Bias in the Effective Bid-Ask Spread*

Björn Hagströmer

First version: March 23, 2017
This version: September 25, 2019

The effective bid-ask spread measured relative to the spread midpoint overstates the true effective bid-ask spread in markets with discrete prices and elastic liquidity demand. The average bias is 18% for S&P 500 stocks in general, and up to 97% for low-priced stocks. Cross-sectional bias variation across stocks and trading venues can mislead stock selection and order routing decisions. The bias also undermines liquidity timing and trading performance evaluations, and can lead to that non-sophisticated investors overpay for liquidity. To overcome these problems, the paper proposes new estimators of the effective bid-ask spread.

*Björn Hagströmer is Associate Professor at Stockholm Business School, Stockholm University, and a visiting researcher at the Swedish House of Finance. Website: hagstromer.org / E-mail: bjh@sbs.su.se. I thank Jonathan Brogaard, Petter Dahlström, Jungsuk Han, Thierry Foucault, Peter Hoffmann, Albert Menkveld, Lars Nordén, Andreas Park, Angelo Ranaldo, Kalle Rinne (discussant), Ioanid Rosu (discussant), Paul Schultz (discussant), Patrik Sandås, and Ingrid Werner, as well as seminar and conference participants at the Bank of America Global Quant Conference, CEPR-Imperial-Plato Market Innovator Conference, ECB, Warwick Frontiers of Finance Conference, Lund University, NBIM, the SEC Annual Conference on Financial Regulation, Stockholm Business School, the Swedish House of Finance, and University of Luxembourg for helpful comments. I am grateful to the Jan Wallander Foundation and the Tom Hedelius Foundation for research funding. A previous version of the paper was circulated under the title *Overestimated Effective Spreads: Implications for Investors*. The article was granted the FESE De la Vega Prize in 2017.

Bias in the Effective Bid-Ask Spread

Abstract

The effective bid-ask spread measured relative to the spread midpoint overstates the true effective bid-ask spread in markets with discrete prices and elastic liquidity demand. The average bias is 18% for S&P 500 stocks in general, and up to 97% for low-priced stocks. Cross-sectional bias variation across stocks and trading venues can mislead stock selection and order routing decisions. The bias also undermines liquidity timing and trading performance evaluations, and can lead to that non-sophisticated investors overpay for liquidity. To overcome these problems, the paper proposes new estimators of the effective bid-ask spread.

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

Conceptually, the effective bid-ask spread measures the cost of immediate execution, defined as twice the difference between the transaction price and the fundamental value. Whereas transaction prices are widely disseminated in financial markets, the fundamental value is unobservable. Empirical implementations of the effective spread instead rely on the quoted bid-ask spread midpoint, the average of the best bid and ask prices (henceforth, the "midpoint"), as its benchmark (Blume and Goldstein, 1992; Lee, 1993). The use of the midpoint as a proxy for the fundamental value goes back to Demsetz (1968), and is also stipulated in Rule 605. I refer to the conceptual definition as the "effective spread" and to its conventional estimator as the "midpoint effective spread".

This paper challenges the use of the midpoint as benchmark for transaction cost measurement. I show that the midpoint effective spread overestimates the illiquidity of US equity markets. The bias varies systematically across stocks and trading venues, and undermines liquidity timing and trading performance evaluations. I propose an alternative estimator, the "micro-price effective spread", which mitigates the bias, and can potentially reduce execution costs for non-sophisticated investors.

The midpoint effective spread bias can be illustrated by a simple example. Consider a stock with a fundamental value of USD 25.0025 that has liquidity supplied at the nearest prices where trading is allowed, USD 25.00 and USD 25.01. The effective spread is then asymmetric. For trades executed at the bid price it is half a cent (2×0.25 cents), whereas for ask-side trades it is three times higher, 1.50 cents (2×0.75 cents). If investors factor in the cost asymmetry in their trading decisions, market orders are in this example 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 one cent).

The problem with the midpoint is that the fundamental value of a security is a continuous variable, but observed prices are discrete. Gradual value changes are thus reflected in the midpoint only to the extent that they trigger a price change. The minimum incremental price change, known as the tick size and equal to one cent for most US stocks, also constrains the ability of market makers to quote prices symmetrically around the fundamental value (Anshuman and Kalay, 1998).

The implication is that the cost of immediacy for a buy market order, like in the example, often differs from that of a sell order. On average, the midpoint may still be a good fundamental value proxy, but at the time of trade it is potentially misleading. Goettler et al. (2005) present a model where rational liquidity traders respond optimally to the cost asymmetry by trading more on the side of the market where the effective spread is tighter. That is, the liquidity demand is elastic. This induces a positive bias in the midpoint effective spread.

Approach. I derive a model-free condition under which effective spread estimators are unbiased. The midpoint effective spread is unbiased only if the direction of trade is uncorrelated to the difference between the midpoint and the fundamental value. Zero correlation can be expected either if the market order flow is unaffected by sub-tick fluctuations in trading costs (i.e., if the liquidity demand elasticity is zero); or if investors are unable to infer fundamental value deviations from the midpoint. Otherwise, the expected midpoint effective spread overestimates the true expected effective spread. Importantly, the expected bias is positive for both buy and sell trades, and is thus not mitigated by averaging across a large set of trades.

I propose the micro-price effective spread as an alternative effective spread estimator. The "micro-price" is a proxy for the fundamental value derived by Stoikov (2018). It adjusts the observed midpoint for expected future midpoint changes, and is thus by construction a martingale. The key advantage of the micro-price effective spread is that it captures the asymmetries between the bid- and ask-side effective costs that emerge when the fundamental value deviates from the midpoint.

My main empirical evaluation is based on a one-week sample (December 7 – 11, 2015) of trades and quotes in the S&P 500 index constituent stocks. To study how the bias differs across investors, I also access a proprietary data set on US equity trading released by Nasdaq, with the distinguishing feature that it flags all trades executed by high-frequency traders (HFTs).[1] I refer to this data set as the "HFT sample". Its categorization of traders allows me to see whether investors with higher market structure sophistication (such as HFTs) are better able to time their liquidity demand.

Results. The condition for the midpoint effective spread to be unbiased does not hold in the data. I find a significant relation between the direction of trade and the fundamental value deviation from the midpoint. For example, when the midpoint is one basis point higher than the fundamental

[1]The data set includes a market-cap stratified sample of 120 stocks and I access the latest available trading week, February 22 – 26, 2010. It is used extensively in research on HFT, see, e.g., Brogaard et al. (2014), Brogaard et al. (2017), and Carrion (2013).

value (as in the example above), only 20% of all trades are buyer-initiated (paying the wide side of the spread) and 80% 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 midpoint effective spread bias is economically significant. I find that the micro-price effective spread, which I use as benchmark in my evaluation, averages 2.73 basis points (bps) for the sample stocks. The midpoint effective spread is 3.22 bps. Though the nominal difference of 0.49 bps may seem small, in relative terms it is 18%. Extrapolating this finding to the annual trading volume in the S&P 500 stocks (which in 2015 amounted to USD 8.7 trillion), the midpoint effective spread overstates the illiquidity costs by USD 213 million annually. The magnitude of the bias is also on par with the effective spread effect of major US equity market structure reforms, such as market fragmentation, algorithmic trading, and tick size decimalization (see p. 19 for details).

The bias has several implications for investors and academic research, including (i) stock selection, (ii) liquidity timing, (iii) order routing, and (iv) trading performance evaluation.

First, the midpoint effective spread bias increases with price discreteness. This is because the asymmetry between bid- and ask-side effective spreads is more prevalent in stocks with high relative tick sizes (Anshuman and Kalay, 1998). With the minimum tick size being fixed at USD 0.01 for most US stocks, those with low share prices have high relative price discreteness. I find that the effective spread overestimation for the lowest priced S&P 500 stocks (below USD 15) has a bias of 97% on average. The bias remains statistically significant for price levels up to USD 115, representing 76% of the S&P500 trading volume.

The systematic variation has potential implications for applications based on the relative illiquidity of stocks. For example, Amihud and Mendelson (1986) show that more illiquid securities are held by investors with longer investment horizons, and Constantinides (1986) and Dumas and Luciano (1991) show that illiquidity influences portfolio rebalancing. I show empirically that the bias can cause misallocations in liquidity portfolios, which are a common feature studies on the effects of liquidity on asset pricing and corporate finance (e.g., Acharya and Pedersen, 2005; Chen et al., 2007, respectively). Forming quintile liquidity portfolios based on the midpoint effective spread, I find that only 56% of the stocks are allocated to the same portfolio as when the micro-price effective spread is used.

Second, overlooking fundamental value fluctuations that do not trigger midpoint changes undermines liquidity timing. I find that 27% of the liquidity variance in the S&P500 stocks goes unnoticed to observers focused on the midpoint effective spread. This problem also applies to real-time liquidity monitoring. A trader looking to sell shares, for example, may track the bid-side quoted spread as the difference between the fundamental value and the best bid price. If such a trader uses the midpoint to proxy the fundamental value, he overlooks a third of the liquidity variation. For low-priced stocks (with a share price lower than USD 50), where the bias is larger, the corresponding number is 83%. In other words, the trader misses five sixths of the action.

Third, the bias varies across trading venues. For example, the average bias for trades at NYSE/Amex is 20%, whereas at Nasdaq BX (formerly known as the Boston Stock Exchange) it is only 7%. The evidence that the bias varies systematically across exchanges is important because it potentially misdirects order routing decisions. The US Securities and Exchange Commission (SEC) notes that when the market structure is fragmented, such as the US equity market, the order routing decision is of critical importance (SEC, 2001). Indeed, the SEC motivation for the Rule 605 reporting requirements is that it facilitates the individual investors' ability to compare execution quality across exchanges.[2] The data for this study do not permit a direct analysis of order routing, but I create hypothetical venue rankings in terms of effective spreads. On average, only half of all stock-day venue rankings based on the midpoint effective spread coincide with the benchmark rankings, and less than one third of the rankings for low-priced stocks.

Fourth, the bias carries over to evaluations of trading performance. It is common to decompose the effective spread into a price impact component, which from a liquidity supplier perspective measures adverse selection costs, and the realized spread. The price impact is defined as the five-minute change in fundamental value following a trade, and is typically measured as the change in midpoints. In my sample, using the midpoint rather than the micro-price leads to an overestimation of price impact by 15% on average, and 37% for low-priced stocks.

The evidence of the effective spread bias shows that some investors are able to infer the fundamental value and adapt their market order submissions accordingly. If noone did, the orders would be randomly distributed between the bid and the ask side, and the average effective spread would be unbiased. A key question is then whether *all* investors are aware of the bias, or if there are systematic differences between investor groups. In the former case, the bias documented here would be merely of academic interest, and could potentially trigger a discussion about the merit of the Rule 605 reporting requirements. In the latter case, however, it may be that the academic and regulatory recognition of the midpoint effective spread as a measure of illiquidity lulls non-sophisticated investors into a false sense of confidence in that metric. This could amplify trading performance differences between investors.

[2]The European Union has a similar rule. According to Directive 2014/65/EU in financial instruments (MiFID II), each trading venue and systematic internaliser should make midpoint effective bid-ask spreads statistics available to the public. See RTS 27, Article 2, available at http://ec.europa.eu/finance/securities/docs/isd/mifid/rts/160608-rts-27_en.pdf.

My analysis of the HFT sample shows that sophisticated traders (represented by the HFTs) are better than other investors at tracking the fundamental value of the security, and to time their trading activity accordingly. For example, the midpoint effective spread shows that the average cost of taking liquidity for HFTs is 2.22 bps, but the "true" effective spread for the same trades is only 1.13 bps, implying a bias of 97%. For Non-HFTs the bias is only 39%.[3]

To level the playing field, regulators could disseminate an estimator of the fundamental value to market participants in real time. This would facilitate liquidity timing for the least sophisticated investors, who are otherwise unable to track sub-tick value fluctuations. For US equities, a fundamental value estimator could be reported along with the national best bid and offer (NBBO) prices by the Securities Information Processors (SIPs). Indeed, all data required for the micro-price estimator, prices and volumes of the NBBO, are already contained in the SIP feed. The micro-price is a nonlinear estimator that is computationally demanding, but virtually all calculations can be pre-processed when the market is closed. Any combination of limit order book prices and volumes can then be mapped into a fundamental value estimate in real time. To facilitate order routing decisions, regulators could also amend the Rule 605 definition of the effective spread to reflect bid-ask asymmetries.

In the absence of a regulatory initiative to provide fundamental value estimates, many investors and academics may consider the micro-price too expensive. For such cases, an alternative estimator that shares many of the micro-price benefits is the *weighted midpoint*. This is a linear estimator, which I show can overcome most of the bias implications presented above, in particular when it is adjusted for order processing costs. Measuring the effective spread relative to the weighted midpoint may thus be viewed as a second-best solution.

Contribution. My findings add to the literature on the measurement of effective spreads, including the early work by Blume and Goldstein (1992), Lee (1993), and Petersen and Fialkowski (1994). The bias is consistent with the simulated limit order book market evidence by Goettler et al. (2005), and empirical evidence for equity options by Muravyev and Pearson (2016). Moreover, the results have implications for the liquidity measurement literature more generally. Roll (1984), Hasbrouck (2009), Holden (2009), Corwin and Schultz (2012), and Abdi and Ranaldo (2017) develop effective spread proxies based on daily equity data. Holden and Jacobsen (2014) show that the use of intraday data from the Monthly Trade and Quote (MTAQ) database results in

[3]The main findings of the S&P500 sample from 2015 are reflected in this sample too, with an average bias of 65%. An added dimension of the Nasdaq sample is that it extends beyond large-cap stocks. Broken down into large-cap, mid-cap, and small-cap segments, the average bias is 72%, 28%, and 14%, respectively. The lower bias in smaller issues is consistent with that the minimum tick size is less constraining in illiquid stocks.

distorted estimates of the effective spread. My findings indicate that the benchmark used for all these illiquidity proxies, the midpoint effective spread, is itself a biased estimator.

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 when information asymmetries are high (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 bid- and ask-side effective spreads.

In addition to the implications described above, overestimation of the effective spread is relevant to applications where illiquidity is analyzed as an outcome variable. For example, the finding that the overestimation problem is increasing with price discreteness is directly relevant to the evaluation of the SEC "tick size pilot" in the US market, where the minimum price increment of randomly selected low-priced stocks is increased from 1 cent to 5 cents.[4] Evaluations of the pilot that do not account for the bias are likely to overestimate the effective spread effect for the treated stocks.

The paper is organized as follows. Section 1 derives the conditions for when the midpoint effective spread estimator is unbiased, and discusses alternative estimators of the fundamental value. Section 2 introduces the data and sample used for the empirical investigation. Section 3 holds the main evidence about overestimation of effective spreads in US equities. Section 4 explores systematic variation in the bias, and shows that it potentially influences stock selection, liquidity timing, order routing, and trading performance evaluation. Section 5 shows that HFTs are more perceptive to nuances of the execution costs, relative to other investors. Section 6 provides a second-best effective spread estimator for computationally constrained analyses. Section 7 concludes.

1 Empirical Framework

In this section, I derive a model-free condition for when effective spread estimators are unbiased, and discuss high-frequency proxies for the fundamental value of a security.[5]

[4] For details, see the SEC press release from May 6, 2015, available at https://www.sec.gov/news/pressrelease/2015-82.html.

[5] As discussed by Hasbrouck (2002), alternative terminology for the fundamental value include "efficient price", "true price", or "consensus price".

1.1 Bias in Effective Bid-Ask Spread Estimators

In the presence of trading frictions, the transaction price P typically differs from the fundamental value X. The effective spread quantifies the difference, and may be viewed as a premium paid for the service of immediacy in securities trading. The nominal effective spread is defined as

$$S = 2D(P - X), \tag{1}$$

where D is a direction of trade indicator taking the value +1 for buyer-initiated trades, and -1 for seller-initiated trades. The multiplication by two is for consistency with the quoted bid-ask spread (defined below for a hypothetical round-trip trade). For ease of exposition, I suppress stock and time subscripts for all variables in this section.

Because the fundamental value at the time of transaction is unobservable, the effective spread is typically measured relative a proxy. I denote the fundamental value proxy $\tilde{X}$, and define the effective spread estimator as

$$\tilde{S} = 2D\left(P - \tilde{X}\right). \tag{2}$$

Various fundamental value estimators are distinguished with the superscript v, $\tilde{X}^v$. For example, I denote the midpoint $\tilde{X}^{mid}$. Similarly, an effective spread estimator utilizing the fundamental value estimator v is denoted $\tilde{S}^v$. The midpoint effective spread as defined by Blume and Goldstein (1992) and Lee (1993), as well as in the RegNMS Rule 605, is thus denoted $\tilde{S}^{mid}$.

An effective spread estimator is unbiased if the expected difference between the expressions in (1) and (2) is zero. The expected difference is

$$\mathrm{E}\left[\tilde{S} - S\right] = 2\mathrm{E}\left[D\left(X - \tilde{X}\right)\right], \tag{3}$$

implying that the effective spread estimator is unbiased if and only if D and $(X - \tilde{X})$ are uncorrelated. This can be expected either if investors are unable to assess the sign of $(X - \tilde{X})$, or if the liquidity demand elasticity is zero.

Consider again the example presented in the introduction. When the fundamental value ($25.0025) is closer to the best bid price ($25.00) than to the best ask price ($25.01), the effective spread for sell market orders is tighter than that for buy market orders. If investors then submit more sell than buy market orders, there is a positive correlation between $\left(X - \tilde{X}^{mid}\right)$ and D. According to (3), such correlation implies that the midpoint effective spread is overestimated.

1.2 Fundamental Value Estimators

The midpoint. The fundamental value of a security is an elusive but central concept in finance, and approximation methods vary widely. In market microstructure, the midpoint is the most common fundamental value estimator, with applications ranging from liquidity measurement (including the effective spread) to price discovery (Hasbrouck, 1995, 2003), realized volatility (Andersen et al., 2003), and returns (Lease et al., 1991).

The midpoint is defined as

$$\tilde{X}^{mid} = \frac{P^A + P^B}{2}, \tag{4}$$

where P^A and P^B are the best bid and ask prices in the limit order book.

The appeal of the midpoint is arguably data availability and simplicity. Data on the best bid and ask prices are publicly available for many asset classes and market types (both auction and dealer markets) and in long time series. In markets where the quotes are valid until canceled, such as limit order book markets, midpoint observations are available *continuously* during trading hours. The midpoint is thus applicable in contexts of either event time or equi-spaced intraday observations. Because the midpoint is simply an arithmetic average of the best bid and ask prices, real-time calculation is straightforward, and the measure is easy to understand for all market participants.

The midpoint has, however, two important shortcomings. First, theoretical evidence shows that liquidity suppliers do not set their quotes symmetrically around the fundamental value when prices are discrete (Anshuman and Kalay, 1998) or when their inventory deviates from the preferred level (Hendershott and Menkveld, 2014). Second, a proxy of the fundamental value should ideally reflect the expectations of future price changes, (see, e.g., the discussion in Hasbrouck, 2002). By definition, the midpoint reflects contemporaneous prices only.

Order book imbalance as a predictor of midpoint changes. Ample empirical evidence shows that price changes are predictable using the difference in market depth posted at the best bid and ask prices, which is known as the *order book imbalance*. For example, Cont et al. (2014) document that their measure of changes in the order book imbalance can explain 65% of the variation in midpoint changes, averaged across 50 randomly selected stocks from the S&P500 index. Gould and Bonart (2016) find in a sample of 10 liquid Nasdaq stocks that a measure of order imbalance is useful to predict the direction of the next midpoint change.

That the volumes posted in the limit order book are potentially useful to track the fundamental value is best understood through the lens of the liquidity suppliers. According to the model by Glosten (1994), the optimal depth posted in the limit order book is based on a tradeoff between the revenues expected from earning the effective spread, and the costs associated with trading with informed market orders. Assuming that adverse selection costs are increasing with the trade size, he shows that the depth posted at a given price level is increasing in the distance to the fundamental value. That is, if the bid-side depth is smaller than on the ask-side, it indicates that the fundamental value is closer to the bid than to the ask price. This conclusion is supported empirically by Sandås (2001), who analyzes Swedish stocks traded in a pure limit order book setting. Kavajecz (1999) shows that NYSE specialists shape their price schedules in a similar manner.

According to Lipton et al. (2013), the signal coming from the order book imbalance is picked up by professional traders:

> “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.” (p. 2)

Taken together, the evidence implies that the approximation of the fundamental value can be improved by adjusting the midpoint using the order book imbalance.

The micro-price. Stoikov (2018) proposes a fundamental value estimator, the *micro-price*, which incorporates expectations of future midpoint changes conditional on the state of the limit order book. The intuition is as follows. Defining the current state as the combination of the midpoint, the quoted spread, and the order book imbalance, the probability of next-period combinations of the same three variables can be estimated using data on what happened when the market was in the same state historically. For any next-period state that does not imply a change in the midpoint, Stoikov (2018) repeats the procedure. In the limit, the result is a probability tree where all branch endpoints are associated with midpoint changes. The micro-price is the weighted average future midpoint implied by the probability tree.

Formally, define the quoted spread as $S^{quoted} = P^A - P^B$, and the order book imbalance as $I = Q^B/(Q^B + Q^A)$, where Q^B and Q^A represent the volumes quoted at the best bid and ask prices, respectively. The micro-price is then given by

$$\tilde{X}^{mic} = \tilde{X}^{mid} + g\left(S^{quoted}, I\right), \tag{5}$$

where $g(S^{quoted}, I)$ is a function that adjusts the current midpoint for expected future midpoint changes. The value of this adjustment function is determined by discretizing the quoted spread and the order book imbalance and treating combinations thereof as a finite state space. To evaluate the adjustment function at infinity, Stoikov (2018) analyzes the state space as a discrete time Markov chain with absorptive states. The absorptive states are given by midpoint changes of different magnitudes, and correspond to the branch endpoints in the probability tree. For micro-price estimation details and examples, see Appendix A.

The micro-price is theoretically appealing in that it is a martingale by construction, and that it allows for quotes to be set asymmetrically around the fundamental value proxy. Relative to the midpoint, the additional data required to calculate the micro-price are the quantities posted at the best bid and ask prices. Such data are available to investors through the SIP consolidated data feeds. For academics, the depth data are available in the major databases used for intraday liquidity analysis, such as the *Daily Trade and Quote* (DTAQ) and *Thomson Reuters Tick History* (TRTH) databases.

The micro-price effective spread. I refer to the effective spread measured relative the micro-price as the micro-price effective spread. Following (2), it is defined as

$$\tilde{S}^{mic} = 2D\left(P - \tilde{X}^{mic}\right). \tag{6}$$

Based on the advantages of the micro-price presented above, I view the micro-price effective spread as the best available effective spread estimator.

A drawback with the micro-price effective spread is that the micro-price is considerably more complex to estimate than the midpoint. Nevertheless, I discuss in Section 5.3 how a regulator can make micro-price estimates available to the public in real time.

In the absence of such centralized initiatives, the micro-price effective spread is likely to be too costly for computationally constrained investors and academics alike. A viable alternative then is to estimate the effective spread relative to the *weighted midpoint*. The weighted midpoint is a linear function of the quoted spread and the order book imbalance that shares the simplicity of the midpoint but overcomes its problem with discrete prices. Though it is not a martingale, I show in Section 6 that the weighted midpoint significantly mitigates the problems associated with the midpoint effective spread. It may thus be viewed as a second-best fundamental value estimator for the purpose of liquidity measurement.

2 Data and Sample

I use the TRTH database to access trades and quotes for US equities.[6] For sample selection purposes, I use stock characteristics available in monthly data from the *Center for Research in Security Prices* (CRSP). In addition, I use a database provided by Nasdaq, described below.

Two samples are considered:[7]

- The baseline sample includes one trading week (December 7 – 11, 2015) for the S&P 500 index stocks. During this sample period, the S&P 500 index consists of 506 stocks, all available in the TRTH. I include trades from all relevant US national securities exchanges.[8] Trades in dark pools and over-the-counter markets are not included. I refer to this data set as the "S&P500 sample".

- To analyze differences across investor groups, I use a proprietary data set provided by Nasdaq, reporting all trades for 120 stocks along with a flag that indicates whether the active and the passive counterparty (or both) of a transaction is a high-frequency trader (HFT) or not (Non-HFT).[9] This data, which I refer to as the "HFT sample", also allows for additional cross-sectional analysis across market capitalization levels, as the stocks are chosen to form a stratified sample across large-, mid-, and small-cap stocks. I use the latest trading week available in the data set: February 22 – 26, 2010. As the proprietary data does not contain NBBO quotes, I match it to trades from TRTH, which are then straightforward to match to TRTH quotes. For details on matching across databases, see Appendix B.

The TRTH data do not contain information on the direction of trade indicator *D*, which is required for effective spread calculations. I create the variable using the Lee and Ready (1991)

[6]The TRTH database is not commonly used for US equity research but it is based on the same data sources as the DTAQ database. The trades come from the consolidated tape, and the quotes from the NBBO feed. For details on the TRTH data sources and quality, see the internet appendix.

[7]In Appendix C, I use a third sample to study the liquidity effect of stock split events. Stock split events are indicated in CRSP as distribution code (`DISTCD`) 5523. I include events in ordinary common stocks with primary listing at NYSE, NYSE MKT, or NYSE Arca, during a 10-year period, Jan. 1, 2006 – Dec. 31, 2015. I refer to this sample as the "Split sample".

[8]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*/*ASE*), and NYSE Arca (*PSE*). The Nasdaq-owned exchange identifiers *NAS* and *THM* are reported together because the two venues trade non-overlapping segments of stocks. The same holds for the ICE-owned venues *NYS* and *ASE* (formerly Amex).

[9]Nasdaq includes 26 trading firms in their definition of HFTs. The HFT flag indicates whether one of those firms is involved in the trade.

algorithm, noting that Chakrabarty et al. (2015) show that the procedure performs well in a recent US equities sample. In the HFT sample the direction of trade is directly observable. Reassuringly, all the conclusions of the HFT sample analysis remain unchanged when using the Lee and Ready (1991) algorithm instead of the observed variable. The two direction of trade indicators are identical in 91% of the trades.

The following screening is applied to all 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 and the last five minutes of the trading day are excluded. I also 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 B. Each trade observation contains information on the date, stock, time, price, volume, and trading venue. The S&P500 sample contains 55.7 million trades, the HFT sample 1.9 million trades, and the Split sample 5.5 million trades.

Retained trades are matched to the last quote observation in force in the preceding millisecond (as recommended by Holden and Jacobsen, 2014). After the quotes have been matched to trades, several screens are applied to exclude invalid and obsolete quotes, see Appendix B.

The retained quote observations contain information on the bid and ask prices and volumes, as well as the trading venue contributing the current quote. 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 limit orders are often cross-posted at several trading venues (van Kervel, 2015). To use NBBO quotes when measuring the effective spread is consistent with Rule 605 in Reg NMS.[10]

All bid-ask spread measures in the paper (unless otherwise noted) are winsorized within each stock by setting observations below the 1% quantile equal to the 1% quantile and observations above the 99% quantile equal to the 99% quantile.

[10]The technical details of the Rule 605 report requirements are in §240.11Ac1-5, available at https://www.sec.gov/rules/final/34-43590.htm. For the effective spread, see section (a)(2).

3 Main Results

In this section, I first confirm that liquidity demand is elastic in the sense that the direction of trade is positively correlated to the difference between the fundamental value and the spread midpoint. As shown in Section 1.1, a non-zero liquidity demand elasticity is a sufficient condition for that the expected midpoint effective spread is biased. Next, I quantify the effective spread overestimation and discuss its economic significance.

3.1 Liquidity Demand Elasticity

I assess the liquidity demand elasticity by investigating how market order arrivals depend on the *fundamental value deviation from the midpoint*, defined as $\log \tilde{X}^{mic} - \log \tilde{X}^{mid}$ and expressed in basis points. I categorize trades in all S&P500 stocks by the fundamental value deviation from the midpoint prevailing just before the trade. I create 21 trade categories using the following breakpoints: -2.1 bps, -1.9, ..., -0.1, +0.1, ..., +1.9, +2.1. The categories are labeled by the midpoint of their interval. For example, all trades where the fundamental value deviation from the midpoint lies within the interval $(1.9, 2.1]$ are put in the *2.0 bps* bucket.[11]

The solid line in Figure 1 shows the frequency of buyer-initiated trades for each trade category, and the bars report the trade category share of the total dollar volume. The null hypothesis of this analysis is that market orders arrive independently of the midpoint deviation, as indicated by the dashed horizontal line.

The results show that the probability of buyer-initiated trades tends to increase with the midpoint deviation. The relation is monotonic for midpoint deviations that do not exceed one basis point. For example, consider the case when the midpoint deviates by -1 bps from the fundamental value. This category corresponds to the example given in the introduction (with a midpoint deviation of -0.25 cents in a stock valued at USD 25.0025). For this case, I find that only 20% of all trades are buyer-initiated. For trades in the *+1.0 bps* category, in contrast, 80% of the trades are buyer-initiated. When the midpoint is close to the fundamental value, the split between buyer- and seller-initiated trades is even. Note that the symmetry around zero on the x-axis of Figure 1 is a feature of the data – it is not imposed by the econometrician.

To assess the relation between direction of trade and fundamental value deviation from the

[11] The interval $(-2.1, 2.1]$ spans 91% of the sample trades.

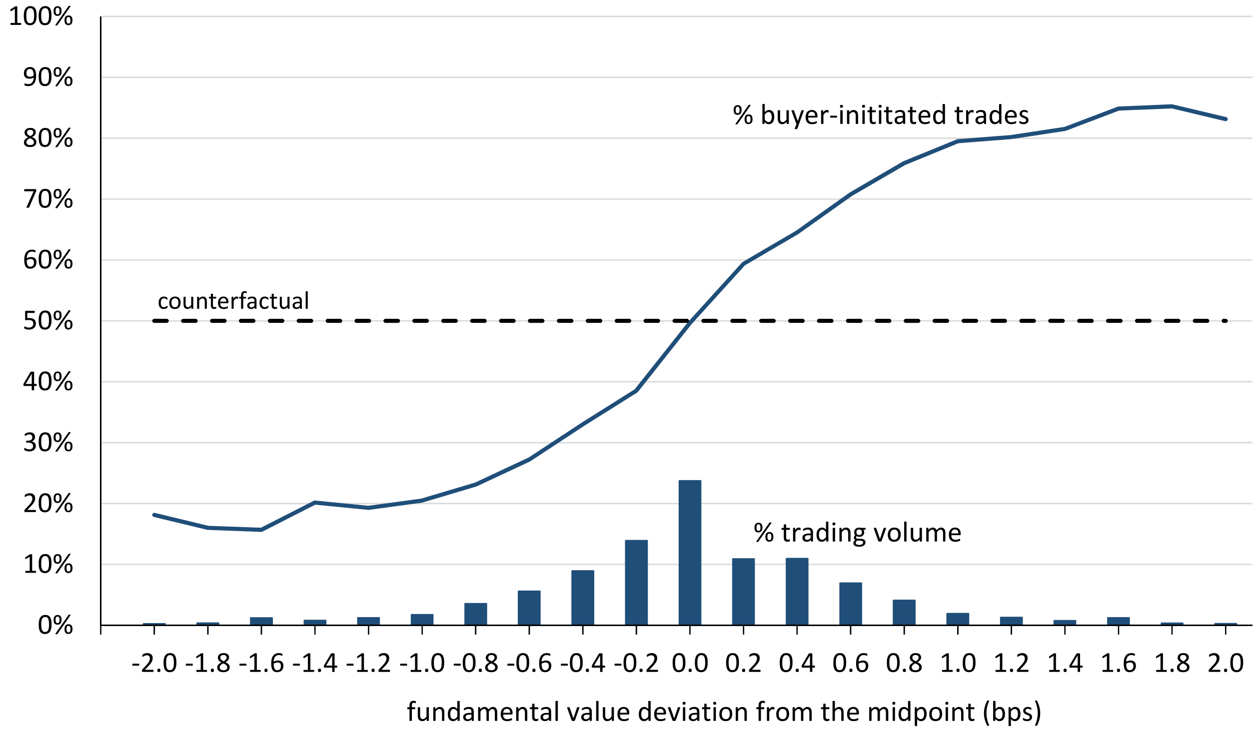


Figure 1: Liquidity demand elasticity in the S&P 500 stocks. This figure shows the frequency of buyer-initiated trades and the dollar volume market shares for different categories of the fundamental value deviation from the midpoint, defined as $log(\tilde{X}^{mic}) - log(\tilde{X}^{mid})$ and expressed in bps. The trade categories are determined by the breakpoints -2.1 bps, $-1.9, , ..., 0.1, 0.1, ..., 1.9, 2.1$, and labeled on the x-axis by the midpoint of each interval. The direction of trade is determined by the Lee and Ready (1991) algorithm. The sample includes all constituents of the S&P 500 index for the five trading days in the period December 7 – 11, 2015.

midpoint formally, I estimate the probit model:

$$\Pr\left(Buy_t\right) = \underset{(-1.01)}{-0.01} + \underset{(5.32)}{0.44}\left(\log \tilde{X}_t^{mic} - \log \tilde{X}_t^{mid}\right) + \varepsilon_t. \tag{7}$$

where t is a trade index, Buy_t equals one for buyer-initiated trades and zero for seller-initiated trades, and variation that is unexplained by the model is captured by the residual term ε_t. The estimated coefficients are reported in (7). The results indicate a positive relation between the direction of trade and the fundamental value deviation from the midpoint. The z-statistic (within parentheses, based on standard errors that are clustered by stock, date, and trading venue following Petersen, 2009) of 5.32 implies that the null hypothesis of zero slope is strongly rejected.

The midpoint effective spread is unbiased when either the liquidity demand elasticity is zero, or when investors are unable to infer the sign of the midpoint deviation (see (3)). The evidence indicates strongly that liquidity demand elasticity is positive, implying that the midpoint effective spread is biased upwards.

Behind the average, of course, lies a range of trading strategies adopted by investors that are diverse in terms of sophistication and holding period. Indeed, the effective spread is often measured for individual strategies and clienteles as part of performance comparisons. What is important to note here is that the conclusion that the effective spread is biased does not depend on the underlying reason for the liquidity demand to be elastic.

For example, several papers analyze the optimal order choice by liquidity demanders in dynamic models of the limit order book (e.g., Parlour, 1998; Foucault et al., 2005; Roşu, 2009). They show that the choice between market and limit orders depends on the tradeoff between execution risk and waiting costs associated with passive trading and the spread cost of active trading. In Parlour's (1998) model, incoming traders assess the probability of a limit order execution by evaluating the depth at both the best bid and the best ask prices. Though that poses an alternative story for the pattern observed in Figure 1, it does not alter the conclusion that such order flow regularities generate a bias in the midpoint effective spread.[12]

3.2 The Midpoint Effective Spread Bias

Table 1 reports properties of the effective spread measured at the stock level. It contains effective spread metrics using either the midpoint or the micro-price as fundamental value proxy. In addition to the effective spread properties, the quoted spread, the trade price, and two measures of the aggregate trading volume are reported. For comparability across price levels, each spread observation is scaled by the prevailing midpoint and reported in basis points. All stock-level observations are dollar volume-weighted averages across all trades of the given stock, and the mean across stocks is in turn dollar volume-weighted across stocks.

The results show that the average midpoint effective spread is 3.22 bps, whereas the micro-price effective spread is 2.73 bps. I define the *Nominal bias* of the midpoint effective spread estimator as its difference to the micro-price effective spread $\left(\tilde{S}^{mid} - \tilde{S}^{mic}\right)$, and report it in bps. The *Nominal bias* is on average 0.49 bps. The t-statistic of 9.76 shows that the *Nominal bias* is statistically significant. To determine its economic importance, it is also interesting to report it in relative terms. I define the *Relative average bias* as the average *Nominal bias* divided by the average micro-price effective spread. For the S&P 500 stocks, it amounts to 18%.

[12] Another aspect of Parlour's (1998) model, as well as the related models by Foucault et al. (2005) and Roşu (2009), is that the order book shape is determined by the patience of liquidity demanders. That assumption, which Foucault et al. (2005) state is for tractability, is in contrast to the theoretical and empirical evidence discussed in Section 1.2.

Table 1: Effective spread properties in the S&P 500 stocks. This table shows the effective spread measured relative the midpoint and the micro-price, respectively, the difference between the two, as well as other characteristics of the constituents of the S&P500 index, measured for December 7 – 11, 2015. The reported statistics are based on stock-level measures of each variable and include the mean, the standard deviation, and the fifth, 25th, 50th, 75th, and 95th percentiles. The effective spread is twice the relative spread between the trade price and the midpoint $\left(\tilde{S}^{mid}\right)$ or the micro-price $\left(\tilde{S}^{mic}\right)$, scaled by the midpoint. The benchmark price is based on quotes prevailing in the microsecond before each trade. The effective spread is measured for each stock as the dollar-weighted average across all trades in the sample, excluding trades occurring in the first or last five minutes of the trading day, as well as block trades. The *Nominal bias* is the difference $\tilde{S}^{mid} - \tilde{S}^{mic}$, reported in bps. The t-statistic corresponding to the null that the value-weighted average *Nominal bias* is equal to zero, based on standard errors that are clustered by stock, date, and trading venue (following Petersen, 2009), is reported within parentheses. The *Relative average bias* is the average *Nominal bias* divided by the average $\tilde{S}^{mic}$. *Quoted spread* is the difference between the national best ask and bid quotes just before each trade, divided by the midpoint and measured for each stock as the dollar-weighted average across trades. *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.	5th	25th	50th	75th	95th
Effective spread							
$\tilde{S}^{mid}$ (bps)	3.22	2.18	1.64	2.35	3.02	4.39	7.78
$\tilde{S}^{mic}$ (bps)	2.73	1.82	1.21	1.85	2.53	3.78	6.99
Nominal bias							
$\tilde{S}^{mid} - \tilde{S}^{mic}$ (bps)	0.49	1.16	-0.01	0.05	0.19	0.66	2.82
(t-stat.)	(9.76)						
Relative average bias	0.18						
Quoted spread (bps)	3.55	2.57	1.71	2.46	3.23	4.92	9.13
Trade price (USD)	119.27	101.06	16.36	37.74	59.60	94.23	186.43
Trade volume (thousands)	102.83	98.57	23.34	44.96	75.54	123.74	266.77
Dollar volume (millions)	687.92	874.92	128.11	271.88	424.31	745.60	2060.58

Is an overestimation by 18% economically significant? The economic magnitude can be illustrated by relating it to findings on the midpoint effective spread in response to major US market structure changes. For example, Bessembinder (2003, Table 3, Panel D) reports that the tick size decimalization in the US leads to reductions in value-weighted effective spreads for large-cap stocks of 33% at NYSE and 5% at Nasdaq. Hendershott et al. (2011) find that a one standard deviation increase in algorithmic trading is associated with a 23% reduction in the large-cap effective spread.[13] O'Hara and Ye (2011) investigate the effects of market fragmentation and report that fragmented stocks have 8% lower spreads than consolidated stocks.[14] In conclusion, the magnitude of the midpoint effective spread bias is on par with the illiquidity effect of major changes to the structure of US equity markets.

The distributional properties reflected by the percentiles reported in Table 1 show that the two effective spread estimators have similar dispersion. The differences for each reported percentile between the two are in the interval 0.36–0.79 bps. This does not tell the whole story, however, because there is also considerable dispersion in the *Nominal bias*. It ranges from -0.01 bps for the fifth percentile to 2.82 bps for the 95$^{\text{th}}$ percentile. In the next section, I explore the cross-sectional as well as the intraday time series variation of the bias, and show how they matter for investors as well as for academic research.

4 Bias Variation and its Implications

In this section, I present several implications of the midpoint effective spread bias, including stock selection, liquidity timing, order routing, and trading performance evaluation.

4.1 Variation Across Stocks

I expect the bias to be increasing with stock liquidity and decreasing with stock price. The reason is that liquid, low-priced, stocks in the US equity market are those where the pricing is most constrained by the minimum tick size, which is fixed at one cent.[15] This leads to greater asymmetry between the bid-side and ask-side effective spreads.

[13] 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 $\frac{0.18 \times 4.54}{3.63} \approx 0.23$. See the corresponding calculation of Hendershott et al. (2011, p.22) for the quoted spread.

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

[15] Stocks priced below one dollar have a tick size lower than one cent, but there are no such stocks in S&P500 sample.

To assess the relation between the overestimation and the relative tick size, I split the sample into trade price groups. The *USD10* group includes all trades in the USD 5.01–15 interval, the *USD20* group includes all trades in the USD 15.01–25 interval, and so on with 10-dollar intervals for each price group. The category with highest priced trades considered is *USD190*, including trades in the USD 185.01–195 interval. In the S&P500 sample, 98% of the trades fall within the price interval USD 5–195. Figure 2 shows the effective spread relative to the midpoint and to the micro-price for each share price group, plotted in Panel (a) as dashed and solid lines, respectively.

The share price groups from *USD30* to *USD120* span the lion's share of the trading activity (74% of the dollar volume and 77% of the trades in the S&P 500 stocks). In that price interval, the effective spreads are around 2.1 bps, on average. Stocks in the *USD10* and *USD20* categories have much 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) that the effective spread bias is concentrated to but not limited to the low-priced stocks.

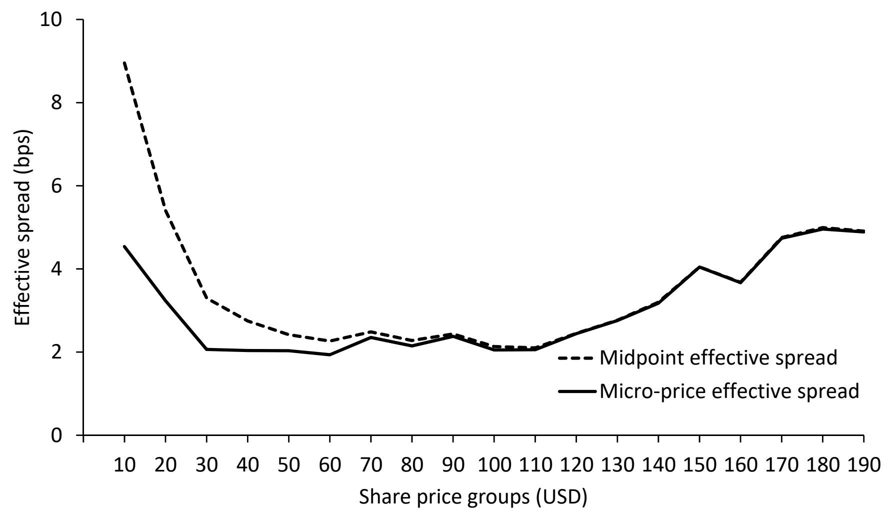
Panel (b) of Figure 2 shows the *Relative average bias* (solid line) and its 95% confidence interval (shaded area).[16] The results support the notion that stocks with lower prices (higher relative tick size) have a more severe overestimation problem. The price groups *USD10* and *USD20* display average overestimations of 96% and 68%. As indicated by the confidence interval, the overestimation is statistically significant for all price groups up to and including *USD110*, corresponding to 76% of the dollar volume and 90% of the trades.

Implications. The cross-sectional variation in the midpoint effective spread has implications for applications where the relative illiquidity of stocks is important. In asset pricing and corporate finance studies, for example, 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). Accordingly, I evaluate to what extent the bias feeds through to liquidity portfolios.

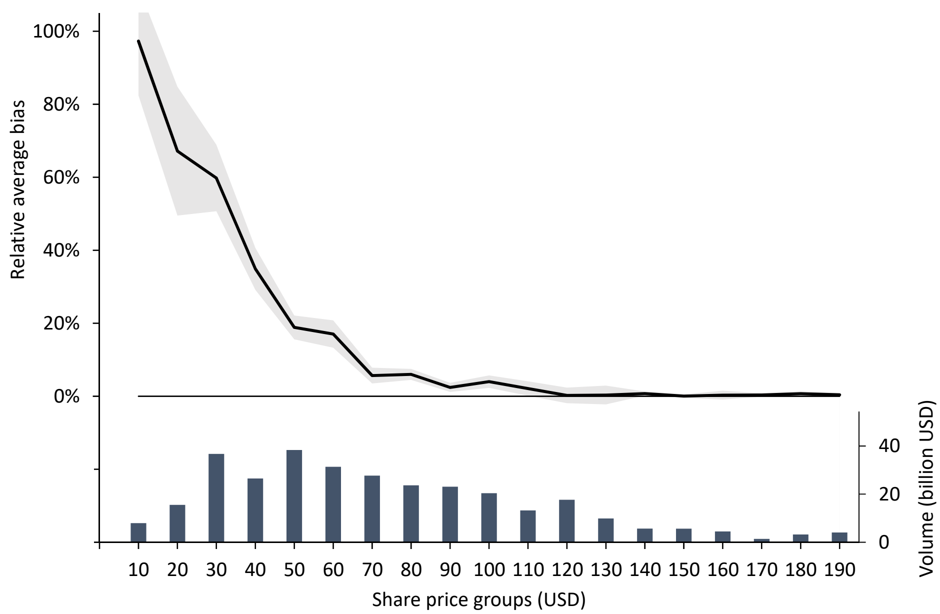
I form two sets of quintile portfolios – one based on the midpoint effective spread and one based on the micro-price effective spread. The sorts are repeated for each of the five days in the S&P500 sample. Following the methodology of Holden and Jacobsen (2014), I then calculate the percentage of stock-day observations where the midpoint effective spread allocates a stock to a lower or higher quintile compared to the micro-price effective spread portfolios.

I find that the stocks sorted by midpoint effective spread end up in the same quintile as when sorted by the micro-price effective spread in no more than 56% of the cases. The midpoint effective

[16]The confidence bounds are calculated for the *Nominal bias* of each price category and divided by the corresponding micro-price effective spread. Standard errors are based on residuals clustered by stock, date, and trading venue (Petersen, 2009).



(a) Effective spreads



(b) *Relative average bias*

Figure 2: Effective spread properties across trade price groups in the S&P 500 stocks. Panel (a) shows the effective spread relative the midpoint $\left(\tilde{S}^{mid}\right)$ and the micro-price $\left(\tilde{S}^{mic}\right)$ averaged across stocks in the same trade price group. Panel (b) presents the *Relative average bias* calculated for all trades in the same price group and its confidence interval, calculated as the *Nominal bias* mean and confidence interval divided by the average micro-price effective spread. Standard errors are based on residuals clustered by stock, date, and trading venue (Petersen, 2009). For variable definitions, see Table 1. Each price group corresponds to a price interval of USD 10. For example, the USD 20 price group includes all trades priced higher than USD 15 and lower than or equal to USD 25. 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 for December 7 – 11, 2015.

spread bias leads to that 27% of the stock-days are put in a more illiquid portfolio than they should, and 17% end up in a less illiquid portfolio.[17]

Systematic cross-sectional variation in the overestimation of the effective spread can also undermine the validity of research where the effective spread is an important outcome variable, such as market design evaluations. In Appendix C, I provide an application showing how the evaluation of liquidity following stock splits is affected by the bias. It shows that a stock split leads to that the midpoint effective spread bias doubles, from around 9% on average to more than 18%.

4.2 Intraday Variation and 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 timing depends on the trader's ability to observe and react to liquidity fluctuations.

Traders who gauge the fundamental value by the spread midpoint effectively overlook any fundamental value variation that does not cause a change in the spread midpoint. To see by how much this undermines such traders' liquidity timing ability, I analyze the components of the micro-price effective spread variance.

The micro-price effective spread variance may be decomposed as

$$\underset{[6.25]}{\mathrm{Var}\left[\tilde{S}^{mic}\right]} = \underset{[4.55]}{\mathrm{Var}\left[\tilde{S}^{mid}\right]} + \underset{[2.36]}{\mathrm{Var}\left[\tilde{S}^{mic} - \tilde{S}^{mid}\right]} + \underset{[-0.66]}{2\mathrm{Cov}\left[\tilde{S}^{mid}, \tilde{S}^{mic} - \tilde{S}^{mid}\right]}, \tag{8}$$

where the first component is the midpoint effective spread variance, the second component is the *Nominal bias* variance, and the third is the covariance between the midpoint effective spread and the *Nominal bias* (multiplied by -1).

I calculate each component of the micro-price effective spread variance using volume-weighted variances across all trades in each stock in the S&P500 sample. I separate the effective spreads paid by buyers and sellers, because the variance would otherwise include switches from ask-side to bid-side market orders, and vice versa. This is consistent with buyers, for example, primarily monitoring ask-side liquidity supply variation. I present the volume-weighted averages across stocks and direction of trade within squared brackets below each component in (8). The results show that an investor who is viewing liquidity variation through the lens of the midpoint effective

[17]Note that the number of artificially low-ranked stocks does not need to match the number of artificially high-ranked stocks. For example, if a stock is assigned to the fifth quintile instead of the third it counts as one case of a higher quintile, while it potentially pushes two other stocks to lower quintiles (one from the fifth to the fourth, and one from the fourth to the third).

spread overlooks 27% of the total variation ($1 - 4.55/6.25 = 27\%$).

Following the finding that the spread asymmetries are stronger for low-priced stocks, I repeat the estimation for stocks priced below USD 50. The results indicate a midpoint effective spread variance of only 2.60 basis points, which is 67% lower than the micro-price effective spread variance (7.80 bps). This indicates that the ability of investors to time their liquidity demand is severely undermined by excluding order book quantities from their information set, in particular when the tick size is binding.

Whereas the effective spread is an ex post measure of trading cost, liquidity timing is typically performed in real time. A relevant measure of *expected* trading cost of a buy market order can be defined as $C = 2(P^A - X)$.[18] This measure is closely related to the *Quoted spread*. The difference is that C is conditional on the trading direction, whereas the *Quoted spread* measures the unconditional cost of a hypothetical roundtrip trade. Decomposing the variance of C, I find that investors who use the midpoint as benchmark overlook about one third of the ex ante liquidity variation across all stocks, and around five sixths in low-priced stocks.[19]

Trading cost variance is also known as liquidity risk. An extensive literature explore how liquidity covaries across stocks (known as commonality in liquidity; see, e.g., Chordia et al., 2000; Hasbrouck and Seppi, 2001) and with systematic factors such as market returns (see, e.g., Acharya and Pedersen, 2005; Pastor and Stambaugh, 2003). I leave for future research to investigate how liquidity commonality and liquidity betas are affected by the midpoint effective spread bias.

4.3 Venue Rankings and RegNMS Rule 605

Current US equity market regulation embraces the midpoint as a fundamental value estimator. According to Rule 605 of RegNMS, all exchanges must publish monthly reports of their execution quality for each security traded. The Rule 605 reports include the average effective spread, defined in the same way as the midpoint effective spread in this paper, using the NBBO midpoint as the point of reference and volume weights when averaging across trades. The SEC (2001, Section I)

[18]The multiplication by two is for comparability to the other spread measures in the paper.

[19]Define an estimator of C as $\tilde{C}^v = 2\left(P^A - \tilde{X}^v\right)$. The decomposition of the buy-side ex ante cost variance, based on the micro-price, is then

$$\underset{[8.76]}{\mathrm{Var}\left[\tilde{C}^{mic}\right]} = \underset{[5.80]}{\mathrm{Var}\left[\tilde{C}^{mid}\right]} + \underset{[3.01]}{\mathrm{Var}\left[\tilde{C}^{mic} - \tilde{C}^{mid}\right]} + \underset{[-0.05]}{2\mathrm{Cov}\left[\tilde{C}^{mid}, \tilde{C}^{mic} - \tilde{C}^{mid}\right]}.$$

where the estimates provided in squared brackets are volume-weighted averages across stocks. The estimates indicate that the midpoint effective spread variance is 33.7% lower than the "true" liquidity variance ($1 - 5.80/8.76 = 33.7\%$). For stocks priced below USD 50, the corresponding number is 83.4%. Similar results are obtained for the ex ante cost of sell market orders.

motivates the disclosure requirement with that order routing decisions across trading venues must be *"well-informed and fully subject to competitive forces."* The bias documented above is thus particularly relevant in this context if it varies across trading venues, as it can then potentially misdirect order routing decisions.

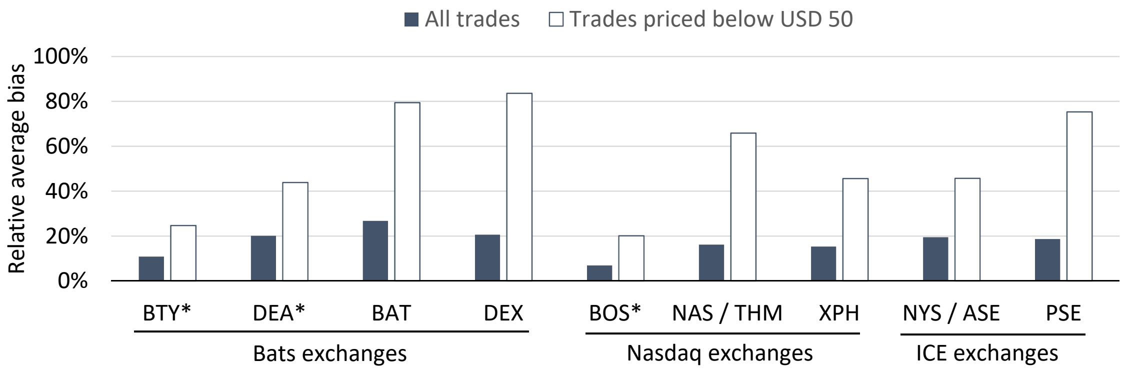
Figure 3, Panel (a), displays the *Relative average bias* for each trading venue in the sample. I exclude *MID* (Chicago Stock Exchange), since it represents only 0.01% of the total trading volume. The remaining nine exchanges are categorized by their corporate ownership (Bats, Nasdaq, and ICE). The bias is reported for all stocks (dark bars), and for all stocks priced below USD 50 (white bars). The results uncover substantial differences across exchanges. When considering all stocks, the bias ranges from 7% for *BOS* (Nasdaq BX) to 28%, on average, for trades executed at *BAT* (Bats BZX Exchange). For stocks priced below USD 50, the lowest bias is again for *BOS* (21%), and the highest is for *DEX* (Bats EDGX Exchange, with 86% bias).

What is the economics behind the bias variation across exchanges? A key distinguishing factor of modern equity exchanges is their fee schedules. Most venues subsidize liquidity suppliers by giving rebates to passively executed trades, and charge fees to actively executed trades (known as maker/taker fees). Some venues, however, do the opposite, which is known as inverted fees. To see why this can potentially influence the midpoint effective spread bias, consider again the logic of Glosten's (1994) model of liquidity supply, discussed in Section 1.2. With a maker rebate, a liquidity provider can make a profit in expectation even when the effective spread equals the expected adverse selection costs. Under the same conditions at an inverted fee venue, liquidity suppliers expect to make a loss. The consequence is that the asymmetry between bid- and ask-side spreads is higher in maker/taker fee venues, and accordingly that their bias is higher.

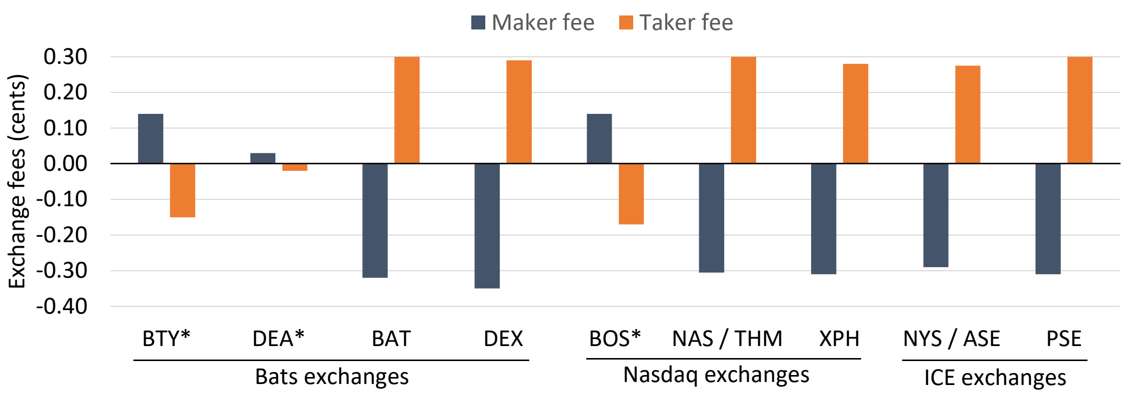
In Figure 3, venues with inverted fees are indicated by an asterisk (*). In addition, Panel (b) reports typical fees for makers and takers of liquidity. The exchanges offer rich variation in fees, depending on the order type and the status and volume traded of the member in question. The fees reported here are for trades executed using visible orders by members with the largest monthly trading volumes.[20]

My findings show that the three venues with inverted fee structures tend to have lower *Relative average bias*. Within the Bats exchange group, for example, I observe a monotonic negative relation between the maker fees and the bias seen for the stocks priced below USD 50 in Panel (a).

[20] I thank Shawn O'Donoghue for sharing the fee information. For data details, see O'Donoghue (2015). The fees presented for NYS/ASE and PSE are for Tape A stocks, which deviate somewhat from Tape B and C fees. All other venues have the same fees for all sample stocks.



(a) *Relative average bias*



(b) Exchange fees

Figure 3: Midpoint effective spread bias across trading venues. Panel (a) shows the *Relative average bias*, defined as in Table 1, for each trading venue in the cross-sectional sample. The sample includes all constituents of the S&P 500 index for the five trading days in the interval December 7 – 11, 2015. The dark bars show the results for all trades, whereas the white bars are conditioned on trades priced below USD 50. Panel (b) shows the fees charged to the liquidity suppliers (Maker fees; dark bars) and the liquidity demanders (Taker fees; light bars) for each exchange, measured in cents per share. The fees represent the amounts paid and received for trades executed using non-hidden orders by users in the large trading-volume brackets. In both panels, exchanges that apply an inverted fee schedule are indicated by *. The exchanges are categorized by their corporate ownership and sorted decreasingly by the maker fee. Exchange names corresponding to the three-letter abbreviations are spelled out in footnote 8.

Implications. A key question for investors' order routing decisions is whether the bias alters the ranking of trading venues in terms of execution quality. To address this issue, I compare venue rankings for the effective spread estimators based on the midpoint and the micro-price, respectively. For each stock-date and each effective spread measure, I rank the exchanges on a scale from one to nine. A rank of one indicates that a venue provides the tightest average effective spread, and a rank of nine shows that the venue has the worst execution quality for the given stock-date. Following Holden and Jacobsen (2014), I then compare the two effective spread estimators by computing their difference in rank for each stock-date. For example, if the exchange *BTY* is ranked 3 for a given stock-date in terms of the micro-price effective spread, but has a rank of 5 in terms of the midpoint effective spread, the rank difference is -2. Because there are nine trading venues in the sample (as above, I exclude MID from the analysis), the rank difference variable can potentially range from -8 to +8.

Table 2 presents the frequency of venue rank differences for each stock exchange in the sample. In Panel (a), which holds results for all S&P500 stocks, the "Average" column shows that the two effective spread estimators yield exactly the same ranking for a given venue in only half of all stock-days. Of all ranking differences that are different from zero, almost half of the cases are off by more than one step. Though rank differences of five steps or more in either direction are rare, it is notable that rank differences of the maximum 8 steps exist. Such cases indicate that a venue that is ranked highest according to one effective spread estimator, is ranked lowest according to the other.[21]

The inverted fee venues (*BTY*, *BOS*, and *DEA*) benefit from the midpoint effective spread bias in terms of higher exchange rankings. The rows marked *Lower Rank* and *Higher Rank* report the sum of rank differences below and above zero. According to these statistics, the inverted fee venues are more likely to be ranked artificially high than to be ranked artificially low. For example, *BTY* benefits from the bias in 26.7% of the stock-days, and suffer from it in only 11.9% of the rankings. The venues that apply maker/taker fees tend to have the opposite pattern.

Table 2, Panel (b), reproduces the analysis above for stocks priced below USD 50. As expected, the bias in venue rankings in this subset of stocks is even stronger. On average, only 31.3% of the cases record no difference between rankings based on the two effective spread estimators, compared to 49.9% for the full sample.

The conclusion from this application is that investors who base their order routing decision on the effective spreads reported by exchanges are potentially misdirected. The result is in sharp

[21] 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 bias is in general not mitigated by averaging across more trades.

Table 2: The difference in venue rankings across effective spread estimators. This table shows how effective spread venue rankings differ based on the effective spread estimator used. For each stock-day and each effective spread estimator, venues are ranked based on the effective spread. The table reports the difference in rankings obtained when using the micro-price effective spread and the midpoint effective spread. A positive (negative) rank difference indicates that a venue is ranked higher (lower) when the ranking is based on the micro-price effective spread, relative to the midpoint effective spread. The columns report the distribution of rank differences for each exchange, as well as a grand average. The venues are categorized by the exchange group and sorted decreasingly by the maker fee. Venues applying inverted fee scheduled are marked by *. The rows *Lower Rank* and *Higher Rank* report the sum of all negative and positive rank differences, respectively. The sample period is December 7 – 11, 2015. Panel (a) shows the results for all constituents of the S&P 500 index, whereas Panel (b) is restricted to stocks with an average trade price below USD 50. For variable definitions, see Table 1. For the full exchange names corresponding to the three-letter abbreviations, see footnote 8.

(a) All stocks

$Rank(\tilde{S}^{mic})$–	Bats exchanges				Nasdaq exchanges			ICE exch.		
$Rank(\tilde{S}^{mid})$	*BTY**	*DEA**	*BAT*	*DEX*	*BOS**	*NAS*	*XPH*	*NYS*	*PSE*	Average
-8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.5%	0.0%	0.1%
-7	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	0.0%	0.1%
-6	0.0%	0.0%	0.0%	0.4%	0.0%	0.1%	0.6%	0.4%	0.0%	0.2%
-5	0.0%	0.0%	0.1%	1.5%	0.0%	0.4%	0.7%	0.8%	0.4%	0.4%
-4	0.0%	0.3%	0.4%	3.5%	0.0%	1.1%	2.2%	1.2%	1.6%	1.1%
-3	0.4%	1.9%	2.6%	5.6%	0.2%	2.3%	4.7%	3.3%	4.1%	2.8%
-2	1.4%	6.4%	7.7%	9.7%	1.2%	5.3%	7.5%	6.3%	9.1%	6.1%
-1	10.1%	11.7%	24.1%	19.0%	12.5%	11.8%	12.8%	14.7%	20.4%	15.2%
0	61.4%	44.3%	45.7%	42.3%	54.7%	48.8%	49.1%	54.2%	49.0%	49.9%
1	18.6%	16.1%	13.0%	12.7%	11.7%	21.1%	8.7%	12.5%	10.1%	13.8%
2	4.6%	9.8%	4.6%	4.0%	7.1%	6.8%	4.7%	3.4%	3.3%	5.4%
3	1.3%	5.0%	1.5%	1.1%	4.6%	1.9%	3.1%	1.3%	1.6%	2.4%
4	1.1%	2.8%	0.3%	0.2%	3.0%	0.3%	2.1%	0.8%	0.4%	1.2%
5	0.6%	1.3%	0.0%	0.2%	2.1%	0.1%	2.5%	0.1%	0.0%	0.8%
6	0.4%	0.4%	0.0%	0.0%	1.6%	0.0%	0.6%	0.0%	0.0%	0.3%
7	0.1%	0.0%	0.0%	0.0%	1.0%	0.0%	0.3%	0.1%	0.0%	0.2%
8	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.2%	0.0%	0.0%	0.0%
Lower Rank	11.9%	20.3%	34.9%	39.7%	13.9%	21.0%	28.6%	27.5%	35.6%	
Higher Rank	26.7%	35.4%	19.4%	18.2%	31.3%	30.2%	22.2%	18.2%	15.4%	

Table 2: The difference in venue rankings across effective spread estimators. Continued from previous page.

(b) Stocks priced below USD 50

$Rank(\tilde{S}^{mic})-$	Bats exchanges				Nasdaq exchanges			ICE exch.		
$Rank(\tilde{S}^{mid})$	*BTY**	*DEA**	*BAT*	*DEX*	*BOS**	*NAS*	*XPH*	*NYS*	*PSE*	Average
-8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	1.1%	0.0%	0.1%
-7	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.3%	0.6%	0.1%	0.1%
-6	0.0%	0.0%	0.0%	1.1%	0.0%	0.3%	1.2%	0.9%	0.1%	0.4%
-5	0.1%	0.1%	0.2%	3.3%	0.0%	0.9%	1.2%	1.7%	0.8%	0.9%
-4	0.0%	0.6%	0.5%	7.9%	0.1%	2.6%	3.7%	2.6%	3.6%	2.4%
-3	0.3%	4.3%	4.6%	11.7%	0.2%	4.7%	7.1%	6.9%	7.5%	5.2%
-2	1.4%	14.1%	9.7%	17.8%	0.9%	11.3%	10.3%	11.7%	13.9%	10.1%
-1	12.1%	12.7%	21.5%	22.1%	9.2%	17.0%	13.3%	19.8%	21.7%	16.6%
0	47.1%	25.2%	30.4%	22.6%	31.5%	30.1%	27.0%	35.9%	32.1%	31.3%
1	23.5%	15.4%	19.7%	9.0%	16.5%	18.4%	8.9%	11.0%	10.4%	14.8%
2	8.5%	12.7%	8.9%	3.3%	14.7%	10.6%	7.1%	3.7%	5.9%	8.4%
3	2.5%	7.2%	3.8%	0.9%	9.5%	3.3%	5.9%	2.0%	3.2%	4.2%
4	2.1%	4.5%	0.7%	0.2%	6.5%	0.6%	5.0%	1.5%	0.7%	2.4%
5	1.3%	2.6%	0.0%	0.2%	4.4%	0.1%	6.2%	0.3%	0.0%	1.7%
6	0.7%	0.6%	0.0%	0.0%	3.6%	0.0%	1.4%	0.0%	0.0%	0.7%
7	0.3%	0.1%	0.0%	0.0%	2.5%	0.0%	0.7%	0.2%	0.0%	0.4%
8	0.0%	0.0%	0.0%	0.0%	0.4%	0.0%	0.5%	0.0%	0.0%	0.1%
Lower Rank	13.9%	31.8%	36.5%	63.9%	10.4%	36.9%	37.2%	45.3%	47.7%	
Higher Rank	38.9%	43.1%	33.1%	13.6%	58.1%	33.0%	35.7%	18.7%	20.2%	

contrast with the regulator's ambition with Rule 605.

4.4 Trading Performance Evaluations

To evaluate the performance of liquidity suppliers, it is common to decompose the effective spread into the price impact and the realized spread. The former is a measure of how much the market maker is losing to liquidity demanders due to trade-induced changes in the fundamental value. The latter is the effective spread net of price impact, which should cover all other costs and potentially leave the market maker with a profit. Just like the effective spread, the realized spread is part of the Rule 605 report requirement. The SEC motivates the reporting requirement with that realized spreads show the extent to which trading venues keep trading at times of stress, and whether exchanges differ in their ability to avoid trading with informed traders (SEC, 2001). Similar arguments for decomposing the effective spread in the evaluation of fast traders are put forward by Hendershott et al. (2011).

The realized spread estimator is defined as

$$\tilde{RS}_s^v = 2D_t\left(P_t - \tilde{X}_{t+s}^v\right), \tag{9}$$

where s is the horizon for the evaluation, which I set to five minutes after the trade time (t). The price impact estimator is denoted $\tilde{PI}_s^v$ and defined as the difference between the effective and the realized spreads, such that

$$\begin{array}{rcccc} \tilde{S}^v & = & \tilde{PI}_s^v & + & \tilde{RS}_s^v. \\ \textit{mid}\text{ (bps)} & & 3.87 & & -0.65 \\ \textit{mic}\text{ (bps)} & & 3.36 & & -0.64 \\ \textit{NomBias}\text{ (bps)} & & 0.50 & & -0.01 \\ \textit{RelBias}\text{ (\%)} & & 15\% & & -2\% \end{array} \tag{10}$$

I measure volume-weighted average price impact and realized spread, scaled by the midpoint, for each stock in the S&P500 sample.[22] The results for $v = [mid, mic]$ are presented below each component of the effective spread in (10), along with their corresponding nominal bias (abbreviated *NomBias*) and relative average bias (*RelBias*). The main takeaway from the decomposition is that the effective spread bias carries over to the price impact metric, which is overestimated by 15%, but not to the realized spread.[23]

Just like for the midpoint effective spread bias, the midpoint price impact bias is worse for low-priced stocks. For stocks priced lower than USD 50, the midpoint price impact is overestimated by 37%, whereas the bias in the midpoint realized spread remains small (-1%).

The bias in price impact is relevant to analyses of investors' ability to predict and trade ahead of price changes. For example, high price impact in active trading is often interpreted as a sign of informed trading (due either to private information or to an early response to public information). If that price impact is based on midpoint changes, however, that skill may reflect "nowcasting" (the ability to estimate the current fundamental value) as much as forecasting. In my results, the nowcasting component of price impact is the *NomBias* of 0.50 bps.

[22]To preserve the equality in (10), I do not winsorize the spread components.

[23]The five-minute horizon is consistent with the specification in Rule 605. On the 10 and 60 seconds horizons, the midpoint price impact relative average bias is 23% and 19%, respectively. The corresponding numbers for realized spread bias are 1% and 6%, respectively.

5 Variation Across Investor Types

Do investors know about the midpoint effective spread bias? Its mere existence is evidence that some investors do, otherwise the relation between the direction of trade and the fundamental value deviation from the midpoint (in Figure 1) would be flat. Though the midpoint estimator of the effective spread is widely used in the academic literature and maintains regulatory support both in the US and in the EU, it may be that investors in general use more sophisticated metrics. If that is the case, regulators should reconsider their support to the metric, but the impact on financial markets of the findings presented here would be limited.

If investors differ in their ability to accurately measure liquidity, however, the bias can potentially drive a wedge between the execution costs of, for example, professional and retail investors. The regulatory support to the midpoint effective spread is then problematic, as it potentially induces confidence in that estimator among non-sophisticated investors. The regulation can then amplify differences in the liquidity timing ability across investors.

In this section, I employ the HFT sample to analyze differences in execution costs across investor types. The HFTs are known to invest heavily in technology in order to monitor and respond to information in real time (Brogaard et al., 2015; SEC, 2001), and to have strong intraday market timing ability (Carrion, 2013). They are also able to predict price changes in the short term (Brogaard et al., 2014), which implies that they, similar to how the micro-price is constructed, factor in probabilities of future midpoint changes in their analysis of the markets. The HFT group may thus be viewed as sophisticated traders in terms of liquidity timing. The Non-HFTs is a more heterogenous group of investors, with trade flows originating from retail and institutional investors, but also highly sophisticated proprietary traders that are not categorized as HFTs. On average, this group is likely to be less sophisticated than the HFTs in terms of limit order book monitoring.

5.1 Differences in Liquidity Demand Elasticity

The first point of the analysis across investor groups is to compare the sensitivity of the direction of trade to the fundamental value deviation from the midpoint. Following the same trade categorization procedure as in section 3.1, I plot the frequency of buyer-initiated trades for each trader group in Figure 4. The dark curve represents trades initiated by HFTs, and the light curve is for trades where Non-HFTs consume liquidity. The curves observed here are not as smooth as for the S&P500 sample, which is likely due to that the HFT sample holds much fewer trades. Nevertheless, the positive relation observed in Figure 1 holds for both investor groups. Furthermore, consistent with that HFTs are better positioned to align their liquidity demand to variation in the

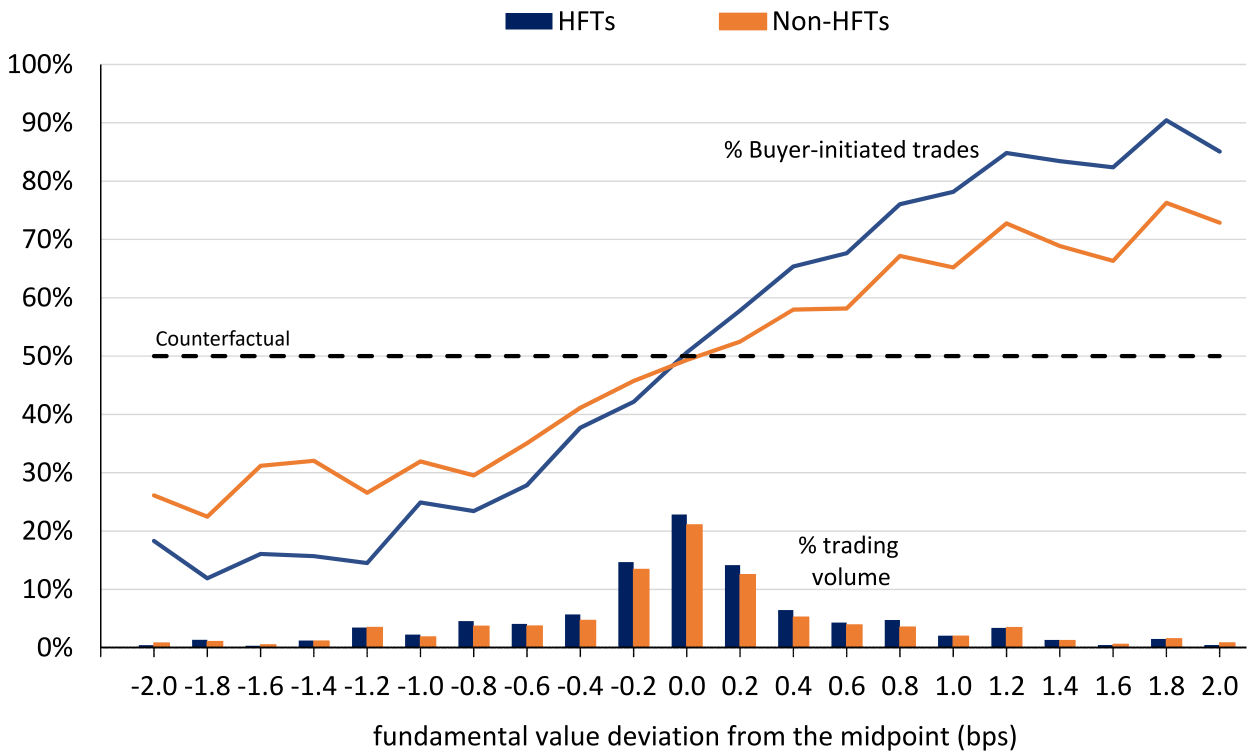


Figure 4: Liquidity demand elasticity across investor groups. This figure shows trade characteristics for two investor groups, HFTs and Non-HFTs, respectively. The trades are categorized by the fundamental value deviation from the midpoint, measured on the x-axis in the same way as in Figure 1. The solid lines display for each investor group the fraction of all non-midpoint trades that are buyer-initiated according to the Lee and Ready (1991) algorithm. The bar plot shows for each investor group the category share of the total dollar volume of non-midpoint trades. The sample includes Nasdaq-executed trades in 120 stocks, covering five trading days in the interval February 22 – 26, 2010.

effective spread, I find that the HFT curve is steeper than the Non-HFT curve.

To assess the statistical significance of the difference in slopes between the two investor groups, I estimate the probit model:

$$\Pr(Buy_t) = \underset{(-2.14)}{-0.04} + \underset{(12.39)}{0.34}\left(\log \tilde{X}_t^{mic} - \log \tilde{X}_t^{mid}\right) + \underset{(2.49)}{0.04}HFT_t\left(\log \tilde{X}_t^{mic} - \log \tilde{X}_t^{mid}\right) + \varepsilon_t. \tag{11}$$

where t is a trade index, Buy_t equals one for buyer-initiated trades and zero for seller-initiated trades, HFT_t is an indicator for trades where HFTs demand liquidity, and variation that is unexplained by the model is captured by the residual term ε_t. The estimated coefficients are reported in (11). The results indicate a positive relation between the direction of trade and the fundamental value deviation from the midpoint, and that the relation is stronger for the HFT group. The z-statistics (within parentheses, based on standard errors that are clustered by stock and date) show that both coefficients are significantly different from zero.[24]

The slope difference is important because it shows that the liquidity timing differs across investor types. I note that this difference is not necessarily due to variation in the liquidity timing *ability*; it could also follow from differences in investors' liquidity demand elasticities. Either way, given that the Non-HFT group is highly diverse, including everything from investment bank trading desks to retail platforms, the reported differences should be viewed as conservative. Next, I quantify how the difference influences the cost of trading.

5.2 Differences in Effective Spread Bias

I calculate the same effective spread properties as in Table 1, but for four partially overlapping trade categories: (i) trades where HFTs consume liquidity; (ii) trades where Non-HFTs consume liquidity; (iii) trades where HFTs act as liquidity suppliers; and (iv) trades where a Non-HFTs act as liquidity suppliers. Table 3, Panel (a), reports the value-weighted average effective spread using the midpoint and the micro-price estimators for each trade category, as well as the bias in nominal and relative terms. As a point of comparison to the S&P500 sample, I also report the unconditional average across all trades. To assess the statistical significance of reported differences between HFTs and Non-HFTs, two-sample dollar-volume weighted *t*-tests with standard errors clustered by stock and date are employed. Furthermore, standard *t*-tests with the same type of clustering are used to test the statistical significance of the *Nominal bias* in each column.

Overall, the *Relative average bias* is higher in the HFT sample (65%) than in the S&P500 sample (18%). Breaking down the HFT sample into market-cap segments (Panel (b)), the bias is highest for large-caps (72%), but remains economically significant for mid-caps (28%) and small-caps (14%).[25]

The effective spread incurred to liquidity demanders categorized as HFTs is vastly overstated by the midpoint effective spread. Measured across all sample stocks, the midpoint effective spread is recorded at 2.22 basis points on average, whereas the micro-price version is about half of that, at 1.13 bps. This confirms the prior that HFTs have a strong ability to time their liquidity demand (Carrion, 2013). My findings indicate that HFTs time their trades by tracking the true value of the

[24]The reason that the standard errors in this section are not clustered on trading venue is that the sample contains trade observations from Nasdaq only.

[25]Appendix D repeats the full analysis of Table 1 using the HFT sample instead of the S&P500 sample. It shows that the HFT sample stocks have higher liquidity on average, due to that the included large-cap stocks are more liquid than the average S&P500 sample stocks, That is also a potential explanation for the higher bias recorded in the HFT sample. The HFT sample effective spreads are however also more dispersed, with a standard deviation of the effective spread measures being about three times higher than for the S&P500 sample. This is due to that the sample is stratified across market capitalization segments.

Table 3: Differences in bias for HFTs and Non-HFTs. This table shows effective spread measures for all traders as well as for the groups of HFTs and Non-HFTs. The group activity is further broken down by liquidity demand and supply. The sample includes 120 stocks, covering five trading days in the interval February 22 – 26, 2010. Panel (a) shows the average effective spread obtained using the midpoint and the micro-price as fundamental value estimators ($\tilde{S}^{mid}$ and $\tilde{S}^{mic}$, respectively), as well as the *Nominal bias* and *Relative average bias* associated with the midpoint effective spread. All variables are defined as in Table 1. Differences between HFTs and Non-HFTs are reported as dollar-volume weighted averages. Statistical significance of the differences is tested using weighted t-tests of averages across stock-day observations, and indicated in the "Diff." columns with ** and * indicating the 95% and 90% confidence levels, respectively. Significance of the *Nominal bias* within each trader group (HFTs and Non-HFTs) is tested for using standard t-tests and is indicated in the same way. The standard errors of all statistical tests are clustered by stock and date, following Petersen (2009). Panel (b) repeats the analysis for different market-cap segments.

(a) Effective spread bias

		Liquidity demand			Liquidity supply		
	All traders	HFT	Non-HFT	Diff.	HFT	Non-HFT	Diff.
Large-caps (N = 40)							
$\tilde{S}^{mid}$ (bps)	2.39	2.22	2.58	0.36**	2.51	2.28	–0.23**
$\tilde{S}^{mic}$ (bps)	1.45	1.13	1.80	0.67**	1.57	1.33	–0.24**
Nominal bias (bps)	0.95**	1.10**	0.78**	–0.31**	0.94**	0.95**	0.01
Relative average bias	0.65	0.97	0.43		0.60	0.72	

asset, rather than the midpoint. This results is consistent across all market cap segments.

Non-HFTs pay higher effective spreads than do HFTs, regardless which effective spread metric is applied. This is unsurprising, as the Non-HFTs include investors with trading needs that are unrelated to the state of the market, whereas the HFTs typically trade only when they can profit from the trading process itself. More importantly for the purpose of this article, the Non-HFTs also have a lower midpoint effective spread bias. The Non-HFT relative average bias is 43%, compared to 97% for the HFTs. This finding is consistent with that many investors in the Non-HFT group overlook trading cost variation that is not reflected by changes in the quoted prices.

In liquidity supply, I find that HFTs earn significantly higher spreads than do Non-HFTs. This difference is however independent of the effective spread estimator used, as there is no difference in *Nominal bias* for the two groups (except for mid-cap stocks, where there is a marginally significant difference in the *Nominal bias*). This is somewhat surprising, as market making is a central strategy to HFTs, and an important part of such activities is to avoid to trade when the price is about to change.

Table 3: Differences in bias for HFTs and NonHFTs. Continued from previous page.

(b) Variation across market capitalization segments

		Liquidity demand			Liquidity supply		
	All traders	HFT	Non-HFT	Diff.	HFT	Non-HFT	Diff.
Large-caps (N = 40)							
$\tilde{S}^{mid}$ (bps)	2.27	2.15	2.40	0.25*	2.43	2.12	−0.31**
$\tilde{S}^{mic}$ (bps)	1.32	1.06	1.62	0.56**	1.49	1.17	−0.32**
Nominal bias (bps)	0.95**	1.09**	0.78**	−0.31**	0.94**	0.95**	0.01
Relative average bias	0.72	1.03	0.48		0.63	0.81	
Mid-caps (N = 34)							
$\tilde{S}^{mid}$ (bps)	4.33	3.87	4.63	0.76**	4.43	4.27	−0.16
$\tilde{S}^{mic}$ (bps)	3.38	2.74	3.80	1.05**	3.58	3.27	−0.31
Nominal bias (bps)	0.95**	1.13**	0.83**	−0.29**	0.85**	1.00**	0.15*
Relative average bias	0.28	0.41	0.22		0.24	0.31	
Small-caps (N = 39)							
$\tilde{S}^{mid}$ (bps)	8.30	7.09	8.77	1.69**	9.20	7.97	−1.23*
$\tilde{S}^{mic}$ (bps)	7.25	5.01	8.13	3.12**	8.23	6.89	−1.33**
Nominal bias (bps)	1.05**	2.07**	0.64*	−1.43**	0.98**	1.07**	0.10
Relative average bias	0.14	0.41	0.08		0.12	0.16	

5.3 Policy Implications

The evidence above indicates that investors differ in the extent to which they adapt their trade flows to the prevailing fundamental value. To protect the least sophisticated investors, policy-makers could make estimates of the fundamental value available to the investor community. In the case of US equities, for example, the SIPs could be given the task to report a fundamental value estimator in real time, along with the NBBO feed. Such dissemination would facilitate liquidity timing for investors who are unable to infer the true value with in-house analysis.

That the micro-price is relatively expensive to compute should not be viewed as an obstacle to real-time dissemination, because the computationally costly operations could be done before the market opens. Specifically, the midpoint adjustment function $g(S^{quoted}, I)$ can be pre-estimated for all relevant combinations of quoted spread and order book imbalance, using historical data. For examples of such estimates, see Figure A.1 in Appendix A. What remains to do in real time is then to simply map the current quoted spread and order book imbalance to the relevant midpoint adjustment.

Given the prevalence of the midpoint effective spread bias, and in particular the bias variation across exchanges (see Section 4.3), what is the value of the Rule 605 reports? One can argue that the midpoint effective spread is the most relevant metric to the non-sophisticated investors. If such investors are unable to distinguish the midpoint from the true fundamental value, their market order submissions will be unrelated to the fundamental value deviation from the midpoint. Then, by the reasoning in Section 1.2, the midpoint effective spread is an accurate metric of their execution cost. For sophisticated investors, who are able to proxy the fundamental value with higher accuracy, the Rule 605 data is not needed, but presumably does no harm. A concern, however, is that the regulators' use of the midpoint may lull non-sophisticated investors into a false sense of confidence in that fundamental value estimator. This could amplify differences between investors.

6 A Second-Best Solution: The Weighted Midpoint Effective Spread

I view the micro-price by Stoikov (2018) as the most appropriate estimator of the fundamental value of security in applications that need to capture high-frequency fluctuations. As discussed above, its main traits are that it is a martingale by definition, that it is unconstrained by price discreteness, and that it can be pre-calculated such that real-time estimates are computationally cheap. For individual researchers and investors alike, however, that pre-calculation may not be feasible. In this section, I discuss computationally cheap estimators that can serve as second-best solutions.

6.1 Weighted Midpoint Estimators of the Fundamental Value

The weighted midpoint is an estimator where the fundamental value deviation from the midpoint is a linear function of the quoted spread and the order book imbalance.

Stoikov (2018) criticizes the weighted midpoint for not being a martingale and for occasionally generating undesirable features. A counterintuitive example, he argues, is that a *lowered* bid-side price can lead the weighted midpoint to indicate a *higher* fundamental value. Furthermore, Stoikov (2018) finds empirically that the weighted midpoint is outperformed by the micro-price in the ability of predicting future midpoint changes.

Harris (2013) proposes that the weighted midpoint should be adjusted to account for order processing costs. I adapt his idea to allow for different fees on the bid and ask sides of the limit order book. Rewriting his Equation (6) in terms of my notation, and allowing the order processing costs on the bid and ask sides to differ, I obtain a general expression for the weighted midpoint,

$$\tilde{X}^{wmid} = \tilde{X}^{mid} + \left(S^{quoted}/2 - \gamma^A\right) I - \left(S^{quoted}/2 - \gamma^B\right)(1 - I), \tag{12}$$

where γ^A and γ^B represent the fees incurred to the liquidity supplier on the bid- and ask-side, respectively.[26]

The fees required for the weighted midpoint estimator are not straightforward to approximate. First, the appropriate fee level to apply on each side of the limit order book is that faced by the liquidity supplier who posts the last order in the queue (following the reasoning by Glosten, 1994), and fee schedules in US equity markets differ across investors, primarily depending on the trading volumes. I address this issue by assuming that the marginal limit order is posted by the trader with the lowest marginal cost of transacting. Accordingly, I apply the fees and rebates associated with the highest volume bracket. Second, there are no consolidated databased of historical fee schedules. I implement the fee levels as presented in Figure 3, Panel (b), except that, where applicable, I also allow fees to vary across stock segments (see footnote 20).

I consider three versions of the weighted midpoint estimator, distinguished by what fees are accounted for.

- **Weighted midpoint without fees.** By imposing the constraint $\gamma^A = \gamma^B = 0$ the standard definition of the weighted midpoint is obtained: $\tilde{X}^{wmid} = \tilde{X}^{mid} + S^{quoted}(I - 0.5)$. Although this estimator overlooks the effect of fees, it has the advantage that it is straightforward to apply to long time series where fee levels are difficult to obtain.

- **Weighted midpoint with exchange fees.** This estimator sets γ^A and γ^B to the maker fees of the exhange from which the best quote on each side is provided to the NBBO. This is akin to the estimator proposed by Harris (2013), adapted to the NBBO setting.

- **Weighted midpoint with exchange, clearing, and Section 31 fees.** In addition to the exchange fees, transaction in the US equity market are subject to clearing fees. In the US, virtually all equity trades are cleared by DTCC. I consider the gross volume fee (known as "into the net fee"), which in 2015 was about 0.003 basis points.[27] Finally, there is a cost known as the "Section 31 fee", which is charged by the SEC to the seller in all transactions. At the time of the S&P500 sample, it amounted to around 0.186 bps. Taken together, for this

[26]To obtain this expression, note that $P^B = \tilde{X}^{mid} - S^{quoted}/2$ and $P^A = \tilde{X}^{mid} + S^{quoted}/2$.

[27]For details on DTCC fees, see https://www.dtcclearning.com/products-and-services/equities-clearing/universal-trade-capture-utc/nscc-equity-trade-capture-fee-descriptions.html.

estimator, I set the order processing cost equal to the exchange maker fees plus the DTCC fee. On the ask-side (γ^A), I also add the Section 31 fees.

To my knowledge, none of these estimators have been applied to effective spread measurement before. Cartea et al. (2015, p. 71) suggest that the weighted midpoint would potentially be a more economically meaningful benchmark than the midpoint when accounting for the effective spread in algorithmic trading strategies, but they do not elaborate further on the issue.

6.2 Effective Spread Bias Across Estimators

In Figure 5, I repeat the analysis of Figure 2 for the weighted midpoint estimators. As a point of reference, I include the corresponding result for the midpoint. All estimators are evaluated with the micro-price effective spread as benchmark.

The results show that using the weighted midpoint estimators of the effective spread overcomes much of the problem for stocks priced below USD 115. For example, stocks in the USD 10 price group have a 23% bias when using the weighted midpoint without fees, which is a substantial reduction compared to the midpoint effective spread bias of 97%. For stocks priced above USD 130, however, the weighted midpoint estimators yield a slight underestimation. The closest match to the micro-price effective spread is obtained when accounting for exchange, clearing, and Section 31 fees. The resulting estimator is close to zero bias for all price groups, with an average of -2%. Accounting for exchange fees only yield an underestimation of the effective spread for low-priced stocks. Overall, the *Relative average bias* for the weighted midpoint estimators range from -4% (adjusted for exchange fees) to 4% (without fees), as compared to 18% for the midpoint.

To assess the impact of the bias in each effective spread estimator across price groups, I evaluate to what extent the bias feeds through to liquidity portfolios. Following the methodology outlined in Section 3.2, I form quintile portfolios based on each the different effective spreads estimator. Table 4 repeats the results for the midpoint effective spread, as reported inSection 3.2, in the first row. The subsequent rows show the corresponding portfolio misclassifications for the weighted midpoint estimators.

The weighted midpoint effective spread mitigates the misallocation problem reported above for the midpoint portfolios. When the effective spread is benchmarked to the weighted midpoint without fees, almost four out of five stocks (78%) are assigned to the same portfolio as when using the micro-price. When accounting for exchange fees (and clearing and Section 31 fees), the agreement between the micro-price and the weighted midpoint portfolio assignments reaches 85%

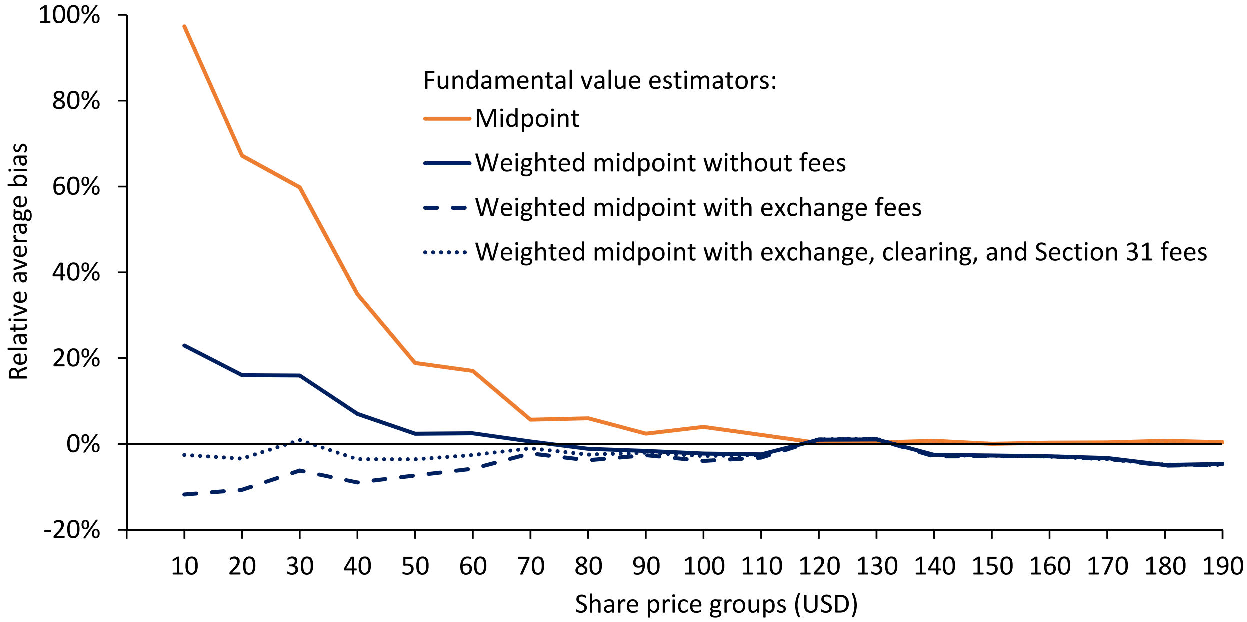


Figure 5: Midpoint effective spread bias across different estimators of fundamental value. This figure shows the *Relative average bias* for a set of effective spread estimators. The effective spread $\tilde{S}^{v}$ is defined as the signed difference between the transaction price and the prevailing fundamental value $\tilde{X}^{v}$, where v denotes the type of fundamental value estimator. The fundamental value estimators include the *Midpoint*, *Weighted midpoint without fees*, *Weighted midpoint with exchange fees*, and the *Weighted midpoint with exchange, clearing, and Section 31 fees*. The *Relative average bias* is defined as the average difference $\tilde{S}^{v} - \tilde{S}^{mic}$ divided by the average $\tilde{S}^{mic}$. The sample stocks are split into price buckets in the same way as in Figure 2. The sample includes all constituents of the S&P 500 index for the date interval December 7 – 11, 2015.

Table 4: Differences in effective spread quintiles. For each stock-day and each effective spread estimator, venues are ranked based on the effective spread and split into quintile portfolios, where portfolio 1 is the most liquid and portfolio 5 is the least liquid. The table reports the difference in portfolio assignments obtained when using the micro-price effective spread and one of the other effective spread estimators (*Midpoint*, *Weighted midpoint without fees*, *Weighted midpoint with exchange fees*, and the *Weighted midpoint with exchange, clearing, and Section 31 fees*). A negative (positive) quintile portfolio difference indicates that a stock is assigned to a more liquid (less liquid) portfolio when the ranking is based on the micro-price effective spread, . The rows report for each effective spread estimator the fraction of stock-days where the portfolio difference is lower, equal, and higher, relative to the micro-price effective spread estimator. The sample includes all constituents of the S&P 500 index for the date interval December 7 – 11, 2015.

	Lower quintiles	Same quintile	Higher quintiles
Midpoint	27 %	56 %	17 %
Weighted midpoint without fees	13 %	78 %	9 %
Weighted midpoint with exchange fees	8 %	85 %	8 %
Weighted midpoint with exchange, clearing, and Section 31 fees	6 %	88 %	6 %

(87%).[28]

6.3 Discussion

I conclude from the analysis in this section that the weighted midpoint effective spread is a viable alternative to the micro-price effective spread. Though the weighted midpoint estimators are not unbiased, they constitute a computationally cheap alternative where most of the bias in the midpoint effective spread is alleviated. The weighted midpoint is a linear function of the spread and the order book imbalance, which makes it feasible to estimate in real time for investors with access to the consolidated tape. Academic researchers can easily calculate it using the NBBO information available in DTAQ, MTAQ, or TRTH. In empirical settings where data on exchange fees are available, my findings show that a fee adjustment can improve the estimator further.

My findings complement those of Holden and Jacobsen (2014). They show that accurate estimation of the effective spread requires data from DTAQ (which is equivalent to TRTH data). I show that the DTAQ midpoint, which they use as their benchmark, should be replaced by a continuous fundamental value estimator such as the micro-price. For analysts that are computationally and financially constrained, their second-best solution based on MTAQ data is straightforward be combine with my second-best solution using a weighted midpoint estimator.

[28] As noted above, the numbers reported in the "Lower quintiles" column are not necessarily equal to those of the "Higher quintiles" column.

7 Conclusion

I show that the midpoint effective spread is a biased estimator of the effective spread. The bias varies systematically across stocks and trading venues, and it is found to undermine liquidity timing and trading performance evaluations. I argue that the bias is driven by price discreteness and differential fee structures across venues. The bias is economically and statistically significant, and robust across market capitalization segments and across continuous fundamental value estimators.

Importantly, I find sizable differences across investor groups in the ability to gauge the fundamental value. The implication is that whereas sophisticated investors understand the midpoint effective spread bias to a large extent, others do not. I argue that regulators may be able to bridge this gap by mandating SIPs to report a more accurate fundamental value estimator, and by changing the Rule 605 report requirements.

It should be noted that the problem with using the midpoint as proxy for the fundamental value is application-specific. The reason that the fundamental value deviation from the midpoint is important for the effective spread is that it *influences* the order flow. It is because the tight side of the spread attracts more flow than the wide side that the bias survives when averaging across trades. In other applications, such as measurement of returns and realized variance, I do not expect the fundamental value deviation from the midpoint to be problematic. The midpoint effective spread bias may, however, influence our understanding of liquidity risk, the merit of low-frequency liquidity estimators, liquidity premia in asset pricing, and corporate finance issues that are related to market liquidity. I leave these issues for future research.

References

Abdi, F. and Ranaldo, A. (2017). A simple estimation of bid-ask spreads from daily close, high, and low prices. *Review of Financial Studies*, 30(12):4437–4480.

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.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

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

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.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306.

Brogaard, J., Hendershott, T., and Riordan, R. (2017). High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics*, 124(1):22–42.

Carrion, A. (2013). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets*, 16(4):680–711.

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

Chakrabarty, B., Pascual, R., and Shkilko, A. (2015). Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the lee-ready algorithm. *Journal of Financial Markets*, 25:52–79.

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

Chordia, T., Roll, R., and Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics*, 56(1):3–28.

Chordia, T., Roll, R., and Subrahmanyam, A. (2001). Market liquidity and trading activity. *The Journal of Finance*, 56(2):501–530.

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

Cont, R., Kukanov, A., and Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, 12(1):47–88.

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.

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.

Goettler, R. L., Parlour, C. A., and Rajan, U. (2005). Equilibrium in a dynamic limit order market. *The Journal of Finance*, 60(5):2149–2192.

Gould, M. D. and Bonart, J. (2016). Queue imbalance as a one-tick-ahead price predictor in a limit order book. *Market Microstructure and Liquidity*, 2(02):1650006.

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

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, 50(4):1175–1199.

Hasbrouck, J. (2002). Stalking the efficient price in market microstructure specifications: an overview. *Journal of Financial Markets*, 5(3):329–339.

Hasbrouck, J. (2003). Intraday price formation in us equity index markets. *The Journal of Finance*, 58(6):2375–2400.

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

Hasbrouck, J. and Seppi, D. (2001). Common factors in prices, order flows, and liquidity. *Journal of Financial Economics*, 59(3):383–411.

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.

Kavajecz, K. A. (1999). A specialist's quoted depth and the limit order book. *The Journal of Finance*, 54(2):747–771.

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

Lease, R. C., Masulis, R. W., and Page, J. R. (1991). An investigation of market microstructure impacts on event study returns. *The Journal of Finance*, 46(4):1523–1536.

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

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

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*.

Muravyev, D. and Pearson, N. (2016). Option trading costs are lower than you think. *Working paper*.

Næs, R., Skjeltorp, J. A., and Ødegaard, B. A. (2011). Stock market liquidity and the business cycle. *The Journal of Finance*, 66(1):139–176.

O'Donoghue, S. M. (2015). The effect of maker-taker fees on investor order choice and execution quality in us stock markets. *Kelley School of Business Research Paper*, 15(44).

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

Parlour, C. A. (1998). Price dynamics in limit order markets. *The Review of Financial Studies*, 11(4):789–816.

Pastor, L. and Stambaugh, R. (2003). Liquidity risk and stock returns. *Journal of Political Economy*, 11:642–685.

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.

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*.

Shkilko, A. V., Van Ness, B. F., and Van Ness, R. A. (2008). Locked and crossed markets on nasdaq and the nyse. *Journal of Financial Markets*, 11(3):308–337.

Stoikov, S. (2018). The micro-price: A high-frequency estimator of future prices. *Quantitative Finance*, 18(12):1959–1966.

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

Appendix

A Micro-Price Estimation Details

This appendix summarizes the micro-price estimation procedure outlined by Stoikov (2018), specifies implementation detail that deviates from his work, and provides an empirical example.

The micro-price estimation essentially amounts to finding the midpoint adjustment $g\left(S^{quoted}, I\right)$ for each combination of the quoted spread $\left(S^{quoted}\right)$ and the order book imbalance (I). To limit the number of states, the state variables are discretized. Furthermore, the midpoint adjustment is assumed to be independent of the midpoint price level $\left(\tilde{X}^{mid}\right)$. Once the adjustment has been estimated for the full state space, the micro-price can be obtained at low computational cost by simply mapping prevailing spreads and order book imbalances to the appropriate midpoint adjustment, and inserting it in the formula given by (5).

A.1 Sample

The input data for the micro-price estimation consists of NBBO quotes. No trade information is considered. I sample the quotes at a 100 millisecond frequency, yielding 24,600 observations per trading day when the first and last five minutes are excluded ((7 hours $\times$ 60 minutes - 10 minutes) $\times$ 60 seconds $\times$ 10 obs. per second).

The micro-price focus on what the probable price change following a given quote is, raises the concern that a trade matched to that quote influences the outcome. To avoid such a forward-looking bias, I base the estimation of $g\left(S^{quoted}, I\right)$ on quotes from the previous week of each sample. That is, for the S&P500 sample, I use data from November 30 to December 4, 2015. For the HFT sample, the previous week is on February 16 – 19, 2010 (February 15, 2010, is a public holiday).

A.2 State Space

The estimation procedure is based on the dynamics of the triplet $\left(\tilde{X}^{mid}, \bar{I}_\tau, \bar{S}_\tau^{quoted}\right)$, where τ is a time index, and the bars above the variable names indicate that they are discrete state variable versions of the continuous variables I_τ and S_τ^{quoted}. The bar is omitted for $\tilde{X}^{mid}$ as no discretization is required for the spread midpoint. The quoted spread is also discrete in nature, and in the application by Stoikov (2018) the spread state variable is simply the number of ticks. In my application, with hundreds of different stocks, a more flexible procedure for the discretization of spreads is required. I outline such a procedure below, along with an approach to define order book imbalance states differently across stocks and spread levels.

Discretizing the quoted spread. I refer to spread levels that are recorded in more than 1% of all quote observations as "common", and spreads that are not common but that have a frequency exceeding 0.01% as "rare". Even less frequent spread levels are disregarded. I form one state for each common spread level. 1% of the quote sample corresponds to more than 1,000 quote observations, which I consider enough to estimate the midpoint adjustment function accurately. For the rare spreads, I do the following:

- If there are rare spreads that are lower than the lowest common spread level, I let them form a new state if they together constitute more than 1% of the quote sample. If they are less frequent than that, I include them in the lowest common spread state.
- If there are rare spreads that are higher than the highest common spread level, I let them form a new state if they together constitute more than 1% of the quote sample. If they are less frequent than that, I include them in the highest common spread state.
- If there are rare spreads that lie between two common spread levels, I include them in the closest lower common spread state.

For example, consider the stock Apple Inc. (`AAPL.O`) in the S&P500 sample. Apple Inc. trades at a 1-cent quoted spread around 88% of the time, and at 2 cents for most of the time otherwise. Spreads at 3 or 4 cents are rare, with 0.21% and 0.03% of the observations, respectively. Accordingly, I form two states: "1 tick" and "2–4 ticks". There are occasional records of higher spreads, at 5 or 6 cents, which I discard.

Discretizing the order book imbalance. Recall the definition of order imbalance in Section 1.2:

$$I = \frac{Q^B}{Q^B + Q^A}, \tag{A.1}$$

where Q^B and Q^A represent the volumes quoted at the best bid and ask prices, respectively. Because the order imbalance is a fraction of quote volumes, I express the state bounds discussed below as fractions of integers, rather than in decimal form.

For each spread state, I form nine order imbalance states, as follows:

- States 1–4 are defined by the quartiles of order imbalance observations that are lower than or equal to 9/20 (if any).

Table A.1: Micro-price estimation example. This table shows for an example stock, Apple Inc., how the spread and order book imbalance states are mapped to the midpoint adjustment function $g\left(\bar{S}^{quoted}, \bar{I}\right)$ There are two spread states: "1 tick" and "2–4 ticks". The order book imbalance I_τ, where τ is an index of quote observations sampled at 100 ms frequency, is defined as $\frac{Q^B}{Q^B+Q^A}$. Each spread state has nine order book imbalance states, defined by the stated intervals on I_τ.

Spread state					
1 tick			2–4 ticks		
Imbalance state		$g\left(\bar{S}^{quoted}, \bar{I}\right)$	Imbalance state		$g\left(\bar{S}^{quoted}, \bar{I}\right)$
1:	$0 < I_\tau \leq 1/6$	-0.0042	1:	$0 < I_\tau \leq 5/19$	-0.0026
2:	$1/6 < I_\tau \leq 1/4$	-0.0030	2:	$5/19 < I_\tau \leq 5/14$	-0.0017
3:	$1/4 < I_\tau \leq 6/17$	-0.0020	3:	$5/14 < I_\tau \leq 4/10$	-0.0012
4:	$6/17 < I_\tau \leq 9/20$	-0.0010	4:	$4/10 < I_\tau \leq 9/20$	-0.0007
5:	$9/20 < I_\tau \leq 11/20$	0.0000	5:	$9/20 < I_\tau \leq 11/20$	0.0000
6:	$11/20 < I_\tau \leq 9/14$	0.0010	6:	$11/20 < I_\tau \leq 10/17$	0.0007
7:	$9/14 < I_\tau \leq 3/4$	0.0020	7:	$10/17 < I_\tau \leq 7/11$	0.0012
8:	$3/4 < I_\tau \leq 16/19$	0.0030	8:	$7/11 < I_\tau \leq 8/11$	0.0017
9:	$16/19 < I_\tau < 1$	0.0042	9:	$8/11 < I_\tau < 1$	0.0026

- State 5 includes order imbalance observations that satisfy $9/20 < I_\tau \leq 11/20$.
- States 6–9 are defined by the quartiles of order imbalance observations that are higher than 11/20 (if any).

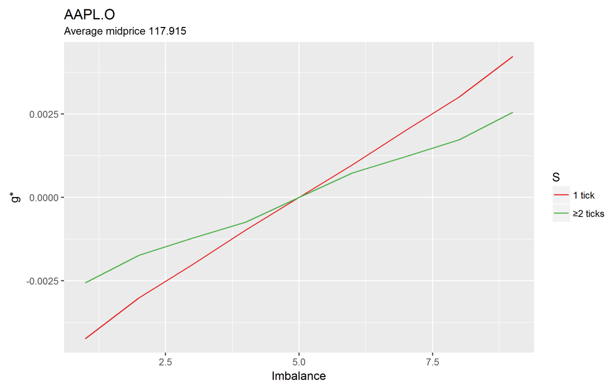
By predefining the State 5 boundaries, I avoid putting a breakpoint at 1/2, which is a very common value in the data, representing a balanced order book. The quantile-defined breakpoints for all other states makes the distribution of observations across states more uniform than with the equi-spaced boundaries used by Stoikov (2018). For the same reason, I use different imbalance breakpoints for each spread state. Nevertheless, due to that imbalance observations cluster at certain fractions, there are infrequent cases in my sample where not all imbalance states are populated. In those cases, the midpoint adjustment can not be estimated for all states.

The order imbalance states for the example stock, Apple Inc., are reported in Table A.1. Notably, the spread state "2–4 ticks" displays less order book asymmetry than the "1 tick" state. This is seen in that the the order imbalance intervals defining the 4$^{\text{th}}$ and 6$^{\text{th}}$ states are relatively tight, whereas the 1$^{\text{st}}$ and 9$^{\text{th}}$ states have relatively wide intervals.

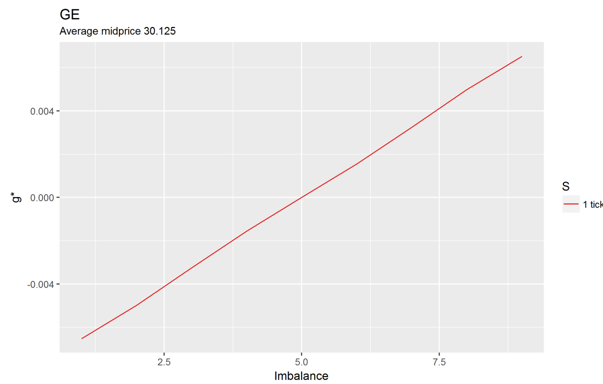
There is also a big difference in the estimated midpoint adjustments across the two spread states. Under the 1-tick spread, an order book imbalance in state 1 leads to an adjustment of -0.42 cents. If the spread is instead 2 cents, the same imbalance state yields an adjustment of -0.26 cents.

The difference is even stronger in relative terms. The 1-tick spread adjustment implies that the bid-side spread is 8% of the total spread, because the micro-price is only 0.08 cents away from the bid price ((0.50 - 0.42) / 1). When the spread is 2 cents, the bid-side spread is 37% of the total spread ((1 - 0.26) / 2).

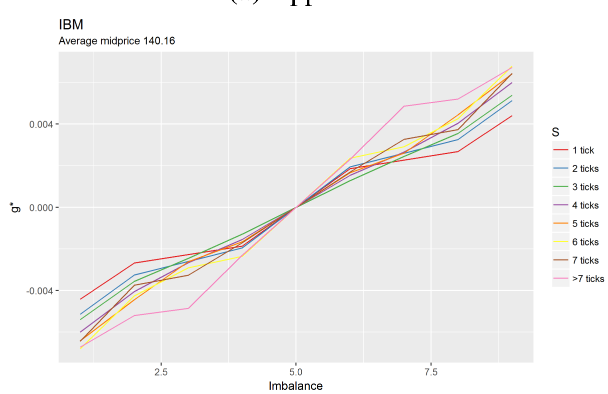
Figure A.1 plots the midpoint adjustment function of Apple Inc. in Panel (a). As illustration of how the number of spread states vary across stocks, the figure also includes the corresponding plots for General Electric, IBM, and Alphabet (see Panels (b)–(d)). Whereas General Electric is virtually always trading at a spread of one tick, the spread of Alphabet tends to be in the range from 20 to 60 ticks. Note also that the scale of the y-axis spans 6 ticks for Alphabet (from -3 to 3 cents), whereas that of Apple is an order of magnitude smaller (from -0.5 to 0.5).



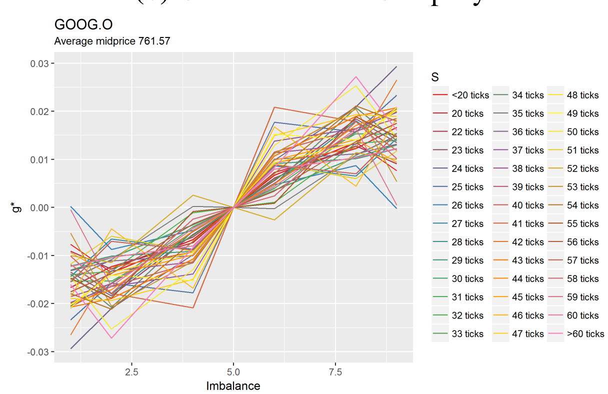
(a) Apple Inc.



(b) General Electric Company



(c) IBM Common Stock



(d) Alphabet Inc Class C

Figure A.1: Midquote adjustment functions for four example stocks. This figure shows the midpoint adjustment functions as estimated for the micro-price for four example stocks. There is one adjustment function for each spread state in each stock. The example stocks are (a) (b) General Electric Company, (c) IBM Common Stock, and (d) Alphabet Inc Class C.

A.3 Estimation

Stoikov's (2018) estimation procedure involves the following steps:

1. Symmetrization. I symmetrize the data such that for each observation ($\bar{I}_\tau$; $\bar{S}^{quoted}_\tau$; $\bar{I}_{\tau+1}$; $\bar{S}^{quoted}_{\tau+1}$; dM), where dM is the midpoint change from τ to $\tau + 1$, I add an observation that is mirrored in the imbalance dimension and has the opposite sign on dM ($10 - \bar{I}_t$; $\bar{S}^{quoted}_t$; $10 - \bar{I}_{t+1}$; $\bar{S}^{quoted}_{t+1}$; $-dM$). The symmetrization of the input data ensures that the micro-price estimation converges. It also leads to the symmetry in the $g\left(\bar{S}^{quoted}, \bar{I}\right)$ estimates seen in Table A.1 (i.e., $g\left(\bar{S}^{quoted}, \bar{I}\right) = -g\left(\bar{S}^{quoted}, 10 - \bar{I}\right)$).

2. Transition probability estimation. The estimation procedure distinguishes transitory and absorbing states. Given the current state, a state is absorbing if it implies a midpoint change, and transitory otherwise. The micro-price estimation may be thought of as a probability tree where branches keep growing until they reach an absorbing state. The midpoint adjustment is then a probability-weighted average of midpoint changes associated with each branch. To analyze the probability tree, the next-period probability of each state, conditional on the current state, is required. The transition probabilities are assumed to equal the historically observed frequencies.

The transition probabilities between transitory states are captured by the square matrix Q. If there are m spread states and n imbalance states, Q is an $mn \times mn$ matrix. For example, the top-left entry of Q may show the probability of staying in the state $\left(\bar{S}^{quoted}, \bar{I}\right) = (1, 1)$, with the spread midpoint unchanged.

The transition probabilities for changes from transitory states to absorbing states are recorded in two matrices, T and R. The former is similar to Q, dimension $mn \times mn$, in that it tracks the transition between spread-imbalance combinations, but it differs in that in only considers the cases where there is also a change in the midpoint. For example, the top-left entry of T may show the probability of staying in the state $\left(\bar{S}^{quoted}, \bar{I}\right) = (1, 1)$ while the midpoint is changing. The matrix R, in turn, captures the magnitude of the midpoint change. Define a vector K of all possible levels of non-zero midpoint changes. The dimension of R is then $mn \times k$, where k is the length of K.

3. Computing the midpoint adjustment. We now have all the ingredients to estimate the one-period ahead midpoint adjustment for each combination of $\bar{S}^{quoted}$ and $\bar{I}$, denoted G^1. Stoikov (2018) shows that $G^1 = (1 - Q)^{-1}RK$. To find the vector G^*, which is the expected midpoint change evaluated at infinity, Stoikov (2018) defines $B = (1 - Q)^{-1}T$ and shows that:

$$G^* = G^1 + \sum_{i=1}^{\infty} B^i G^1. \tag{A.2}$$

I consider ten iterations of the sum in (A.2), but the value of G^* typically converges after 2–3 iterations.

B Data Matching and Screening

B.1 Matching CRSP and TRTH Identifiers

To my knowledge, this is the first study that matches data from the CRSP and TRTH databases. The issue identifier in TRTH is called the Reuters Instrument Code (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. In order to match all securities, however, the following adjustments are considered:[29]

- When `TSYMBOL` is empty, the CRSP field `TICKER` is used instead.
- Before January 1, 2012, share class information is not included in `TSYMBOL`. Then, when `TSYMBOL` cannot be matched to a 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`).
- After January 1, 2012, `TSYMBOL` and `TICKER` differ when there is a share class suffix for `TSYMBOL`. I make the `TSYMBOL` share class suffix lowercase to match the 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`).

B.2 TRTH Data Screening

Each trade and quote observation in TRTH includes additional information in the `Qualifier` field. I use that information to screen trades and quotes, using the following criteria:

(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 [GVx_TEXT] (where x can be a number from 1 to 4) 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`).

[29] The RICs in TRTH change over time. To track a security over time, a viable strategy is to access the CRSP time series, where the security identifier `PERMNO` is permanent. The time-varying `TSYMBOL` can then be matched to RICs as described here. This procedure is not necessary for the samples considered here.

(T3) Trades with any of the following conditions indicated in the [PRC_QL2] information are excluded: *agency cross-trade* (`AGX`), *stock option trade* (`B/W`), *not eligible for last* (`NBL`), *derivatively priced* (`SPC`), and *stopped* (`STP`).

(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 or as coinciding with changes in the limit up–limit down (LULD) price bands are retained, unless any of the criteria (Q2)–(Q4) are satisfied. This screening utilizes the [PRC_QL_CD] and [PRC_QL3] information and excludes everything but the following entries: `R__`, `___`, `LPB`, and `RPB`. For example, quotes with non-positive bid-ask spread, associated with trading halts, or marked as slow due to a liquidity replenishment point, are thus excluded. Quotes coinciding with changes in the LULD price bands are retained because LULD limit updates do not influence the validity of the current quotes.

(Q2) Quotes marked as *non-executable* are excluded (`A`, `B`, or `C`, in the `[GV1_TEXT]` field).

(Q3) Quotes with non-regular conditions indicated by the `[CTS_QUAL]` information (taking the value `TH_`, `IND`, or `O__`) are excluded.

(Q4) Quotes where the bid-ask spread is either negative ("crossed"), zero ("locked"), or exceeding USD 5 are excluded.

The effects of the different screening criteria are presented in Table B.1. The trade screening criteria disqualify a negligible number of trades for both data sets.

Among the quote screening criteria, (Q2) and (Q3) each affect less than 0.01%. The criteria specified in (Q1) and (Q4), however, disqualify a substantial number of quotes. In the S&P500 sample, they affect 1.04% and 5.01% of the quotes, respectively. For the HFT sample the corresponding filters capture 4.17% and 8.71% of the quote observations, respectively. For the Split sample the corresponding numbers are 0.05% and 4.17%.

Virtually all excluded quotes are locked, meaning that the bid and ask prices are equal. It is well-known that locked quotes are common in the NBBO data (Shkilko et al., 2008). Locked quotes cannot exist within an exchange. In the NBBO feed, however, they can appear due to that price changes are not simultaneous across venues, for example. Around 4.89% of all trades in the S&P 500 sample, 8.41% in the Nasdaq HFT sample, and 3.46% in the Split sample are matched to such quotes (see the rightmost column of Table B.1, Panel (b)). Excluding the locked quotes is consistent with Holden and Jacobsen (2014).

Table B.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 sample, Dec. 7 – 11, 2015	< 0.01%	< 0.01%	< 0.01%	< 0.01%	< 0.01%	55.7 million
HFT sample, Feb. 22 – 26, 2010	< 0.01%	< 0.01%	0.58%	< 0.01%	0.58%	1.9 million
Split sample, Jan. 1, 2006 – Dec. 31, 2015	< 0.01%	< 0.01%	< 0.01%	< 0.01%	< 0.01%	5.5 million

(b) Quote screens

Sample	(Q1)	(Q2)	(Q3)	(Q4)	All filters combined	Locked quotes
S&P 500 sample, Dec. 7 – 11, 2015	1.04%	< 0.01%	< 0.01%	5.01%	5.01%	4.89%
HFT sample, Feb. 22 – 26, 2010	4.17%	< 0.01%	< 0.01%	8.71%	8.71%	8.41%
Split sample, Jan. 1, 2006 – Dec. 31, 2015	0.05%	< 0.01%	< 0.01%	4.12%	4.17%	3.46%

C Event Study: Stock Splits

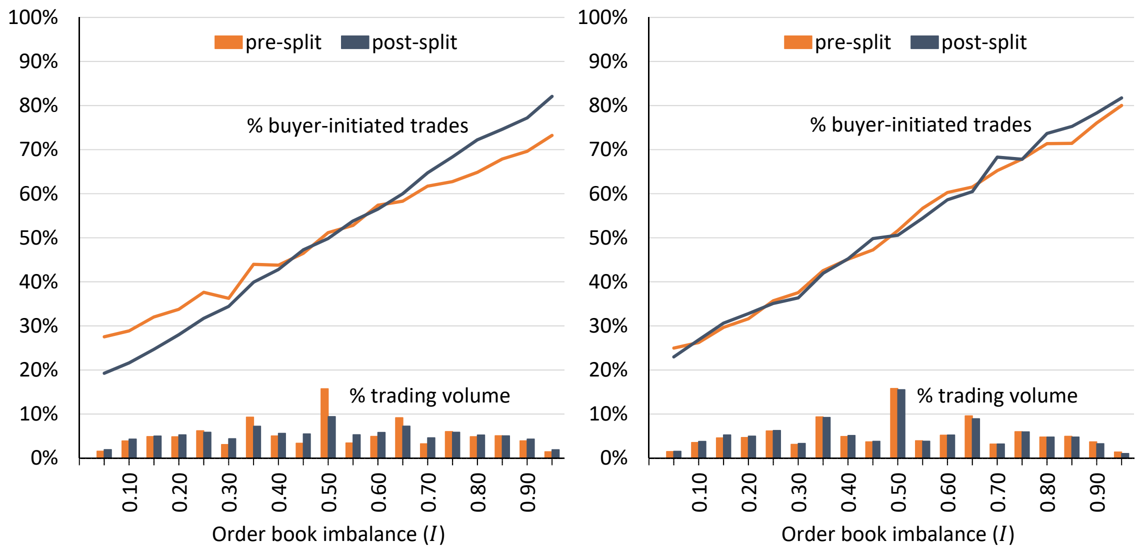
In this appendix I illustrate the role of price discreteness in the midpoint effective spread bias through an event study around stock splits. When shares are split their price falls, and the split thus works as an exogenous shock to the bindingness of the tick size. For this analysis, I employ the Split sample described in Section 2.

I refer to the closest Wednesdays before and after regular splits as the *pre-split* and *post-split* dates, respectively. To make reverse splits comparable to regular splits, I set the closest Wednesday *after* the event as the pre-split date, and vice versa. By using Wednesdays only, I avoid influence of any seasonality related to the day of the week, as documented by Chordia et al. (2001). For an event to be retained, I require both the pre-split and post-split dates to be valid trading days with at least 10 trades per day recorded in TRTH. Moreover, to focus on split events that lead a stock to move from a non-binding to a binding minimum tick size, I only retain events where the average quoted bid-ask spread on the post-split date is below USD 0.015 (i.e., 1.5 ticks).

For each treatment stock, I choose a control stock among all stocks that satisfy the same inclusion criterion as the treatment stocks, but that do not have a stock split event in the same or the previous month, and that is in the same market capitalization decile as the event stock. The control stock is defined as the stock in that set that has the price closest to that of the treatment stock in the end of the previous month. If there is no control stock priced less than USD 5 away from the event stock price, the event is excluded. In the ten-year sample period, I obtain 79 split events that satisfy all criteria.

Figure C.1, Panel (a), shows how investors respond to the stock split events. Similar to Figure 1, the plot measures the probability of buyer-initiated trades on the y-axis. Aligned to the order choice prediction, the x-axis shows trade categories based on the order book imbalance, I (instead of the fundamental value deviation from the midpoint, used in Figure 1). The trade categories are determined by the following breakpoints for I: $0.025, 0.075, ..., 0.925, 0.975$, and labeled by the midpoint of each interval.

The light curve shows the results for the pre-split dates. The curve is clearly upward-sloping, consistent with both the order choice literature and a positive elasticity of liquidity demand. In the post-split dates, displayed by the dark curve, the slope becomes steeper. That is, investors become more sensitive to the order book imbalance when the minimum tick size becomes more binding and the bid-ask spread widens. This result is consistent with a positive liquidity demand elasticity. Importantly, the evidence does not invalidate the trade-off described by Foucault et al. (2005) and Roşu (2009). It shows, however, that the liquidity demand elasticity channel in this setting dominates the order choice channel.



(a) Split stocks **(b)** Control stocks

Figure C.1: Liquidity demand elasticity around stock splits. This figure shows the frequency of buyer-initiated trades and the dollar volume market shares for different categories of the order book imbalance. The direction of trade is determined by the Lee and Ready (1991) algorithm. The order book imbalance is defined as $I = Q^B/(Q^B + Q^A)$, where Q^B and Q^A represent the volumes quoted at the best bid and ask prices, respectively. The trade categories are determined by the following breakpoints for I: $0.025, 0.075, ..., 0.925, 0.975$, and labeled by the midpoint of each interval. Transactions recorded exactly at the midpoint, or when I is outside the interval $0.025 : 0.0975$, are not included. The sample includes 79 stock split events in the period Jan. 1, 2006 – Dec. 31, 2015. Eligible split events are for ordinary common stocks where the bid-ask spread on the post-split date is below 1.5 ticks on average. The pre-split date is the last Wednesday before the split is effective, and the post-split date is the first Wednesday when the split is effective. Control stocks are selected based on market capitalization decile (which should be the same as for the treatment stock) and share price, both in the end of the previous month. Results for stocks with split events are in Panel (a), while the control stocks are analyzed in Panel (b).

Panel (b) shows that the control stocks are virtually unaffected by the events. To confirm that the change in slope for the treatment group relative to that of the control group, I set up a probit model,

$$\Pr(Buy_t) = \alpha + \beta_1 I_t StockSplit_t + \beta_2 I_t Post_t + \beta_3 I_t StockSplit_t Post_t + \beta_4 I_t + \varepsilon_t, \tag{C.1}$$

where $StockSplit_t$ and $Post_t$ are binary variables equal to one for event stocks and post-split periods, respectively, and zero otherwise, and ε_t is the residual. The interaction terms with the order book imbalance I_t creates a difference-in-difference setup. The main parameter of interest is β_3, showing the event slope effect while accounting for variation in the control stocks. The model estimates, presented in the left column of Table C.1, confirm that the change in slope observed in Figure C.1 is positive and statistically significant.

Next, I verify that the minimum tick size becomes more binding on the post-split dates, and that the relative quoted spread becomes wider. I use ordinary least squares (OLS) to estimate a difference-in-difference regression model,

$$y_{di} = \alpha + \beta_1 StockSplit_{di} + \beta_2 Post_{di} + \beta_3 StockSplit_{di} Post_{di} + u_{di}, \tag{C.2}$$

where y_{di} is the dependent variable, measured for each date d and stock i, and u_{di} is the residual. I consider the *Tick spread* and the *Quoted spread* as dependent variables. The *Tick spread* is the average nominal bid-ask spread measured in number of ticks, which in the US equity context is equivalent to measuring the quoted spread in cents. As expected, the difference-in-difference coefficient β_3 is significantly negative for the *Tick spread*, and significantly positive for the *Quoted spread*.

To test whether the stock split events increase the bias in the midpoint effective spread estimator, I use the same regression model with *Relative average bias* as the dependent variable. Over the full sample period, I find that the bias increases by more than nine percentage points (9.061%) following stock splits. With a t-statistic of 18.4, based on standard errors clustered on stocks and dates, the result shows strong statistical significance. The effect is also economically significant, as the bias more than doubles following the events (to see this, note that β_3 exceeds the sum of the intercept α and the β_1 and β_2 coefficients).

Furthermore, the effect is robust to splitting the time series into two five-year periods, 2006–2010 and 2011–2015. The two right-most columns of Table C.1 confirm that the stock split effect on the *Relative average bias* is positive and statistically and economically significant.

Table C.1: Stock split difference-in-difference regressions. This table presents how stock splits influence two measures of the bid-ask spread, as well as the direction of trade sensitivity to order book imbalances. The left column displays results of a probit model where the binary variable Buy_t, equal to one for buyer-initiated trades and zero for seller-initiated trades, is related to the order book imbalance I_t, defined as in Figure C.1. Midpoint trades are not included. The right part of the table holds results for OLS regressions where the dependent variables are the *Tick spread* (the quoted bid-ask spread measured in ticks), the *Quoted spread* (bps, defined as in Table 1), and the *Relative Average Bias* in the midpoint effective spread (defined as the average difference between the midpoint effective spread and the weighted midpoint effective spread, divided by the average weighted midpoint effective spread, measured in %). The *Tick spread*, the *Quoted spread*, and *Relative Average Bias* are calculated on stock-day frequency. The split events and pre-split and post-split periods are defined as in Figure C.1. The *Relative Average Bias* model is estimated for the full sample as well as for the first and second five-year periods separately, see the three rightmost columns. The variable *StockSplit* is one for event stocks and zero for control stocks, and *Post* is one for post-split dates and zero for pre-split dates. *StockSplit*, *Post*, and *StockSplit* × *Post* are interacted with I_t in the probit regression, and with the intercept in the OLS regressions. Each estimate is reported along with z-statistics for the probit model and t-statistics for the OLS model (within parentheses), and the superscripts * and ** indicate significance at the 10% and 5% confidence levels, respectively. The standard errors in the probit model are clustered on stock, date and trading venue, following Petersen (2009). The standard errors of the OLS model are clustered on dates and stocks. The R^2 statistic reported for the probit model is the McFadden pseudo-R^2.

	Probit trade-level model	OLS stock-day models				
Dependent variable	Buy_t	*Tick spread*	*Quoted spread* (bps)	*Rel. Av. Bias* (%)	*Rel. Av. Bias* (%)	*Rel. Av. Bias* (%)
Sample period	2006–2015	2006–2015	2006–2015	2006–2015	2006–2010	2011–2015
Interaction variable	I_t	Intercept	Intercept	Intercept	Intercept	Intercept
Intercept	−0.80**	2.18**	7.16**	9.128**	7.279**	10.676**
	(−7.80)	(15.47)	(4.96)	(5.367)	(3.327)	(4.404)
StockSplit	−13.19**	−0.23*	−3.06**	−1.967	0.649	−4.156**
	(−2.03)	(−1.68)	(−2.68)	(−1.549)	(0.374)	(−2.162)
Post	1.88	−0.02	−0.28	0.504	1.485	−0.317
	(0.43)	(−0.17)	(−0.23)	(0.335)	(0.989)	(−0.129)
StockSplit × *Post*	8.28**	−0.72**	5.19**	9.061**	7.934**	10.005**
	(2.55)	(−15.68)	(6.81)	(18.4)	(10.381)	(12.361)
Order book imbalance, I_t	168.58**					
	(7.56)					
Observations	2,553,612	316	316	316	144	172
R^2	0.07	0.14	0.02	0.08	0.10	0.07

D Further Descriptives for the HFT Sample

Table D.1: Effective spread properties in the HFT sample. This table replicates Table 1 using the HFT sample instead of the S&P500 sample. All definitions are the same as in Table 1.

	Mean	Std. Dev.	Percentiles				
			5th	25th	50th	75th	95th
Effective spread							
$\tilde{S}^{mid}$ (bps)	2.42	6.02	1.44	2.59	4.43	8.39	16.95
$\tilde{S}^{mic}$ (bps)	1.47	5.89	0.93	1.79	3.05	7.13	14.42
Nominal bias							
$\tilde{S}^{mid} - \tilde{S}^{mic}$ (bps)	0.95	1.66	-0.29	0.30	0.71	1.90	4.24
Average relative bias	0.65						
Rel. quoted spread (bps)	2.89	7.55	1.72	3.20	5.24	11.65	21.12
Trade price (USD)	130.88	60.53	7.32	13.29	26.65	43.98	92.23
Trade volume (thousands)	4.32	5.81	0.08	0.43	1.40	7.71	14.89
Dollar volume (millions)	41.76	114.33	0.20	0.90	5.14	51.15	147.43

Bias in the Effective Bid-Ask Spread

Internet appendix

TRTH Data Quality

For US equities, the TRTH contains consolidated instruments that merge trades taken from the consolidated tape and quotes taken from the official NBBO feed.[30] 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.

Figure IA.1 confirms that the TRTH data conform to those of DTAQ. It displays 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 the problems with canceled quotes.

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.

[30] The Thomson Reuters support staff confirms in 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/.

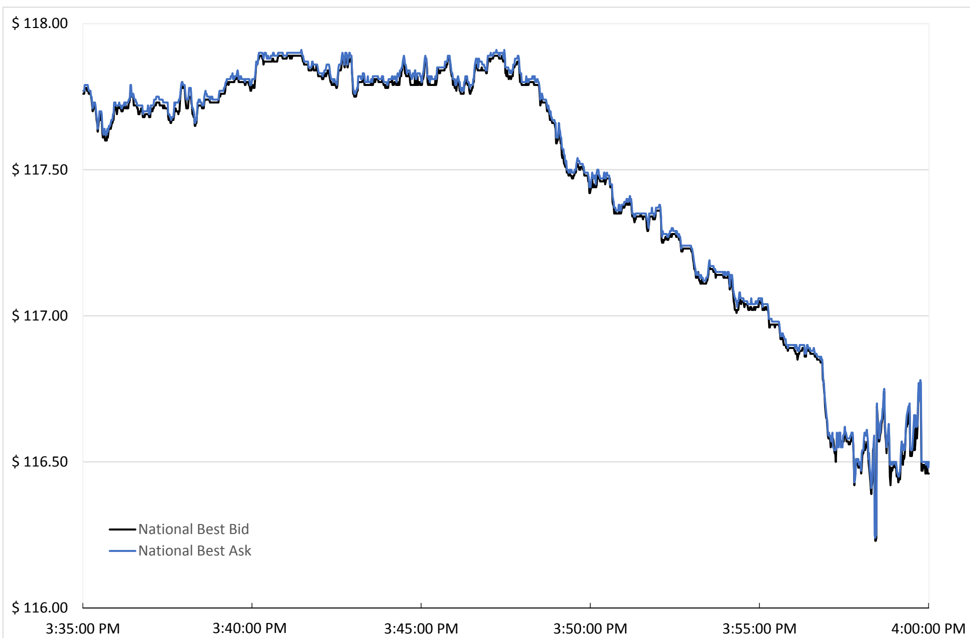


Figure IA.1: NBBO accuracy for TRTH data. This figure shows the NBBO prices for IBM on April 1, 2008, between 3:35 PM and 4:00 PM, as reported for the TRTH consolidated instrument `IBM`.