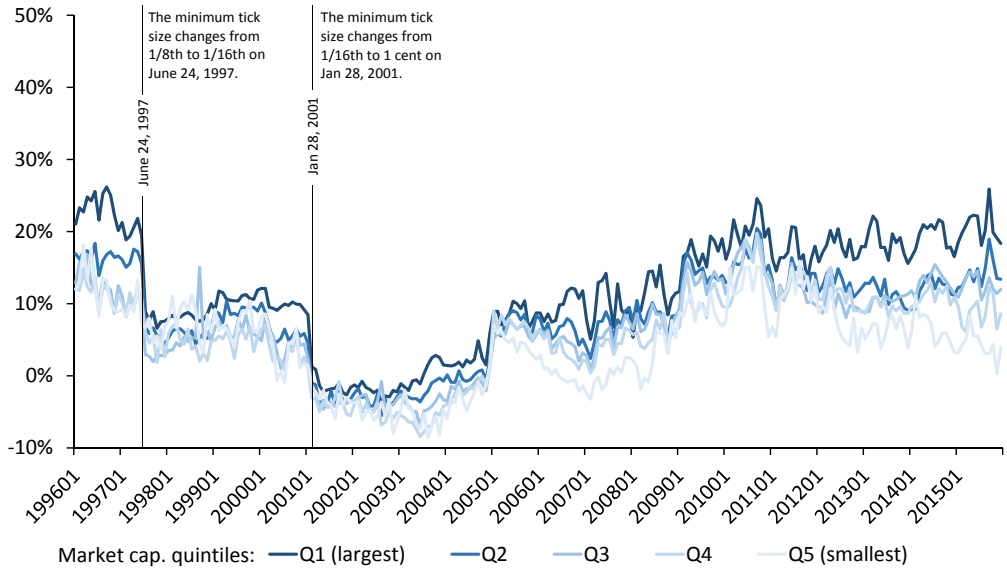
3.4 Time Series Determinants of the Effective Spread Overestimation

To study the effective spread bias over time, I calculate the monthly overestimation for each stock in the time series sample. For each month, I split the sample stocks into size quintiles based on their market capitalization at the end of the previous month. I exclude stocks that have an average trade price lower than USD 5 or higher than USD 1000. Figure 4, Panel (a), plots the overestimation for each time series, where *Q1* and *Q5* are the quintiles with the largest and smallest stocks.

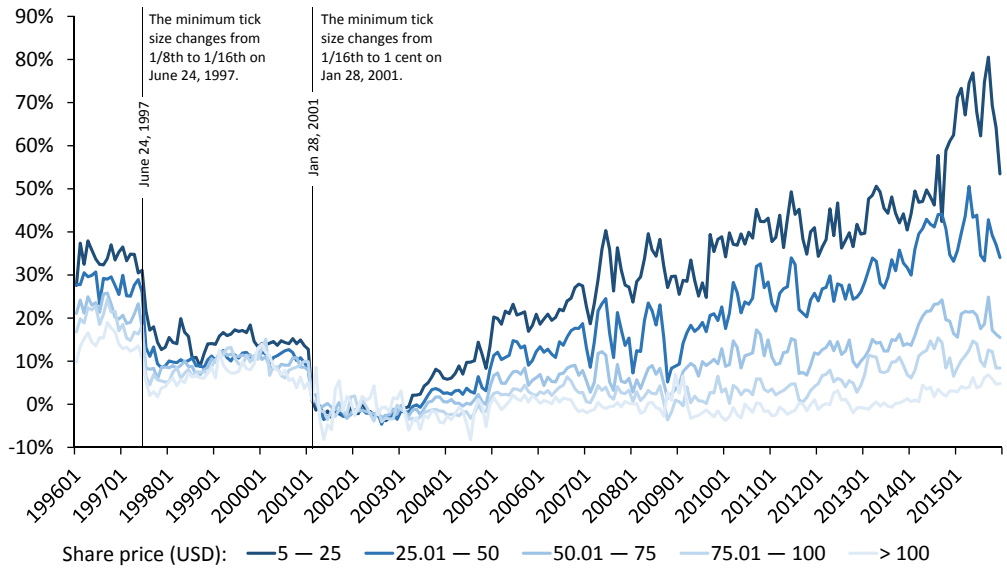
Two observations stand out in the time series plots in Figure 4. First, as indicated by the vertical bars, there are two tick size reforms in the sample: the change from one-eighth increments on the dollar to 1/16 increments on June 24, 1997, and the decimalization implemented on January 28, 2001. These cuts in the minimum tick size coincide with sharp declines in the overestimation.[14]

Second, since early 2003, there is an upward trend in the average overestimation. From being

[14]Liquidity effects of the tick size reforms are analyzed by Goldstein and Kavajecz (2000) and Bessembinder (2003).



(a) Market capitalization quintiles



(b) Price groups within the top market cap. quintile

Figure 4: Effective spread overestimation in the time series, 1996–2015. This figure shows the effective spread overestimation (see the definition in Table 1) for market capitalization quintiles (Panel (a)) and price groups within the quintile of the largest stocks (Panel (b)). Cross-sectional averages are calculated using dollar volume weights. The sample includes all common stocks with primary listings on the NYSE, NYSE MKT (or AMEX), or NYSE Arca, from January 1996 to December 2015.

slightly negative in 2001–2002, the average overestimation of the top market capitalization quintile peaks in September 2015, at 26%. The start of the upward trend may be associated with the introduction of the Autoquote system at NYSE, which, according to Hendershott et al. (2011), leads to increasing liquidity due to enhanced algorithmic trading. Autoquote was introduced from January 29 to May 27 in 2003. The upward trend indicates increasing attention to the asymmetry of the effective spread, particularly, for large-cap stocks.

Next, I split the sample into five share price groups: *USD 5.01–25*, *USD 25.01–50*, *USD 50.01–75*, *USD 75.01–100*, and *USD >100*. Since the effective spread overestimation is more prevalent in the largest stocks, the results presented in Figure 4, Panel (b), are restricted to the top quintile of stocks (*Q1*). Throughout the sample, I find that stocks with lower average trade prices tend to have higher overestimation in the effective spread. The overestimation of the price groups *USD 5–25* and *USD 25.01–50* peak at 81% and 51%, respectively. This is consistent with the results on relative tick size presented above.

The results presented so far establish that the use of the spread midpoint as a proxy for the equilibrium implies a substantial effective spread overestimation. In the next section, I propose that the microprice effective spread is a viable solution to overcome the bias.

4 An Alternative Effective Spread Measure

What should investors and researchers do to overcome the overestimation of the effective spread? The unobservable nature of the fundamental value makes exact measurement infeasible and the evaluation of proxies difficult.

I argue that the microprice effective spread is an attractive alternative to the midpoint effective spread. My motivation is threefold. First, the microprice is already established as a fundamental value proxy in the financial industry. Second, the computational cost is on par with that of the midpoint effective spread and the data required are readily available. Third and most importantly, the model in this paper provides theoretical support for taking the depth asymmetry into account when measuring the effective spread.

Accepted in the financial industry. Few academic papers discuss the microprice as a proxy for the fundamental value. Harris (2013) provides an alternative theoretical motivation. Lipton et al. (2013) and Avellaneda et al. (2011) study the depth asymmetry as a determinant of trade price movements. Nevertheless, the concept is not new to market practitioners. For example, Lipton et al. (2013, p. 2) write,

> A common intuition among market practitioners is that the order sizes displayed at the top of the book reflect the general intention of the market. When the number of shares available at the bid exceeds those at the ask, participants expect the next price movement to be upwards, and inversely, for the ask.

Related to the ideas presented in this paper, Cartea et al. (2015, p. 71) suggest that the microprice would potentially be a more economically meaningful benchmark than the midpoint when accounting for the effective spread in algorithmic trading strategies.

Cheap. Relative the midpoint effective spread, the additional data required to calculate the microprice effective spread are the depth posted at the best bid and ask prices. Such data are available to investors through the SIP consolidated feeds. For academics, the depth data are available in the major databases used for intraday liquidity analysis, such as the DTAQ and TRTH databases. In terms of computational costs, the additional operation required to adjust the midpoint for depth asymmetry is negligible with modern hardware.

Theoretically founded and self-correcting. The use of depth asymmetry to proxy for the fundamental value is supported by the model in this paper. A key question, however, is whether use of the effective spread lower bound is superior to that of the upper bound. Does this not merely lead to risk of underestimation instead of overestimation? I argue that this concern is mitigated by a "self-correcting property" of the microprice effective spread.

The microprice is not an exact measure of the fundamental value. For example, the higher the order processing costs are, the noisier is the microprice as a fundamental value proxy. If investors use the microprice to infer the fundamental value, the noise induces bias in the microprice effective

spread, which then underestimates the true effective spread. If, on the other hand, investors use other sources of information to gauge the fundamental value, they deemphasize the microprice in environments where it is less accurate.

For example, consider a security where the midpoint equals the fundamental value but the depth asymmetry is positive due to factors unrelated to the fundamental value. Investors who use the microprice to gauge the fundamental value are then more inclined to submit market orders on the ask side, where the spread appears to be tighter. They underestimate the spread and so does the econometrician who measures the average microprice effective spread. In contrast, if investors are able to correctly estimate the fundamental value, the depth asymmetry does not influence their trading decision. In the example, they are then equally likely to submit market orders to buy and to sell. Accordingly, the microprice effective spread is equal to the midpoint effective spread. The misleading depth asymmetry in the microprice is effectively overcome by the market order arrivals, mitigating the risk of underestimating the effective spread.

To investigate empirically whether investors rely on the microprice in their trading decisions or if they factor in other information sources to estimate the fundamental value, I consider the differences in fee schedules across trading venues. According to the theoretical expression in (9), the higher the fees charged to liquidity suppliers, the less accurate is the microprice as a proxy for the fundamental value. The quote data that I access identify the venues contributing the orders at the NBBO. I use this to analyze if the trading decisions are less related to the microprice when venues that charge higher fees to liquidity suppliers are represented at the NBBO.

Specifically, I record for each trade whether the prevailing NBBO quotes are provided by one of the exchanges with inverted fee schedules (*BTY*, *DEA*, and *BOS*). I create a dummy variable $InvertedFeeQuotes$ that indicates trades matched to NBBO quotes where the inverted fee venues are present on either the bid or the ask side or both. I hypothesize that the direction of trade has a weaker relation to the observed depth asymmetry when the quotes matched to the trade are supplied by venues that apply inverted fees.

The $InvertedFeeQuotes$ dummy flags about 5% of all trades in my sample. This is low relative to the aggregate dollar volume market share of the inverted fee venues, which is around 13%.

Given that liquidity provision is relatively costly at the such venues, however, the low incidence of $InvertedFeeQuotes$ is not surprising.

I plot the probability of buyer-initiated market orders for different categories of observed depth asymmetry in Figure 5, with the sample split by the $InvertedFeeQuotes$ dummy. Except for the sample split, this figure is exactly the same as in Figure 1. The dark line is for trades matched to quotes where inverted fee venues are present on at least one side of the book and the light line represents all the other trades.

The results are in line with my hypothesis. When either side of the NBBO quotes are contributed by a venue applying an inverted fee schedule, investors adhere less to the microprice in their market order submissions. For example, consider the observed depth asymmetry category *DA –0.5*, where the ask-side spread is three times larger than the bid-side spread. In this category, 34% of the trades flagged by the $InvertedFeeQuotes$ dummy are buyer initiated, meaning that they pay the ask-side spread (the wide side). For the rest of the trades in the same depth asymmetry category, only 26% are buyer initiated.

The evidence points to investors factoring in other sources than depth asymmetry in their estimates of the fundamental value. When the microprice is less accurate, the investors are not misled. By submitting market orders more independently of the microprice when it is noisier, investors counteract the effective spread underestimation. I refer to this as a self-correcting property of the microprice effective spread. No such mechanism is in place to mitigate overestimation of the midpoint effective spread.

I consider a probit model to assess the significance of the self-correcting property. I extend the model in (15), which relates to the direction of the trade and the observed depth asymmetry, by adding an interaction term of the observed depth asymmetry and the $InvertedFeeQuotes$ dummy. To control for differences across stocks, I also add interactions between the observed depth asymmetry and the same stock-specific variables as considered above, $RelativeTickSize$ and $QuotedSpread$. I estimate the following model:

$$Buy_{ij} = \alpha + \beta DA_{ij} + \sum \phi_k S_{k,i} DA_{ij} + \nu InvertedFeeQuotes_{ij} DA_{ij} + \varepsilon_{ij}, \tag{19}$$

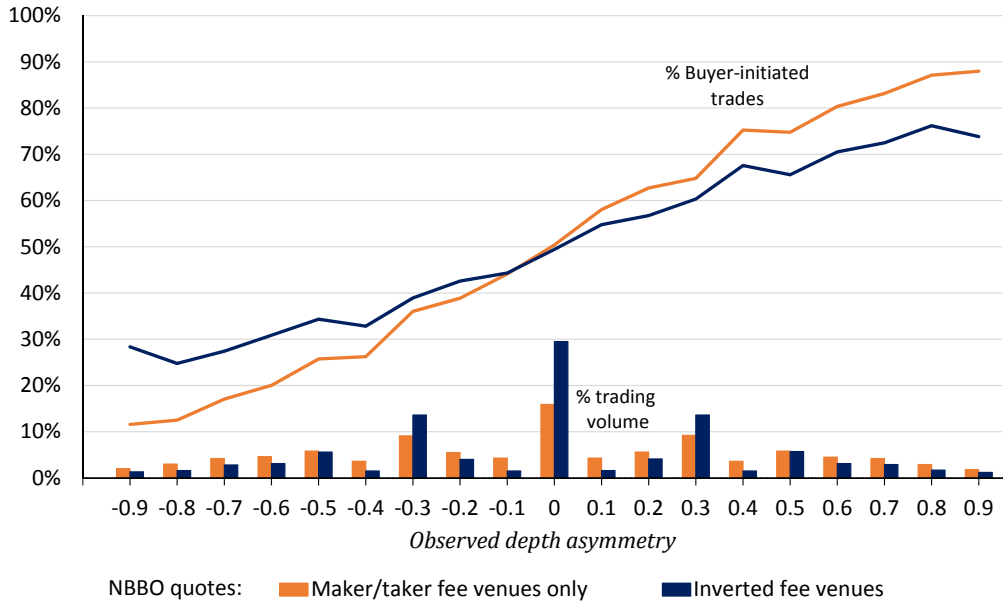


Figure 5: Liquidity demand elasticity depending on the venue types that supply the NBBO quotes. This figure shows the frequency of buyer-initiated trades and dollar volume market shares for different categories of observed depth asymmetry (*DA*) for two categories of trades: those matched to NBBO quotes where at least one side is contributed by a venue that applies inverted fees and those matched to quotes where only maker/taker fees are represented. The *DA* categories are determined in the same way as in Figure 1. The line charts show the percentage of trades for the given *DA* category that are buyer initiated. The bar charts show the fraction of all dollar trading volume that falls in the given *DA* category. The sample includes trades in all S&P 500 stocks for the five trading days from November 30 to December 4, 2015, excluding midpoint trades.

where Buy_{ij}, DA_{ij}, and $InvertedFeeQuotes_{ij}$ are at the stock trade frequency and $S_{k,i}$ varies across stocks only. I exclude midpoint trades. Variation that remains unexplained by the model is captured by the residual term ε_{ij}.

The probit regression estimates are presented in Table 3. Specification [1] is the model outlined in Section 3.1, which here works as a benchmark model with the effect of the interaction terms fixed at zero. Specification [2] controls for stock-level variation by including the interaction terms of *DA* with *QuotedSpread* and *RelativeTickSize*. The coefficient estimates of the interaction terms have the expected sign and are highly statistically significant. Due to the high number of observations in this analysis, 47.5 million trades, the standard errors are low and the z-statistics are high.

The specification of primary interest is specification [3], which includes the $InvertedFeeQuotes$ dummy. The coefficient of $InvertedFeeQuotes$ is negative and significantly different from zero, indicating that traders adhere less to the microprice when it is expected to be less accurate. This result supports the self-correcting property of the microprice effective spread.

5 Implications for Investors

The overestimation of effective spreads has important implications for investors. In this section, I show that (i) traders who overlook the effective spread asymmetry ignore a substantial share of the total liquidity variation and, as a result, overpay for liquidity; (ii) the effective spread reports mandated by Reg NMS may be misleading for investors' order routing decisions; (iii) the use of midpoints in effective spread decompositions leads to an overstatement of the adverse selection risk incurred to liquidity suppliers; and (iv) investors who optimize their portfolios weights and re-balancing frequency with respect to transaction costs or form portfolios based on liquidity sortings are influenced by the overestimation.

5.1 Liquidity Timing

A trader with elastic liquidity demand submits more market orders when it is cheap to do so and less when it is expensive. Successful liquidity timing thus depends on liquidity varying with time

Table 3: Determinants of the direction of trade in relation to observed depth asymmetry. This table shows the probit regression estimates of the cross-sectional determinants of the relation between the direction of trade and observed depth asymmetry. The dependent variable in all specifications is a dummy variable *Buy* that is one for buyer-initiated trades and zero for seller-initiated trades. The observed depth asymmetry, *DA*, is defined as in (14). The variables *Buy* and *DA* are measured for all non-midpoint trades in the S&P 500 stocks for the five trading days from November 30 to December 4, 2015. The first specification estimates the general relation between *Buy* and *DA* and the other two add interaction terms for *DA*. The interaction variables include *QuotedSpread* and *RelativeTickSize*, defined as in Table 2, and the *InvertedFeeQuotes* dummy, which indicates when either the best bid or ask volume at the NBBO (or both) originates from a venue that applies an inverted fee schedule. Each estimate is reported along with z-statistics (in parentheses) and the superscripts *, **, and *** indicate statistical significance at the 10%, 5%, and 1% confidence levels, respectively. The pseudo-R^2 value is based on McFadden (1974).

	[1]	[2]	[3]
[intercept]	0.01^{***}	0.01^{***}	0.01^{***}
	(55.4)	(56.1)	(56.3)
DA	1.35^{***}	1.04^{***}	1.06^{***}
	(2989.6)	(1244.0)	(1252.4)
DA interact. terms			
QuotedSpread		-3.37^{***}	-3.25^{***}
(cents)		(-310.1)	(-300.1)
RealtiveTickSize		14.39^{***}	14.12^{***}
$\times 100$		(576.2)	(564.9)
InvertedFeeQuotes			-0.33^{***}
(dummy)			(-150.2)
Pseudo R^2	0.153	0.163	0.163
Millions of obs.	47.5	47.5	47.5

and the trader's ability to observe and react to such variation.

A trader who gauges the fundamental value by the spread midpoint effectively overlooks any fundamental value variation that does not cause a change in the spread midpoint. To see by how much this undermines the trader's liquidity timing ability, I analyze the components of the microprice effective spread variance.

The microprice effective spread variance may be decomposed as follows:

$$var(s_{ij}^{\mu}) \approx var(s_{ij}^{M}) + var(s_{ij}^{\mu} - s_{ij}^{M}) + 2cov(s_{ij}^{M}, s_{ij}^{\mu} - s_{ij}^{M}), \tag{20}$$

where the first component is the midpoint effective spread variance, the second component is the variation due to the difference between the microprice and the midpoint effective spread, and the third is the covariance between the midpoint effective spread and the difference between the microprice and the midpoint effective spreads.[15]

I calculate each component of the microprice effective spread variance using volume-weighted variances across all trades in each stock in the cross-sectional sample. In Table 4, I report the volume-weighted averages across the stocks for each component, as well as the total variance. I separate the effective spreads paid by buyers and sellers (Panels (a) and (b), respectively), because the variance would otherwise include switches from ask-side to bid-side market orders and vice versa that potentially face different effective spreads. This result is also consistent with buyers, for example, primarily monitoring ask-side liquidity variation; they are not directly influenced by bid-side trades.

The first line of each panel in Table 4 reports the results for all stocks in the S&P 500 index. The results show that an investor who ignores the deviations from the midpoints overlooks about one-third of the total effective spread variance (see the ratio reported in the rightmost column). Such an investor trades at the same rate when the microprice indicates that liquidity is cheap as when it indicates liquidity as expensive. The consequence is that the investor overpays for liquidity.

[15]The decomposition is approximate because $s_{ij}^{M} = \frac{D_{ij}(p_{ij}-M_{ij})}{M_{ij}} \approx \frac{D_{ij}(p_{ij}-M_{ij})}{\mu_{ij}}$. I empirically find that the difference between $var(s_{ij}^{M})$ and $var(\frac{D_{ij}(p_{ij}-M_{ij})}{\mu_{ij}})$ is negligible (they differ by less than 0.1%). Furthermore, the difference $s_{ij}^{\mu} - s_{ij}^{M}$ can be written equivalently as the signed difference between the microprice and the spread midpoint relative to the microprice. In accordance with (14), the difference between the microprice and the midpoint equals $0.5QS_{ij}DA_{ij}$.

Table 4: Effective spread variance decomposition. This table shows the variance of the microprice effective spread, $var(s_{ij}^{\mu})$, as well as its components, namely, the variance of the midpoint effective spread, $var(s_{ij}^{M})$; the variance of the difference between the microprice and midpoint effective spreads, $var(s_{ij}^{\mu} - s_{ij}^{M})$; and twice the covariance between s_{ij}^{M} and $s_{ij}^{\mu} - s_{ij}^{M}$, $2cov(s_{ij}^{M}, s_{ij}^{\mu} - s_{ij}^{M})$. The variance measures are calculated separately for ask-side and bid-side trades for each stock, using volume weights across trades. The covariance term is calculated as $var(s_{ij}^{\mu}) - var(s_{ij}^{M}) - var(s_{ij}^{\mu} - s_{ij}^{M})$. The stock-level observations are then reported as equal-weighted averages across all stocks, as well as for stocks grouped by their share prices. Five share price groups are considered: *USD ≤25*, *USD 25.01–50*, *USD 50.01–75*, *USD 75.01–100*, and *USD >100*. The ask-side results are presented in Panel (a) and the bid-side results are in Panel (b). For each stock segment, the ratio of the average $var(s_{ij}^{\mu})$ and the average $var(s_{ij}^{M})$ is also reported. The analysis is based on five trading days for the S&P 500 stocks, from November 30 to December 4, 2015.

(a) Ask-side market orders

	$var(s_{ij}^{\mu})$	$var(s_{ij}^{M})$	$var(s_{ij}^{\mu}-s_{ij}^{M})$	$2cov(s_{ij}^{M}, s_{ij}^{\mu} - s_{ij}^{M})$	$\frac{var(s_{ij}^{\mu})}{var(s_{ij}^{M})}$
All stocks	2.56	1.68	1.03	-0.16	65.9%
Share price groups (USD)					
< 25	3.77	1.31	2.71	-0.25	34.8%
25.01–50	1.82	1.14	0.66	0.02	62.8%
50.01–75	1.78	1.31	0.54	-0.06	73.3%
75.01–100	1.98	1.47	0.67	-0.16	74.2%
>100	4.25	3.09	1.59	-0.44	72.9%

(b) Bid-side market orders

	$var(s_{ij}^{\mu})$	$var(s_{ij}^{M})$	$var(s_{ij}^{\mu}-s_{ij}^{M})$	$2cov(s_{ij}^{M}, s_{ij}^{\mu} - s_{ij}^{M})$	$\frac{var(s_{ij}^{\mu})}{var(s_{ij}^{M})}$
All stocks	2.85	1.95	1.06	-0.17	68.5%
Share price groups (USD)					
< 25	4.46	2.07	2.70	-0.32	46.5%
25.01–50	1.79	1.23	0.65	-0.09	68.8%
50.01–75	1.99	1.48	0.54	-0.03	74.1%
75.01–100	2.29	1.74	0.69	-0.13	75.9%
>100	4.85	3.48	1.75	-0.39	71.8%

Because the results above show that the difference between the midpoint and the microprice spread is largest in low-priced stocks, I also break down the variance decomposition by share price. The results are consistent with the evidence above. In low-priced stocks (USD <25), the investor who overlooks microprice deviations from the midpoint misses 53.5% (65.2%) of the total bid-side (ask-side) liquidity variation. Notably, for ask-side market orders, the variance due to differences between the midpoint and the microprice is more than twice the variance that is due to changes in the spread midpoint.[16]

5.2 Order Routing and Rule 605

The merit of the effective spread as a measure of execution quality is reflected by the US market regulation Reg NMS. According to Rule 605, all exchanges must report the effective spread for each security monthly. The SEC (2001, Section I)) motivates the disclosure requirement as follows:

> In a fragmented market structure with many different market centers trading the same security, the order routing decision is critically important, both to the individual investor whose order is routed and to the efficiency of the market structure as a whole. The decision must be well-informed and fully subject to competitive forces.

The definition of the midpoint effective spread used in this paper is consistent with that of the average effective spread mandated by Rule 605, using the NBBO midpoint as a reference point and volume weights when averaging across trades.

But how useful is the midpoint effective spread for investors' order routing decisions? Given the results presented above, showing that differences in the fee schedule across venues influence the accuracy of the midpoint effective spread, the effective spreads reported according to Rule 605 may be misleading.

I compare the ranking of exchanges in terms of effective spreads measured using the midpoint and the microprice. I present the results in the "ladder charts" in Figure 6. When the two measures

[16]Liquidity timing may alternatively be performed by monitoring the quoted spread, which is an ex ante measure of the cost of immediacy. The reasoning in this ection applies for a quoted spread defined relative to the microprice too. The main result of the paper, however, does not extend to the quoted spread, because the quoted spread is time weighted, whereas the effective spread is trade weighted.

produce the same ranking of the exchanges, the chart consists of a stack of horizontal lines, like a ladder.

Panel (a) reports the results averaged across all stocks in the S&P 500 index. It shows that, whereas the THM (part of Nasdaq) is the best venue in terms of the midpoint effective spread, ASE (NYSE MKT) is the top contender according to the microprice effective spread. Note that this result isaveraged across all S&P 500 stocks. The average effective spread reports mandated by Reg NMS are at a lower level of aggregation (stock–month frequency) and are thus likely to show even more differences in rankings.[17]

I disaggregate the venue rankings by share price groups. Panel (b) of Figure 6 reports the rankings averaged across stocks with an average price lower than USD 25, a group with a high overestimation of the effective spread. The result is striking: ASE is ranked as the worst of all exchanges in terms of the midpoint effective spread but, again, as the venue with the highest execution quality according to the microprice effective spread. Out of the 11 venues considered, only two have the same ranking for both measures. The rankings are more in sync for higher-priced stocks, but discrepancies prevail in all share price groups (see Panels (c)-(f)).

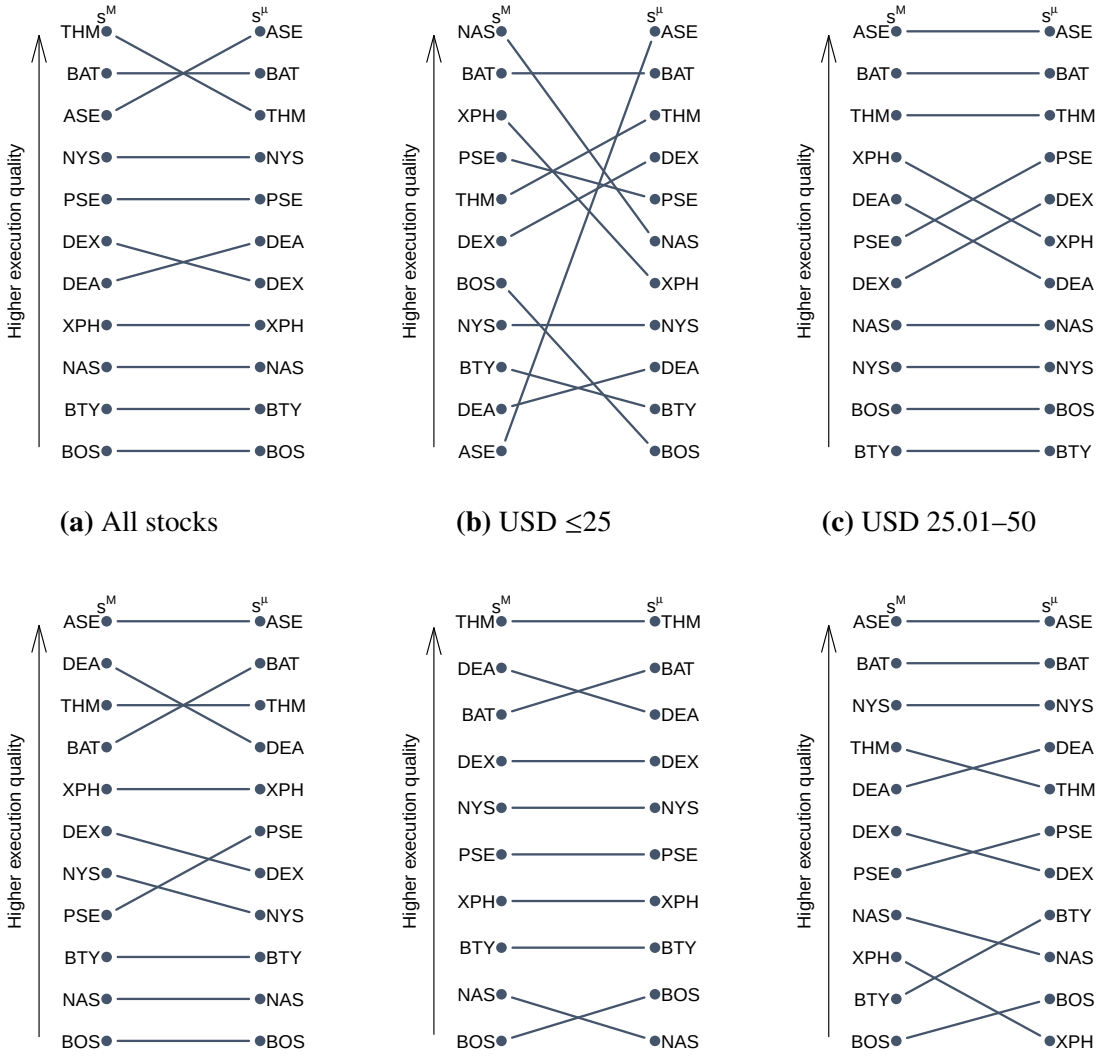
The conclusion from this application is that investors who base their order routing decision on the effective spreads reported by exchanges in accordance with Rule 605 are potentially misguided. The result is in sharp contrast with the ambition of Rule 605, that the order routing decision must be well informed (see the SEC quote above).

5.3 Liquidity Supplier Performance Evaluation

From the liquidity supplier point of view, the effective spread is a source of revenue that should cover the costs associated with market making. As modeled in Section 1, the costs include adverse selection costs and order processing costs. Outside the model, for example, inventory costs should also be covered.

To evaluate the performance of liquidity suppliers, it is common to decompose the effective spread into the realized spread and the price impact. The price impact is a measure of how much

[17]The sample used here spans five trading days, whereas the Rule 605 are for a month. The difference is unlikely to influence the results, because the overestimation is in general not mitigated by averaging across more trades.



(a) All stocks **(b)** USD ≤25 **(c)** USD 25.01–50

(d) USD 50.01–75 **(e)** USD 75.01–100 **(f)** USD >100

Figure 6: Effective spread venue ranks across price groups of S&P 500 stocks. This figure shows how venues are ranked based on their effective spread, depending on whether the midpoint or microprice is used as a benchmark for the effective spread. In each panel, the rank based on the midpoint effective spread is shown on the left-hand side, with the highest-ranked venue (lowest effective spread) at the top of the rank and the lowest ranked venue at the bottom. The ranking based on the microprice effective spread is shown on the right-hand side. Panel (a) shows the results for all stocks in the cross-sectional sample, with the average effective spreads calculated using dollar volume weights. Panels (b) to (f) show the corresponding results for subsets of stocks based on their volume-weighted average price levels: *USD ≤25*, *USD 25.01–50*, *USD 50.01–75*, *USD 75.01–100*, and *USD >100*. For variable definitions, see Table 1. The sample includes all constituents of the S&P 500 index from November 30 to December 4, 2015.

the market maker is losing to the liquidity demander due to changes in the fundamental value in a period after the trade. The realized spread is the remaining part of the effective spread, which should cover the costs of market making, once the price impact is accounted for.

Following the principles of Huang and Stoll (1996), I decompose the microprice effective spread according to

$$s_{ij}^{\mu} = \underbrace{\frac{D_{ij}\left(p_{ij} - \mu_{ij}^{+5\text{ min}}\right)}{\mu_{ij}}}_{\text{Microprice realized spread } (rs_{ij}^{\mu})} + \underbrace{\frac{D_{ij}\left(\mu_{ij}^{+5\text{ min}} - \mu_{ij}\right)}{\mu_{ij}}}_{\text{Microprice price impact } (pi_{ij}^{\mu})}, \tag{21}$$

where $\mu_{ij}^{+5\text{ min}}$ represents the microprice prevailing five minutes after trade j in stock i. I denote the *microprice realized spread* as rs_{ij}^{μ} and the *microprice price impact* as pi_{ij}^{μ}.

I now analyze how overestimation of the effective spread carries over to the realized spread and the price impact. I measure the microprice effective spread components as in (21) and the components of the midpoint effective spread using the same formula with spread midpoints in place of the microprice. I use the trades in the cross-sectional sample, excluding those where no order book information is available five minutes after the trade. The results are reported, on average, for all stocks, as well as for the five share price groups used above (see Table 5).

The results for all the stocks (presented in the first row of Table 5) show that overestimation of the effective spread carries over to the price impact. The nominal difference between the upper and lower bounds of the effective spread is 0.17 bps and exactly the same difference is seen for the price impact. In relative terms, the overestimation is lower for the price impact (10.4%) than for the effective spread (15.7%), because the magnitude of the price impact is higher.[18] For stocks priced lower than or equal to USD 25, the nominal overestimation for both the effective spread and the price impact is about 1 bps (51.8% and 31.0%, respectively, in relative terms). As for the effective spread, the evidence indicates that overestimation of the price impact is negatively related to the share price.

The price impact overestimation is likely related to the possibility that an incremental funda-

[18] Subsequent to the spread decomposition, I winsorize the realized spread and the price impact in the same way as described above for the effective spread. For this reason, the spread components do not sum exactly to the effective spread.

Table 5: Effective spread decomposition. This table shows the average effective spread, realized spread, and price impact calculated using either the microprice or the spread midpoint. It also reports the percentage overestimation resulting from using the midpoint instead of the microprice when calculating the effective spread and the price impact. The results are reported as volume-weighted averages across all stocks, as well as for stocks grouped by their share price. Five share price groups are considered: *USD ≤25*, *USD 25.01–50*, *USD 50.01–75*, *USD 75.01–100*, and *USD >100*. The analysis is based on five trading days for the S&P 500 stocks, from November 30 to December 4, 2015.

	Microprice			Midpoint			Overestimation	
	s_{ij}^{μ} (bps)	rs_{ij}^{μ} (bps)	pi_{ij}^{μ} (bps)	s_{ij}^{M} (bps)	rs_{ij}^{M} (bps)	pi_{ij}^{M} (bps)	$\frac{s_{ij}^{\mu} - s_{ij}^{M}}{s_{ij}^{\mu}}$	$\frac{pi_{ij}^{\mu} - pi_{ij}^{M}}{pi_{ij}^{\mu}}$
All stocks	1.30	-0.39	1.70	1.47	-0.38	1.87	15.7%	10.4%
Share price groups (USD)								
≤ 25	2.20	-1.45	3.67	3.24	-1.41	4.68	51.8%	31.0%
25.01–50	1.12	-0.65	1.78	1.46	-0.65	2.13	34.3%	22.6%
50.01–75	1.01	-0.33	1.35	1.13	-0.33	1.47	15.9%	11.1%
75.01–100	1.03	-0.29	1.33	1.09	-0.29	1.40	8.2%	6.3%
> 100	1.58	-0.11	1.71	1.58	-0.12	1.71	0.9%	0.5%

mental value change can trigger a discrete change in the spread midpoint. This happens when all limit orders posted at the best bid or ask price are either consumed or withdrawn. Consider an example where the fundamental value is close to the best bid price. According to the model, the bid-side depth is then relatively low and the bid-side market order arrival rate is relatively high. It is then relatively likely that all the volume at the best bid price is consumed. The ensuing discrete change in the spread midpoint can make the price impact of the trade appear dramatic. The microprice, in contrast, adjusts gradually to the change in fundamental value.

I find that the realized spread is virtually unaffected by the choice of fundamental value proxy.[19] The reason is probably that it benchmarks the transaction price to the fundamental value five minutes later. For the effective spread, the driving force behind the overestimation is that the trade flows depend strongly on the distance between the current fundamental value and the transaction price. For this effect to carry over to the realized spread, the difference between the microprice and the midpoint must persist over the horizon used. My evidence indicates that the difference between

[19]This is seen by comparing the nominal results. Because the realized spread is frequently close to zero, I do not report the percentage overestimation, which can be strongly influenced by outliers.

the microprice and the midpoint does not persist over a five-minute horizon.

The takeaway from this analysis is that the ability to avoid adverse selection costs (minimize price impact) is overestimated when measured as the change in spread midpoints over a given horizon. If, however, the evaluation is focused on the revenue net of price impact, the realized spread, the use of spread midpoints is less misleading.

5.4 Portfolio Selection

The overestimation in the effective spread has implications for portfolio selection and rebalancing. Amihud and Mendelson (1986) identify a clientele effect where assets with higher illiquidity are held by investors with longer investment horizons. Investors may thus underweight assets with overestimated illiquidity. Constantinides (1986) and Dumas and Luciano (1991) show how illiquidity influences portfolio rebalancing. According to their models, more illiquid assets are allowed to deviate more from the optimal allocation. If illiquidity is overestimated, it can thus erroneously discourage investors from rebalancing their portfolios to the efficient frontier.

In asset pricing and corporate finance studies, it is common to study the properties of portfolios that are sorted by stock illiquidity (e.g., Acharya and Pedersen, 2005; Chen et al., 2007). Following the methodology of Holden and Jacobsen (2014), I form quintile portfolios based on the midpoint and the microprice versions of the effective spread and report the extent to which the stocks are put in different quintiles depending on the measure used. I repeat the portfolio sort for each trading day in the cross-sectional sample as well as in the time series sample. I report in Table 6 the percentage of stock–day observations where the quintiles differ for the two effective spread measures. For the cross-sectional sample, I aggregate the result across all stock–days and, for the time series sample, I aggregate the results for four five-year periods.

In the cross-sectional sample, the stocks end up in the same quintile in about two-thirds of the cases and, in the time series sample, the corresponding number is between 74% and 88%. The stronger disagreement between the two effective spread measures observed for the S&P 500 stocks is not surprising, because the overestimation is concentrated to the most liquid stocks.

It is interesting to compare the results in Table 6 to the findings of Holden and Jacobsen (2014).

Table 6: Differences in effective spread quintiles. This table shows the difference in effective spread quintiles by stock–day when using the midpoint effective spread instead of the microprice effective spread. The first row is for the cross-sectional sample, which includes five trading days for the S&P 500 stocks, from November 30 to December 4, 2015. The following four rows are for the time series sample, which includes all common stocks with primary listings on the NYSE, NYSE MKT (or AMEX), and NYSE Arca. Stock–day observations for the time series sample spanning January 1996 to December 2015 are aggregated into five-year periods: 1996–2000, 2001–2005, 2006–2010, and 2011–2015.

	Lower quintiles	Same quintile	Higher quintiles
Cross-sectional sample			
S&P 500 stocks, Nov 30–Dec 4, 2015	17%	66%	16%
Time series sample			
NYSE/NYSE MKT/NYSE Arca, 1996–2000	13%	74%	12%
NYSE/NYSE MKT/NYSE Arca, 2001–2005	7%	86%	7%
NYSE/NYSE MKT/NYSE Arca, 2006–2010	6%	88%	6%
NYSE/NYSE MKT/NYSE Arca, 2011–2015	7%	86%	7%

The authors report quintile portfolio differences emerging when comparing different methods to calculate the midpoint effective spread. In their analysis, the DTAQ midpoint effective spread is used as the benchmark. In my analysis, I regard that measure as potentially biased. The quintile portfolio differences that I find are smaller than what Holden and Jacobsen (2014) report. Nevertheless, the results here indicate that the precision in the portfolio formation can be improved by using the microprice effective spread in asset pricing and corporate finance studies.

6 Concluding Remarks

I present a model showing that the midpoint effective spread overstates the true effective spread if the liquidity demand is elastic, as well as empirical evidence of such elasticity. My estimates show that the overestimation of the effective spread is 16%, on average, for the S&P 500 stocks and up to 60% for low-priced stocks.

I propose the microprice effective spread as a viable alternative metric. It utilizes the asymmetry in depth posted at the best bid and ask prices to proxy for deviations between the spread

midpoint and the fundamental value.

Investors who use the microprice instead of the spread midpoint as a benchmark for the effective spread are better positioned to minimize illiquidity costs through liquidity timing and order routing, evaluate the adverse selection cost in liquidity supply, and optimize portfolio weights and rebalancing frequency.

References

Acharya, V. and Pedersen, L. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77(2):375–410.

Amihud, Y. and Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, 17(2):223–250.

Anshuman, V. R. and Kalay, A. (1998). Market making with discrete prices. *Review of Financial Studies*, 11(1):81–109.

Avellaneda, M., Reed, J., and Stoikov, S. (2011). Forecasting prices from level-i quotes in the presence of hidden liquidity. *Algorithmic Finance*, 1(1):35–43.

Bessembinder, H. (2003). Trade execution costs and market quality after decimalization. *Journal of Financial and Quantitative Analysis*, 38(04):747–777.

Blume, M. E. and Goldstein, M. A. (1992). Displayed and effective spreads by market. *Working paper, Rodney L. White Center for Financial Research, The Wharton School, University of Pennsylvania*.

Brogaard, J., Hagströmer, B., Nordén, L., and Riordan, R. (2015). Trading fast and slow: Colocation and liquidity. *Review of Financial Studies*, 28(12):3407–3443.

Cartea, Á., Jaimungal, S., and Penalva, J. (2015). *Algorithmic and high-frequency trading*. Cambridge University Press.

Chao, Y., Yao, C., and Ye, M. (2017). Why discrete price fragments u.s. stock exchanges and disperses their fee structures. *Working paper, available at SSRN.*

Chen, Q., Goldstein, I., and Jiang, W. (2007). Price informativeness and investment sensitivity to stock price. *Review of Financial Studies*, 20(3):619–650.

Colliard, J.-E. and Foucault, T. (2012). Trading fees and efficiency in limit order markets. *Review of Financial Studies*, 25(11):3389–3421.

Constantinides, G. (1986). Capital market equilibrium with transaction costs. *Journal of Political Economy*, 94(4):842–862.

Corwin, S. A. and Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance*, 67(2):719–760.

Demsetz, H. (1968). The cost of transacting. *The Quarterly Journal of Economics*, pages 33–53.

Dumas, B. and Luciano, E. (1991). An exact solution to a dynamic portfolio choice problem under transactions costs. *The Journal of Finance*, 46(2):577–595.

Epps, T. W. (1976). The demand for brokers' services: The relation between security trading volume and transaction cost. *Bell Journal of Economics*, 7(1):163–194.

Fang, V. W., Noe, T. H., and Tice, S. (2009). Stock market liquidity and firm value. *Journal of Financial Economics*, 94(1):150–169.

Foucault, T., Kadan, O., and Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4):1171.

Glosten, L. R. (1994). Is the electronic open limit order book inevitable? *The Journal of Finance*, 49(4):1127–1161.

Goldstein, M. A. and Kavajecz, K. A. (2000). Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the nyse. *Journal of Financial Economics*, 56(1):125–149.

Goyenko, R., Holden, C., and Trzcinka, C. (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92(2):153–181.

Harris, L. (2013). Maker-taker pricing effects on market quotations. *Working paper. University of Southern California, San Diego, CA*.

Hasbrouck, J. (2009). Trading costs and returns for us equities: Estimating effective costs from daily data. *The Journal of Finance*, 64(3):1445–1477.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1):1–33.

Hendershott, T. and Menkveld, A. J. (2014). Price pressures. *Journal of Financial Economics*, 114(3):405–423.

Holden, C. W. (2009). New low-frequency spread measures. *Journal of Financial Markets*, 12(4):778–813.

Holden, C. W. and Jacobsen, S. (2014). Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions. *The Journal of Finance*, 69(4):1747–1785.

Huang, R. and Stoll, H. (1996). Dealer versus auction markets: A paired comparison of execution costs on nasdaq and the nyse. *Journal of Financial Economics*, 41(3):313–357.

Korajczyk, R. and Sadka, R. (2008). Pricing the commonality across alternative measures of liquidity. *Journal of Financial Economics*, 87:45–72.

Lee, C. (1993). Market integration and price execution for nyse-listed securities. *The Journal of Finance*, 48(3):1009–1038.

Lipton, A., Pesavento, U., and Sotiropoulos, M. G. (2013). Trade arrival dynamics and quote imbalance in a limit order book. *Working paper, available at arXiv.or*.

Malinova, K. and Park, A. (2015). Subsidizing liquidity: The impact of make/take fees on market quality. *The Journal of Finance*, 70(2):509–536.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

O'Hara, M., Saar, G., and Zhong, Z. (2015). Relative tick size and the trading environment. *Working paper, available at SSRN*.

O'Hara, M. and Ye, M. (2011). Is market fragmentation harming market quality? *Journal of Financial Economics*, 100(3):459–474.

Panayides, M. A., Rindi, B., and Werner, I. M. (2017). Trading fees and intermarket competition. *Working paper, available at SSRN*.

Petersen, M. and Fialkowski, D. (1994). Posted versus effective spreads: Good prices or bad quotes? *Journal of Financial Economics*, 35(3):269–292.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, 22(1):435–480.

Ranaldo, A. (2004). Order aggressiveness in limit order book markets. *Journal of Financial Markets*, 7(1):53–74.

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*, 39(4):1127–1139.

Roşu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641.

Sandås, P. (2001). Adverse selection and competitive market making: Empirical evidence from a limit order market. *Review of Financial Studies*, 14(3):705–734.

Sarkar, A. and Schwartz, R. A. (2009). Market sidedness: Insights into motives for trade initiation. *The Journal of Finance*, 64(1):375–423.

SEC (2001). Disclosure of order execution and routing practices. *Release No. 34-43590; File No. S7-16-00*.

Seppi, D. J. (1997). Liquidity provision with limit orders and a strategic specialist. *Review of Financial Studies*, 10(1):103–150.

van Kervel, V. (2015). Competition for order flow with fast and slow traders. *Review of Financial Studies*, 28(7):2094–2127.

Werner, I. M., Wen, Y., Rindi, B., Consonni, F., and Buti, S. (2015). Tick size: theory and evidence. *Working paper, available at SSRN.*

Yao, C. and Ye, M. (2015). Why trading speed matters: A tale of queue rationing under price controls. *Working paper, available at SSRN.*

Yueshen, B. Z. (2016). Uncertain market making. *Working paper, available at SSRN.*

Appendix

A Derivation of the Equilibrium Price

By solving (8) and (7) for λ and setting the resulting expressions equal to each other, I obtain

$$\frac{P_A - X - \gamma}{Q_A + \phi + \delta(P_A - X)} = \frac{X - P_B - \gamma}{Q_B + \phi + \delta(X - P_B)}. \tag{A.1}$$

By taking cross-products,

$$(P_A - X - \gamma)(Q_A + \phi + \delta(P_A - X)) = (X - P_B - \gamma)(Q_B + \phi + \delta(X - P_B)), \tag{A.2}$$

and solving for X, I retrieve

$$X = \frac{(P_A - \gamma)Q_B + (P_B + \gamma)Q_A + (\delta\gamma + \phi)(P_A + P_B)}{Q_B + Q_A + 2\phi + 2\delta\gamma}. \tag{A.3}$$

Given the definitions of the spread midpoint and the midpoint effective spread, the bid and ask prices can be written as $P_B = M - s^M$ and $P_A = M + s^M$, respectively. Substituting these into (A.3), I obtain

$$X = \frac{(M + s^M - \gamma)Q_B + (M - s^M + \gamma)Q_A + 2M(\delta\gamma + \phi)}{Q_B + Q_A + 2\phi + 2\delta\gamma}. \tag{A.4}$$

This can be rewritten as

$$X = \frac{M(Q_B + Q_A + 2\phi + 2\delta\gamma)}{Q_B + Q_A + 2\phi + 2\delta\gamma} + \frac{(s^M - \gamma)(Q_B - Q_A)}{Q_B + Q_A + 2\phi + 2\delta\gamma}, \tag{A.5}$$

which simplifies to the expression in (9).

B The Midpoint Effective Spread is the Upper Bound for the True Effective Spread

Denote the expected effective spread for trades on the wide and tight sides of the spread as $E_W(s)$ and $E_T(s)$, respectively. Let the difference to the midpoint effective spread be given by a ($a \geq 0$), such that $E_W(s) = E_W(s^M + a)$ and $E_T(s) = E_T(s^M - a)$. Under positive liquidity demand elasticity ($\delta > 0$) and $a > 0$, the frequency of market order arrivals on the wide side of the spread (W) must be lower than the frequency on the tight side ($W + T$). Because the density function for market order quantities absent transaction costs is the same for both sides of the limit order book, the market order arrival probabilities are $\pi_{wide} = \frac{W}{2W+T}$ and $\pi_{tight} = \frac{W+T}{2W+T}$.

Inserting the expressions for the effective spread and market order probabilities in (11) yields

$$E(s) = \frac{WE_W(s^M + a) + (W + T)E_T(s^M - a)}{2W + T} = E(s^M - \frac{Ta}{2W + T}), \tag{B.1}$$

which implies $E(s) \leq E(s^M)$.

C TRTH Data Matching and Screening

Matching CRSP and TRTH identifiers. To my knowledge, this is the first study that matches data from the CRSP and TRTH databases. In the CRSP database, the issue identifier `PERMNO` has the advantage of being permanent. The issue identifier in TRTH is called the Reuters Instrument Code (RIC) and it is not permanent. To track an issue over time, a viable strategy is thus to access the time series from the CRSP and then match each CRSP observation to the relevant RIC.

The CRSP field with closest correspondence to the RICs is the ticker symbol at the primary exchange, TSYMBOL. For most issues, `TSYMBOL` is identical to the RIC of the consolidated instrument in TRTH. To maximize the matching performance, however, the following adjustments are considered:

- When `TSYMBOL` is empty, the CRSP field `TICKER` is used instead.

- Before January 1, 2012, share class information is not included in TSYMBOL. For this period, when TSYMBOL cannot be matched to an RIC and the CRSP field SHRCLS is equal to A or B, I add a lowercase share class suffix (e.g., the TSYMBOL entry AIS is set to AISa).

- January 1, 2012, onward, TSYMBOL and TICKER differ when there is a share class suffix for TSYMBOL. In these cases, the TSYMBOL share class suffix is made lowercase to match TRTH identifier conventions (e.g., the TSYMBOL entry VIAB is set to VIAb). Other four-letter TSYMBOL entries are given a suffix *.K*, in line with TRTH consolidated instrument conventions (e.g., the TSYMBOL entry ADGE is set to ADGE.K).

TRTH data screening. Each TAQ observation in TRTH includes additional information in the Qualifier field. I use that information to screen TAQs, as follows:

(T1) Trades marked as regular, odd lots, or due to intermarket sweep orders are retained, unless any of the criteria (T2)–(T4) are satisfied. This screening utilizes the [GV1_TEXT] and [LSTSALCOND] information and excludes everything but the following entries: @F_I (where _ represents a space), @__I, @F__, @___, _F__, _F_I, and ___I.

(T2) Trades with any of the following conditions indicated in the [CTS_QUAL] information are excluded: *derivatively priced* (DPT), *stock option related* (SOT), *threshold error* (XSW, RCK, XO), *out of sequence* (SLD), and *cross-trades* (XTR).

(T3) Trades with any of the following conditions indicated in the [PRC_QL2] information are excluded: *stopped* (STP), *agency cross-trade* (AGX), or *not eligible for last* (NBL).

(T4) Trades flagged as corrected are excluded. Corrections are entered as separate observations in TRTH and linked by an order sequence number (Seq..No.) to the trade in question.

(Q1) Quotes marked as regular are retained, unless any of the criteria (Q2) to (Q5) are satisfied. This screening utilizes the [PRC_QL_CD] information and excludes everything but the following entries: R__ and ___. This excludes, for example, crossed and locked quotes.

(Q2) Quotes marked as anything other than regular or coinciding with changes in the limit up–limit down (LULD) price bands in the [PRC_QL3] information are excluded. That is, [PRC_QL3] entries other than R__, ___, LPB, and RPB are excluded. The updates of LULD limits are merely indicated in Qualifier; they do not influence the validity of the quotes. This filter, however, excludes quotes associated with trading halts and quotes marked as slow due to a liquidity replenishment point or gap quote.

(Q3) Quotes marked as *non-executable* are excluded (A, B, or C, in the [GV1_TEXT] field).

(Q4) Quotes with non-regular conditions indicated by the [CTS_QUAL] information (taking the value TH_, IND, or O__) are excluded.

(Q5) Quotes where the bid–ask spread is either negative or exceeding USD 5 are excluded.

The effects of the different screening criteria are presented in Table C.1. The trade screening criteria disqualify a negligible number of trades for both the cross-sectional data set and the time series data set. Among the quote screening criteria, (Q5) is the one disqualifying the most quotes. Untabulated results show that the vast majority of the quotes captured by (Q5) are observations where the spread is equal to zero.

Table C.1: Data screening statistics. This table shows the extent to which different screening criteria filter out trade observations (Panel a) and set quotes matched to trades to missing (Panel b). Prior to the trade screening, trades that are time stamped within five minutes of the opening or closing time are excluded, as well as trades recorded in the alternative display facility.

(a) Trade screens

Sample	(T1)	(T2)	(T3)	(T4)	All filters combined	Remaining # obs.
S&P 500, Nov. 30–Dec. 4, 2015	< 0.01%	< 0.01%	< 0.01%	< 0.01%	< 0.01%	52.0 million
NYSE/AMEX, 1996 – 2015	< 0.01%	< 0.01%	< 0.01%	< 0.01%	< 0.01%	4337.6 million

(b) Quote screens

Sample	(Q1)	(Q2)	(Q3)	(Q4)	(Q5)	All filters combined
S&P 500, Nov. 30–Dec. 4, 2015	1.09%	1.09%	< 0.01%	< 0.01%	4.98%	4.98%
NYSE/AMEX, 1996 – 2015	0.06%	0.02%	< 0.01%	< 0.01%	5.61%	5.66%

D NBBO Example: IBM, April 1, 2008

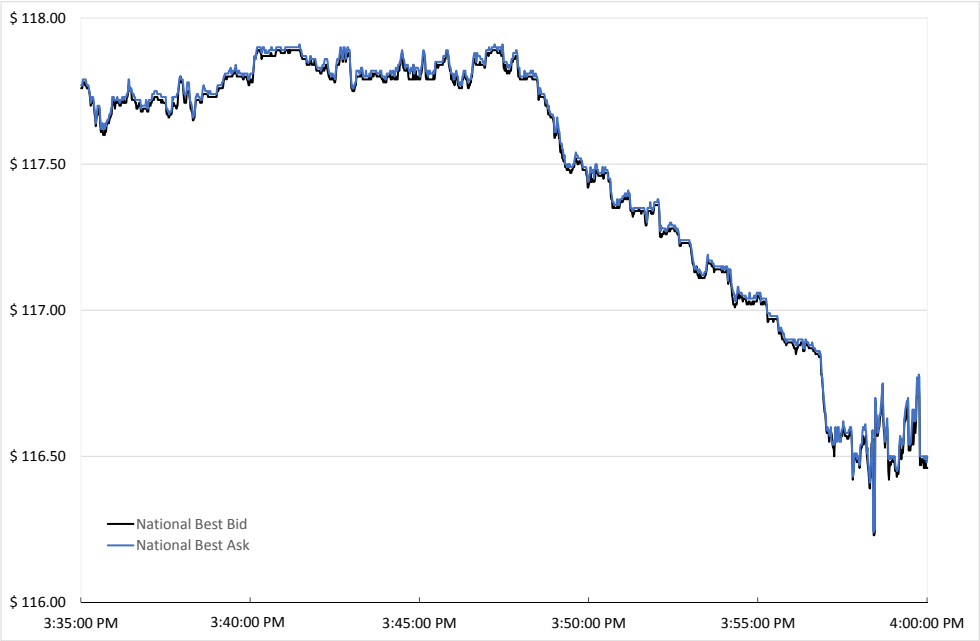


Figure D.1: NBBO accuracy for TRTH and MTAQ. This figure shows the NBBO prices for IBM on April 1, 2008, between 3:35 PM and 4:00 PM, as reported for the TRTH consolidated instrument `IBM`. The figure corresponds to Figure 2 from Holden and Jacobsen (2014), based on NBBO data constructed from DTAQ data. A visual comparison of the two figures confirms that TRTH NBBO data do not suffer from problems associated with canceled quotes.

E Alternative Explanation for the Liquidity Demand Results

In this appendix, I consider an alternative rationale for why the direction of trade is related to the observed depth asymmetry.

In the model of Foucault et al. (2005), the liquidity demanders' choice between limit orders and market orders depends on trader patience. For a trader needing to buy shares, in their model, a long queue for limit buy orders implies a high waiting cost for an added limit order. If the waiting cost of the limit order exceeds the cost of crossing the bid–ask spread, the trader is inclined to submit

a market order. This reasoning poses an alternative rationale for market orders arriving more frequently on the tight side of the spread. Empirical evidence supporting this view is available from, for example, Ranaldo (2004).

To rule out the alternative story, I rely on changes in the bid–ask spread. According to the model of Foucault et al. (2005), a wider bid–ask spread implies that higher waiting costs are required to trigger a market order submission. If the liquidity demanders' order choice drives the patterns observed above, a widening bid–ask spread should lead to a weaker relation between the observed depth asymmetry and the market order arrivals.

The evidence presented above indicates that limit order books with wider relative bid–ask spreads display stronger links between the observed depth asymmetry and the market order arrivals. Though this evidence speaks against the order choice story, it may be spurious. An omitted variable influencing both the trade price and the relation of interest could generate the same pattern.

To verify that a widening relative bid–ask spread leads to a stronger relation between the observed depth asymmetry and the market order arrivals, I run an event study around stock splits. Stock splits are suitable events because they imply an arguably exogenous cut in the stock price, which, in turn, implies a positive change in the relative tick size. If the minimum tick size becomes binding after the split, it also implies exogenous widening of the relative bid–ask spread.

For a stock split to be considered as an event, I require the stock to have its primary listing on the NYSE, NYSE MKT/AMEX, or NYSE Arca and to be in the top market capitalization quintile at the time of the split. In addition, the split date should be within the last 10 years of my sample (January 2006 to December 2015). The restrictions are to ensure that the stock is liquid enough for the minimum tick size to become binding after the stock split. I refer to the closest Wednesdays before and after the split as the *pre-split* and *post-split* dates, respectively. For an event to be retained, I require both the pre-split and post-split dates to be valid trading days. Over the 10 years considered, 88 stock split events satisfy these criteria.

For each treatment stock, I choose a control stock among all stocks that satisfy the same inclusion criterion as the treatment stock which does not have a stock split event within the pre-split and post-split dates. The control stock is the stock in that set that has the average price closest to

that of the treatment stock on the pre-split date.

For each treatment stock and each control stock, I measure the effective spread overestimation for the pre-split and post-split dates. The recorded data are entered into a difference-in-difference regression specified as

$$Overestimation_{ijt} = \alpha + \beta_1 Post_{ijt} + \beta_2 Treatment_{ijt} + \beta_3 Post_{ijt} Treatment_{ijt} + u_{ijt}, \qquad \text{(E.1)}$$

where $Overestimation_{ijt}$ is the effective spread overestimation for each event j, stock i, and date t; $Post_{ijt}$ is a dummy that is equal to one for all post-split dates; $Treatment_{ijt}$ is a dummy that is equal to one for all treatment stocks; and u_{ijt} comprises the residuals. I report the coefficient estimates in Table E.1, along with t-statistics based on standard errors clustered by stock and date.

The regression coefficient of primary interest is β_3. I find that β_3 is positive and strongly significant, with a t-statistic of 3.742. The estimates imply that the pre-split treatment group overestimation is $0.046 - 0.014 = 3.2\%$, increasing to $0.046 - 0.014 - 0.011 + 0.050 = 7.1\%$ on the post-split date. The overestimation of the treatment stock effective spread thus more than doubles after the stock split and the change is statistically significant relative to the change in the control stock.

The event study results support the model presented in this paper and are inconsistent with the alternative story.

Table E.1: Event study: Stock splits and the overestimation of the effective spread. This table shows the coefficient estimates for difference-in-difference regressions around stock split events. The treatment stocks are those that belong to the top market capitalization quintile of NYSE, NYSE MKT/AMEX, and NYSE Arca stocks, have a stock split in the interval January 2006 to December 2015, and have normal trading days on the Wednesdays before and after the split date (the pre-split and post-split dates). The control stock for each event is the one that belongs to the top market capitalization quintile of NYSE, NYSE MKT/AMEX, and NYSE Arca stocks and that have the average trade price closest to the average trade price of the treatment stock on the Wednesday before the split date. The regression specification is $Overestimation_{ijt} = \alpha + \beta_1 Post_{ijt} + \beta_2 Treatment_{ijt} + \beta_3 Post_{ijt} Treatment_{ijt} + u_{ijt}$, where $Overestimation_{ijt}$ is the effective spread overestimation (defined as in Table 1) of each event j, stock i, and date t; $Post_{ijt}$ is a dummy that is equal to one for all post-split dates, $Treatment_{ijt}$ is a dummy that is equal to one for all treatment stocks, and u_{ijt} comprises the residuals. Standard errors are clustered by stock and date and t-statistics are reported within parentheses. Statistical significance at the 1% and 5% levels is indicated by ** and *, respectively.

	Coefficient estimates
int.	0.046^{**}
(t-stat.)	(7.093)
$Treatment$	-0.014
(t-stat.)	(-1.809)
$Post$	-0.011
(t-stat.)	(-1.684)
$Treatment \times Post$	0.050^{**}
(t-stat.)	(3.742)
R^2	0.052
N	352