Latency and the Look-Ahead Bias in Trade and Quote Data*

Robert H. Battalio, Craig W. Holden[†], Matthew Pierson,
John J. Shim, and Jun Wu[‡]

February 28, 2026

Abstract

The NYSE Trade and Quote (TAQ) dataset reports events sequentially out of order. Trades and quote changes are clustered but recorded with differential latency, partly due to geography, effectively mixing up the order of events. Consequently, quote changes after, and in response to, a trade are regularly reported before it, creating a look-ahead bias. This bias creates errors in signing and spread measurement when using the TAQ NBBO. Errors intensify with latency: around 20% of high-latency trades are signed incorrectly and effective spreads are understated by over 40%. We propose a more accurate latency-free signing methodology, grounded by exchange rules.

Keywords: TAQ, SIP, Latency, Trade Signing, Effective Spread, Price Impact
JEL Classification: C81, G12, G14, G20

*This paper is heavily revised from a previous draft circulated under the title, "In the Blink of an Eye: Exchange-to-SIP Latency and Trade Classification Accuracy." The revisions are substantial enough to interpret this as a new paper. We thank audiences at the Holden Conference, University of Notre Dame, University of Virginia, and Vanderbilt University, as well as Eric Budish, Bob Jennings (discussant), Craig Lewis, Tim Loughran, Phil Mackintosh, Albert Menkveld, Gideon Saar, Yajun Wang (discussant), and Mao Ye.

†The late Craig Holden was at Kelly School of Business, Indiana University

‡Battalio and Shim are at the University of Notre Dame Mendoza College of Business (rbattali@nd.edu and jshim2@nd.edu). Pierson and Wu are at Wharton Research Data Services (WRDS) at the Wharton School, University of Pennsylvania (mpiers@wharton.upenn.edu, jun5@wharton.upenn.edu).

1 Introduction

Trade and quote data are ubiquitous in finance, economics, and accounting research, and are commonly used in regulation-mandated reports and analyses that support rule-making. The New York Stock Exchange's Trade and Quote (TAQ) dataset, which contains every trade and top-of-book quote update for all U.S. publicly-traded securities on U.S. stock exchanges, is the most commonly used trade and quote data. TAQ data are generated from Securities Information Processors (SIPs), which aggregate trade and quote updates across exchanges at central locations and disseminate that data in real-time to millions of subscribers. There are two SIPs: the Nasdaq SIP aggregates and disseminates data for all Nasdaq-listed securities, and the NYSE SIP does the same but for all other securities (mostly NYSE-listed).[1] While a significant number of market participants rely on the SIP feed for market data, the most latency-sensitive traders use so-called direct-feed data from each exchange, which is the fastest and most comprehensive data source for exchange activity and is significantly more expensive.

The TAQ data are relatively comprehensive but they do not report whether a trade is buyer or seller initiated (i.e., the trade sign). As a result, TAQ users infer trade direction as a necessary first step to study other objects of interest, such as effective and realized spreads, price impact, and order imbalances. The classic methodological paper by Lee and Ready (1991) relies on the midpoint of the prevailing quotes at the time of the trade to sign trades. The idea is that the prevailing midpoint captures the market's view of the security's "fair" price at the time of the trade. For example, if a trade's price is above the prevailing midpoint, the trade is classified as buyer initiated and the effective spread is computed as the percentage premium over the midpoint.

At the time of Lee and Ready (1991), NYSE-listed securities largely traded at a single location (the NYSE), which made the choice of which midpoint to use for signing obvious. Over the past two decades, the proliferation of securities exchanges, each displaying quotes and offering trading in every symbol, means that there are now more than a dozen midpoints to choose from. The de facto convention in the literature has been to use the midpoint of the SIP National Best Bid and Offer (NBBO), which is computed as the highest bid price and the lowest offer price across all exchanges for a given security. This approach is used by thousands of papers in the academic literature, as well as by regulators to inform rulemaking,

[1] The Nasdaq SIP is officially called the Unlisted Trading Privileges (UTP) SIP. The NYSE SIP is officially called the Consolidated Tape Association (CTA) SIP. We simply label them Nasdaq and NYSE SIPs instead of UTP and CTA SIPs to convey which symbols each SIP is responsible for.

distribute SIP revenues, and construct execution quality statistics.

In this paper, we show that the SIP NBBO is subject to a systematic look-ahead bias that is particularly pronounced around trades. That is, the prevailing SIP NBBO midpoint just before a trade regularly incorporates information in quote changes that occur *after* the trade. In fact, some of the quote changes that the SIP erroneously reports before a trade are very likely *responses* to the trade itself. The SIP look-ahead bias is a consequence of two facts: (1) trades and quote changes have heterogeneous latency in traveling from the originating exchange to the SIP because of geographical dispersion and variation in SIP processing, and (2) trades and quote changes tend to be clustered together in time because participants act in response to common signals and each other's actions. The implication of these facts is that quote changes that occurred after a trade, and even in response to that trade, will be reported by the SIP first and affect the prevailing SIP NBBO if the quote change has a lower latency than the trade.

For context, exchange-to-SIP latencies range from around 15 to 560 microseconds. To clarify the units, a microsecond (abbreviated μs) is a millionth of a second and is much finer than a millisecond (abbreviated ms), which is a thousandth of a second. For context, it takes around 200,000 microseconds to blink an eye. This range of latencies alone does not deliver a look-ahead bias. If a trade occurs but nothing else happens within 500 μs, there is nothing to rearrange. In fact, we document that in nearly all 500 μs intervals for every symbol, there is no activity. However, *conditional* on a trade, we show that there is a pronounced clustering of events, with thousands of times as many events in the hundreds of microseconds around the trade than we would expect if events were uniform and i.i.d. throughout the trading day. In addition, this clustering is more pronounced *after* a trade in that there tends to be more activity immediately following a trade. The combination of these two facts – events are clustered within 500 μs and events are assigned latencies that can range from 15 to over 500 μs – means the SIP will regularly scramble the order of events.

The combination of clustering and heterogeneous latencies means that the SIP may be prone to both a look-ahead bias, where quote changes after a trade are reported before the trade, and a stale quote issue, where quote changes that occur before the trade are disseminated after the trade. While we find evidence of both types of errors, the look-ahead bias has much more bite because it causes the prevailing SIP quote at the time of the trade to incorporate information from after the trade, and that quote is then used to sign trades and measure spreads. The stale quote issue, while still prevalent, is less distortionary for signing and spread measurement because it is less likely to alter the SIP NBBO.[2]

Because the SIP is sequentially out of order and is prone to a look-ahead bias, the ubiquitous practice of using the SIP NBBO midpoint to sign trades and measure effective spreads/price impact will lead to errors. We propose a new signing methodology called the Latency-Free (LF) signing methodology, which consists of two rules. Both rules utilize exchange timestamps, available in TAQ since 2015, which are assigned by the exchange itself just after the matching engine updates the order book and are not prone to latencies from traveling from the exchange to the SIPs.[3] Our main signing rule is motivated by the rules of the limit order book. We simply use the prevailing exchange best bid and offer (EX BBO) midpoint from the exchange where the trade occurred to sign the trade. To the extent that TAQ records the best available limit orders without data errors, this method *must* sign trades with 100% accuracy: in a limit order book, marketable orders trade against the best available outstanding orders. We confirm this logic in the data.

One reason why the EX BBO rule does not achieve 100% accuracy for all trades is because TAQ has incomplete data on limit orders: odd lot limit orders – orders for less than 100 shares – are not reported to the SIP. That means that if an exchange's best bid or offer comes from an odd lot limit order, a marketable order will trade at that limit order's price. According to TAQ, it may appear that the trade occurred inside of the exchange's BBO, but only because odd lot limit orders are not reported to the SIP. To handle these trades, we add a supplementary rule based on the prevailing latency-free NBBO (LF NBBO), or the NBBO constructed instantaneously using exchange timestamps, but from 1 *millisecond* (ms) before the trade. We use the delayed LF NBBO to sign odd lot trades that occur inside the TAQ-observed EX BBO, which accounts for less than 23% of trades and 9% of dollar volume. The 1 ms delay helps capture the state of the book closer to when the odd lot limit order was submitted, and before the cluster of activity that typically surrounds a trade.[4]

Thus, our proposed LF signing methodology uses the Lee-Ready logic with the EX BBO midpoint to sign all round lot trades and odd lot trades at or outside the EX BBO, and the

[2]In fact, Lee and Ready (1991) suggested adding an intentional delay to ensure reference quotes used for signing are from before the trade at the cost that these quotes may be somewhat stale.

[3]These timestamps are officially called "participant timestamps" in TAQ, but we use the term exchange timestamp to clarify to readers that these timestamps are assigned by the exchange itself.

[4]We actually document *three* reasons why the EX BBO rule might be imperfect. The first is because of odd lots, as described above. The second is because of "price slide" trades, which are working in the book at one price but displayed at another to comply with Reg NMS. The third is trades against hidden or non-displayed limit orders. The latter two represent special cases and require direct-feed data to detect. As a result, we do not adapt our signing methodology for these cases, but highlight them to understand where our signing methodology does not achieve 100% accuracy. We discuss all cases in detail in Section 5.2.1.

delayed LF NBBO for all odd lot trades inside the EX BBO. Using NYSE Arca direct-feed data, which can be used to infer the true direction of a trade, we find that our LF signing methodology accurately signs 95.3% of trades and 96.6% of dollar volume. This represents an improvement in signing accuracy over the standard Lee-Ready method using the SIP NBBO for this sample of trades by more than 7 percentage points (pp).[5]

The look-ahead bias in the SIP NBBO midpoint also impacts the measurement of effective spreads and price impact. As an alternative, we propose the LF NBBO midpoint. Because the LF NBBO never has the sequence of events wrong, it has no look-ahead bias but aggregates information in a way that actual market participants cannot perceive (because even the fastest participants are subject to some latency). We argue that the LF NBBO midpoint is useful because it aggregates the best available information in the market at a point in time with no latency and is a better approximation of the market's view of "fair" value than the SIP NBBO midpoint.[6] We find that the SIP NBBO midpoint moves in a way that *anticipates* a trade, i.e., the SIP NBBO midpoint moves towards the trade price before the trade occurs. This is because some of the price impact occurring after the trade is incorporated into the prevailing SIP NBBO midpoint before the trade (i.e., the SIP look-ahead bias). As a result, estimates of the effective spreads and price impact using the SIP NBBO midpoint are biased downwards. We estimate that the SIP effective spread is biased downward by 13.92% for the average trade and 14.38% for the average dollar. These numbers triple to above 40% when we examine trades with high exchange-to-SIP latency.

To illustrate how trade and quote changes can be reported out of sequence by the SIP, consider the following stylized example. Suppose an order to buy 100 shares of JP Morgan (JPM) arrives at the Nasdaq exchange at **10:00:00.000000** and trades, taking all 100 shares at Nasdaq's best offer of $200.00, making the new best offer $200.01. Because JPM is listed on NYSE, all trades and top-of-book quote updates are reported to the NYSE SIP. Thus, notification of the trade and quote update must travel from the Nasdaq data center in Carteret, NJ, to the NYSE data center 35 miles north in Mahwah, NJ, where the NYSE SIP operates. Figure 1 shows the locations of the three main cities where exchange data centers are located. For this particular transmission, we find that it typically takes approximately **560 μs** from the time the trade is recorded on Nasdaq to the time it is disseminated by the NYSE SIP.

[5]For this sample, we can also confirm that our signing methodology outperforms all other proposed rules for U.S. securities. See Section 2.1 for details.

[6]We emphasize that the LF NBBO is meant to be a benchmark, representing spreads paid relative to the most timely market quotes. We do not propose the LF NBBO as part of a trading strategy or for real-time signals, which would make it invalid because market participants cannot observe it.

While notification of the Nasdaq trade is en route to the SIP, high-frequency traders (HFTs) and algorithmic traders (ATs), having also observed the Nasdaq trade using Nasdaq direct-feed data, react by changing their quotes on NYSE Arca, another major exchange in Mahwah, NJ.[7] For example, HFTs and ATs can observe the Nasdaq trade in Carteret, then send modify order requests from Carteret to the Arca matching engine in Mahwah with a (state-of-the-art) latency of ~190 μs. We estimate that it will take ~30 μs for Arca to process and update its own order book, then another ~110 μs for those quote changes to be sent to and published by the NYSE SIP. In total, it takes **~330 μs** for HFTs and ATs to observe the Nasdaq trade, update their quotes on Arca, and have those quote changes reflected on the SIP. The implication is that trader *responses* (i.e., price impact from the trade) are published by the SIP at `10:00:00.000330`, while the trade that triggered the initial sequence is reported by the SIP at `10:00:00.000560`. That is, the prevailing SIP NBBO does not reflect the state of the market at the time the trade actually occurred, but has already incorporated information from events that occurred as a direct response to the trade. This is often exacerbated by the fact that top-of-book quote updates on Nasdaq, triggered by the trade itself and subsequent responses, will be recorded on the SIP before the trade.[8]

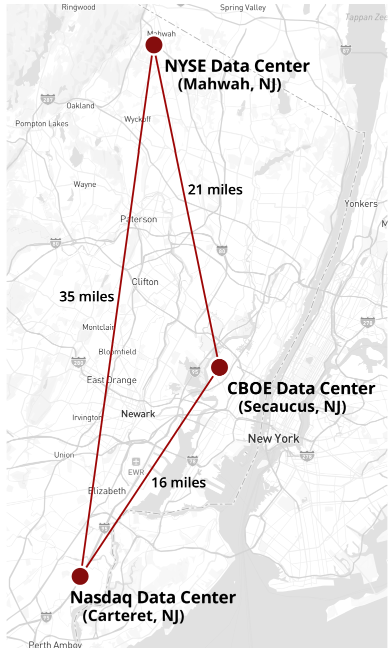
This example provides a concrete sense of how trades and quotes are reported out of order on the SIP. We build on this example to show that SIP sequence issues occur regularly. For each JPM trade that occurs on a Carteret exchange, we count the number of JPM "events" – top-of-book quote updates or other trades – that occur on exchanges in Mahwah (e.g., NYSE Arca) in the 1,000 μs before and after the trade according to exchange timestamps. We then average across all trades to get a sense of the activity in Mahwah around the time of a Carteret trade.

Figure 2 shows a histogram-like figure, where the x-axis is time relative to the Carteret trade in microseconds according to exchange timestamps. The y-axis is the average number of events. The blue bars depict the average number of Mahwah events in each 10 μs window around a trade in Carteret. The darker red bars show how many of these Mahwah events are reported out of order on the SIP in the following sense: red bars after the trade (and to the right of zero) signify that the event occurred at a specific point in time *after* the trade

[7] Glosten and Milgrom (1985) provide an interpretation where bid and offer quotes are conditional expected values, and a trade provides information that changes these expected values, causing liquidity providers to (instantaneously) adjust their quotes. That logic, applied to a low-latency world, means that HFTs and ATs will use the technology they have to update quotes as quickly as possible to avoid adverse selection and sniping (Budish et al., 2015).

[8] We provide two detailed examples that we extract directly from the data. The examples are much more involved but provide a more granular level of detail and exactly reflects real-world events. See Appendix A for these examples.

Figure 1: New Jersey Triangle: Exchange and SIP Data Center Locations



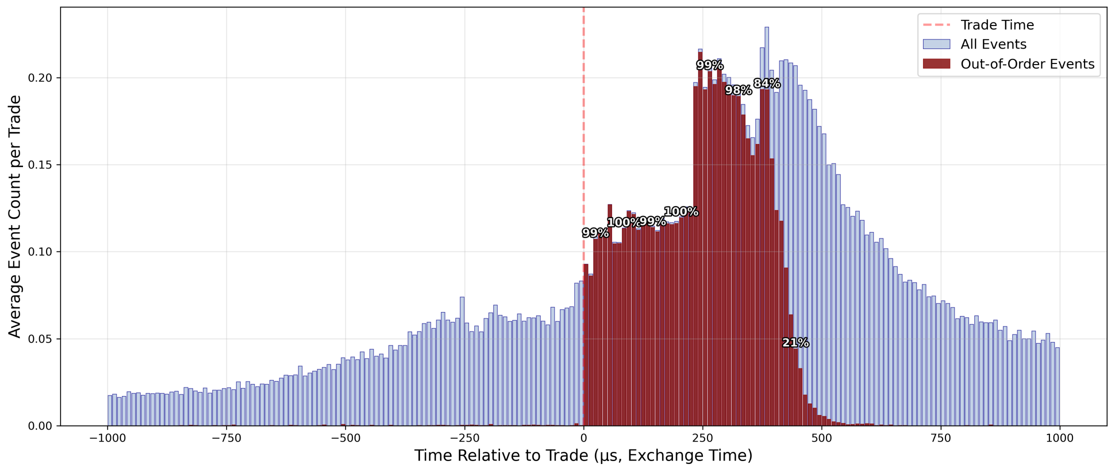
Notes. This figure shows the locations of the data centers where most U.S. securities exchanges and the SIPs are operated out of. The NYSE-owned exchanges (e.g., NYSE and NYSE Arca) and the NYSE SIP are housed in the NYSE data center in Mahwah, NJ at the top of the figure. The Nasdaq data center is in Carteret, NJ at the bottom of the figure, and operates the Nasdaq stock exchange and the Nasdaq SIP. The CBOE exchanges (BZX, BYX, EDGA, and EDGX) are run out data centers in Secaucus, NJ.

in exchange time but was reported *before* the trade by the SIP. Similarly, a red bar before the trade indicates that those events occurred before the trade according to exchange time (based on the location of the bar on the x-axis), but the events were reported at any point after the trade on the SIP.

The first observation from the plot is that there is a spike in Mahwah JPM activity around 220 to 230 microseconds after a trade on an exchange in Carteret, which is the time it takes for low-latency traders to send order requests from the Nasdaq data center in Carteret to the NYSE data center in Mahwah.[9] We interpret the jump in events around 220 μs as a *response* by HFTs and ATs to the Carteret trade – after observing a Nasdaq trade, they adjust quotes on Mahwah exchanges (and all other exchanges for that matter) as quickly as their technology allows. The figure also shows that nearly all of the Mahwah events in the hundreds of μs after a Carteret trade are out of order on the SIP, including the events that are very likely a response to the trade itself. This is how the look-ahead bias manifests: JPM trades in Carteret will be reported to the NYSE SIP with a median latency of 560 μs.

[9]From the example above, ~190 μs of travel time + ~30 μs of exchange processing time.

Figure 2: Average Number of Mahwah Events Around a JPM Trade in Carteret



Notes. We plot the average number of events (quote changes and trades) around each trade in JP Morgan (JPM) on June 20, 2019. For each JPM trade on a Carteret exchange (e.g., Nasdaq), we count events on Mahwah exchanges (e.g., NYSE Arca) in 10 μs bins using exchange timestamps, then report the average over all trades (blue bars). We also plot in red the average number of events that are out of order: events that occur after (before) the trade in exchange time but before (after) the trade according to the SIP. For every 6th bar, we also report the percentage of events that are out of order in white if a non-trivial percentage of events is out of order.

However, the trade generates responses in Mahwah after around 220 to 250 μs and is then reflected on the SIP in another 110 μs (a total of around 330 μs), which means the responses to the trade appear on the SIP before the trade itself.

The out-of-order examples from JPM are representative of patterns found across all trading days, exchanges, and symbols, even for symbols with much less trading activity than JPM. We even find that quote updates after a trade on the same exchange are sequentially out of order on the SIP. Over all trades in all symbols, 53% of trades have at least one event that occurred after the trade but is reported by the SIP as occurring before. The out-of-order rate in the opposite direction – an event occurs before a trade but is reported on the SIP after – is 28%. We also show that trades with longer latencies have more severe out-of-order issues. For example, 76% of all trades on Nasdaq that report to the NYSE SIP have at least one event that occurs after the trade according to exchange timestamps but is reported before on the SIP. On average, there are 7.3 events per trade that are out of order in this direction. High-latency trades also represent a significant fraction of trades – high-latency trades are not a function of economic activity per se but about geography and listing exchange. In fact, Nasdaq has nearly twice as many trades as any other exchange for all symbols that report updates to the NYSE SIP. For the quotes and trades from the same exchange that are out of order, the errors are less frequent – occurring only in about 20% of trades – but are arguably more egregious in that when a strict sequential process like a limit order book is reported out of order, the result will often be nonsensical.

Since the SIP look-ahead bias will often change a trade's prevailing SIP NBBO midpoint, it will also affect the two most basic methodological steps in using TAQ data: trade signing and effective spread/price impact measurement. We next provide more details on our latency-free signing and spread measurement methodologies[10]

Our latency-free (LF) signing methodology uses exchange timestamps and starts with the EX BBO rule, which signs each trade using the prevailing best bid and offer from the exchange where the trade occurred. As expected (based on the rules of the limit order book), we find that trades against an exchange's best bid or offer (according to TAQ) are signed with 100% accuracy in our sample where we definitively know the trade direction from direct-feed data. In comparison, the SIP methodology signs this sample of trades with 94% accuracy. Moreover, this set of trades, where the EX BBO rule signs trades perfectly, accounts for 68% of all trades in our sample and 80% of all dollar volume.

The most prominent reason why trades do not execute against an exchange's best bid or offer (as recreated from TAQ data) is that odd lot limit orders are not reported to the SIP. Odd lot limit orders can sit inside the exchange's BBO, but since they are not broadcast to the SIP, they will not be accounted for in TAQ's view of the EX BBO. When a marketable order is submitted, the more aggressively priced odd lot limit orders will trade first (following the rules of the limit order book), allowing for buyer-initiated trades to occur at a price lower than the TAQ-observed EX BBO midpoint (similarly, seller-initiated trades above the EX BBO midpoint). We present evidence that odd lot trades inside the EX BBO are often with stale odd lot limit orders (likely from retail traders). As a result, we supplement our signing methodology for these problematic trades with the latency-free (LF) NBBO that was prevailing 1 millisecond before the trade, then using the Lee-Ready logic.

The reason why a stale LF NBBO works better than other methods is that many incorrectly signed odd lot trades result from stale limit orders that do not adjust to recent quote changes. Thus, stale quotes better capture market conditions prevailing closer to when the order was submitted by ignoring the cluster of activity leading up to the trade. The stale LF NBBO signing rule is more accurate by 15 pp over the SIP method (91.8% vs. 76.6% accuracy). To be clear, the stale LF NBBO rule is based on a pattern we observe in the data – odd lot trades inside the EX BBO tend to be stale with stale limit orders. We emphasize that

[10]For our analysis, we only consider trades on exchange (i.e., "lit" trades). We provide some suggestions for signing off-exchange trades in Section 5.2.1.

the foundation of our LF methodology – the EX BBO signing rule – has no discretion and is a function of the rules of the limit order book. This is the sense in which the EX BBO rule offers a certain type of stability, involves no guesswork, and is applicable as long as the limit order book is the predominant market design used by exchanges. On this note, we also highlight that the SIP feed will include odd lot limit orders at the best bid and offer starting in May 2026. While this does not change our signing methodology per se, it does mean that more trades will be signed with the EX BBO rule because fewer trades will occur inside the EX BBO. We estimate that if odd lots were reported to the SIP in our Arca signing sample, the accuracy rate would rise to 97%.

We also test our LF signing methodology on trades on all exchanges and compare it to the SIP signing approach. Since we do not have the true sign on all exchanges, we use the percentage of trades where the latency-free (EX BBO + delayed LF NBBO) and SIP methodologies assign different trade signs as a proxy for signing accuracy. Using our Arca direct-feed sample, we verify that this proxy is highly correlated with differences in signing accuracy, and most of the variation in our proxy is driven specifically by SIP signing inaccuracy. With this proxy, we present evidence that SIP signing errors occur more frequently on other exchanges, especially for trades with high exchange-to-SIP latency. These high-latency trades (e.g., Nasdaq to the NYSE SIP) have more than double the trade signing differences that we see on NYSE Arca. With longer exchange-to-SIP latencies, more quote updates from after the trade will reach the SIP before the trade, which makes it more likely that the SIP NBBO midpoint will move enough to flip the sign of a trade.

The last part of our paper compares effective spread and price impact measures computed using the SIP NBBO midpoint with measures computed with the LF NBBO midpoint. The LF midpoint, which is similar to the NBBO midpoint introduced in Bartlett and McCrary (2019), has no look-ahead bias and aggregates information as it is produced with zero latency. We think of the LF NBBO as a good proxy for the most informationally efficient price at any point in time, better than what market participants could construct themselves. We find that 19% of trades and 22% of dollar volume are assigned a smaller effective spread when computed with the SIP NBBO midpoint compared to using the LF NBBO midpoint. This percentage is greater than the rate of signing differences. The reason for this is that, while the SIP may have a look-ahead bias that moves the SIP NBBO, it may not move enough to flip the inferred direction of the trade. However, these less extreme cases will still move the SIP NBBO midpoint and alter the effective spread. Averaging over all trades, we estimate that SIP effective spreads are understated by an average of around 14% of trades and dollar volume. Price impact results are similar because almost all of the variation in both effective spreads and price impact measures comes from differences in the prevailing midpoint used at the time of the trade.

While these numbers are large, they disguise how extreme the bias in spreads and price impact can be. For example, with Nasdaq trades that report to the NYSE SIP, we find that 43% of trades have a SIP effective spread smaller than their LF counterpart. The understatement of the effective spread is not just widespread, but also severe in magnitude. We estimate that the effective spread of these Nasdaq trades is understated by 48%. In general, we find that as latency increases, the effective spread and price impact bias become larger in magnitude because the look-ahead bias becomes more severe.

The rest of the paper is organized as follows. Section 2 reviews the related literature and describes important institutional details that are necessary background information. Section 3 describes the data. Section 4 estimates latencies between exchanges and the two SIPs, and shows that these latencies are relevant because trades and quote updates tend to be clustered, which leads to a sequentially inaccurate SIP. Section 5 explains why a sequentially inaccurate SIP leads to signing errors and a bias in the effective spread, then estimates how often signing errors occur and the degree of the effective spread bias. This section also introduces our exchange BBO signing methodology and latency-free NBBO spread measure. Section 6 concludes.

2 Related Literature and Institutional Details

2.1 Related Literature

Latency and the NBBO

Ding et al. (2014) compare the SIP NBBO and a synthetic NBBO computed using direct data from five exchanges – BYX, BZX, EDGA, EDGX, and Nasdaq – and the SIP top of book for all other stock exchanges. They find that the two NBBOs have frequent price dislocations, but last for only about one to two milliseconds each. They argue that the cost of dislocations is small for investors who trade infrequently.

Hasbrouck and Saar (2013) develop a measure of low-latency activity using "strategic runs," which can be interpreted as a form of clustering, though over a much longer horizon of 100 ms (100,000 μs) than our clustering analysis over 500 μs intervals. Ernst et al. (2021) document that there is a burst of activity following the dissemination of *off-exchange* trades by the SIP.

Aquilina et al. (2022) use proprietary message-level matching engine data to identify races to snipe stale quotes (in the spirit of Budish et al. (2015)). They examine attempts to trade or cancel limit orders that have timestamps within the "information horizon," or the period where messages could not have been responses to each other. Our paper finds that many of the trades and quote changes that fall into this window are often out of order on the SIP. In addition, we also show that trades and quote changes that fall *outside* of the information horizon are also out of order.

Trade Signing

Lee and Ready (1991) (henceforth, Lee-Ready) is the classic methodological paper that has guided decades of research using the NYSE TAQ dataset specifically and trade and quote data more generally. They identify two problems in assigning a direction to a trade using the prevailing best bid and offer quotes at the time of the trade. The first is that quote changes and trades can be out of order, similar to what we document in this paper.[11] To combat this issue, Lee-Ready recommend using prevailing quotes 5 seconds before the trade. This recommendation is similar to what we recommend in signing odd lots. While the intuition for using stale quotes is the same, the reasoning is quite different (see Section 5.2.1 for details).

The second problem Lee-Ready identify is that trades often occur inside the best bid and offer, even when quotes and trades are not out of order. Lee-Ready suggest classifying trades above the midpoint as buys, trades below the midpoint as sells, and trades at the midpoint using the tick test. The tick test assigns a direction by comparing a trade's price to the last trade at a price different from the current trade's price. If the last trade price is below the trade price, the tick test assigns the direction as a buy; otherwise, the trade is marked as seller initiated.

In some ways, our proposed signing methodology comes full circle back to Lee-Ready thirty-five years later. The standard practice today is to use Lee-Ready with the SIP NBBO midpoint. Our primary signing rule takes a different route: we use the prevailing exchange

[11] The reason is quite distinct and is a relic of human-based markets: Lee-Ready note that if the system handling the specialist's quotes is faster than the clerk handling the trade, the quote updates generated by the trade can be entered and recorded before the trade. They write on page 734 of their paper: "The problem of quotes recorded ahead of trades has always existed, but has increased substantially with the widespread use of 'electronic books' by specialists. We show that misclassifications can be greatly reduced by comparing the trade to the quote in effect 5 seconds earlier."

BBO with exchange timestamps, which, perhaps ironically, returns to the spirit of Lee-Ready by using the *relevant* quotes (i.e., the prevailing exchange quotes) to sign trades instead of *all* quotes (i.e., the prevailing NBBO). In a further parallel to the ideas in Lee and Ready (1991), we use stale quotes to sign trades against unobserved odd lot limit orders.

Holden and Jacobsen (2014) find significant differences in liquidity measures between the so-called monthly NYSE TAQ (timestamped to the second) and the daily NYSE TAQ (timestamped to the millisecond), and offer a solution for monthly TAQ users. Looking ahead to the introduction of sub-millisecond timestamps, they describe a Relative Best Bid and Offer (RBBO), which constructs an NBBO based on the BBO from different market centers but with a lag that accounts for the travel latency in receiving updated data from different geographical locations.[12]

Schwenk-Nebbe (2022) documents that a marketable order that interacts with one or more limit orders will lead to a mechanically-generated quote update, and all trades and quote updates have the same participant timestamp. This fact can be used to generate a marketable order identifier, linking several trades to one marketable order. In addition, Schwenk-Nebbe (2022) shows that the mechanical quote update (MQU) can be reported before the trade from which it originated, leading to errors in signing and spread/price impact measurement. These results are a special case of our same-exchange out-of-order results. Our paper focuses on the broader phenomenon of out-of-order messages, some of which are responses to a trade, and shows that the out-of-order SIP is due to geographic latency and SIP processing latency. We discuss automatically-generated quote updates further in Appendix B.2.

Holden et al. (2023) document that Lee-Ready signing using the SIP NBBO introduces signing errors and a downward bias in the effective spread because the SIP is constructed with latency. Holden et al. (2023) proposes a relative BBO (RBBO) method and a simpler latency timestamp adjusted (LTA) method, each of which reduces signing errors by applying a latency adjustment that varies based on observed exchange-to-SIP latency.[13] Both of these adjustments are in the spirit of what Holden and Jacobsen (2014) suggest. This paper is distinct for at least two reasons. First, our signing methodology, which is built on the rules of the limit order book, is conceptually trivial, simpler than applying a latency adjustment, outperforms latency adjustment signing rules, and will further improve once

[12]Holden and Jacobsen (2014) write on page 1782, "In an Einsteinian universe, a bid and offer update message could not travel faster than the speed of light, so high-speed traders face immutable lag times in receiving bid and offer updates from remotely located exchanges."

[13]Holden et al. (2023) is an earlier working paper version of this paper. This paper takes a completely different approach from Holden et al. (2023) and focuses on *why* latency causes the SIP to be out of order. Because the papers are so different, we leave Holden et al. (2023) available as a permanent working paper.

odd lot limit orders are added to the SIP. Second, our paper documents *why* signing errors occur: responses to a trade may come before the trade itself because low-latency market participants can respond to a trade and have those responses recorded by the SIP before the SIP records the original trade. Together, these insights help clarify that, while a latency adjustment helps capture the perspective of traders in a particular location, it does not capture the most relevant information for signing trades.

2.2 Important Institutional Details

The U.S. Securities Market Landscape

The U.S. securities market is highly fragmented and has been that way since the mid-to-late 2000s. During the period of our data in 2019, there were 13 registered securities exchanges (as of this writing, there are 17 exchanges). All exchanges offer electronic trading, and their matching engines operate out of data centers in Northeast New Jersey.

Every U.S. listed security trades on every exchange because of a regulation called Unlisted Trading Privileges (UTP). For example, Apple (AAPL), which is listed on Nasdaq, can be traded on any national securities exchange. So shares of AAPL can be bought on Nasdaq, then sold on CBOE BZX, then purchased again on NYSE Arca.[14]

While the number of exchanges has fluctuated over the years, as of 2019, there are 9 exchanges that are responsible for more than 95% of all exchange trades. These 9 exchanges are owned by four distinct companies: (1) Nasdaq operates the Nasdaq and Nasdaq BX exchanges in Carteret, NJ; (2) the NYSE runs NYSE and NYSE Arca, operated in Mahwah, NJ; (3) CBOE has four exchanges – BZX, BYX, EDGX, and EDGA – which are run in Secaucus, NJ; and (4) IEX, the subject of the popular book *Flash Boys*, has its point of-presence (where market participants place their servers) in Secaucus, NJ. See Figure 1 to get a sense of the geographic locations of these exchanges.

Nearly all exchanges use the electronic limit order book market design.[15] The basic building blocks of a limit order book are (displayed) limit orders, which specify a direction, price, and quantity. Outstanding limit orders – limit orders that cannot execute when they initially arrive at an exchange – constitute the order book, and the displayed buy order(s) with the highest price is the best bid and the displayed sell order(s) with the lowest price is the best

[14] See Budish et al. (2024) for more details on the industrial organization of U.S. securities exchanges.

[15] IEX uses a limit order book design, but with exchange-managed sliding orders as an additional feature. The NYSE arranges trades with parity, sharing trades with potentially many orders at the best price.

ask or offer. A trade occurs when an incoming order is willing to sell at or below outstanding orders waiting to buy, or willing to buy at or above outstanding orders waiting to sell. See Harris (2002) for more details on the limit order book design.

In 2005, the SEC enacted Regulation National Market System (Reg NMS), which formally created the notion of a national best bid and offer (NBBO), or the best bid and best offer across all exchanges in a given symbol. Rules 610 and 611 in Reg NMS have shaped the way exchanges operate. Rule 610, or the Access Rule, prevents an exchange from displaying limit orders that would lock or cross the best bid and offer on other exchanges (i.e., "protected quotes" on other exchanges). That is, it *attempts* to prevent a locked or crossed NBBO. Rule 611 requires exchanges to prevent a "trade-through" of protected quotes, where a marketable order trades with an outstanding limit order at a price that is inferior to a protected quote on another exchange.

The Securities Information Processor

The securities information processor (SIP) collects trade and top-of-book quote updates on every exchange in every symbol and disseminates that information to subscribers in real time. The historical archive of the SIP feed generates the NYSE TAQ data, which is used in this paper and is the primary source of trade and quote data in the academic literature.

There are two SIPs, one operated by NYSE and the other operated by Nasdaq. The NYSE SIP, formally called the Consolidated Tape Association (CTA) SIP, reports quotes and trades for all securities listed on NYSE, NYSE Arca, and NYSE American. The NYSE SIP is operated in the NYSE data center in Mahwah, NJ, the same data center where the NYSE family of exchanges operates their matching engines. All Nasdaq-listed securities are reported on the Nasdaq SIP, formally called the Unlisted Trading Privileges (UTP) SIP. The Nasdaq SIP is operated in a data center in Carteret, NJ, the same data center where Nasdaq exchanges run their matching engines. The CBOE exchanges are operated in Secaucus, NJ.

The SIPs aggregate all trades and BBO updates in a given symbol from all exchanges and report the updates from a central location. From these updates, subscribers can observe all trades from a single feed, get a view of the BBO on each exchange throughout the trading day, and construct the NBBO. For example, all trades and top-of-book quote updates across all exchanges in JP Morgan (JPM), an NYSE-listed stock, are reported to and disseminated by the NYSE SIP. A trade in JPM can occur on Nasdaq, and after it is recorded by Nasdaq, the exchange will send notification of the trade to the NYSE SIP. Once it is received by the SIP, it will be processed, and SIP subscribers will then be notified of the trade. The process is the same for Nasdaq securities, but trade and quote updates are sent to the Nasdaq data center in Carteret, NJ, and disseminated by the Nasdaq SIP from there.

Complying with Reg NMS requires keeping track of the best bid and offer (BBO) on every exchange in every symbol. Since Reg NMS is enforced by exchanges, each exchange must monitor protected quotes from all exchanges and construct the NBBO to (1) ensure no limit orders are displayed that lock or cross the NBBO, and (2) prevent executions that would trade through a protected quote on another exchange.[16] Reg NMS elevated the role of the SIP, which offered a way to get the BBO of every exchange and construct the NBBO without subscribing to the direct-feed data from all exchanges.[17]

Many exchanges initially used the SIP to comply with Reg NMS. Today, many exchanges use direct-feed data, or data directly from the exchange on limit order book activity, to comply with Reg NMS. The direct-feed data is the lowest latency and most comprehensive trade and order book data available, and it is sold directly by exchanges.

When the SIP receives updates from securities exchanges, it processes the data and assigns a SIP timestamp. This was the first timestamp that was available in the TAQ data. Starting in 2015, the SIP was updated to include so-called participant timestamps, which are assigned by the *exchange* after limit order book activity occurs but before the update is sent to the SIP. This participant timestamp is key to understanding the actual sequence of events, since this timestamp is assigned by the venue when the activity occurred. Also note that TAQ timestamps had millisecond granularity starting in 2003, moved to microsecond granularity in 2015, and nanosecond granularity in 2016 and 2017 for UTP and CTA symbols, respectively.

3 Data and Sample

We use two datasets throughout the paper: the NYSE Arca direct-feed data and NYSE's Daily TAQ dataset. We describe each dataset below.

[16]Market participants can send intermarket sweep orders (ISOs), which take the burden of Reg NMS compliance away from the exchange and onto the participant. These orders are used to trade on multiple exchanges, clearing all protected quotes in the process. These orders are necessary to trade through the book at multiple exchanges because, without them, exchanges would prevent trades because of the order protection rule without knowing that the trader is in flight to clear protected quotes on other exchanges.

[17]The SIP is still used by market participants to get a sense of what the market for a security is. Since many participants are not latency sensitive, they do not pay for direct-feed data that can be many multiples more expensive than SIP data (Budish et al., 2024). HFTs, ATs, and market participants that operate at a large scale (e.g., large broker-dealers) tend to use direct-feed data because the SIP is less comprehensive and, as we document in Section 4.1, is much slower than direct-feed data.

NYSE Arca Direct-Feed Data

NYSE Arca's Integrated Feed data, or the so-called *direct-feed data*, are an archive of real-time direct feeds, which are typically used by HFTs, ATs, and other sophisticated market participants to get the most timely and complete information on trade and quote activity on NYSE Arca. The direct-feed data provide the most comprehensive trade and limit order book data, including order submissions, cancellations, and executions, all timestamped to the nanosecond. These data also allow us to infer the direction of the marketable-side of each trade, which is key in the trade-signing analysis in Section 5.2. Specifically, the data allow us to identify the specific limit order (via an order ID) that was the contra-side of the marketable order. Since every limit order must specify a direction and is added to the order book as a bid or an ask, we know that if the limit order was, say, a buy, the marketable order must have been a sell. This allows us to ascertain the true direction of the trade initiator and objectively measure a methodology's signing accuracy for NYSE Arca trades.

There is a subset of trades that we cannot sign using the Arca direct-feed data. These trades, labeled as hidden order executions, are a result of a marketable order interacting with a non-display (i.e., hidden) limit order that is invisible to all participants except for the exchange itself. This is distinct from the issue of odd lots discussed in Section 5.2. While odd lot limit orders are not reported to the SIP (and thus not visible in TAQ), they are still identified as regular orders in the direct feed (unless the submitter specifically marks them as non-display). As a result, we can still infer the sign of these trades with the direct-feed data if they are purely odd lot trades and do not have any additional hidden modifications.

The reason we cannot sign hidden order executions with the direct-feed data is that the limit order that interacted with a marketable order and led to a trade is not identified. For example, if someone submits a hidden midpoint order that is bid at 10.03 but not displayed, if that order trades, the direct-feed data shows notice of a hidden order execution, but there is no record of the hidden bid submitted at 10.03 or an identifier for that order.[18] Since we cannot identify the contra-order for hidden order executions, we cannot identify the marketable order direction and are forced to omit hidden order executions from our sample used in Section 5.2.1. These trades represent 12.14% of all Arca trade observations and 12.86% of all Arca dollar volume. We discuss these trades in detail in Appendix B.1. We conjecture that many of these trades will be signed accurately by our LF signing methodology, while others (like true midpoint trades) are inherently difficult to sign using any methodology.

[18] Identifying hidden limit orders in the direct-feed data would defeat the purpose of submitting a hidden order. As a reminder, the direct-feed data is an archive of the real-time direct-feed data, so if these hidden limit orders were disseminated, market participants would know of their presence in real time.

NYSE Daily TAQ

The NYSE's Daily TAQ dataset is a historical archive of the SIP feeds. We access this dataset through Wharton Research Data Services (WRDS) Cloud. We use both SIP timestamps as well as "participant" timestamps. SIP timestamps are assigned by the SIP after it finishes processing updates from exchanges. SIP timestamps have a granularity of a microsecond starting in 2015, and a nanosecond (a billionth of a second) starting in 2016 for UTP symbols and 2017 for CTA symbols. Participant timestamps are assigned by the exchange when the trade or quote change occurred and, like SIP timestamps, have nanosecond resolution after 2015. These timestamps indicate when a trade or quote update was processed by the exchange's matching engine, and before the update left the exchange's network and began traveling to the SIP. We will refer to participant timestamps as *exchange timestamps* to highlight that these timestamps are directly from the exchange.

For most of our analyses, we use TAQ data for the month of June 2019. In some supporting analyses, we use a one day sample from June 20, 2019. We filter out trades that occur outside of regular trading hours or have a trade price of less than $1.00, and we remove official market opening and closing trade status messages. We restrict our sample to symbols that are classified as a common stock or an exchange-traded product according to the "Security Type" field in the TAQ Master File (codes "A," "H," "ETF," "ETN," and "ETV"). In analysis that aggregates across symbols (e.g., exchange or day/month level), we include all symbols that satisfy the above filters and have at least 100 trades. For any analysis aggregated to the symbol level, we require at least 1,000 trades to be included in the sample to get meaningful averages.

We use the TAQ data to expand trade signing to all exchanges, not just NYSE Arca trades in the direct-feed data. We also use the TAQ data to measure spreads for all exchange trades using the SIP NBBO and an LF NBBO benchmark. We do not extend our analysis to off-exchange trades because we view these trades as economically distinct from lit trades. However, we do recommend using our 1 ms delayed LF NBBO rule to sign off-exchange trades because of the long delay in reporting to the SIP, which makes SIP-timestamped quotes particularly stale. See Section 5.2.1 and Appendix G.1 for more details.

Reconciling the Datasets

The direct-feed data are much more comprehensive than the TAQ data in that they contain details about every displayed order (order addition, modify, and cancel), including odd lots and orders outside the BBO. The direct-feed data also provide more detailed information about every trade (which allows us to infer the direction of the marketable order). The TAQ data show only top-of-book quote changes for every exchange and basic details for every trade. However, it is important to note that the timestamps in the direct-feed data are an exact match with the exchange timestamps in the TAQ data. The perfect timestamp match further confirms that these timestamps are generated by the exchange and are exactly what a direct-feed subscriber would have received.

Because the TAQ data only has top-of-book quotes, quote messages are often generated "automatically" because of a trade, simply to inform the SIP that the BBO price and/or quantity has changed as a result of the trade. For example, a marketable order that takes all shares at an exchange's best offer will generate an automatic quote update to notify the SIP that the best offer price has increased. These quote updates also have the same exchange timestamp as the trade. However, when the automatic quote updates reach the SIP, they will almost always have different SIP timestamps and will often be out of order. This alone can create sequencing errors in the TAQ data, and was documented by Schwenk-Nebbe (2022). Much of our analysis in Section 4 considers sequence errors for events not including the automatically-generated quote update. We think of this as a conservative approach that will understate out-of-order issues. See Appendix B.2 for more details on automatically-generated quotes.

Constructing the Latency-Free NBBO

We construct the LF NBBO for a security in the same way the NBBO is constructed using the TAQ except we first sort all updates by exchange timestamps. More specifically, we begin by selecting an exchange and symbol, and we use the exchange timestamp to sort quote updates based on when they occurred on that exchange. Using these sorted quote updates, we record each exchange's best bid and offer (i.e., the EX BBO) for each security on the exchange throughout the day. We repeat this process for each exchange that traded the security on that day. Finally, we compute the LF NBBO for each security-day by taking the highest bid and the lowest offer displayed by an exchange at each point in the trading day according to exchange timestamps. The LF NBBO is the same as the Direct NBBO used in Bartlett and McCrary (2019), though the use cases are quite different.[19]

Note that the LF NBBO must be completely reconstructed. Using the SIP NBBO, as commonly available on WRDS or from existing code, and simply rearranging the quotes

[19] Bartlett and McCrary (2019) use the Direct NBBO as a type of upper bound in market information for the purpose of latency arbitrage. Our paper uses it as an objective benchmark for the market's estimate of the "fair" price at a point in time.

using exchange timestamps does not actually construct the LF NBBO. The reason is that what the SIP NBBO views as the NBB and NBO depends on the sequence of events across all exchanges. Because the sequence is often out of order, as we document in Section 4.3, simply rearranging the SIP NBBO with exchange timestamps will lead to errors.

4 Latency Magnitudes and Why Latency Matters

In this section, we report the distribution of exchange-to-SIP latencies across all U.S. equities. Although latencies are unconditionally small, we show that latencies still influence the SIP because latencies are heterogeneous and trades and quotes tend to be tightly clustered together. This combination implies that the SIP will frequently report trades and quotes out of chronological order, which we confirm in the data.

4.1 Latency Magnitudes

We measure latencies of every exchange quote change and trade for one day in June 2019. We compute the latency of every trade by taking the difference between the SIP timestamp and the exchange timestamp. We report the median latencies for trades and quotes for each exchange. In addition, we organize the numbers based on the geographic distance between the exchange on which the trade/quote occurred and the SIP to which the trade/quote was reported. This sorting allows us to get a sense of latencies from travel time (i.e., geographic latency) and those due to the SIP itself. We refer to the latter source of latency generically as "SIP processing latency," and it includes SIP processing logic and server/gateway latencies.

Table 1 presents exchange-to-SIP latencies. The median latency generally falls within a range between 15 μs and 550 μs.[20] We highlight several patterns. First, the latencies are larger as the distance between the exchange and the SIP increases. For example, messages from exchanges in Carteret to the NYSE SIP in Mahwah, or from Mahwah exchanges to the Nasdaq SIP in Carteret, have latencies that are greater than 350 μs. The distance from Carteret to Mahwah (and vice versa) is the longest of any of the connections between the three cities in the New Jersey triangle (~35 miles). For context, the physical limit of sending

[20]There are a few outliers, even for median latencies, which involve NYSE Chicago. The Chicago Stock Exchange was purchased by the NYSE, and the matching engine was moved from the Chicago area to the NY/NJ area. This transition was staggered by symbol, which meant that during the transition, some symbols were traded in a data center in Chicago and others were traded out of New Jersey.

Table 1: Median Latency from Exchanges to NYSE/Nasdaq SIP

		NYSE SIP (Mahwah)		Nasdaq SIP (Carteret)	
Exchange	Location	Quote	Trade	Quote	Trade
Nasdaq	Carteret	546	561	17	21
Nasdaq BX	Carteret	544	556	16	19
Nasdaq PSX	Carteret	545	568	18	24
Cboe BYX	Secaucus	407	437	192	208
Cboe BZX	Secaucus	405	443	191	210
Cboe EDGA	Secaucus	413	448	198	217
Cboe EDGX	Secaucus	413	451	199	220
IEX	Secaucus	454	460	216	233
NYSE Arca	Mahwah	106	133	373	377
NYSE	Mahwah	106	130	367	370
NYSE American	Mahwah	110	129	371	374
NYSE National	Mahwah	216	233	103	121
NYSE Chicago	Mahwah	10251	1029	8036	560

Notes. This table presents median latencies in microseconds for trades and quote changes sent from 13 U.S. stock exchanges to the two SIPs. All symbols listed on Nasdaq report trades and top-of-book quote changes to the Nasdaq SIP in Carteret, NJ. All other symbols report to the NYSE SIP in Mahwah, NJ. We use all trades and quotes from June 20, 2019 during the regular trading day. We measure latency for each trade and quote as the difference between the SIP-assigned timestamp and the exchange timestamp.

a message between the two locations based on the speed of light is ~181 μs, and cutting edge HFTs and ATs have a latency of around 190 μs.[21]

Second, the NYSE SIP has greater processing latencies than the Nasdaq SIP. Since the NYSE exchanges operate in the same data center as the NYSE SIP, we can infer that the latency between the Mahwah exchanges and the NYSE SIP is entirely due to SIP processing latency. Similarly, the latency between Carteret exchanges and the Nasdaq SIP can be interpreted as SIP processing latency. Using these pairs, we infer that the NYSE SIP has a latency of around 100 to 130 μs, while the Nasdaq SIP has a smaller latency of 15 to 20 μs.

Using these SIP processing latencies, we can also infer that exchange-to-SIP travel time latencies are much greater than state-of-the-art latencies. For example, the total latency from Nasdaq (Carteret) to the NYSE SIP (Mahwah) is ~550 μs. Given that NYSE SIP processing latencies are around 110 μs, we infer that the travel time latency is 450 μs, which is 260 μs slower than the state-of-the art latency of 190 μs cited above. In fact, the 450 exchange-to-SIP latency is large enough that an HFT could observe a trade on Nasdaq, adjust quotes on Mahwah exchanges, and have those adjustments reflected by the SIP, all

[21] Anova Financial Networks is a provider of low-latency connections, and they report a latency of 187 μs from Carteret to Mahwah and vice-versa on their website.

before the original trade reaches the SIP.

Third, trades have greater latencies than quote updates for each exchange-SIP pair. The discrepancy ranges from 15 to 30 μs for the NYSE SIP, and from 3 to 20 μs for the Nasdaq SIP. In Appendix C.1, we also report the interquartile range (IQR) of quote and trade latencies for each exchange-SIP pair. Table C.1 shows that the IQR values are small relative to the overall range of median latencies: there is much more variability from geography and SIP processing than from randomness within an exchange-SIP pair.

4.2 Why Latency Matters: Clustering of Trades and Quotes

By nearly any measure, a latency of, say, 500 μs is extremely small. For context, it takes about 250,000 μs to blink an eye – even the high end of exchange-to-SIP latencies is tiny in comparison. It is quite sensible to conjecture that these types of latencies do not matter.

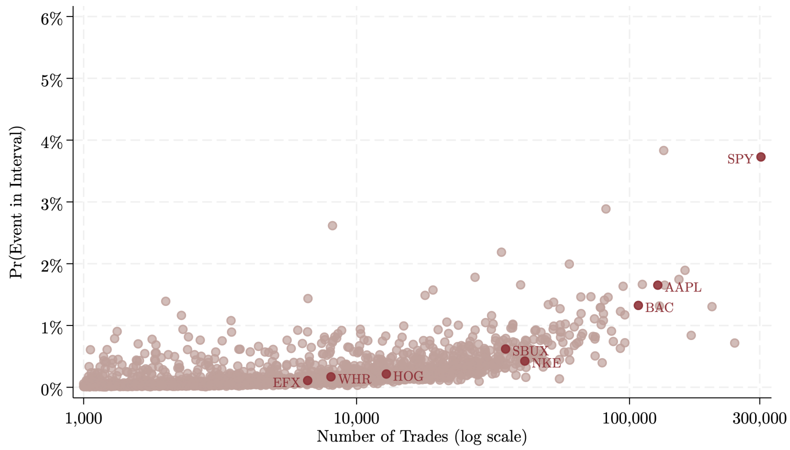
Empty Intervals

Latencies of up to 550 μs are only relevant if events tend to occur within 550 μs of each other. As an example, consider a trade occurs and is reported to the SIP with a 500 microsecond delay. The standard practice is for researchers and practitioners to use the prevailing SIP NBBO midpoint at the time the trade is broadcast by the SIP to sign the trade and compute the effective spread/price impact. If no other trades or quote changes occur in the 500 μs around the trade, there is nothing for the SIP to get out of order, and there is no change to the prevailing SIP NBBO midpoint.

As a starting point, we divide the trading day into non-overlapping 500 μs intervals, then count how many of those intervals have at least one BBO quote update or trade (i.e., an "event") according to exchange timestamps. Figure 3 reports the probability that a 500 μs interval contains at least one event. We compute the probabilities separately for each symbol for one trading day, and require at least 1,000 trades in a day for a symbol to be included. The figure shows that almost all 500 μs intervals for *every* symbol are empty. Even the most active symbol in the U.S. stock market, SPY, has nothing happening in more than 96% of intervals. In addition, almost all symbols have nothing happening in more than 99% of intervals, including large-cap stocks like Starbucks (SBX) and Nike (NKE).

Even if we were to hypothetically take every trade and quote change in a given symbol and spread them out to cover as many 500 μs intervals as possible, we would still leave more

Figure 3: How Many 500 Microsecond Intervals have a Quote Change or Trade?



Notes. This figure examines non-overlapping 500 microsecond (μs) intervals during the regular trading day on June 20, 2019. For each symbol, we compute the probability that an interval has at least one quote update or trade. We plot the percentage on the y-axis, and the number of trades on the x-axis. We highlight several example symbols that span a range of trading activity: Equifax (EFX), Whirlpool (WHR), Harley Davidson (HOG), Starbucks (SBUX), Nike (NKE), Bank of America (BAC), Apple (AAPL), and the SPDR S&P 500 ETF (SPY). See Table C.2 for specific numbers on the example symbols.

than 99% of intervals empty for nearly all symbols, and more than 99.8% of intervals for the average symbol (by trading volume). So even though the electronic U.S. stock market is characterized by frequent trading and quote changes (Angel et al., 2011), there is simply not enough activity in any symbol to cover the 46.8 million 500 μs intervals in a trading day.

We provide summary statistics and exact numbers for Figure 3 in Table C.2 in Appendix C.2. That table includes the number of quote and trade messages observed in a typical symbol-day, the various event probabilities, and the hypothetical maximum probability for each of the example symbols highlighted in Figure 3.

Trades and Quotes are Clustered

Given that nearly all 500 microsecond intervals are empty, it seems sensible to conclude that a latency of between 15 and 550 μs might have little to no impact on using the SIP to sign trades and measure spreads. While it is true that most intervals are empty, we document that trades and quotes are also extremely clustered. That is, *conditional on a trade* occurring in an interval, there tend to be many other quotes and trades within 500 μs.

Before showing data on how clustered trades and quotes are, we provide an illustration of the clustering observed in the data. For this illustration, we examine all trades and all quote changes in JP Morgan (JPM) on one trading day (June 20, 2019). For each trade, we examine all events that occur within 1,000 μs (1 ms) before or after the trade, all according to exchange timestamps. For each 10 μs bin before and after the trade, we count the number of events that occur in that bin. We then take an average in each bin over all trades, which gives us the expected number of events in each 10 μs bin around a trade.

Figure 4 presents clustering plots for JPM. Panel A includes all trades and events across all exchanges. Panel B restricts attention to Carteret: it considers only JPM trades on Carteret exchanges and counts surrounding events that also occurred on Carteret exchanges. Panel C uses the same Carteret trades but instead counts events on Mahwah exchanges. For the clustering activity, we show the events in the *blue* bars. The red bars show how many events are out of order, which will be explained in Section 4.3.

The overall clustering figure in Panel A shows that there is a significant amount of activity immediately surrounding the trade. Moreover, activity is steadily increasing leading up to the trade, then there is a spike of activity just after the trade, followed by a gradual decline starting around 250 μs after the trade. Between the time of the trade to 250 μs after the trade, we see that there are about 0.55 events per 10 μs bar. That means that there are an average of 13 to 14 JPM events in the 250 μs immediately after a JPM trade. Panel A also shows that JPM events are absolutely not uniform and i.i.d. throughout the day.[22]

Panel B, which considers only Carteret trades and Carteret events in JPM, provides additional insight into how events are clustered around trades. At 30 μs after a trade, there is a large spike in the expected number of events on the same exchange. That spike is extremely pronounced, lasts for about 30 to 50 μs, and then steeply declines in the 50 μs that follow. We interpret this spike as a *response* to the trade, where market participants observe the trade in Carteret using direct-feed data, which causes them to update quotes and submit marketable orders. The timing of the spike in activity is particularly suggestive that it is a response: 30 μs is about the time it would take for a participant to observe the trade, respond, and have the response recorded by an exchange in the same data center.[23]

The 30 μs response time is similar to what Aquilina et al. (2022) observe in their data when determining their "information horizon," which is the interval of time where messages in the

[22]A uniform and i.i.d. data-generating process would appear as bars around the trade that are of the same height. Moreover, the height would be extremely low – there simply are not enough events to spread out in all 500 μs intervals, much less 10 μs bins.

[23]The processing time for market participants to make a trading decision and submit a request to the exchange is, depending on the complexity of the algorithm, much faster than 30 μs. However, there are additional latencies that affect the gap between the trade and the responding event. This could include latency from exchange servers (gateways), network latencies associated with distributing and receiving direct-feed data, and processing by the matching engine.

horizon could not have been a response to the initial trade. Aquilina et al. (2022) define this horizon because they want to examine messages that are responding to a common public signal, not each other. We are more interested in finding responses to a trade: if the SIP reports responses to the trade before the trade itself, these errors represent a meaningful look-ahead bias. Note that the exact response time is likely to vary between exchanges because of differences in how the direct feed is broadcast, gateway differences, etc.

In Panel C, we show the Mahwah JPM events around a JPM trade in Carteret. As a reminder, we show all events around the trade using exchange time. If a market participant is responding to a JPM trade in Carteret, it will take a minimum of 183 μs based solely on the speed-of-light travel time from the Carteret data center to the Mahwah data center. Real world latencies will be much higher due to frictions in transmission and other latencies described above. It is still possible that events in Mahwah can occur less than 183 μs after a Carteret trade if one or more traders act on a common signal and send messages to multiple exchanges, possibly from a third location.

Panel C shows that there are indeed Mahwah events leading up to the Carteret trade and a very slight uptick in events in the 20 to 200 μs following the trade. But, most notably, there is a sharp spike in Mahwah events about 230 μs after the trade. This spike is similar to the sharp jump in Panel B and occurs around the time it takes for the lowest-latency market participant to respond to a Carteret trade in Mahwah.[24] We also find that these patterns exist for much less actively traded stocks.[25]

Figure 4 shows that, not only are events clustered around trades, but there is a "shape" to clustering. That is, there seem to be more events after a trade than before because at least some of them are likely responses to the trade. To measure clustering in our entire sample of symbols, we examine a 500 μs window before and after every trade and count the probability that a trade has at least one other event in the 500 μs before and after the trade.

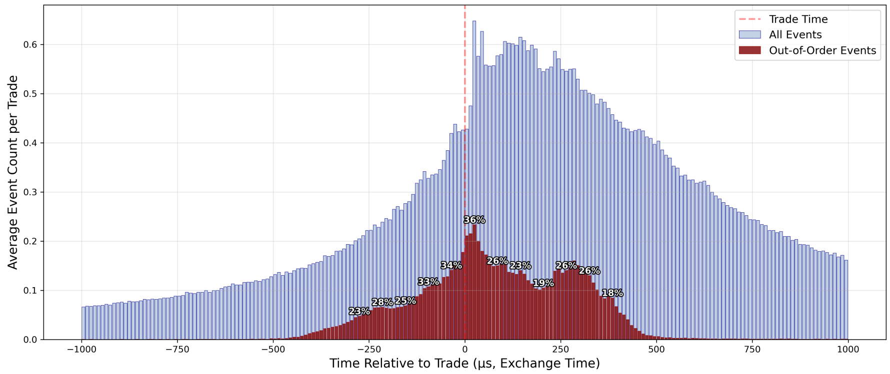
We plot both the probability of observing at least one event around a trade in Figure 5 computed for each symbol in our data. The left and right plots show the probability in the 500 μs before and after the trade, respectively. The probabilities are computed over all trades in a given symbol, and the symbols are plotted with the probability on the y-axis and the number of trades on the x-axis. The plots also show the probabilities if all events are distributed uniformly and i.i.d. throughout the trading day (in gray).

[24]The fastest connection from Carteret to Mahwah is about 190 μs. We saw that the Carteret to Carteret response is around 30 μs, which puts the fastest Carteret-to-Mahwah response at around 220 μs.

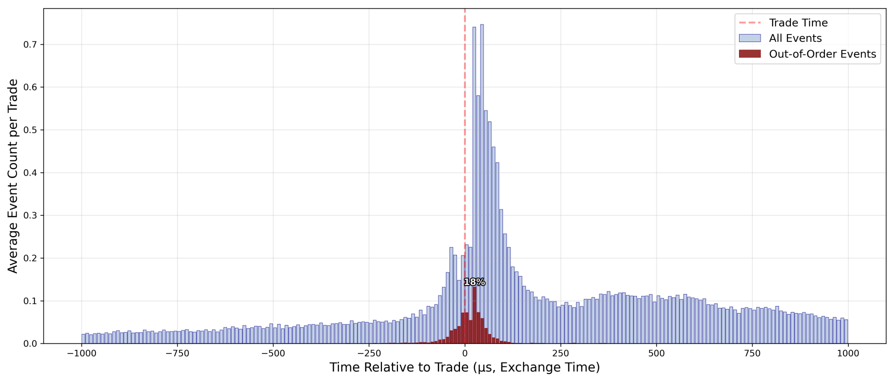
[25]See Figure C.2 in Appendix C.3 for an additional example with Equifax (EFX).

Figure 4: Clustering and Out-of-Order Events Around a Trade: JP Morgan (JPM) Example

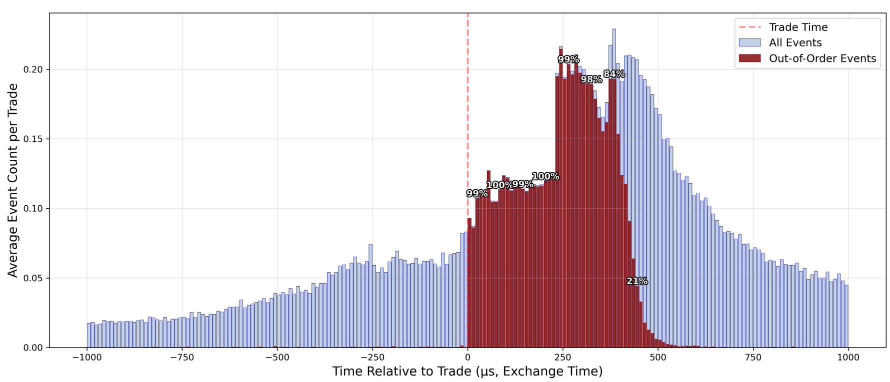
Panel A: All Trades and Events



Panel B: Trades and Events on Carteret Exchanges

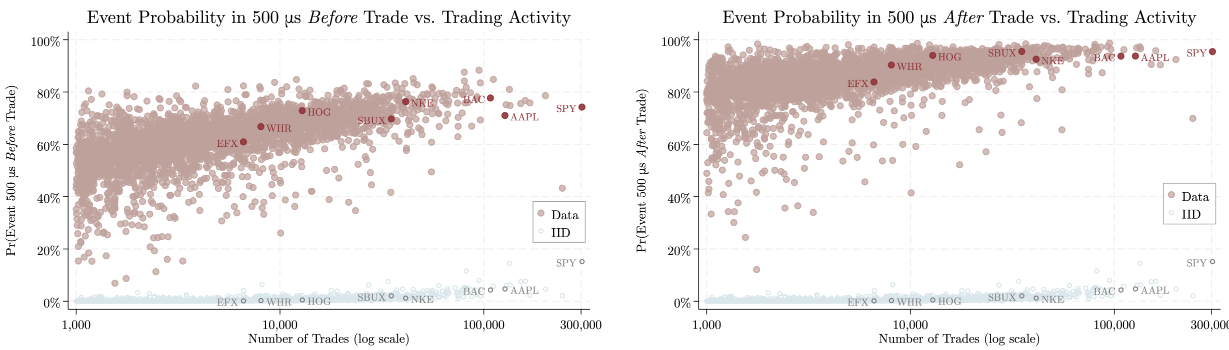


Panel C: Trades on Carteret Exchanges and Events on Mahwah Exchanges



Notes. We plot the expected number of events (quote changes or other trades) in 10 microsecond (μs) intervals around each trade of JPM on June 20, 2019. We use exchange timestamps to assign events to bins (excluding events in the same nanosecond as the trade). The blue bars show the expected event count per interval averaging over all trades in the day. The red bars show the expected number of events that are out of order. An event is counted as out of order if it occurs after (before) the trade in exchange time but reported before (after) the trade on the SIP. Panel A uses all trades and events. Panel B restricts both trades and events to Carteret exchanges. Panel C uses Carteret trades but only counts events on Mahwah exchanges.

Figure 5: Event Clustering within 500 μs of a Trade: Probabilities



Notes. This figure computes the probability of at least one event (quote change or trade) occurring within 500 microseconds (μs) before (left panel) or after (right panel) a given trade for each symbol on June 20, 2019. We plot the probability on the y-axis and the number of trades on the x-axis. The gray dots show the probability if all events were distributed uniformly and i.i.d. throughout the trading day. We highlight several example symbols that span a range of trading activity: Equifax (EFX), Whirlpool (WHR), Harley Davidson (HOG), Starbucks (SBUX), Nike (NKE), Bank of America (BAC), Apple (AAPL), and the SPDR S&P 500 ETF (SPY). See Table C.3 for specific numbers on the example symbols.

The figure shows that there are a significant number of events before and after the trade, and the probabilities far exceed what would be expected in an i.i.d. world. The pattern holds for *every* symbol in our data, and even symbols with very little trading volume (e.g., 2,000 trades in a day) have pronounced clustering. There is also a greater degree of clustering after a trade than before. This is consistent with the JPM example above, where the trade might provide some information to market participants, and they adjust quotes or submit marketable orders shortly after the trade.

The figure also highlights eight example symbols in bold. These symbols represent a wide range of symbols with very different trading activity, and span from small-cap stocks like Equifax (EFX) and Whirlpool (WHR) to the SPDR S&P 500 ETF (SPY), which is the most active symbol in the market. We highlight each of these symbols in Figure 5, and provide more detailed numbers in Appendix C.2 and Table C.3. For example, the number of Whirlpool trades/quotes is just under 2% of the number of SPY trades/quotes. Despite the enormous gap in activity, the probability of an event within 500 μs before and after a trade is 67% and 90%, respectively, which is similar to SPY's numbers of 74% and 95%.

Figure C.1 in the appendix shows the expected number of other events in the 500 μs window before and after the trade instead of probabilities. We find a very similar pattern, where even symbols with very little activity typically have around 5 and 7 events in the 500 μs before and after the trade, respectively. See Appendix C.2 for more details.

4.3 Out-of-Order SIP Events

The clustering of events in the 500 μs around a trade, combined with latencies that range from 15 to 550 μs, points to the possibility that events may often be reported out of order on the SIP. For example, consider a trade, which is followed by a quote change in response to that trade on a different exchange 300 μs later. If the trade has a high exchange-to-SIP latency of 500 μs and the quote change has a 100 μs latency to the SIP, the SIP will publish the quote change before the trade. If the SIP mixes up the sequence of the trades and quote changes in this fashion, the SIP will incorporate some of the price impact from the trade into the prevailing SIP NBBO before the trade and erroneously move the NBBO midpoint in the direction of the trade price. That is, the SIP NBBO will have a look-ahead bias.

Out-of-Order Examples

We return to the clustering JPM example in Section 4.2 but now focus on the fraction of clustered events that are out of order on the SIP (i.e., sequentially inaccurate relative to exchange timestamps). Events are considered out of order if an event occurs after the trade in exchange time but is reported *before* the trade using SIP timestamps. Similarly, if an event occurs before the trade based on exchange timestamps but is reported after the trade using SIP timestamps, it will also be marked as out of order. We count the expected number of out-of-order events and plot them as red bars (overlaying the blue bars) as a function of exchange timestamps in Figure 4.

The easiest way to read the out-of-order bars is as follows: a red bar to the right of the trade represents events that took place strictly after the trade according to exchange timestamps, but the SIP reported them at some time strictly before the trade. Conversely, if a red bar is before the trade, those events took place before the trade according to exchange timestamps, but the SIP reports them after the trade. The maximum height of a red bar is the blue bar it overlaps (i.e., the maximum expected out-of-order events is equal to the total expected events). We place the bars on the x-axis using exchange timestamps for two reasons: (1) exchange times give a sense of the actual temporal sequence of events, and (2) plotting by SIP timestamps is less important because it is irrelevant "how much" the SIP timestamps are missequenced by, just that the event did or did not affect the prevailing SIP NBBO.

The plot in Panel A of Figure 4 shows that around 25% of clustered events are reported out of order. The expected number of events is slightly higher after the trade than before the trade, but there are out-of-order events both before and after the trade. The out-of-order events start around 500 μs before the trade and end around 500 μs after the trade. There are two types of errors that might arise. First, there is a *look-ahead bias*: events that took place, say, 300 μs after the trade (in exchange time) are sometimes reported before the trade on the SIP. The other error is where the SIP is *stale*: events that took place, say, 300 μs before a trade are reported after the trade by the SIP.

Figure 4 Panels B and C produce the same out-of-order plots as Panel A, except they use the subset of *Carteret trades* and count the expected number of out-of-order *Carteret events* (Panel B) and out-of-order *Mahwah events* (Panel C). Panel B shows that of the spike in Carteret events, which are likely a response to a Carteret trade, relatively few of these events are out of order. Of the events in the initial spike about 30 μs after the Carteret trade, 18% of them are out of order. As a reminder, JPM is an NYSE-listed stock, which means trades *and* top-of-book quote updates in Carteret must travel to the NYSE SIP in Mahwah, where they are then reported by the SIP feed. Panel B suggests that the majority of Carteret events around a Carteret trade are reported by the SIP in the correct order (at least with respect to the trade as a reference point).

Panel C shows the Mahwah events that occur around a Carteret trade – this is repeated from Figure 2 in the introduction. Panel C shows a spike in Mahwah events around 230 μs that are seemingly a response to the Carteret trade, and illustrates that nearly *all* of the Mahwah events that occur within 350 μs after the trade are reported before the trade on the SIP. As described above, this is characteristic of a potentially serious look-ahead bias – to the extent that these Mahwah events move the SIP NBBO, some of the price impact *from the trade itself* will be incorporated into the SIP NBBO midpoint ex-ante, which will then be used for trade signing and spread measurement.[26]

Out-of-Order Event Probabilities

Figure 4 highlights a potentially serious out-of-order problem that can lead to a look-ahead bias and a stale NBBO. We now examine all trades for every symbol in our one-day sample. For each trade, we count the number of events in the same symbol that occurred on any exchange after the trade according to exchange timestamps but before the trade using SIP timestamps. That is, the SIP reports the event as happening before the trade, but it actually occurred after. We label these events as out of order before the trade ("OoO Before"). These

[26]As stated for the clustering results, the out-of-order results hold for stocks with less trading activity. We provide an example with Equifax (EFX) – see Figure C.2 in Appendix C.3. We also provide a slightly modified version of the JPM figure where we only examine trades that were the first in a potential sequence of trades – see Figure C.3 in Appendix C.3 and supporting text for details.

OoO Before events correspond to the red bars to the right of the trade in Figure 4. We do the same for events that come before a trade by exchange time but after the trade with SIP timestamps ("OoO After"). The OoO After events correspond to the red bars to the left of the trade in Figure 4.[27] Note that OoO Before events are logically consistent with the look-ahead bias described above, and OoO After events align with a stale SIP NBBO.

We report the probability that a trade has at least one OoO Before and OoO After event for each symbol in Figure 6. The figure shows that every symbol has a positive out-of-order probability, both before and after. The figure also shows that the OoO Before issue (left plot) is much more prevalent than the OoO After issue (right plot). Most symbols have an OoO Before probability between 15% and 80%, compared to a range of 10% to 40% for OoO After. The implication is that the out-of-order issue consistent with a look-ahead bias is much more prevalent. In addition, there is a slight increase in out-of-order probabilities as the number of trades increases.

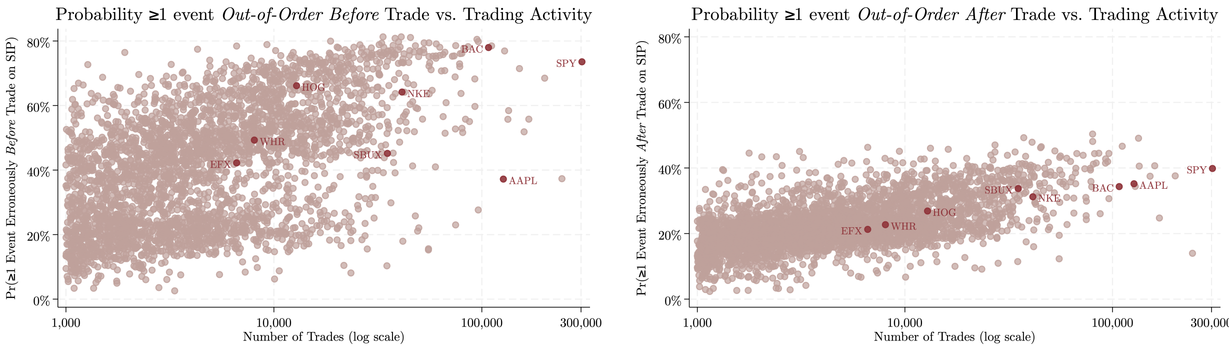
We provide additional out-of-order event details in Appendix C.3. Specifically, we present Figure C.4, which is similar to Figure 6 above but reports the average number of events that are out of order instead of the probability. We also provide summary statistics and exact numbers for several example symbols in Table C.4.

One observation from Figure 6 is that there is a noticeable cluster of symbols on the left plot that have relatively low OoO Before probabilities (between 10% and 25%). We can report that these are mostly Nasdaq-listed symbols. This is significant because these symbols all report events to the Nasdaq SIP, which operates with lower latency than the NYSE SIP. Figure 4 also suggests that out-of-order events seem to be associated with exchange-to-SIP latency. As a result, we present out-of-order statistics for trades grouped by exchange and SIP. This will help group trades based on geographic latency and SIP processing latency to better understand how out-of-order effects and latency are related.

Table 2 reports the out-of-order probabilities and expected events. We provide overall numbers for all trades reported by each SIP, but we also show the statistics for each of the nine highest-volume exchanges that account for more than 95% of volume. We include the total number of trades for context on how much activity is on each exchange. Note that we separate the trades by the location where they occurred but, for each trade, we use all events on all exchanges to compute the out-of-order numbers.

[27] One useful way to remember what OoO Before represents is to think of an event that the SIP reports *before* the trade but *should not have*. Similarly, a convention for OoO After is an event that the SIP reports after the trade but it should not have.

Figure 6: Out-of-Order Events within 500 μs of a Trade: Probabilities



Notes. This figure computes the probability of at least one event (quote change or trade) that is within 500 microseconds (μs) of a trade according to exchange timestamps but is reported out of order using SIP timestamps for each symbol on June 20, 2019. The left (right) plot shows the probability of an event erroneously reported by the SIP as occurring before (after) the trade. We plot the probability on the y-axis and the number of trades on the x-axis. We highlight several example symbols that span a range of trading activity: Equifax (EFX), Whirlpool (WHR), Harley Davidson (HOG), Starbucks (SBUX), Nike (NKE), Bank of America (BAC), Apple (AAPL), and the SPDR S&P 500 ETF (SPY). See Table C.4 for specific numbers on the example symbols.

Panel A provides the probabilities and expected counts for symbols that report to the NYSE SIP. Panel A shows interesting heterogeneity by exchange. The exchange with the most trade activity for NYSE symbols is Nasdaq, and it also has the most OoO Before activity: 85.8% of all Nasdaq trades have at least one event that the SIP reports before the trade but actually occurred after. The expected number of OoO Before events is 11.68. Just as observed in the JPM example in Figure 4 Panel C, which represents a Nasdaq trade for a NYSE symbol, we see that the potential look-ahead bias is large in that there are many events that occur after a trade but are reported before the trade by the SIP. Panel A also shows that the Secaucus exchanges, such as BZX and EDGX, have only slightly smaller out-of-order probabilities and expected event counts. We interpret the slightly smaller numbers as attributable to the fact that these exchanges are closer to the NYSE SIP in Mahwah and have lower exchange-to-SIP latency.

The NYSE Arca and NYSE trades in Panel A provide additional insights. These trades occur in the same data center as the NYSE SIP. Because these trades have low exchange-to-SIP latency, we see that the OoO patterns are flipped: the stale quote issue from OoO After seems to be more severe. Since the median latency from NYSE Arca to the NYSE SIP is relatively small, events that occur at more distant exchanges will have longer exchange-to-SIP latencies and will be reported after the trade on the SIP, even if they occurred before according to exchange timestamps. In other words, trades that occur on exchanges in the

Table 2: Out-of-Order Events by Exchange and SIP

Panel A: NYSE SIP (Mahwah)

Exchange	Location	# Trades	Pr(≥1 OoO Event) Before	Pr(≥1 OoO Event) After	E[#Events OoO] Before	E[#Events OoO] After
All Exchanges		19,557,153	62.17%	28.81%	7.08	3.88
NASDAQ	Carteret	5,286,140	85.80%	10.38%	11.68	0.68
NASDAQ BX	Carteret	704,031	80.99%	11.32%	9.05	0.95
BZX	Secaucus	2,093,681	77.32%	26.07%	9.49	2.46
BYX	Secaucus	1,931,203	78.35%	23.49%	7.18	1.59
EDGX	Secaucus	1,504,301	80.41%	26.21%	9.52	2.53
EDGA	Secaucus	1,233,410	69.52%	18.64%	5.24	1.33
IEX	Secaucus	966,867	33.52%	11.39%	2.74	1.20
NYSE Arca	Mahwah	2,321,615	32.11%	62.36%	3.31	12.23
NYSE	Mahwah	2,792,464	17.34%	51.33%	0.31	7.67

Panel B: Nasdaq SIP (Carteret)

Exchange	Location	# Trades	Pr(≥1 OoO Event) Before	Pr(≥1 OoO Event) After	E[#Events OoO] Before	E[#Events OoO] After
All Exchanges		9,221,786	33.42%	27.96%	3.10	1.81
NASDAQ	Carteret	3,831,340	7.04%	43.84%	0.22	3.13
NASDAQ BX	Carteret	265,614	5.26%	42.90%	0.08	3.51
BZX	Secaucus	1,036,584	56.32%	22.99%	5.05	1.05
BYX	Secaucus	685,681	49.82%	22.04%	2.79	0.88
EDGX	Secaucus	717,208	62.66%	23.09%	5.95	1.04
EDGA	Secaucus	478,122	41.85%	16.12%	2.31	0.57
IEX	Secaucus	426,704	21.15%	12.70%	1.68	0.46
NYSE Arca	Mahwah	1,346,439	66.47%	0.66%	8.65	0.02
NYSE	Mahwah	196,920	68.93%	12.88%	8.80	0.42

Notes. This table presents out-of-order event statistics in the 500 microseconds (μs) before and after each trade for all symbols on June 20, 2019. An event (quote update or trade) is out of order before a trade if it occurs within 500 μs after the trade using exchange timestamps but is reported before it on the SIP. Similarly, an event is out of order after a trade if it occurs within 500 μs before the trade using exchange timestamps but is reported after it on the SIP. We exclude events in the same nanosecond as the trade. For before and after, we report the probability of at least one out-of-order event and the expected number of out-of-order events. Panel A shows the statistics by exchange for NYSE SIP symbols. Panel B does the same but for Nasdaq SIP symbols.

same location as the SIP may be more prone to the stale quote issue than the look-ahead bias issue, at least from the pure out-of-order numbers.[28]

Panel B, which shows out-of-order numbers for symbols that report to the Nasdaq SIP in

[28] We find in Sections 5.2 and 5.3 that there is still a look-ahead bias when the exchange and SIP are in the same location. This is because, while there are fewer OoO Before than OoO After events for these trades, the OoO Before events are more likely to move the SIP NBBO than OoO After events. See the conceptual exercise in Section 5.1 for details.

Carteret, shows similar patterns. The highest OoO Before percentages come from distant exchanges in Mahwah and Secaucus, which also have the largest expected OoO event counts. The most OoO After events are for Nasdaq and Nasdaq BX trades, which are in the same data center as the Nasdaq SIP. We note that the probabilities and expected events are generally smaller for Nasdaq symbols than for NYSE symbols. This is also consistent with the latency story: Table 1 shows that the NYSE SIP has a much greater SIP processing latency than the Nasdaq SIP (about 100 μs).

The general theme that emerges from Table 2 is that latency generates a sequentially inaccurate SIP. Because events tend to be clustered and some of those events have longer exchange-to-SIP latencies than others, events that occur closer to the SIP are likely to be reported by the SIP first. The underlying reason is that traders can act in multiple places around the same time or even observe a trade and respond at time horizons much shorter than many exchange-to-SIP latencies.

In Appendix C.3, we provide supporting results that show *same exchange* out-of-order statistics, which only consider events around a trade that are on the same exchange.[29] Appendix C.3 shows that there are consistently same-exchange out-of-order events. The same-exchange out-of-order probabilities and expected events are noticeably smaller than the all-events results above. We highlight two patterns. First, the same-exchange sequence issues are primarily OoO Before (i.e., the look-ahead bias). For example, 25% of NYSE Arca trades have at least one event on NYSE Arca that occurred after the trade according to Arca's own timestamps, but is reported by the NYSE SIP as happening before the trade. This is compared to only 3% OoO After. Second, the same-exchange sequence issues are concentrated on the NYSE SIP. We conjecture that the sequence issue arises from the way the NYSE SIP processes messages from exchanges, possibly prioritizing the dissemination of quotes over trades. Note that the NYSE SIP in our sample also has a much longer SIP processing latency than Nasdaq (nearly an order of magnitude larger – see Section 4.1).[30]

While the prevalence of same-exchange out-of-order issues is much lower compared to counting all events, each occurrence is likely more problematic for trade signing and spread measurement. For example, a trade for all units at the exchange's best bid or offer will *mechanically* move the top-of-book quote. Moreover, events following a trade on the same exchange are much

[29]We also provide an example from the data of this type of out-of-order error in Appendix A.2.

[30]We do not count the quote update mechanically generated by a trade (and assigned the same exchange timestamp) as an event when considering out-of-order events (see Section 3 for more details). This choice is a subtle but important difference between our same exchange results and Schwenk-Nebbe (2022), who documents sequence errors caused by "mechanical quote updates" or MQUs appearing on the SIP before the trade that mechanically created the quote update to begin with.

more likely to be a response to the just-observed activity. Like the BAC example in Appendix A.2 shows, same-exchange out-of-order events by themselves can easily flip the inferred sign of the trade or lead to a drastically smaller effective spread.

5 Latency-Induced Errors and Solutions

A sequentially out-of-order SIP feed has direct implications for trade signing and estimating effective spreads/price impact, and one or both of these practices are required in almost every application of studying trades in the TAQ data. We first conceptually show how an out-of-order SIP leads to changes in the prevailing SIP NBBO midpoint at the SIP time of the trade in Section 5.1. This midpoint is a key object precisely because the literature almost universally uses Lee-Ready with the SIP NBBO midpoint as the reference price to sign trades, and the SIP NBBO midpoint to measure effective spreads and price impact. To what extent an out-of-order SIP changes the NBBO midpoint and, in turn, affects signing and spreads is an empirical question. We document signing errors in Section 5.2 and the spread/price impact bias in Section 5.3.

5.1 How SIP Sequence Errors Affect Signing and Spreads

Our out-of-order results show it is common to have (1) at least one quote change after a trade will be reported by the SIP as occurring before (i.e., "OoO Before"), and/or (2) at least one quote change before a trade will be reported by the SIP as occurring after ("OoO After"). These changes affect the prevailing SIP NBBO midpoint just before a trade by (1) incorporating changes into the NBBO that should not have been considered, and/or (2) failing to incorporate changes that should have been considered. We provide a detailed conceptual example of each case in Appendix D.1. In this subsection, we show how the midpoint is the key object in signing and spread measurement.

The Midpoint is Key

Discrepancies between the SIP NBBO midpoint and the true NBBO/BBO midpoint generate all signing and spread differences. (We mention the BBO midpoint here alongside the NBBO because, as we explain in the next section, our main signing rule relies on the BBO midpoint of the exchange where the trade occurred.)

To see why all variation in signing and spreads comes from the midpoint, consider a generic example of a buyer-initiated trade with a trade price of P^*. The correctly-sequenced NBBO has a midpoint of M^*. The trade direction is a buy if $P^* > M^*$ and the simple effective half spread for a buy is $P^* - M^*$ and is labeled EHS_{M^*}.[31] We provide a graphical depiction of this example in Figure 7.

The figure also presents four possible examples of where the SIP NBBO midpoint might land, which can differ from the true midpoint because of SIP sequence errors. The first example shows when the SIP midpoint, M_{SIP1}, is greater than the true midpoint M^*. This can occur when price impact causes quotes to be revised upwards after the trade, but the SIP reports those quote changes before the trade (i.e., OoO Before). More generally, many quote updates can be out of order, with some OoO before (look-ahead bias) and others OoO after (stale quotes), but if the OoO before quotes have a greater impact on the SIP NBBO than the OoO after quotes, we could still see a case like this first example. In this example, if we use the SIP midpoint to sign the trade and compute the effective spread, we would label the trade as a sell ($P* < M_{SIP1}$) and the effective spread would be negative ($EHS_{SIP1} = P^* - M_{SIP1}$). Note that the effective spread requires a trade direction. Here, we have assigned the true sign in computing the spread.[32]

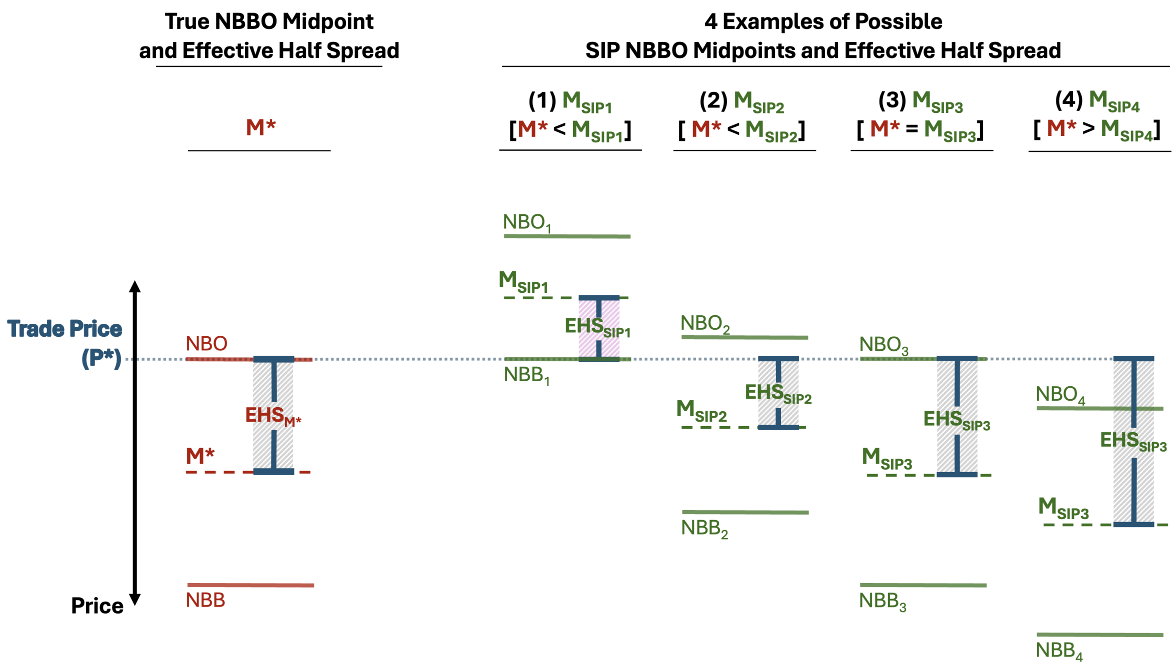
Example 2 illustrates a look-ahead-style sequence error, but is less extreme. The midpoint (M_{SIP2}) is still greater than the trade price, indicating that some of the price impact from the trade may have been incorporated by the SIP, but the price impact is small and/or not enough has been mis-sequenced to flip the sign of the trade (i.e., P^* is still greater than M_{SIP2}). Still, this example shows that the effective spread will be biased down. In the third example, the SIP midpoint is equal to M^*. This does not necessarily mean that the SIP was sequentially accurate, since out-of-order messages can change the NBB and/or NBO size without moving the actual NBBO midpoint. In these cases, the SIP will be equivalent to the true midpoint from a signing and spread perspective.[33]

[31] As is common in the literature, we will normalize by the midpoint to compare across symbols when actually going to the data. We describe the exact calculation in Section 5.3.

[32] If we were to (erroneously) infer that the trade is seller initiated, we would compute the midpoint minus the trade price instead of the trade price minus the midpoint.

[33] Hagströmer (2021) proposes a weighted midpoint, with the SIP NBB and NBO quantities as weights, as a better estimate of the effective spread. Our paper shows that the effective spread is biased because of SIP sequence errors, which move SIP *prices* towards the trade price. This critique also applies to NBBO *quantities* as much as it does to prices. There could be mis-sequenced events on the SIP that lead to a look-ahead bias, but those errors may affect quantities without affecting prices (as shown in Example 3), which would affect the weighted midpoint but not the simple, equal-weighted midpoint. That implies that the weighted midpoint would be even more sensitive to out-of-order SIP errors: spreads computed with the weighted midpoint will be more biased than spreads using a simple midpoint.

Figure 7: Possible SIP Midpoint Discrepancies



Notes. This figure compares the prevailing true NBBO midpoint (M^*), when a trade occurs at price P^*, with four possible SIP midpoints ($M_{SIP1} - M_{SIP4}$) that can arise from an out-of-order SIP. For simplicity, effective half spreads are shown as the difference between the trade price and midpoint (fixing the direction as a buy) and are labeled EHS. In Examples 1 and 2, the SIP midpoint has moved toward the trade price and has incorporated quote changes from after the trade, reducing the effective half spread (or even producing a negative spread – see red shaded area). Example 3 shows a case where the SIP midpoint equals M^* despite out-of-order events (e.g., events that only affect quantities). Example 4 provides a case where the SIP has not received updates from some exchange and shows a stale NBBO.

The last example shows what might occur if the SIP is stale. That is, if there was some price impact before the trade, causing quotes to increase leading up to a trade, but the SIP does not observe those quote changes, the SIP midpoint might be stale. The fourth example shows that this can cause the true midpoint to be greater than the SIP midpoint for a buyer-initiated trade. This would still lead to accurate signing, but the SIP effective half spread (EHS_{SIP4}) will be larger than the true effective spread.

We highlight several important elements from the cases above. First, the cases are not necessarily direct mappings from the SIP sequence errors. For example, it is possible that there is a trade at the ask, and then quotes are immediately revised downwards after the trade. That being said, we think that look-ahead errors (OoO Before) are logically most likely to lead to the cases in examples 1 and 2, and stale errors (OoO After) lead to example 4. Even if there are more stale errors or look-ahead errors for any given trade, what really matters is how much these types of errors ultimately move the SIP NBBO and, thus, the midpoint. In that sense, we will simply let the data tell us how often SIP out-of-order issues flip the sign, how much they bias the spread, and in what direction.

Second, note that only the most extreme midpoint differences lead to signing errors. In the examples above, only Example 1 causes the SIP NBBO midpoint to flip the true direction of the trade. However, there can be distortions in the spread without necessarily flipping the sign. That is, signing errors are indicative of *severe* out-of-order problems, and less severe out-of-order issues can still lead to a bias in the spread.

Third, signing errors are much more likely to occur when the SIP NBBO moves in the direction of the trade, consistent with a look-ahead bias. A stale SIP will often preserve the trade direction. This point dates as far back as Lee and Ready (1991) – staleness was a feature, not a bug, to ensure signing was done with a midpoint from before the trade.

Lastly, *all* of the variation in signing and spread issues comes from variation in the SIP NBBO midpoint. All signing and effective spread measures use only two variables: the trade price and the midpoint. Since the SIP always reports the trade price accurately, signing and spread differences *must* come from the midpoint.[34]

This logic holds for a measure like price impact, which is computed just like the effective half spread, except that the trade price is replaced with the midpoint at some time in the future (e.g., 1 minute). Even though price impact is effectively computed using the difference between two midpoints (the future midpoint and the prevailing midpoint), the future midpoint is *effectively* fixed. The reason for this is that, for every symbol, if we examine a random point in time, there is likely no activity, as documented in Section 4.2. That is, in most 500 μs intervals in every symbol, nothing happens (much less a single point in time, i.e., a microsecond or nanosecond). So, while the prevailing SIP midpoint is likely to have sequence issues because of the clustering of events around a trade, examining the SIP NBBO at a point 60,000,000 microseconds (1 minute) after the trade will almost surely land in an empty interval where nothing is happening in the several hundred microseconds around it. In this sense, sequence errors are extremely unlikely to affect the future midpoint used to compute price impact.

As a result, effective spreads and price impact are similar in that all variation in price impact comes from the prevailing SIP midpoint relative to the prevailing latency-free midpoint. The only difference is that price impact will change P^* in Figure 7 from the trade price to the

[34]As mentioned above, computing spreads requires a trade sign. This is why we attempt to hold signing fixed when assessing spreads in Section 5.3.

midpoint price at some point in the future. But, as described above, this future price will almost always be the same for the true NBBO midpoint and the SIP NBBO midpoint, leaving only variation in the prevailing SIP and true midpoint before the trade.

5.2 Trade Signing

The first and most basic step in utilizing trades in the TAQ data is assigning a trade direction. The SIP only reports the trade price and quantity, not the direction. We introduce a new signing methodology and compare it to the Lee-Ready SIP NBBO methodology in Section 5.2.1 for NYSE Arca trades. We focus on Arca because we can compute the true signing accuracy of each methodology using the direct-feed data. In Section 5.2.2, we expand our signing to all trades on all exchanges. While we only have direct-feed data for NYSE Arca, we use a simple proxy for accuracy that can be utilized for trades on all exchanges.

5.2.1 Signing Methodology and Arca Trade Signing

We propose a new latency-free (LF) signing methodology. The foundation of the methodology uses the midpoint of the prevailing best bid and offer of the exchange on which the trade occurred, according to exchange timestamps, to sign trades (i.e., the exchange best bid-offer or EX BBO). This methodology leverages the rules of the limit order book and, in theory, should have 100% accuracy as long as limit orders are displayed. For example, if an exchange is 10.01 bid and 10.05 offered, all marketable buy orders *must* trade at 10.05 or higher, and all marketable sell orders *must* trade at 10.01 or lower. Thus, using the prevailing exchange BBO midpoint must sign trades accurately if the underlying limit order book data are reported accurately.[35] This approach is latency free in that it is not susceptible to the latency in transmission from the exchange to the SIP.

We find that the TAQ data with exchange timestamps are highly accurate. While the direct-feed data are more comprehensive and delivered in a more timely fashion from the exchange itself, the data that TAQ reports exactly matches what we see in the direct-feed data and, importantly, has identical exchange-assigned timestamps. So, timestamp and data accuracy are not issues in the TAQ data. Still, the EX BBO rule cannot achieve 100% accuracy because some outstanding limit orders are not visible in the TAQ data or are displayed at an inferior price to where they are actually working.

[35]We use exchange timestamps for exactly this reason: it preserves the sequence of events on an exchange because the timestamps come from the matching engine itself and the limit order book is a serial process.

There are three specific scenarios that lead to unobserved limit orders in the TAQ order book: (1) odd lot limit orders, or orders for less than a round lot of 100 shares, (2) "price-slide" orders, which are displayed at one price but exist in the actual limit order book at a different price, and (3) hidden or non-displayed limit orders. We briefly describe each case below, and then we introduce our complete LF signing methodology, which includes a supplementary signing rule to handle one of these special cases.

Odd Lot Trades To understand how odd lots present a challenge to the EX BBO method, consider a continuation of the previous example. A symbol on a particular exchange has a BBO of 10.01 and 10.05 with 100 shares at each best quote. Just as before, if a marketable buy order were to arrive for 100 shares, it must trade at 10.05 and can be accurately signed using the EX BBO midpoint. Now, imagine a trader submits a sell limit order of 40 shares at a limit price of 10.02. This changes the actual EX BBO to 100 at 10.01 and 40 at 10.02. However, since the SIP does not report any odd lot limit orders, it will appear that the EX BBO is unchanged with 100 at 10.01 and 100 at 10.05. If a market buy order for 100 shares were to arrive at this point, an odd lot trade would occur, with 40 shares purchased at 10.02 and 60 shares bought at 10.05. If we use the EX BBO method with TAQ data (even with exchange timestamps), the prevailing EX BBO would be 10.01/10.05 with a midpoint at 10.03, which would sign the odd lot trade at 10.02 as a sell. If we evaluate the trade at 10.02 with the actual BBO of 10.01/10.02, the EX BBO method correctly signs the trade as a buy.

As this example shows, the only reason why trades against displayed odd lots might be incorrectly signed is because they interact with odd lot limit orders that are *effectively* non-display on the SIP. Because odd lot trades result from interacting with odd lot limit orders not observable on TAQ, we introduce a modified signing rule below, which supplements the EX BBO approach with a rule specific to odd lot trades with a trade price inside the EX BBO. If the odd lot limit orders are displayed, they are incorporated into the top-of-book quotes, and the EX BBO signing rule should work with 100% accuracy for trades against all limit orders. This is exactly the change that the SIP will implement starting in May 2026.[36]

Price Slide Trades The second exception that leads to systematic signing errors is "price slide" trades. These trades appear to behave contrary to the rules of the limit order book. For example, consider a symbol on NYSE Arca with a prevailing EX BBO of 10.01 and 10.03 and a round lot trade occurs at 10.02. Using the direct-feed data, we can see additional

[36]From that point, the EX BBO rule will end up signing a much larger fraction of trades because it will weakly narrow the spread, leaving fewer prices inside the EX BBO where trades can occur and would be signed with our supplemental rule. We discuss this SIP development further below.

details on the trade, including the direction of the marketable order that initiated the trade and the order ID of the limit order that participated in the trade (i.e., the resting contra-side order). When we look at the direct-feed data, we see the trade was seller initiated, and the limit order that took the other side of the trade was a buy order with a limit price at 10.01. However, this is impossible given the rules of the limit order book – a seller-initiated trade at 10.02 cannot be matched with a buy limit order with a limit price of 10.01.

The answer lies in the way exchanges operate and adjust to comply with Reg NMS. If there is another exchange, say BZX, that NYSE Arca perceives as having the best offer at 10.02, NYSE Arca cannot allow a bid at 10.02 on its own exchange because that order would create a locked NBBO (10.02 bid on Arca and 10.02 ask on BZX) – a violation of Rule 610 in Reg NMS. Exchanges offer an order type (which can sometimes be the default order type) where the exchange will accept the bid and it will work the order at 10.02 but *display* it at 10.01 to comply with Reg NMS. If a trader submits a marketable order to sell, it will trade at 10.02, not 10.01, despite the exchange showing the best bid at 10.01.[37,38]

We do not apply a different methodology to these trades. The primary reason is that direct-feed data is required to identify these price slide trades. We point out these trades because they help understand why our EX BBO method might not achieve 100% accuracy.

Hidden Order Trades Some trades result from interacting with hidden or non-displayed limit orders. As described in Section 3, there is no way to know the sign of these trades in the direct-feed data. As a result, they are not included in this analysis. However, we provide an analysis and discussion of these trades separately in Appendix B.1. In short, we argue that a non-trivial fraction of these trades is likely to be signed more accurately with our signing methodology than with the SIP approach. Still, some of these trades (i.e., midpoint trades) are likely to be difficult to sign with any signing rule.

Latency-Free Signing Methodology

We now provide the exact details of our signing methodology, which we name the Latency-Free (LF) methodology. The LF methodology combines two separate signing rules, both using exchange timestamps (and are therefore not exposed to latency).

The main rule is the EX BBO rule. For all round lot and mixed lot trades ($\geq$100 shares), we

[37] As an example of one type of price slide order, see the "OUCH - Price Slide" order description in the appendix of the Nasdaq Port Request Form.

[38] See Section 2.2 for more details on Reg NMS.

use the prevailing EX BBO midpoint according to exchange timestamps to sign trades. In addition, all odd lot trades (<100 shares) with a trade price at the prevailing EX BB or BO, we also use the prevailing EX BBO midpoint. Specifically, for these trades, (1) if the trade price is strictly greater than the EX BBO midpoint, the trade is marked as buyer initiated; (2) if the trade price is strictly less than the EX BBO midpoint, the trade is marked as seller initiated; and (3) if the trade price is equal to the EX BBO midpoint, we use the tick test of Lee-Ready.[39]

The supplementary rule uses the 1 ms delayed latency-free NBBO, or delayed LF NBBO, to sign all odd lot trades with a trade price inside the EX BBO. The delayed NBBO is constructed by chronologically walking through each update using exchange timestamps and keeping track of the best bids and best offers across all exchanges for each symbol (see the end of Section 3). This construction is similar to the instantaneous NBBO in Bartlett and McCrary (2019). Once we have this LF NBBO, we then adjust it by 1 ms (1,000 μs) to make it delayed or "stale." The midpoint of this delayed LF NBBO is then used to sign trades just as above: (1) buy if the trade price is above the delayed LF NBBO midpoint, (2) sell if the trade price is below this midpoint, and (3) use the tick test if equal to this midpoint.

To be clear, the delayed NBBO rule for odd lots is based on examining empirical patterns in the data. It is not motivated by limit order book rules but is necessitated by the fact that odd lot trades often result from odd lot limit orders not reported to the SIP. The reason the delayed LF NBBO rule works reasonably well for odd lot trades is due to a particular scenario that usually leads to signing errors. As an example of this scenario, imagine that an odd lot sell limit order at 10.02 is placed when the BBO is 9.99/10.03. That is, the odd lot limit order is closer to the best ask. If the market were to move down – say, the BBO becomes 9.95/10.00 – the odd lot limit order is unlikely to trade. If the market were to move up – the BBO becomes 10.01/10.05 – the odd lot order is much more likely to trade (odd lots are visible in feeds sold by the exchange). Most signing rules incorrectly sign trades that interact with these "stale" odd lot orders, i.e., orders that do not move as the market moves. The delayed LF NBBO is much better at signing these trades because it uses an intentionally stale LF NBBO.[40]

Section 4.2 shows that event activity is clustered before and after trades. If there were price movements that preceded a trade involving a non-displayed, stale order, it makes sense

[39] We use the sequence of trades on any exchange according to exchange timestamps whenever we must rely on the tick test.

[40] While this application of stale quotes is novel, the idea of using stale quotes dates back to Lee and Ready (1991), where delayed quotes were used to, ironically, ensure no future information was used to sign trades.

to go back to a point before the cluster of activity that preceded the trade. Given that the clustering is concentrated within 500 μs around a trade, 1 ms (1,000 μs) is a sensible increment to avoid the cluster related to the trade. Going back before the cluster of activity is also likely to lead to an interval where there is no activity. That is, the 1 ms stale LF NBBO is from the last point of "stability" in the NBBO.[41]

As noted above, the LF signing methodology is likely to significantly improve in May 2026, when the SIP feeds will include all odd lot limit orders that might be at the best bid and offer. After this point, the EX BBO method will be used for a larger fraction of trades, and those trades will be signed with 100% accuracy (price slide trades notwithstanding).

While we do not attempt to sign off-exchange trades, we suggest using the delayed LF NBBO rule to sign these trades. The reason is that there is often a delay from when the trade is executed to the time the trade is reported to the SIP, much like the delay from when an exchange reports to the SIP for lit trades. That means using prevailing quotes based on an off-exchange trade's SIP timestamp to sign trades will be using quotes from *after* the trade, conceptually identical to the look-ahead bias we document in this paper. Our delayed LF NBBO rule addresses this concern because it uses the quotes prevailing 1 ms before the trade according to participant-assigned timestamps.[42]

We discuss off-exchange trades further in Appendix G.1, where we also summarize the LF methodology. In addition, we note that Battalio et al. (2026) use the LF methodology to sign retail trades from one or more proprietary wholesalers where the true direction of the trade is known. They find that the LF methodology improves signing accuracy over using the SIP NBBO by more than 7 pp, and is only marginally worse than using the NBBO as constructed directly by the wholesaler(s) using direct-feed data sources. See Section III.B, Figure 3, and Footnote 11 in Battalio et al. (2026) for more details.

Trade Signing Accuracy

We sign every NYSE Arca trade for every symbol in our sample in June 2019. We measure accuracy for the LF and SIP methods by counting the percentage of trades that match the sign given by the NYSE Arca direct-feed data. We also measure accuracy in dollar

[41] Alternative delays of 500 μs or 5 ms work similarly well for signing. In general, the accuracy of the delayed LF NBBO rule marginally increases in the size of the delay but levels off after around 1 ms.

[42] Using TAQ/SIP timestamps for these trades is particularly problematic because they have extremely long latency from the execution venue to a trade-reporting facility and then to the SIP. The median latency, inferred from the difference between SIP time and participant time for all off-exchange trades on June 20, 2019, is 2,375 μs with an interquartile range of more than 5,900 μs. These numbers dwarf the exchange-to-SIP latencies we see, with a high end of around 550 μs.

volume terms (relevant for economic activity). We report overall accuracy numbers and also decompose the numbers into different subsamples. The decomposition first splits the sample based on round lot trades, odd lot trades at the prevailing EX BBO, and odd lot trades inside the EX BBO. Within these three groups, we further decompose trades based on whether they are price slide trades. Table 3 presents the results.

We show that the LF method has an almost 8 pp improvement in signing accuracy over the ubiquitous Lee-Ready SIP method. In terms of overall accuracy in trade terms, our LF approach signs trades with 95% accuracy. While the accuracy number is reasonably high, it would be even greater if not for the two exceptions we described above: odd lot and price slide trades. In trade terms, round lots and odd lots at the EX BBO are 100% accurate with the EX BBO rule.[43] Our LF method is also significantly better than the SIP for odd lot trades inside the exchange BBO, with a gap of about 15.2% for these trades (when not price slide trades). The LF method has relatively poor accuracy for all three price slide groups, though the SIP is worse in every case.

Table 3 also shows that trade-level aggregation implicitly increases the weight of odd lot trades in the overall accuracy numbers. In trade terms, there are about as many odd lot trades as round lot trades, roughly consistent with O'Hara et al. (2014). However, odd lots account for only around 16% of dollar volume, though the accuracy numbers are similar when aggregating at the trade or dollar level. The fact that odd lots will be included in the SIP feed and in TAQ starting in May 2026 further emphasizes the importance of the EX BBO rule that is used to sign round lots and odd lots at the exchange BBO. In fact, we estimate that all non-price slide trade groups in Table 3 will be signed with 100% accuracy once the SIP feed includes odd lot limit orders. That would move the LF accuracy rate for trades and dollars to over 97%.[44]

[43]The EX BBO accuracy for round lots is not literally 100%. In trade terms, the accuracy is 99.9983%. These rare errors are generated by trades with non-displayed, round lot limit orders that are still included in our direct-feed data. The accuracy number for non-price slide odd lot trades at the EX BBO is also not literally 100% – it is 99.9998%.

[44]We also note that our LF signing methodology outperforms methodologies that apply a latency adjustment, like those proposed in Holden et al. (2023). For the Arca sample where we can definitively measure signing accuracy, the LTA method in Holden et al. (2023) achieves an accuracy of 92.3%, compared to the 95.3% we find above. For detailed comparisons on overall accuracy between other methods, see Wu and Pierson (2025).

Table 3: NYSE Arca Signing Accuracy

	Trades				Dollars			
	LF	SIP	Diff.	*% Obs.*	LF	SIP	Diff.	*% Obs.*
Overall Accuracy	95.3%	87.4%	7.9%		96.6%	89.3%	7.3%	
Decomposition:								
Round Lots	100.0%	94.4%	5.6%	*42.3%*	100.0%	93.6%	6.4%	*71.7%*
Round Lots (Price Slide)	71.7%	55.5%	16.2%	*7.2%*	79.6%	67.5%	12.1%	*12.1%*
Odd Lots (At EX BBO)	100.0%	96.3%	3.7%	*26.1%*	100.0%	95.9%	4.1%	*8.7%*
Odd Lots (At EX BBO + Price Slide)	79.6%	74.5%	5.1%	*2.1%*	90.7%	85.1%	5.7%	*0.5%*
Odd Lots (Inside EX BBO)	91.8%	76.6%	15.2%	*21.5%*	88.8%	76.2%	12.7%	*6.6%*
Odd Lots (Inside EX BBO + Price Slide)	40.6%	36.7%	3.9%	*0.8%*	50.6%	47.4%	3.3%	*0.3%*

Notes. This table presents the signing accuracy for our Latency-Free ("LF") signing methodology and the Lee-Ready methodology using the SIP NBBO midpoint ("SIP"), as well as the accuracy difference between the two approaches ("Diff."). We provide the accuracy percentages separately at the trade and dollar level. We split the sample of trades by round lot, odd lots at the EX BBO, and odd lots inside the EX BBO. Within each of these three groups, we split on whether trades are or are not "price slide" trades, or trades with orders that are working at one price but displayed at an inferior price. We also report the percent of observations in each sample subset.

5.2.2 Signing Non-Arca Trades

We apply our LF methodology to the full sample of trades across all exchanges. Since we only have direct-feed data for NYSE Arca, we cannot estimate the true accuracy of the LF and SIP methods across all exchanges. We introduce an alternative measure – the percentage of trades assigned different signs from the LF and SIP methods – as a proxy for the difference in signing accuracy of the two rules.

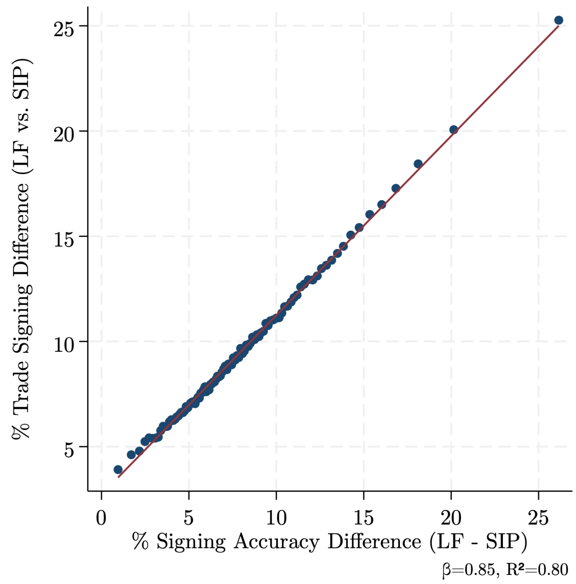
A Signing Accuracy Proxy

For each symbol-day, we measure the percentage of trades where the LF and SIP methods assign different signs to a trade. To be clear, these differences could be because the LF method is correct and the SIP is wrong or vice versa. That is, this percentage does not explicitly tell us the difference in accuracy, only when the two methods disagree. We compare the signing difference percentage to the difference in signing accuracy in the NYSE Arca sample, where we can measure the true accuracy for each method.

Figure 8 provides a binscatter plot of the percent of trades signed differently between the two methods (y-axis) against the true signing accuracy difference (x-axis). The unit of observation is at the symbol-day level, and we weight each symbol-day by the number of trades to capture how our proxy will do when aggregated across all trades.[45]

[45]For this analysis, we only consider cases where the accuracy difference is positive. There are rare cases where the accuracy difference is negative but they are for extremely thinly traded symbols that we would

Figure 8: Signing Difference vs. Signing Accuracy Difference: NYSE Arca



Notes. This figure shows a binscatter plot of the difference in true signing accuracy between the LF and SIP methodologies (x-axis) against the percent of trades assigned different signs by the two methodologies (y-axis). The unit of observation is at the symbol-day level, and observations are weighted in proportion to the number of trades. We provide the beta and R^2 from a univariate regression.

The plot shows that there is a clear and strong positive relation between trade signing differences and accuracy differences. That is, the more accurate the LF method is relative to the SIP, the greater the signing difference between the two methods. We can also report that the beta and R^2 from a univariate regression are 0.85 and 0.80, respectively. The slope is somewhat flatter than 1 because, when signing accuracy differences are very small, there are still some trade signing differences between the two methodologies. Starting from around a 10% accuracy difference, the relationship between the two variables is close to one-to-one. The high R^2 further ensures that the two variables are tightly connected. The patterns are similar when we construct both measures in dollar terms.

We also further investigate whether variation in our trade signing difference proxy is better explained by our LF method's accuracy or by SIP accuracy. We do this by estimating a regression of the percentage signing difference (y variable from Figure 8) on either the LF accuracy or the SIP accuracy (instead of the difference). We find that most of the explanatory power comes from SIP accuracy. This indicates that variation in trade signing differences comes almost entirely from variation in SIP accuracy, further supporting the use of simple signing differences as a proxy for the gap in accuracy between our two measures.

not change our methodology to accommodate.

See Appendix E.2 and Table E.2 for more details.

Trade Signing Differences

We compute our trade signing difference measure across all trades on all exchanges. Table 5 shows the overall signing difference percentages, and decomposes the sample by round lots, odd lots at the EX BBO, and odd lots inside the EX BBO (we cannot identify price slide trades in TAQ). We also further decompose the sample based on which SIP a given symbol reports to (NYSE vs. Nasdaq) and when the SIP is locked or crossed. The last sample decomposition serves as a simple diagnostic for cases where the prevailing SIP NBBO appears to be most affected by latency-induced sequence errors.

The table shows that the trade signing differences are, on average, larger than what we observe on Arca. The round lot, odd lot at the EX BBO, and odd lot inside the EX BBO signing differences aggregated over trades are 13.95%, 6.96%, and 13.85%, respectively. Notably, the round lot difference is around 7 pp larger than the accuracy difference in the LF and SIP rules in Table 3.[46] This gap suggests that the SIP may have even worse signing accuracy on other exchanges. This would be consistent with the high latency and out-of-order numbers for the large volume of Nasdaq trades in Carteret that report to the slower NYSE SIP in Mahwah. The results are similar when aggregating over dollars instead of trades. As a rough back of the envelope estimate, if we assume that the trade difference proxy overstates accuracy differences by a somewhat conservative 2 pp (based on the pattern in Figure 8), Table 5 suggests that at least 10% of trades across the U.S. stock market are incorrectly signed with the SIP methodology *and* could be corrected by our LF methodology.

The second decomposition suggests that signing issues are worse for NYSE SIP symbols than Nasdaq SIP symbols. These numbers are also consistent with the latency explanation – Section 4.1 shows that the NYSE SIP has a considerably greater SIP processing latency than the Nasdaq SIP. The table also shows that round lot trades when the SIP is locked or crossed have a much larger signing difference percentage.[47] More than 20% of all round lot trades are assigned different signs between the two methodologies when the SIP is locked or crossed. In general, a locked or crossed market is nonsensical – in a limit order book market, it can literally never occur because an order that would lock or cross the market would lead

[46] The accuracy difference for round lots between the LF and SIP methods in Table 3 can be computed as a weighted average of the two "Round Lot" rows. The weighted average accuracy difference in round lots is around 7.1%. For odd lots, taking a weighted average across all four odd lot categories, is 8.6%

[47] To sign trades in a locked or crossed market, we still use the midpoint of the bid and ask. We see these trades as indicative of underlying latency problems, and so we must pick some SIP reference price to try to sign these trades so they are included in the sample. We do the same with effective spreads in Section 5.3.

to a trade. Since the SIP aggregates quotes across many exchanges located in different cities, it can have locked or crossed markets, but it is a sign of latency issues or other potential inefficiencies. In addition, Reg NMS prevents an exchange from displaying a limit order that would lock or cross its view of the NBBO, so these locked/cross states are indicative of what the SIP sees, not what the exchanges see.[48]

We highlight trades with a locked or crossed benchmark SIP NBBO for two reasons. First, many papers in the literature simply drop these observations from their samples. That is, they drop up to 11% of trading activity. This may be because a locked or crossed market appears nonsensical, and some may conclude that there are data errors. Trades in these market conditions are not data errors. Rather, they are caused by latency and an out-of-order SIP. Second, a locked or crossed SIP can serve as a *diagnostic* for when out-of-order problems are most severe. For example, a locked or crossed market can occur if the NBO comes from a stale or distant exchange at, say, a price of 50.50, with an update that would raise the NBO to 50.52 en route. Meanwhile, closer exchanges have updates that raise the NBB from 50.49 to 50.50, and these updates hit the SIP while the NBO update is still en route. Because the NBO updates are so slow to arrive, the NBB can rise to 50.50, equal to the NBO at 50.50.[49]

5.2.3 Signing by Geographic and SIP Latency

We also report trade signing differences by exchange and SIP combinations. Table 5 reports the trade difference percentages and, for context, the percentage of trades executed by exchange within all symbols on each SIP. We only include the top 9 exchanges by trading activity (which account for more than 95% of trades), so the percentage of observations columns do not sum to 100%. We separately report the numbers for NYSE-listed and Nasdaq-listed symbols (i.e., by SIP).

Table 5 exhibits several interesting patterns related to latency. For NYSE SIP symbols, trades that originate from exchanges further from the NYSE SIP in Mahwah, and thus have greater geographic latency, show larger trade signing differences. Nasdaq has the highest signing difference at 22.3%, which is much higher than the overall average of 13.69% for all NYSE SIP trades. Recall from Section 4 that the Nasdaq to NYSE SIP path has the highest

[48] We also provide a signing accuracy breakdown when the SIP is locked or crossed with the more granular NYSE Arca signing sample in Table E.1. See Appendix E.1 for more details.

[49] In fact, the exchange with the bid at 50.50 will allow this while still complying with Reg NMS if it uses direct-feed data from all exchanges. That means the exchange would know that the distant exchange is no longer offered at 50.50, even though the SIP has yet to learn that.

Table 4: LF vs. SIP Signing Differences: All Exchanges

	Trades		Dollars	
	Signing Difference %	*% Obs.*	Signing Difference %	*% Obs.*
All Trades	12.45%		12.33%	
Decomposition 1:				
Round Lots	13.95%	*56.00%*	12.71%	*83.65%*
Odd Lots (At EX BBO)	6.96%	*21.15%*	7.79%	*7.70%*
Odd Lots (Inside EX BBO)	13.85%	*22.85%*	12.76%	*8.65%*
Decomposition 2:				
NYSE SIP	13.69%	*65.54%*	13.71%	*64.00%*
Nasdaq SIP	10.09%	*34.46%*	9.88%	*36.00%*
Decomposition 3:				
SIP Not Locked/Crossed	11.93%	*93.28%*	11.61%	*92.85%*
SIP Locked/Crossed	19.63%	*6.72%*	21.71%	*7.15%*

Notes. This table presents signing differences for our Latency-Free (LF) signing methodology vs. the Lee-Ready methodology using the SIP NBBO midpoint ("SIP"). We show the percent of trades and dollar volume assigned a different sign between the two methodologies. We use trade signing difference percentage as a proxy for signing accuracy differences. We further decompose signing differences three ways. First, we split the sample by round and mixed lot trades, odd lot trades at the EX BBO, and odd lot trades inside the EX BBO. These subsets correspond to parts of the LF rule (EX BBO used to sign the first two, delayed LF NBBO used for the third). Second, we separately report signing differences for symbols that report to the NYSE or Nasdaq SIP (i.e., listing exchange). Third, we separate trades based on whether the prevailing SIP NBBO was or was not locked/crossed. We report the size of each subset within each decomposition.

latency and the most severe out-of-order numbers. Secaucus has the next highest geographic latency, and we see that the traditional maker-taker venues have the highest signing difference numbers (IEX has the highest, but it incorporates additional delays into their market design, which adds an additional latency of 350 μs). The two taker-maker exchanges, BYX and EDGA, both have relatively lower signing difference percentages. This might mean that trades on these exchanges occur when markets are not "in motion" (Battalio et al., 2016). So, despite the fact that there are still many out-of-order events for these exchanges on average, those events do not seem to move the NBBO midpoint as much. Of the five major maker-taker exchanges, NYSE and NYSE Arca have the lowest signing difference numbers.

Note that when we separate trades by exchange, volume itself is not a great predictor of signing differences. Rather, the signing difference percentages make more sense when viewed through the lens of latency and out-of-order propensity, the things we know will cause a look-ahead bias in the SIP.

Results for the Nasdaq SIP follow a similar pattern, though the overall numbers are somewhat smaller than for the NYSE SIP. This pattern is also attributable to latency: the NYSE SIP operates with much greater latency than the Nasdaq SIP. In addition, IEX, with its

Table 5: Latency-Free vs. SIP Signing Differences: By Exchange and SIP

		NYSE SIP (Mahwah)		Nasdaq SIP (Carteret)	
Exchange	Location	Sign Diff.	*% of Obs.*	Sign Diff.	*% of Obs.*
All Exchanges		13.69%		10.09%	
NASDAQ	Carteret	22.30%	*26.54%*	7.94%	*41.34%*
NASDAQ BX	Carteret	7.95%	*3.80%*	8.46%	*3.23%*
BZX	Secaucus	16.01%	*10.51%*	13.87%	*10.74%*
BYX	Secaucus	4.44%	*9.81%*	6.05%	*7.51%*
EDGX	Secaucus	13.23%	*7.73%*	11.98%	*7.76%*
EDGA	Secaucus	5.53%	*6.53%*	6.08%	*5.60%*
IEX	Secaucus	23.59%	*4.94%*	22.72%	*4.65%*
NYSE Arca	Mahwah	8.77%	*11.34%*	10.50%	*13.60%*
NYSE	Mahwah	8.52%	*14.72%*	14.47%	*2.22%*

Notes. This table reports the percentage of trades where the Latency-Free (LF) and Lee-Ready SIP NBBO signing methodologies assign different signs. We report the numbers separately for symbols that report to the NYSE SIP in Mahwah and the Nasdaq SIP in Carteret. The top row includes all trades for each SIP. The other rows separately show results by by exchange, as well as each exchange's share of trade observations within a SIP (in italics). We report figures for the top 9 exchanges, which account for more than 95% of trades. Note that NYSE SIP symbols have almost twice as many trade observations as Nasdaq SIP symbols.

intentional delay, has the highest signing difference, and the taker-maker exchanges have lower differences than their maker-taker counterparts (BYX vs. BZX and EDGA vs. EDGX). We find it somewhat interesting that NYSE and NYSE Arca do not have very large signing differences for Nasdaq SIP symbols, despite the fact that these trades have the biggest geographic latency. This may also suggest that, at least for Nasdaq SIP symbols, NYSE and NYSE Arca trades are a response to activity on other exchanges.

Table E.3 in Appendix E.3 shows the same signing difference numbers as Table 5, but it aggregates by dollar volume within each exchange-SIP combination. The patterns are largely similar. In addition, we report a more granular version of the table above, where we separate results by round lots, odd lot trades at the EX BBO, and odd lot trades inside the EX BBO. We do this to better align with our LF signing methodology. The results are largely similar across exchanges and SIPs. Within the three trade-lot categories, the biggest differences are in round lots (signed with the EX BBO), and odd lots inside the EX BBO (signed with the delayed LF NBBO). Interestingly, odd lots at the EX BBO (signed with the EX BBO) have noticeably smaller signing differences, indicating that these trades may occur when the market is relatively "static," i.e., when there are not many quote changes before or even after the trade, which means there will be fewer out-of-order events that move the prevailing SIP NBBO midpoint. See Table E.4 in the appendix for details.

Signing Differences and Characteristics In Appendix E.4, we study whether variation in trade signing differences is related to other variables that are used as conditioning or outcome variables in research. At a high level, we find that many symbol characteristics are correlated with signing differences. We find that symbol share price and dollar volume are positively related to signing differences, and the quoted spread, intraday volatility, absolute intraday return, and variance ratio are all negatively related to signing differences. See Appendix E.4 for more details. We also provide a sense of how much order imbalance measures are distorted in Appendix E.5.

5.3 Spreads

For many analyses, trades need to be assigned a direction and magnitude: was the trade buyer or seller initiated (i.e., trade signing), and what was the premium/discount at the time of the trade (i.e., effective spread). In this section, we examine the latter – effective spreads – measured using the SIP NBBO midpoint, as nearly all of the academic literature does, and compare it to the LF NBBO midpoint as a new benchmark. We also highlight that spread measurement is likely more sensitive to issues presented by an out-of-order SIP, as described in Section 5.1.

5.3.1 Measuring the SIP Effective Spread Bias

We use the LF NBBO midpoint as a new reference price to compute the effective spread. This approach is similar to the NBBO constructed by Bartlett and McCrary (2019). As stated in Section 3, the LF NBBO is similar to the SIP NBBO, except that we construct the NBBO after each quote update sorted by exchange timestamps instead of SIP timestamps. We think of the LF NBBO as incorporating all available information from dispersed exchanges instantaneously and with no look-ahead bias. In doing so, our LF NBBO will also act as if it can consolidate information on quotes and trades without incurring any latency penalty (i.e., it can get information from different locations faster than the speed of light). However, it will only receive information as it occurred and will never be out of order with respect to when actual order book updates occurred (even if market participants would not have been able to react to them).[50]

[50]Holden and Jacobsen (2014) refer to a relative best bid and offer (RBBO), which aggregates information as fast as theoretically possible, assuming the only latency is from traveling from market center to market center at the speed of light. This requires computing an RBBO at every location since latency is a function of the distance between where the update occurred and the data center of the information aggregator. The

Our goal is to compare the SIP effective spread to the LF effective spread by measuring how often they differ and the difference between the spreads. We compute the effective spread as

$$\text{ES}_{i,j,d}^{\text{type}} = 2 \times \text{Sign}_{i,j,d} \times \log\left(\frac{p_{i,j,d}}{m_{i,t}^{\text{type}}}\right), \tag{1}$$

for the trade at time t on exchange j in symbol i. The "type" is either SIP or LF, and determines which midpoint $m_{i,t}^{\text{type}}$ is used to compute ES. The midpoint is the prevailing midpoint at the time of the trade (i.e., the prevailing NBBO according to the SIP or the LF benchmark). We use the LF signing methodology described in Section 5.2 to sign all trades, regardless of what midpoint we use to compute spreads, to isolate the role of the midpoint on the effective spread measurement.

We further decompose the effective spread into two pieces based on Glosten (1987), Hasbrouck (1988), Glosten and Harris (1988), and Stoll (1989):

$$\text{RS}_{i,j,d}^{\text{type}} = 2 \times \text{Sign}_{i,j,d} \times \log\left(\frac{p_{i,j,d}}{m_{i,t+\ell}^{\text{type}}}\right) \tag{2}$$

$$\text{PI}_{i,j,d}^{\text{type}} = 2 \times \text{Sign}_{i,j,d} \times \log\left(\frac{m_{i,t+\ell}}{m_{i,t}^{\text{type}}}\right). \tag{3}$$

RS is interpreted as the realized spread (i.e., what the limit-order submitter received in trading revenues) and PI is the price impact of the trade. We use 500 ms, 1 min, and 5 minutes for ℓ, the horizon over which we examine RS and PI. Note that these two pieces are a decomposition of the effective spread: $ES_{i,j,d} = RS_{i,j,d} + PI_{i,j,d}$. This decomposition also holds when averaging each component over sets of trades (e.g., after taking averages over all trades for a symbol on an exchange).

We leave the results for the decomposed pieces to Appendix F because, as articulated in Section 5.1, all variation in the effective spread and price impact comes from differences in the prevailing midpoint. As a result, the price impact findings in particular will look very similar, with some variation across symbol-days associated with the percentage of the spread attributable to the price impact component. In Appendix F.2, we provide the decomposition results, focusing on price impact, and find that they are indeed very similar to what we show for effective spreads. See Appendix F.2 for more details.

Table 6 reports the percentage of trades and dollar volume that have a SIP ES that differs

LF NBBO is simpler in that we aggregate updates instantaneously – it has the most timely information – and does not need a complicated set of latency adjustments depending on locations.

from the LF ES. We report the overall percentages when the LF ES is strictly greater than the SIP ES (i.e., the look-ahead bias) and when the LF ES is strictly less than the SIP ES (i.e., the stale NBBO scenario). The headline result, regardless of whether we look at trades or dollars, is that the LF ES is much more likely to be greater than the SIP ES than the other way around.

We also separately report how often the LF and SIP ES differ by odd lots and round lots (Decomposition 1), symbols that report to the NYSE vs. Nasdaq SIP (Decomposition 2), and when the SIP NBBO is and is not locked or crossed (Decomposition 3). There is little change in the overall pattern for round lots vs. odd lots. There are more significant differences when looking at NYSE and Nasdaq symbols. The NYSE SIP is more likely to have an LF ES greater than the SIP ES by about 5 pp of all trade observations. The differences are slightly smaller for the dollar volume numbers. The percentage of trades where the LF ES is smaller than the SIP ES is similar for both NYSE and Nasdaq symbols. Since the NYSE SIP has greater latency, it is more likely to lead to greater out-of-order issues, which accounts for the spread differences. Interestingly, the greater latency in the NYSE SIP over the Nasdaq SIP does not seem to materialize in differences in how often the LF ES is smaller than the SIP ES.

The most severe issues arise when the SIP is locked and crossed. Decomposition 3 shows that more than half of all trades and dollar volume have a LF ES that is greater than the SIP ES when the SIP is locked and crossed. The numbers are large, but given what was observed in Sections 5.2.1 and 5.2.2, perhaps they are not surprising. A locked or crossed SIP is a sign of severe out-of-order issues – significant market movements may have occurred but the SIP NBBO is constructed by a combination of some exchanges from which it has received updates and others where it has not. As for the NYSE and Nasdaq SIP decomposition, the fraction of trades where the LF ES is smaller than the SIP ES does not seem to change much based on whether the SIP is or is not locked or crossed.

We take the results from Table 6 as supporting the idea that the SIP has a look-ahead bias. While the problem of quote updates that occur before a trade but show up on the SIP after the trade (i.e., the stale quote problem) is still commonplace, it does not have as much bite as the reverse (the look-ahead bias). This may suggest that the types of updates that occur after a trade and are incorporated into the SIP before the trade are much more likely to move the SIP NBBO. This seems logical – a trade presents new information and is likely to induce some market response and price impact. If that price impact is incorporated into the SIP NBBO before the trade, it will reduce the effective spread. Another complementary

Table 6: Latency-Free vs. SIP: Effective Spread Comparison

	Trade Weighted		Dollar Weighted	
Sample	LF > SIP	LF < SIP	LF > SIP	LF < SIP
All Trades	19.09%	4.41%	22.19%	5.05%
Decomposition 1:				
Round Lots	19.57%	3.16%	21.91%	4.61%
Odd Lots (At EX BBO)	21.38%	2.59%	29.64%	3.70%
Odd Lots (Inside EX BBO)	15.79%	9.14%	18.25%	10.47%
Decomposition 2:				
NYSE SIP	20.92%	4.61%	23.11%	4.93%
Nasdaq SIP	15.59%	4.03%	20.55%	5.25%
Decomposition 3:				
SIP Not Locked/Crossed	16.46%	4.44%	19.68%	5.09%
SIP Locked/Crossed	55.56%	3.92%	54.65%	4.46%

Notes. This table presents the percentage of trades and dollar volume where the effective spread (ES) measures differ when using the latency-free (LF) NBBO midpoint vs. the SIP NBBO midpoint. We sign every trade using the LF signing methodology. We report the percentage of trades and dollars where the LF ES is strictly less than or greater than the SIP ES. We further decompose the sample three ways: (1) round and mixed lot trades, odd lot trades at the EX BBO, and odd lot trades inside the EX BBO, (2) NYSE SIP and Nasdaq SIP symbols, and (3) when the SIP is or is not locked or crossed.

interpretation is that an overly stale NBBO is not nearly as big of a problem as an NBBO with a look-ahead bias.

What are the relative magnitudes of the SIP ES vs. the LF ES? We compute the effective spread using the LF NBBO midpoint and SIP NBBO midpoint across all trades in our sample. This covers more than 560 million trade events and 3.7 trillion dollars exchanged. We also report the same decompositions as above in Table 6: by lot type, SIP, and locked/crossed markets. Table 7 presents the spread estimates.

We find an economically large difference between the LF and SIP effective spreads. The average trade has a SIP ES of 5.81 bps, 13.9% smaller than the LF ES of 6.74 bps. The magnitude of the LF effective spread decreases when considering the average dollar to around 4.1 bps, undoubtedly dominated by higher priced and more actively traded symbols that tend to have narrower spreads. However, the ES difference between the SIP ES and LF ES is slightly larger in dollar terms than in trade terms – the SIP ES is 14.4% smaller than the LF ES. At a high level, these findings suggest that not only does latency affect trade signing, but it also leads to the effective spread being understated by more than 13 percentage points.

The decompositions reveal more detail. Round lots, NYSE SIP symbols, and locked and

Table 7: Latency-Free vs. SIP: Effective Spread Magnitudes

	Trade Weighted			Dollar Weighted		
Sample	LF	SIP	% Diff	LF	SIP	% Diff
All Trades	6.744	5.805	-13.92%	4.121	3.529	-14.38%
Decomposition 1:						
Round Lots	7.181	6.034	-15.96%	4.181	3.579	-14.38%
Odd Lots (At EX BBO)	6.957	5.985	-13.98%	4.008	3.205	-20.05%
Odd Lots (Inside EX BBO)	5.476	5.078	-7.28%	3.644	3.326	-8.72%
Decomposition 2:						
NYSE SIP	5.268	4.355	-17.33%	3.144	2.548	-18.95%
Nasdaq SIP	9.551	8.564	-10.34%	5.858	5.272	-10.01%
Decomposition 3:						
SIP Not Locked/Crossed	6.946	6.209	-10.60%	4.123	3.772	-8.51%
SIP Locked/Crossed	3.940	0.188	-95.22%	4.102	0.382	-90.69%

Notes. This table presents the effective spread (in basis points) computed using the latency-free (LF) NBBO midpoint vs. the SIP NBBO midpoint for the average trade and dollar. We sign every trade using the LF signing methodology. We also report the percentage difference between the two measures (SIP ES / LF ES - 1). We further decompose the sample three ways: (1) round and mixed lot trades, odd lot trades at the EX BBO, and odd lot trades inside the EX BBO, (2) NYSE SIP and Nasdaq SIP symbols, and (3) when the SIP is or is not locked or crossed.

crossed SIP markets have even smaller SIP effective spreads compared to the LF benchmark. The NYSE SIP has a rather pronounced effect – the degree to which the SIP ES is understated is 7 percentage points greater for symbols that report to the higher-latency NYSE SIP than the Nasdaq SIP (17.3% vs. 10.3%). This goes up to around 9 percentage points in dollar terms. The most striking results come from when the SIP is locked or crossed, but in a way that is understandable. When the SIP NBBO is locked or crossed, the effective spread is almost zero. We also highlight that in this scenario, the LF ES is typically smaller than what we see throughout the rest of the sample. While these observations only account for a single-digit percentage of all observations, they are important to highlight because they are particularly egregious, are likely associated with signing errors, and highlight the bias that exists in the literature – many academic studies simply remove observations where the SIP is locked or crossed from their tests. These trades clearly do not represent a random sample – they are more likely to exist for high-volume symbols with lower spreads on average.

Effective Spreads by Geographic and SIP Latency

The link between latency and an out-of-order SIP suggests that latency will also be related to differences in the LF and SIP ES. We examine how often the SIP ES and LF ES differ for all trades within an exchange-SIP pair. We separate all trades in our sample by the exchange where the trade occurred and the SIP that the trade is reported to. This repeats the exercise in Table 6, but grouping trades by exchange helps to understand how spread differences align with latency. The numbers are reported in Table 8. We focus on the 9 highest volume exchanges, but also include the overall numbers for each SIP (which includes trades on all exchanges, not just the 9 highest volume ones).[51]

We highlight several observations from Table 8. First, nearly every exchange-SIP combination is more likely to have a SIP ES that is smaller than the LF ES, indicative of a look-ahead bias in the SIP. The only notable exceptions are for the NYSE Arca and NYSE trades that report to the NYSE SIP. Recall from the out-of-order results in Table 2 that these "local" trades (Mahwah to Mahwah and Carteret to Carteret) are more likely to have events out of order after (i.e., a stale SIP NBBO) than before (i.e., a look-ahead bias). However, NYSE and NYSE Arca have similar percentages for LF ES > SIP ES and LF ES < SIP ES. This implies that, despite a more severe out-of-order issue that would, on the surface, point to a bigger stale quote problem, the look-ahead and stale-quote issues materialize in different spreads at about the same rate. This also implies that each look-ahead bias out-of-order message is more likely to move the SIP NBBO than a stale out-of-order message. Interestingly, for Nasdaq trades that go to the Nasdaq SIP, we observe the look-ahead bias issue in spreads dominates, where the LF ES > SIP ES, despite the out-of-order numbers from Table 2 suggesting the opposite. This further reinforces the idea that quote changes after a trade but erroneously reported before the trade by the SIP are more likely to move the SIP NBBO.

Second, there is significant variation across exchanges in the probability that the SIP ES is less than the LF ES, even when exchanges are in the same city and for trades that report to the same SIP. For example, in Secaucus, there are five exchanges that all have substantial volume in NYSE symbols, but only two exchanges (BZX and EDGX) have around 20% or more of trades with the SIP ES less than the LF ES. Generally speaking, across both SIPs, the 5 venues with the most total trading volume (BZX, EDGX, NASDAQ, NYSE, and NYSE Arca) have the biggest percentage of trades where the SIP ES is smaller than the LF ES. This may suggest that not all out-of-order issues are created equally, and trades on the most active exchanges have out-of-order issues that move the SIP NBBO midpoint the most because they are more likely to occur in moving markets.

We also report the average effective spread magnitudes of all trades for every exchange-SIP combination. Table 9 reports the averages, as well as the percentage difference of the SIP

[51] To get a sense of what fraction of trades for each SIP happen on each exchange, see Table 5.

Table 8: Latency-Free vs. SIP: Effective Spread Comparison by Exchange

		NYSE SIP (Mahwah)		Nasdaq SIP (Carteret)	
Exchange	Location	LF > SIP	LF < SIP	LF > SIP	LF < SIP
All Exchanges		20.92%	4.61%	15.59%	4.03%
NASDAQ	Carteret	43.25%	2.13%	15.67%	6.28%
NASDAQ BX	Carteret	11.27%	1.75%	4.10%	3.94%
BZX	Secaucus	25.71%	2.55%	24.10%	2.55%
BYX	Secaucus	5.33%	1.08%	7.11%	1.44%
EDGX	Secaucus	19.79%	2.16%	18.79%	2.57%
EDGA	Secaucus	7.19%	1.32%	8.46%	1.43%
IEX	Secaucus	5.10%	2.13%	3.21%	2.63%
NYSE Arca	Mahwah	11.95%	11.90%	22.55%	2.31%
NYSE	Mahwah	11.12%	11.57%	17.20%	2.98%

Notes. This table reports the percentage of trades where the effective spread (ES) measures differ when using the latency-free (LF) NBBO midpoint vs. the SIP NBBO midpoint. We sign every trade using the LF signing methodology. We show the percentage of trades where the LF ES is strictly less than or greater than the SIP ES. We compute the percentages for all trades on each SIP ("All Exchanges"), and within each exchange-SIP pair for the nine highest-volume exchanges (which account for more than 95% of trades).

effective spread relative to the LF effective spread. There are several new observations. First, the percentage differences between the LF and SIP ES are generally largest among the highest volume exchanges (BZX, EDGX, Nasdaq, NYSE Arca, and NYSE). This pattern is consistent with Table 8 and confirms that the SIP ES is much smaller in magnitude than the LF ES, and is not limited to just the count of trades with different ES measures. The exceptions are for NYSE Arca and NYSE for NYSE SIP symbols, likely because the LF ES is about as often larger than the SIP ES as it is smaller (see Table 8).

The second observation is that there is a significant amount of heterogeneity across exchanges. The single-largest exchange for NYSE SIP symbols, Nasdaq, has a nearly 50% difference in the effective spread. Given that the SIP is ubiquitous in measuring the effective spread, these numbers suggest that the literature may significantly understate the effective spread for an enormous swath of trades. Third, we see that the less active exchanges (BX, BYX, EDGA, IEX) have very little difference in effective spread magnitudes, suggesting that the out-of-order issues there are more likely to move SIP NBBO quantities and less likely to move SIP NBBO prices.

The general theme from Tables 8 and 9 is that latency serves as a useful guide in how often the LF and SIP spread measures disagree and by how much. The higher latency NYSE

Table 9: Latency-Free vs. SIP: Effective Spread by Exchange

		NYSE SIP (Mahwah)			Nasdaq SIP (Carteret)		
Exchange	Location	LF ES	SIP ES	% Diff	LF ES	SIP ES	% Diff
All Exchanges		5.27	4.35	-17.33%	9.55	8.56	-10.34%
NASDAQ	Carteret	4.45	2.29	-48.55%	10.21	9.09	-10.98%
NASDAQ BX	Carteret	6.58	6.17	-6.30%	9.47	9.41	-0.66%
BZX	Secaucus	4.37	3.16	-27.69%	7.81	6.32	-19.17%
BYX	Secaucus	7.13	6.90	-3.20%	11.32	10.86	-4.11%
EDGX	Secaucus	5.42	4.37	-19.51%	9.88	8.51	-13.84%
EDGA	Secaucus	6.79	6.54	-3.62%	10.83	10.35	-4.46%
IEX	Secaucus	2.06	1.88	-8.60%	2.87	2.84	-1.09%
NYSE Arca	Mahwah	5.12	4.92	-4.03%	9.84	8.66	-12.06%
NYSE	Mahwah	5.31	5.24	-1.31%	6.55	5.79	-11.60%

Notes. This table presents the effective spread (in basis points) computed using the latency-free (LF) NBBO midpoint vs. the SIP NBBO midpoint for the average trade and dollar. We sign every trade using the LF signing methodology. We also report the percentage difference between the two measures (SIP ES / LF ES - 1). We show the average ES and difference for all trades on each SIP ("All Exchanges"), and within each exchange-SIP pair for the nine highest-volume exchanges (which account for more than 95% of trades).

SIP has more severe differences between the LF and SIP measures, as well as trades with greater geographic latency. We also note that while the out-of-order results in Section 4.3 are more uniform across exchanges, the spread results skew towards exchanges with higher volumes. We also show Tables 8 and 9 with dollar volume instead of over trade observations in Appendix F.1; the results are very similar.

Effective Spread Differences and Characteristics We study how spread biases are related to symbol characteristics in Appendix F.3. We find that the SIP ES has a greater downward bias as symbol share price and dollar volume increase. We also find that the SIP ES has a smaller downward bias for symbols with greater quoted spreads, intraday volatility, absolute intraday return, and variance ratio. To be clear, the SIP is still biased downwards on average across all symbol-days, but we describe correlations between the extent of the bias and characteristics. Unsurprisingly, these patterns are similar to what we observe for signing differences and characteristics (because all variation comes from the SIP NBBO midpoint as articulated in Section 5.1). See Appendix F.3 for more details.

6 Conclusion

The SIP feeds, which generate the TAQ data used throughout finance research and in securities regulation, systematically report trades and quotes out of sequence. Latencies ranging from 15 to 550 microseconds, combined with tight clustering of market events, cause quote changes that occur after a trade according to exchange timestamps to be reported before the trade by the SIP. The result is a look-ahead bias: the prevailing SIP NBBO midpoint, used as the de facto standard to sign trades and measure effective spreads, includes price impact from the trade itself.

We propose two simple solutions that eliminate latency-related SIP errors. For trade signing, we introduce the Latency-Free (LF) signing methodology, which uses only exchange timestamps and consists of two rules: the EX BBO rule, which uses the midpoint of the prevailing best bid and offer from the exchange where the trade occurred, and a delayed LF NBBO rule, which uses the midpoint of a latency-free NBBO that is delayed by 1 ms (1,000 μs). We use the EX BBO rule to sign round/mixed lot trades and odd lot trades at the exchange BBO, and the delayed LF NBBO rule to sign all other trades (odd lots inside the EX BBO and off-exchange trades). For effective spreads, we use the LF NBBO (without a delay) as a latency-free reference price instead of the SIP NBBO midpoint.

We estimate that the LF methodology signs at least 10% of trades more accurately than the standard SIP methodology, and corrects a SIP effective spread bias of -13.9%. Moreover, we find that the errors are significantly greater with trades that have high exchange-to-SIP latency. For example, for trades on Nasdaq, which has the most volume of any exchange, the LF methodology more accurately signs at least 20% of trades compared to the SIP, and the LF effective spread corrects a bias in the SIP effective spread of -43%.

The SIP sequence errors are severe and likely impact at least a significant fraction of the thousands of studies using TAQ. In Appendix G.3, we provide a guide for what types of analyses are likely affected and in what direction. Since many studies using the TAQ data take similar methodological approaches, we think of the guide as useful for researchers evaluating past studies and future research using TAQ data.

Understanding *why* latency causes signing and effective spread measurement issues helps to better understand possible solutions. While the TAQ data can always be reconstructed to be sequentially accurate using exchange timestamps, the real-time SIP feed may appear distorted as data comes in to subscribers. A small delay of, say, 0.01 seconds or even 0.001 seconds, could be used to rearrange updates that the SIP receives according to exchange timestamps. Given that latency-sensitive market participants typically use direct-feed data, we think the costs of a delay are likely small for nearly all SIP subscribers (and the SIP is effectively delayed anyway). To the extent that execution algorithms or other applications rely on the pattern of quotes and trades, the benefits from a more sequentially accurate SIP could be large.

More generally, the proliferation of securities and derivatives exchanges has led to the creation of other consolidated feeds like the SIP. One application of our paper is that, if the actual trade sign provided by exchanges themselves cannot be provided, these projects should provide at least the following two objects (in order of importance): (1) timestamps assigned by the exchange or venue, and (2) comprehensive data on the best bid and offer quote, including from all types of displayed orders (e.g., odd lots). These objects allow for the reconstruction of events on the venue where the trade occurred, and capture the necessary and relevant quotes to sign trades. That is, this information allows for the use of the EX BBO rule to sign as many trades as possible.

We provide two examples. The first is the Options Price Reporting Authority (OPRA), which aggregates trades and quotes from 13 equity options exchanges in the US. Just like in the US securities market, options exchanges are scattered throughout New Jersey and OPRA is operated out of Mahwah, NJ. That means that the out-of-order issues almost surely extend to OPRA data. Unlike the SIP, OPRA does not include an exchange timestamp. which implies that studies using OPRA data are prone to the same issues we document in this paper (and in Appendix G.3), but with no clear way to correct them. The second example is the EuroCTP, or the European Consolidated Tape, which will consolidate trade and quote activity from European securities exchanges and launch in Q3 2026. Our understanding is that the EuroCTP project plans to include exchange timestamps but not odd lots, though design details are not finalized. We expect that the SIP sequence issues we document will also be a concern for the EuroCTP, but at least the included exchange timestamps will allow researchers to correct for these errors using our LF methodologies.

References

Angel, J. J., Harris, L. E., and Spatt, C. S. (2011). Equity trading in the 21st century. *The Quarterly Journal of Finance*, 1(01):1–53.

Aquilina, M., Budish, E., and O'neill, P. (2022). Quantifying the high-frequency trading "arms race". *The Quarterly Journal of Economics*, 137(1):493–564.

Bartlett, R. P. and McCrary, J. (2019). How rigged are stock markets? evidence from microsecond timestamps. *Journal of Financial Markets*, 45:37–60.

Battalio, R., Corwin, S. A., and Jennings, R. (2016). Can brokers have it all? on the relation between make-take fees and limit order execution quality. *The Journal of Finance*, 71(5):2193–2238.

Battalio, R. H., Jennings, R. H., Saglam, M., and Wu, J. (2026). Difficulties in obtaining a representative sample of retail trades from public data sources. *Available at SSRN 4579159*.

Budish, E., Cramton, P., and Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621.

Budish, E., Lee, R. S., and Shim, J. J. (2024). A theory of stock exchange competition and innovation: Will the market fix the market? *Journal of Political Economy*, 132(4):1209–1246.

Ding, S., Hanna, J., and Hendershott, T. (2014). How slow is the nbbo? a comparison with direct exchange feeds. *Financial Review*, 49(2):313–332.

Easley, D. and O'hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial economics*, 19(1):69–90.

Ernst, T., Sokobin, J., and Spatt, C. (2021). The value of off-exchange data. *preprint*.

Ernst, T., Spatt, C. S., and Sun, J. (2024). Why did retail liquidity programs fail? *Available at SSRN 4749232*.

Glosten, L. R. (1987). Components of the bid-ask spread and the statistical properties of transaction prices. *The Journal of Finance*, 42(5):1293–1307.

Glosten, L. R. and Harris, L. E. (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics*, 21(1):123–142.

Glosten, L. R. and Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1):71–100.

Hagströmer, B. (2021). Bias in the effective bid-ask spread. *Journal of Financial Economics*, 142(1):314–337.

Harris, L. (2002). *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford University Press.

Hasbrouck, J. (1988). Trades, quotes, inventories, and information. *Journal of Financial Economics*, 22(2):229–252.

Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4):646–679.

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.

Holden, C. W., Pierson, M., and Wu, J. (2023). In the blink of an eye: Exchange-to-sip latency and trade classification accuracy. *Available at SSRN 4441422*.

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

O'Hara, M., Yao, C., and Ye, M. (2014). What's not there: Odd lots and market data. *The Journal of Finance*, 69(5):2199–2236.

Schwenk-Nebbe, S. (2022). The participant timestamp: Get the most out of taq data. *PhD Dissertation, Aarhus University*, page 1.

Stoll, H. R. (1989). Inferring the components of the bid-ask spread: Theory and empirical tests. *The Journal of Finance*, 44(1):115–134.

Wu, J. and Pierson, M. (2025). A latency commentary: Why dynamic rbbo outperforms fixed latency adjustment. *Available at SSRN 5185422*.

Internet Appendix for "Latency and the Look-Ahead Bias in Trade and Quote Data"

Robert H. Battalio, Craig Holden, Matthew Pierson,
John J. Shim, and Jun Wu

List of Appendices

A Latency Examples from TAQ

We provide two detailed examples from the data that highlight how latency can distort the SIP NBBO midpoint and result in incorrect trade signing and spread measurement.

A.1 Example 1: Geographic Latency

The first example uses Apple (AAPL). We examine the limit-order books from a subset of exchanges, although AAPL trades on all 13 U.S. securities exchanges that were operational at the time of the trade. AAPL is listed on Nasdaq, which means it uses the Nasdaq SIP, operated in Carteret, NJ. The trade we examine occurs on NYSE Arca on June 7, 2019 and has an exchange-assigned participant timestamp of `10:00:21.777579008 ET`. From that point in time, the trade was transmitted to the SIP and was disseminated by the SIP feed with a SIP timestamp of `10:00:21.777953208 ET`. The overall latency for this trade was 0.0003742 seconds or 374.2 μs. We outline the sequence of events before and after this trade as reported by the SIP data and using participant and SIP timestamps. We do not include every event around this trade for brevity, but the events we show are as they appear in the data. We make some slight tweaks to the SIP quantities to reflect the subset of exchanges we consider in this example.

We analyze the limit-order books at three locations: (1) NYSE Arca, located in Mahwah, NJ; (2) BZX, located in Secaucus, NJ; and (3) Nasdaq, located in Carteret, NJ. We also include the view of the NBBO from the Nasdaq SIP, operated out of Carteret, NJ. See Figure 1 to get a sense of the geographic locations of each of these exchanges.

- **Prevailing Best Bids and Offers** We report the prevailing best bid and ask on each of the three exchanges (i.e., the EX BBOs) and the Nasdaq SIP NBBO at the time of the trade. For the exchanges, we use exchange time (i.e., at the moment in time just before the trade, what was the limit-order book from the perspective of the exchange itself). For the SIP, we use the SIP timestamp. Thus, the EX BBOs are free from latency effects, while the SIP NBBO is not – since it aggregates information from different locations, it can only be constructed with some latency). The prevailing best bids and asks (and sizes) are provided below:

Arca:	188.57 / 188.58 (100 x 100)
BZX:	188.56 / 188.58 (100 x 100)
Nasdaq:	188.56 / 188.58 (100 x 400)
Nasdaq SIP:	188.57 / 188.58 (100 x 600)

- **Arca Trade** A trade occurs on Arca at `10:00:21.777579008` at a price of 188.57 with a size of 100 shares. Immediately after, and with the same exchange timestamp, the Arca best bid goes from 188.57 to 188.55. The trade is clearly seller-initiated: it is a round-lot trade and crossed the spread to trade with a displayed buy limit order. This trade and quote update are transmitted to the Nasdaq SIP. As of the time just after the trade, these updates must traverse about 35 miles and are "in flight" over communication networks. The state of the book at each location just after the trade is:

NYSE Arca:	188.55 / 188.58 (400 x 100)
BZX:	188.56 / 188.58 (100 x 100)
Nasdaq:	188.56 / 188.58 (100 x 400)
Nasdaq SIP:	188.57 / 188.58 (100 x 600)

- **Arca Trade in Flight** While the Arca trade and quote change (which was a direct result of the trade) are in flight to the Nasdaq SIP, many other changes occur on each exchange. 374 microseconds (0.000374 seconds) elapses from when the trade occurred on Arca to when the SIP ultimately broadcasts the trade. Over that time, Arca has 4 quote changes, Nasdaq has 6 quote changes, and BZX has 5 quote changes. The overall changes for the books at BZX and Nasdaq result in a lower best bid and ask, possibly in response to the trade observed on Arca. This can happen if the latency is long relative to what high-frequency and algorithmic traders can observe using state-of-the art, low-latency technology. We provide the state of the book at each location just before the Arca trade is published by the SIP. In addition, we also show the state of the SIP NBBO, which is based on the updates that it has already received and published from the three exchanges.

NYSE Arca:	188.55 / 188.58 (100 x 300)
BZX:	188.55 / 188.57 (100 x 400)
Nasdaq:	188.55 / 188.57 (100 x 400)
Nasdaq SIP:	188.57 / 188.57 (100 x 500)

The SIP presents a locked NBBO, with the NBB price equal to the NBO price. This occurs because the SIP has not yet received the quote change from Arca that shows that the previous Arca best bid of 188.57 is no longer there. That limit order traded, and the new best bid on Arca is 188.55, which is reflected in Arca's book. In the mean time, the SIP has received and published updates from both BZX and Nasdaq, both of which are geographically closer to the SIP (and Nasdaq is in the same data center as the Nasdaq SIP). These changes could have been contemporaneous changes by market participants to adjust their quotes from some outside signal or even a response to the trade itself from traders utilizing faster technology than the SIP.

At this point in time, the SIP has received the final update of Nasdaq's best ask at 188.57 x 400, which occurred hundreds of microseconds after the original trade on Arca. Some of BZX's best ask updates have also been published – the SIP sees BZX's best ask at 188.57 x 100 (it has not received the last update which increases the best ask size from 100 to 400). The BZX adjustments also occurred hundreds of microseconds after the Arca trade. So the SIP NBBO shows a stale best bid from Arca and updated best offers from Nasdaq and BZX, which causes the SIP NBBO to be locked.

- **Arca Trade is Published by the SIP** The trade transmitted from Arca to the Nasdaq SIP more than 300 microseconds ago is published by the SIP. It reports the same exact details as Arca would in its direct-feed data: a trade of 100 shares at a price of 188.57. If we were to use the common practice of signing this trade based on the SIP NBBO midpoint (using the last available book data in the previous entry), it would be equal to the NBB, NBO, and the NBBO midpoint, necessitating the tick test, which ends up signing this trade as a buy. This occurs exactly because the trade has arrived so late and, in the mean time, the NBBO has updated the midpoint with quotes that occurred *after* the trade and may have even been a response to the trade. That is, the SIP NBBO is utilizing information that participants could not have known at the time of the trade. Note that the effective spread in this example is also zero, though the actual spread paid as of the time of the trade was $0.005.

A.2 Example 2: SIP Processing Latency

The second example uses Bank of America (BAC), which also trades on all 13 U.S. securities exchanges and is listed on the NYSE (meaning it uses the NYSE SIP operated in Mahwah, NJ). This example highlights an instance when quotes and trades from the *same venue* are reported out of order on the SIP. That is, even though quote changes happened after the trade on the same exchange, the SIP will report those quote changes before the trade, changing the SIP NBBO midpoint using post-trade information.

This example highlights the role of SIP processing latency. We take the same approach as for AAPL, but we only examine the order book on Arca, EDGX, and the SIP. Since BAC is listed on the NYSE, trades and quote updates are sent to the NYSE SIP operated in Mahwah, NJ. To be clear, there is little to no geographical latency in play for this example – both NYSE Arca and the NYSE SIP are operated in the same data center in Mahwah. The trade we examine occurs on NYSE Arca on June 7, 2019 and has an exchange-assigned participant timestamp of `11:48:40.901345792 ET`. From that point in time, the trade was sent to and disseminated by the SIP feed with a SIP timestamp of `11:48:40.901539821 ET`. The overall latency for this trade was 0.000194 seconds or 194 microseconds. We describe the sequence of events below and, like for AAPL, use events as actually reported in the data but exclude some superfluous events on other exchanges for brevity and clarity. We adjust some quantities to reflect the subset of exchanges we consider here and keep the example internally consistent. We report the sequence of events from an observer that can see all events as they happen (i.e., is not subject to latency or speed-of-light limitations).

- **Prevailing Best Bids and Offers** We report the prevailing best bid and ask on Arca, EDGX, and the NYSE SIP at the time of the trade. The prevailing best bids and asks (and sizes) are:

NYSE Arca:	27.55 / 27.56 (15,000 x 3,100)
EDGX:	27.55 / 27.57 (3,000 x 4,000)
NYSE SIP:	27.55 / 27.56 (18,000 x 3,100)

- **Arca Trade** A trade occurs on Arca at `11:48:40.901345792` at a price of 27.56 with a size of 3,100 shares. Immediately after, and with the same exchange timestamp, the Arca best offer goes from 27.56 to 27.57. This order book update is a result of the trade executing against a displayed sell limit order at 27.56, which removes it from the book since they are not longer outstanding. Thus, the new best offer is posted by Arca. The trade is clearly buyer-initiated: it was a round-lot trade and crossed the spread to trade with an existing sell limit order. This trade and quote update are transmitted to the NYSE SIP. As of the time just after the trade, the two updates are on the way to the SIP, run out of the same data center, where they will be disseminated to SIP subscribers. The state of the book at each location just after the trade is below. Note that, at the moment just after the trade, the SIP has not updated its NBBO or reported the trade, even though Arca and the NYSE SIP is in the same data center. That is, there is still a very small amount of travel latency as well as the longer SIP processing latency.

NYSE Arca:	27.55 / 27.57 (15,000 x 8,700)
EDGX:	27.55 / 27.57 (3,000 x 4,000)
NYSE SIP:	27.55 / 27.56 (18,000 x 3,100)

- **Another Arca Update** 75 microseconds after the Arca trade, another Arca book update occurs. That update increases the best bid from 27.55 to 27.56. The update occurs on Arca before the original trade and resulting quote change is reflected in the SIP feed despite having been transmitted to the SIP. The new state of each book is

NYSE Arca:	27.56 / 27.57 (11,800 x 8,700)
EDGX:	27.55 / 27.57 (3,000 x 4,000)
NYSE SIP:	27.55 / 27.56 (18,000 x 3,100)

- **SIP Posts Update from Arca** After another 46 microseconds (121 microseconds after the original trade), the SIP disseminates an update from Arca. The update is not the trade from Arca but, instead, the SIP posts the change in the best offer that was a direct result of the trade. This means that while the trade and resulting quote change left Arca at around the same time (reflected by the fact that the two updates were assigned the same participant timestamp by the exchange). The status of each order book and the SIP NBBO as of this moment is

NYSE Arca:	27.56 / 27.57 (11,800 x 8,700)
EDGX:	27.55 / 27.57 (3,000 x 4,000)
NYSE SIP:	27.55 / 27.57 (18,000 x 12,700)

- **SIP Posts Another Update from Arca** After another 66 microseconds (187 microseconds after the trade), the SIP disseminates an update from Arca. The update is, yet again, not the trade from Arca. Instead, the SIP posts the quote update that occurred 75 microseconds after the trade (two bullet points above). This means that both quote changes, one an order book update that is a direct implication of the trade and the other as a genuine order book update that occurred after the trade, are disseminated by the SIP feed before the trade is reported. The status in all books as of this moment is

NYSE Arca:	27.56 / 27.57 (11,800 x 8,700)
EDGX:	27.55 / 27.57 (3,000 x 4,000)
NYSE SIP:	27.56 / 27.57 (11,800 x 12,700)

- **SIP Reports the Arca Trade** After another 5 microseconds (197 microseconds after the trade), the SIP feed disseminates the Arca trade. It reports the correct details – a trade of 3,100 shares at a price of 27.56. However, since the SIP has already updated from quote updates that occurred because of and after the trade, the SIP NBBO has moved from the initial state: the book is now 27.56/27.57 instead of what the book was at the time of the trade (27.55/27.56). Based on the SIP NBBO when the trade is disseminated by the SIP, the trade is signed as a sell (the trade price is at the NBB). If we were to use the prevailing exchange BBO on Arca *or* the SIP NBBO at the actual time of the trade, either would sign the trade correctly.

NYSE Arca:	27.56 / 27.57 (11,800 x 8,700)
EDGX:	27.55 / 27.57 (3,000 x 4,000)
NYSE SIP:	27.56 / 27.57 (11,800 x 12,700)

This example highlights a stark inconsistency: the SIP can mix up the sequence of events *from the same exchange*. This comes from the way the SIP processes updates – trades typically have a longer processing latency than quotes. This pattern is true independent of where trades originate (i.e., the geographical latency). We will show that this particular issue of out-of-order updates from the same exchange is less prevalent than out-of-order issues that seem to be a result of geography. But, SIP processing latencies are particularly notable because there may be a flurry of activity right after a trade on the same exchange. For example, co-located servers that run trading logic will update quotes and initiate trades immediately after observing a trade. These responses can be measured on the order of tens of microseconds or less.

B Additional Data Details

B.1 Arca Direct-Feed Data Sample

We are not able to validate our signing methodology for trades that execute against hidden orders, categorized as "hidden order executions" in the Arca direct-feed data. These trades, that execute against hidden or non-displayed limit orders, are separated in the direct-feed data. They are not separately identified by the SIP and lumped in with all lit trades.

In the direct-feed data, a trade message will report the order ID of the limit order that participated in the trade. Identifying this limit order with the order ID will reveal whether the limit order was a buy or sell, which allows us to identify if the marketable order was a sell or buy. These hidden order executions in the direct-feed data do not provide the order ID of the (hidden) limit order (since that would defeat the purpose of a hidden or non-displayed order in the first place), so there is no way to definitively ascertain whether the trade was buyer or seller initiated. We provide additional details on these executions, including where these trades tend to fall relative to the prevailing Arca BBO and LF NBBO at the time of the trade.

We provide a summary of all hidden order executions from the Arca direct-feed data for one day in our data, June 20, 2019, in Table B.1. Our approach is to identify what types of trades occur at prices that are likely to be difficult to sign.

The first notable pattern is that most odd lot hidden trades occur at a penny price increment. Of these, about a third are at or outside the NBBO (9% of all hidden trades), and these are likely to be signed accurately (we also confirm that most of these trades do occur at or outside the EX BBO, which means they would be signed using the EX BBO rule and have high signing accuracy for the reasons provided in the paper). The other two thirds of hidden odd lot trades (17% of all hidden trades) occur inside the NBBO. It seems reasonable to conclude that these trades might be similar to trades against displayed odd lot limit orders because we can identify the sign in direct-feed data. Though we acknowledge that the reason for making an odd lot limit order hidden might be related to these orders being less "stale" compared to those studied in the paper. Nevertheless, it seems difficult to come up with a scenario that would make our signing methodology perform measurably worse than existing methodologies.

Second, we examine round/mixed lot trades against hidden liquidity. We note that about

40% of all round lot trades (26% of all hidden trades) occur at the NBBO midpoint with a half-penny increment. These are notable because it is likely that any signing rule will have difficulty signing these trades. For example, see Footnote 11 in Battalio et al. (2026), which uses a proprietary sample of retail trades executed off exchange with known trade signs from one or more wholesalers. Regardless of the signing methodology used (including using direct-feed data from the wholesaler(s)), they find an accuracy rate for these trades around 50%. There are another 4.5% of all hidden trades at the midpoint but at a penny increment. These trades are also likely to have low accuracy by our methodology, though it seems likely that any methodology would perform poorly.

Third, another quarter of hidden round lot trades (a little less than a fifth of all hidden trades) are at or outside the NBBO. Just as for odd lots, these trades are likely signed with high accuracy by our methodology – they are similar to most of our round-lot trades in our sample, which are signed with extremely high accuracy since these hidden trades occur against a hidden order at or outside the NBBO. In Table B.2, we confirm that most of these trades are at/outside the NBBO and also at or outside the EX BBO, which means that these trades are also very likely to be signed with the EX BBO portion of our rule and signed accurately.

There are some trades that are at a sub-penny price but not at an increment of 0.005. These are relatively rare – 1.54% for round lots and 1.05% for odd lots – and may come from retail liquidity programs (RLPs), which allow for price improvement in increments of $0.001. The fact that they occupy a small fraction of these hidden order executions is consistent with their limited use, as noted by Ernst et al. (2024). These trades are also likely to be signed accurately using our LF methodology given the signing results in Battalio et al. (2026) for trades with de minimis price improvement.

Lastly, almost another 10.70% of all hidden trades occur inside the NBBO but are not at the midpoint, suggesting that there is sometimes round lot hidden liquidity inside the NBBO. These trades are different from other trades in our sample and merit further examination. Table B.2 focuses on round/mixed lot trades that are at a penny price increment (essentially all trades in the first column in Table B.1).

The table shows that most hidden round lot trades at a full-penny-increment price are at the EX BB *and* NBB, the EX BO *and* NBO, at some point between the bid/mid or ask/mid. We do not see many cases where a hidden execution is at the NBO but between the EX BB and EX BBO midpoint, or at the NBB but between the EX BO and EX BBO midpoint. These two cases describe something that might occur in lieu of using a price slide order. For

Table B.1: Where are Hidden Order Executions Priced?

	Round/Mixed Lots			Odd Lots			
	At $0.01	At $0.005	Other	At $0.01	At $0.005	Other	Total
At/Outside NBBO	18.93%	0.22%	0.00%	9.32%	0.08%	0.00%	28.55%
NBBO Midpoint	4.55%	26.66%	0.00%	2.96%	2.80%	0.00%	36.97%
Inside NBBO (Non-Mid)	10.70%	0.40%	1.54%	17.38%	0.16%	1.05%	31.23%
When NBBO Locked	1.92%	0.56%	0.00%	0.71%	0.06%	0.00%	3.25%
Total	36.10%	27.84%	1.54%	30.37%	3.10%	1.05%	100.00%

Notes. This table examines all trades that are labeled as hidden order executions in the NYSE Arca direct-feed data for June 20, 2019. The table reports the percentage of these trades that have a trade price at certain intervals relative to the NBBO BBO, are round/mixed or odd lots, and whether the price is at the penny, half-penny, or other sub-penny price increment.

Table B.2: Where are Hidden Order Executions Priced? Trade Price vs. EXBBO/NBBO

	P<NBB	P=NBB	NBB<P<NBBO MID	P=NBBO MID	NBBO MID<P<NBO	P=NBO	P>NBO	NBBO Locked
P<EX BB	0.47%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
P=EX BB	0.28%	10.85%	0.00%	0.00%	0.00%	0.00%	0.00%	1.43%
EX BB<P<EX MID	0.65%	3.89%	16.58%	3.81%	3.79%	2.12%	0.23%	0.14%
P=EX MID	0.05%	2.81%	0.41%	3.68%	0.43%	2.69%	0.05%	0.45%
EX MID<P<EX BO	0.19%	1.96%	3.43%	3.81%	17.60%	3.98%	0.53%	0.10%
P=EX BO	0.00%	0.00%	0.00%	0.00%	0.00%	11.06%	0.27%	1.80%
P>EX BO	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.41%	0.02%

Notes. This table examines all trades that are labeled as hidden order executions in the NYSE Arca direct-feed data for June 20, 2019. For this table, we only examine trades at a full penny and no sub-penny price value. The table reports the percentage of these trades where the trade price falls in certain intervals relative to the EX BBO and LF NBBO.

example, the EX BBO is 10.02/10.05 and the NBBO is 10.02/10.03. If someone wants to be bid on the exchange at 10.03, the exchange prohibits them from doing so because it would lock the NBBO. One option is to submit a price slide order, but the other is to submit a hidden bid at 10.03. However, since we do not see many cases of this type, we think it is safe to rule out that this is the reason for non-displayed orders.

The most plausible reason for hidden orders, especially those inside the NBBO, is to hide the desired quantity. If these orders are stale, like the retail odd lot orders that trade inside the NBBO that we see in the paper, then the 1 ms delayed LF NBBO is likely the most appropriate quote to sign these trades. If these trades are not stale and represent an improvement of the NBB or NBO without a desire to display the order, then they are most likely signed better with the EX BBO rule.

Given that these orders are more likely to come from institutions, we think it is less likely that these orders are stale. That implies that the EX BBO is likely superior and ultimately why we believe our signing rule – to sign all round/mixed lot trades with the EX BBO – is appropriate. However, we are open to an amendment to our signing methodology, which simply uses whether the trade price is at the EX BBO to sign trades with the EX BBO, and all trades inside the EX BBO are signed with the delayed LF NBBO. That is, instead of using the distinction of round/odd lots, use only the trade price relative to that exchanges EX BBO. We have tried this signing methodology with the full Arca direct-feed sample and we find that it is marginally worse than our proposed methodology, likely because of the way it handles price slide trades and other types of round lot trades just inside the EX BBO. However, we are open to the possibility that the amended signing rule performs worse for our known signing sample, but that disimprovement might be outweighed by signing these hidden order executions better (though, as this appendix points out, we cannot test this). We suspect that this alternative does not improve signing, but we raise it as a source of possible uncertainty and a sensible future consideration if a researcher were to obtain proprietary data on hidden round lot trades in the future.

B.2 Automatically Generated SIP Quote Updates

When there is a change in the top-of-book bid or ask on an exchange, the exchange notifies the SIP. This excludes any odd lot limit orders, which are not currently reported to the SIP. Whenever there is a trade with a top-of-book quote, the exchange reports the trade to the SIP, but also reports the resulting change in the top-of-book quote. The latter could be a simple change in quantity or, if the trade took all size at the top-of-book (or left less than 100 shares), the exchange will report the next best bid or offer to the SIP. This quote update, which results automatically from a trade, is reported to the SIP to inform market participants of a change in the current best bid and offer on the exchange. What we find in the TAQ data is that the automatically-generated quote update resulting from a trade have the same exact exchange timestamp as the trade itself. Note that these updates do not exist in the direct-feed data and do not actually reflect separate actions by market participants to modify or cancel orders. Rather, they simply reflect that the price and or quantity at the exchange's BBO have changed because of the trade.

For example, suppose there are 100 shares at the best offer at a price of 50.01 and 100 shares at the second best offer at 50.02 on NYSE Arca. If a trader submits a marketable buy order for 100 shares, causing a trade at 50.01, two updates will be sent to the SIP to report this market activity. The first is that a trade occurred at a price of 50.01 and a size of 100 shares. The second is that the best offer is now 50.02 for 100 shares, which is not based on any actual limit order changes but only because the best offer is now gone and the new best offer is 50.02. These two updates always have the same exact exchange timestamps, reflecting the fact that they are from the same "event."[52] However, when these two messages get to the SIP, they will almost always have different SIP timestamps and will often be out of order. This is exactly why, for many analyses, we do not include updates with an exchange timestamp at the same nanosecond. This is a catchall to ensure we do not count an automatically generated quote update, which does not reflect a limit-order submission, modification, or cancellation, as an event that is clustered around a particular trade. See Schwenk-Nebbe (2022) for more details on these automatically generated messages and how they can be used to identify all trades that came from a single marketable order.

[52]Schwenk-Nebbe (2022) uses these automatically-generated BBO updates to create a marketable order identifier.

C Latency Magnitudes and Why Latency Matters

C.1 Additional Latency Magnitude Results

We provide supporting materials for our latency results. Table C.1 shows the interquartile range for latencies by exchange and SIP. There is relatively little variability in latencies for exchanges that submit messages to the Carteret SIP. The IQR of latencies for the NYSE SIP is an order of magnitude larger than for the Nasdaq SIP. In addition, trades have more variability across all exchanges for both SIPs. We note that the IQRs are relatively small compared to the level of latencies shown in Table 1. Median latencies range from 15 to 550 microseconds, but IQRs mostly range from 1 to 50 microseconds. This pattern indicates that most of the variation in latencies comes not from randomness of any one transmission but from consistently different latencies based on geography and which SIP processes the update (NYSE vs. Nasdaq).

C.2 Additional Clustering Results

In this appendix subsection, we provide additional results on the clustering of trades and top-of-book quote changes.

Table C.2 reports the total number of trades and BBO quote changes for several sample symbols on June 20, 2019 in Panel A. We also divide the day into non-overlapping 500 microsecond intervals and report the probability that an interval contains a quote change, a trade, or either of the two using exchange timestamps. In Panel B, we report the same numbers for each symbol at various percentiles in the distribution of trading activity (based on the number of trades). The table shows that even the most active symbols have fewer than 10 million events, and most symbols have far fewer than 1 million events. This is in contrast to 46.8 million non-overlapping 500 μs intervals for a 6.5 hour trading day. That is, even though modern markets are characterized by hyper-frequent trading and quote updates, a 500 μs interval is so small that there are 100 times as many 500 μs intervals as there are events for most symbols.

As a stark example, the S&P 500 ETF (SPY), regularly the single most active symbol in the U.S. market, had only 3.71% of 500 microsecond intervals populated with at least one trade or one quote change on this particular trading day. That means more than 96% of 500 μs intervals for SPY had literally no trade or top-of-book activity. Apple (AAPL), also

Table C.1: Latency IQR from Exchanges to to NYSE/Nasdaq SIP

		NYSE SIP (Mahwah)		Nasdaq SIP (Carteret)	
Exchange	Location	Quote	Trade	Quote	Trade
Nasdaq	Carteret	45	49	2	7
Nasdaq BX	Carteret	38	46	1	2
Nasdaq PSX	Carteret	40	53	2	6
Cboe BYX	Secaucus	40	58	8	12
Cboe BZX	Secaucus	40	59	9	12
Cboe EDGA	Secaucus	39	53	9	11
Cboe EDGX	Secaucus	40	57	11	13
IEX	Secaucus	53	51	31	33
NYSE Arca	Mahwah	23	51	3	12
NYSE	Mahwah	23	32	3	3
NYSE American	Mahwah	22	38	4	3
NYSE National	Mahwah	31	33	23	35
NYSE Chicago	Mahwah	76	9527	7839	63

Notes. This table presents the interquartile range (IQR) of latencies for trades and quote changes sent from each of the 13 U.S. stock exchanges to the two SIPs. All symbols listed on Nasdaq report trades and quote changes to the Nasdaq SIP in Carteret, NJ. All other symbols report to the NYSE SIP in Mahwah, NJ. We use all regular-hour trades and quotes from June 20, 2019. We measure latency for each trade and quote as the difference between the SIP-assigned timestamp and the exchange-assigned participant timestamp, then take the IQR (75th - 25th percentile values) for each exchange-SIP pair.

one of the most active symbols and well above the 99th percentile of symbols in terms of events on this day, has more than 98% of 500 μs intervals that have nothing happening. This is especially striking for the average and 75th percentile symbols, which have nothing happening in more than 99.8% of 500 μs intervals. Even if we were to hypothetically spread out all quotes and trades to cover as many intervals as possible, the average symbol would still have 99.5% of 500 μs intervals with nothing happening ("Hypothetical Max" column in each panel).

We also present supplementary plots to Figure 5 in the main text, which show the probability of having at least one event in the 500 μs before and after a trade at the symbol-day level. Here, we repeat the plot, but instead of a probability, we count the expected number of events in the 500 μs window before and after a trade. Figure C.1 presents the plots, with the left and right plots showing the expected counts before and after the trade, respectively. Note that the scale of the y-axis in Panel B is twice that in Panel A, indicating that there are many more events after a trade than before.

The figure shows that there is quite extreme clustering, and the average number of events is loosely correlated with trading activity (x-axis). Even for thinly traded stocks, there is significant clustering. For the example symbols highlighted in dark red, there is still a

Table C.2: How Many 500 Microsecond Intervals have Quotes and Trades?

Panel A: Example Symbols

			500 μs Intervals			
	# Quotes	# Trades	Pr(Quote)	Pr(Trade)	Pr(Q or T)	Hypoth. Max
S&P 500 ETF (SPY)	7,419,631	303,039	3.71%	0.23%	3.73%	16.50%
Apple (AAPL)	2,145,479	126,924	1.64%	0.12%	1.65%	4.86%
Bank of America (BAC)	1,974,496	107,726	1.31%	0.08%	1.32%	4.45%
Starbucks (SBUX)	956,450	35,156	0.62%	0.03%	0.62%	2.12%
Nike (NKE)	531,718	41,338	0.42%	0.04%	0.42%	1.22%
Harley Davidson (HOG)	239,395	12,854	0.21%	0.01%	0.21%	0.54%
Whirlpool (WHR)	129,618	8,056	0.17%	0.01%	0.17%	0.29%
Equifax (EFX)	83,902	6,618	0.11%	0.01%	0.11%	0.19%

Panel B: Distribution (Based on Trade Count)

			500 μs Intervals			
	# Quotes	# Trades	Pr(Quote)	Pr(Trade)	Pr(Q or T)	Hypoth. Max
Mean	203,469	9,042	0.17%	0.01%	0.17%	0.45%
25th Pctile	15,673	2,105	0.02%	0.00%	0.02%	0.04%
50th Pctile	48,947	4,304	0.05%	0.00%	0.05%	0.11%
75th Pctile	160,231	10,011	0.11%	0.01%	0.11%	0.36%
90th Pctile	388,293	20,381	0.32%	0.02%	0.32%	0.87%
99th Pctile	1,392,357	73,556	0.82%	0.06%	0.82%	3.13%

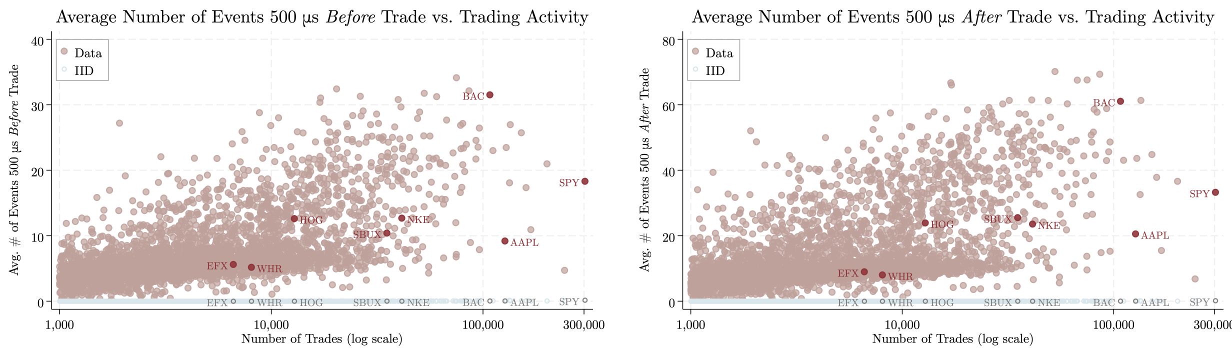
Notes. This table presents the number of quotes and trades for several example symbols (Panel A) and for symbols at various percentiles, where the ranking is based on the number of trades (Panel B). The data come from one full day of trading on June 20, 2019. We also report the probability that a quote update or trade occurs within a 500 microsecond interval using exchange timestamps (if the 6.5 hour trading day were broken up into distinct 500 microsecond intervals). The last column in each panel reports the total number of trades and quotes divided by the total number of distinct 500 microsecond intervals in a day (46.8 million), or the hypothetical maximum number of 500 microsecond intervals that could be covered if trades and quotes were spread out as thinly as possible.

significant degree of clustering, with around 5 events before and 10 events after for non-large-cap stocks like Equifax and Whirlpool (EFX and WHR). Just as with probabilities, there are many more events on average after a trade than before.

The light gray dots at the bottom of each plot represent the expected clustering if all events were distributed uniformly and i.i.d. While the dots form a line at the bottom of each plot, this underscores just how small these numbers are. In fact, the expected number of events that we observe in the data (red dots) is thousands of times greater than what we would expect if events were completely random throughout the trading day.

To complement Figures 5 and C.1, we report numbers for the exact amount of clustering. Table C.3 presents the probability of at least one event and the expected event count before and after a trade for several example symbols, as well as several symbols that fall at various points in the trade count distribution. In Panel A, we show the numbers for the example

Figure C.1: Event Clustering within 500 μs of a Trade: Average Number of Events



Notes. This figure examines the 500 microseconds (μs) before or after a trade (left and right plot, respectively). We compute the expected number of events in the 500 μs before or after a given trade. We compute the expected counts for each symbol on June 20, 2019. We plot the expected number of events on the y-axis, and plot it against the number of trades on that day on the x-axis. We also plot the expected events if all quote changes and trades for a given symbol were distributed uniformly and i.i.d. throughout the trading day in gray. We highlight several example symbols in bold that represent a range of trading activity: Equifax (EFX), Whirlpool (WHR), Harley Davidson (HOG), Starbucks (SBUX), Nike (NKE), Bank of America (BAC), Apple (AAPL), and the SPDR S&P 500 ETF (SPY). See Table C.3 for specific numbers on the example symbols.

symbols. We provide the statistics for various symbols in the distribution of all symbols that had at least 1,000 trades on the day. Panels A and B show that events are highly clustered around trades. For very active symbols, like SPY and AAPL, there is an extremely high probability that there is at least one other event in the 500 μs before and after a trade. What might be somewhat more surprising is that even less active symbols, like WHR and EFX, are also very likely to have another event within 500 μs, both before and after. Given how little overall activity there is in these two stocks, which have total trade and quote activity that is less than 6% of AAPL's total number of events, it seems reasonable to expect that trades and quotes in these symbols are more isolated. The fact that there is still a high degree of clustering with such few events speaks to the data-generating process for events in modern securities markets.

Panel B shows that even for a symbol at the 10th percentile of the trade count distribution, there is more than an 80% chance that a trade is followed by at least one update within 500 μs after the trade. In addition, the expected number of events is above 6.[53] It is also worthwhile to note that the expected number of events is many multiples larger than what we would expect in an i.i.d. world. Based on this measure, the clustering is rather extreme.

[53] On average, the degree of clustering generally increases with the activity of a symbol. Note that the pattern is not monotonic in Table C.3 because we take a symbol at a point in the distribution, and other variables besides pure activity can generate more or less clustering (e.g., share price).

Table C.3: How Clustered are Trades and Quotes? Other Events within 500 μs of a Trade

Panel A: Example Symbols

	Pr(≥1 Event)			E[#Events]		
	Before	After	IID	Before	After	IID
S&P 500 ETF (SPY)	74.25%	95.46%	15.21%	18.32	33.26	0.17
Apple (AAPL)	71.02%	93.75%	4.74%	9.20	20.55	0.05
Bank of America (BAC)	77.72%	93.71%	4.35%	31.51	61.06	0.04
Starbucks (SBUX)	69.71%	95.50%	2.10%	10.40	25.52	0.02
Nike (NKE)	76.34%	92.55%	1.22%	12.69	23.56	0.01
Harley Davidson (HOG)	72.87%	94.01%	0.54%	12.61	23.91	0.01
Whirlpool (WHR)	66.77%	90.33%	0.29%	5.18	8.03	0.00
Equifax (EFX)	60.92%	83.83%	0.19%	5.61	8.99	0.00

Panel B: Distribution (Based on Trade Count)

	Pr(≥1 Event)			E[#Events]		
	Before	After	IID	Before	After	IID
Mean	61.48%	85.35%	0.45%	7.77	15.08	0.00
10th Pctile	59.21%	82.58%	0.05%	4.22	6.63	0.00
25th Pctile	48.62%	68.84%	0.09%	2.59	4.06	0.00
50th Pctile	63.84%	92.80%	0.17%	7.42	14.92	0.00
75th Pctile	72.72%	94.24%	0.43%	16.24	43.60	0.00
90th Pctile	73.02%	94.74%	1.11%	7.80	15.48	0.01

Notes. This table presents statistics that describe how clustered events are around trades. We report the probability of observing at least one event in the 500 microseconds before and after the trade. In addition, we also provide the expected number of events in the 500 microseconds before and after the trade. Lastly, we show the probability of at least one event and the expected number of events in any 500 microseconds interval if all of a symbol's events were distributed uniformly and i.i.d. throughout the trading day. Panel A shows the numbers for a set of sample symbols. Panel B reports the same statistics for symbols throughout the distribution of trading activity (measured by the number of trade events in the day).

C.3 Additional Out-of-Order Results

Just as in Figure 4 in Section 4.3, we present the clustering and out-of-order expected event counts, but for a less-heavily traded stock (Equifax (EFX) instead of JP Morgan (JPM). EFX provides a better illustration of how the clustering and out-of-order counts change as the total activity decreases. For context, JPM on this day had more than 74K trades and 1.56M top-of-book quote changes, whereas EFX had around 6.6K trades and a little less than 84K top-of-book quote changes.

Figure C.2 presents the clustering and out-of-order plot. Panel A plots the average number of events around an EFX trade, Panel B focuses on Carteret trades and plots the expected number of Carteret events around a trade, and Panel C examines Mahwah events around a Carteret trade. The blue bars indicate events, and the red bars indicate the expected number of events that are out of order according to the SIP relative to the actual sequence of events based on exchange timestamps.

Figure C.2 is strikingly similar to Figure 4 for JPM. We highlight two differences. First, the expected number of events is much smaller, as indicated by the scale of the y-axis. Second, there is much more noise in the picture, especially Panel C where there is a less clear jump at around 220 μs after the trade, though it is still identifiable. In addition, the picture is more jagged in general, with less consistency in the average event count numbers. Even with these differences, the out-of-order issue is still observable, and the most severe issues come from the trades with the longest latency (EFX is listed on the NYSE, which means Carteret trades travel to the NYSE SIP in Mahwah and have a long exchange-to-SIP latency).

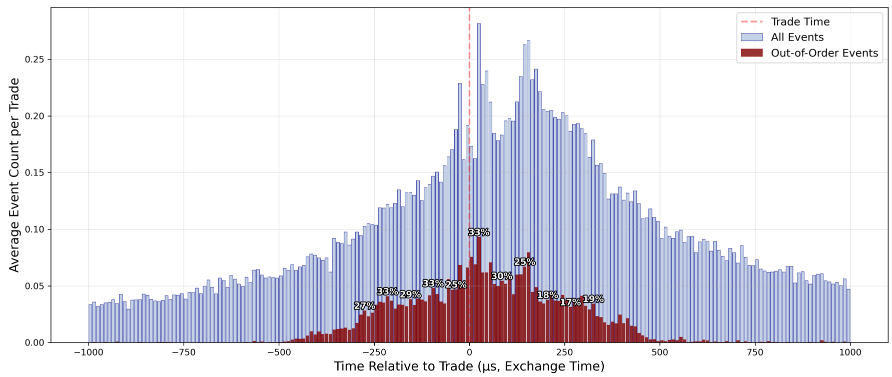
Another approach to highlighting the severity of the out-of-order problem and better understanding the clustering of events is to only examine events around trades focusing on trades that are the first in a potential series of trades. We know from the clustering results that conditional on a given trade, there are many other quote changes and other trades that are likely to surround it. However, we also know that most 500 μs intervals are empty – nothing happens. As a result, the JPM clustering example in Figure 4, as well as the EFX example above, may be examining events around a particular trade when the focal trade itself is a response to another trade that occurred before. That is, some of the events surrounding a trade could be overlapping many trades.

We slightly tweak Figure 4 from the paper by considering only the set of trades where no trade occurred in the 1 ms before the trade. That means we will only count the events surrounding the first trade in a sequence without considering trades that are within 1 ms. We conduct this procedure sequentially – when encountering the first trade in a sequence, and include it in the sample by counting the events around it. We then find the next trade that is at least 1 ms after the first trade and do the same.

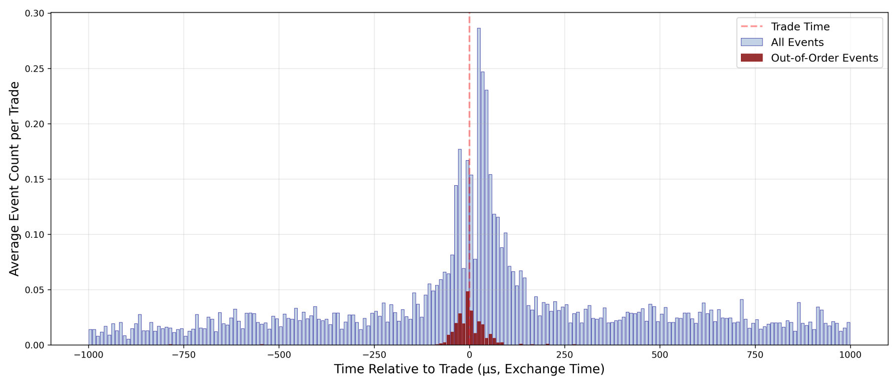
Figure C.3 shows the revised clustering plots with out-of-order events highlighted. The three plots are largely similar but, if anything, more starkly highlight the patterns described in Figure 4. Panel A shows that there is not much of an increase in the number of quotes leading up to a trade, followed by a flurry of activity after the trade. This suggests that much of the increase in trades and quotes that we saw before the trade is a result of including trades that were second or third in a sequence of trades, i.e., the “run-up” before the trade was driven completely by events that may have caused the trade in the first place. Panels B and C paint a mostly similar picture as their counterparts in Figure 4 from the main text.

Figure C.2: Clustering and Out-of-Order Events Around a Trade: Equifax (EFX) Example

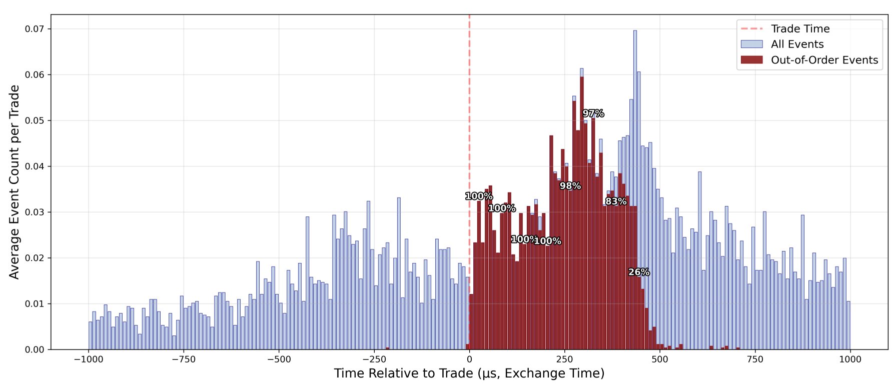
Panel A: All Trades and Events



Panel B: Trades and Events on Carteret Exchanges



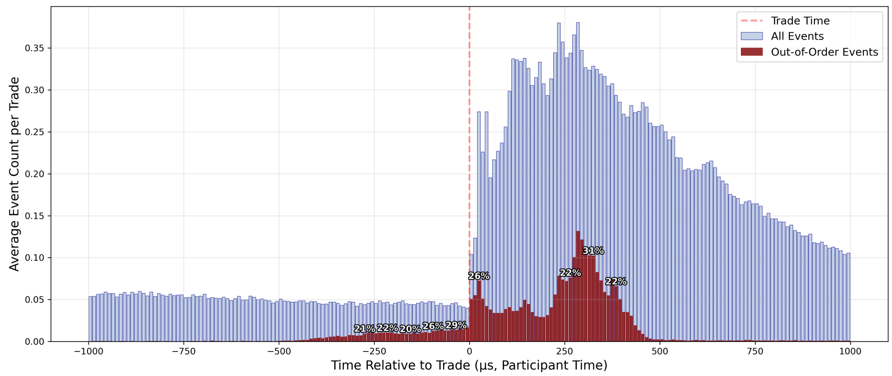
Panel C: Trades on Carteret Exchanges and Events on Mahwah Exchanges



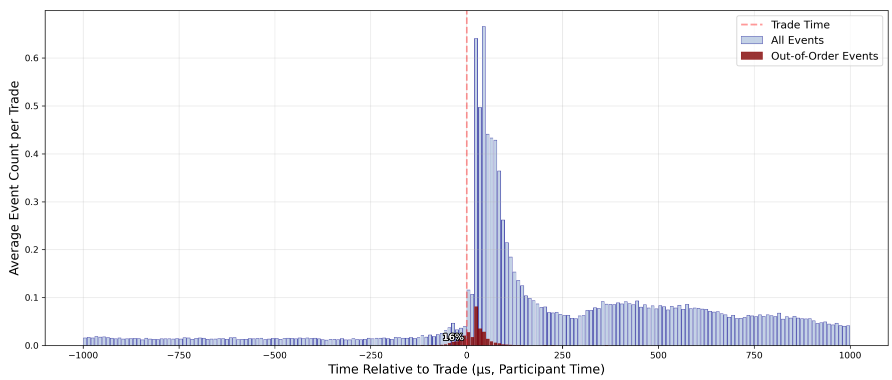
Notes. We plot the expected number of events (quote changes or other trades) in 10 microsecond (μs) intervals around each trade of EFX on June 20, 2019. We use exchange timestamps to assign events to bins (excluding events in the same nanosecond as the trade). The blue bars show the expected event count per interval averaging over all trades in the day. The red bars show the expected number of events that are out of order. An event is counted as out of order if it occurs after (before) the trade in exchange time but reported before (after) the trade on the SIP. Panel A uses all trades and events. Panel B restricts both trades and events to Carteret exchanges. Panel C uses Carteret trades but only counts events on Mahwah exchanges.

Figure C.3: Out-of-Order Events Around the First Trade in a Cluster: JP Morgan (JPM) Example

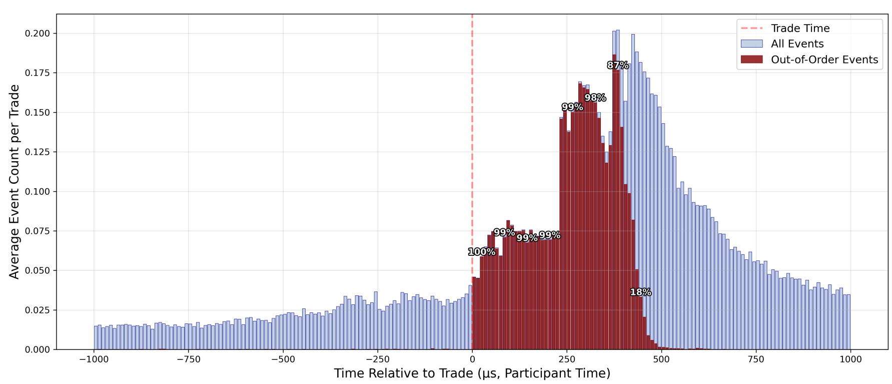
Panel A: All First Trades and Events



Panel B: First Trades and Events on Carteret Exchanges



Panel C: First Trades on Carteret Exchanges and Events on Mahwah Exchanges



Notes. We plot the expected number of events (quote changes or other trades) in 10 microsecond (μs) intervals around each trade of JPM on June 20, 2019. We only examine trades that do not have another trade in the 1 ms before (i.e., the "first" trade in a potential sequence). We use exchange timestamps to assign events to bins (excluding events in the same nanosecond as the trade). The blue bars show the expected event count per interval averaging over all trades in the day. The red bars show the expected number of events that are out of order. An event is counted as out of order if it occurs after (before) the trade in exchange time but reported before (after) the trade on the SIP. Panel A uses all trades and events. Panel B restricts both trades and events to Carteret exchanges. Panel C uses Carteret trades but only counts events on Mahwah exchanges.

We also provide supplementary results for Figure 6, which shows the probability of at least one event out of order before and after a trade at the symbol-day level. First, we provide the same plots but for the expected number of out-of-order events before and after (i.e., labeled OoO Before and After in Section 4.3) in Figure C.4. We show in the left plot the expected number of events at the symbol-day level that are reported before the trade on the SIP but actually occurred after the trade according to exchange timestamps. Similarly, the right plot shows the same variable, except it counts events that occurred after the trade on the SIP but took place before the trade based on exchange timestamps.

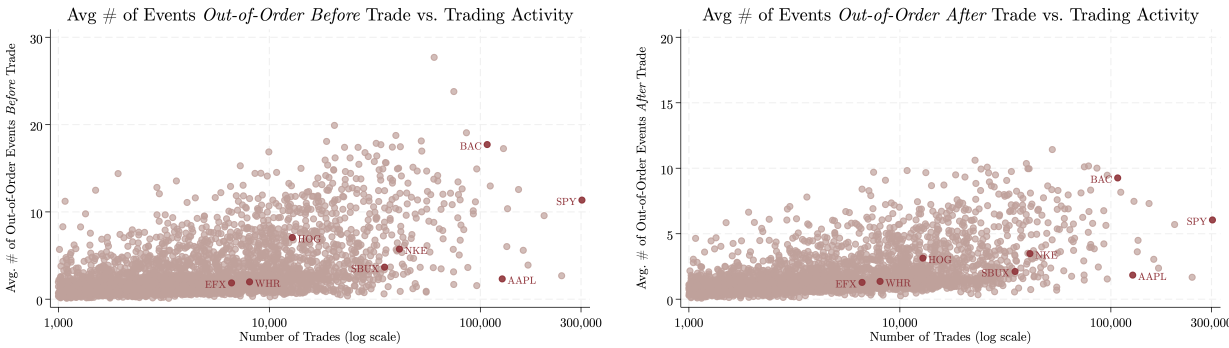
Figure C.4 shows a familiar picture, both to the out-of-order probabilities plots in Figure 6 and the clustering plots in Figure C.1 above. This similarity indicates that both probabilities and counts are closely related, and that the greater the degree of clustering, the higher the expected out-of-order event counts. Just as with clustering, the out-of-order counts are increasing with the number of trades. In addition, OoO Before events are more common than OoO After events, which also corresponds to clustering. As a reminder, out-of-order before events occur after the trade according to exchange timestamps but are reported by the SIP before the trade. Also note that the y-axis scale is much greater on the left panel than on the right.

We provide concrete numbers for the out-of-order results in Figures 6 and C.4. Table C.4 reports the numbers for two sets of statistics. First, we report the probability that there is at least one OoO Before event and one OoO After event, corresponding to Figure 6. We also provide the expected number of out-of-order events, both OoO Before and OoO After, mapping to Figure C.4. We do this for our set of example symbols in Panel A. We show the values for symbols at various percentiles in the trade count distribution in Panel B.

Panel A shows that the example symbols have at least one event OoO Before around 30% to 70% of the time. The OoO After probabilities are lower, ranging from 20% to around 35%. The expected number of out-of-order events also shows a similar pattern – there are more out-of-order events before than after. Moreover, out-of-order probabilities and expected events decline as trading activity decreases (just as before, the example symbols are sorted by activity). Panel B also shows a similar pattern: symbols at higher percentiles in terms of trade activity have higher OoO Before and After numbers.

As mentioned in the main text, we also find that sometimes trades and quote changes from the *same exchange* are out of order on the SIP. See the BAC example in Appendix A.2 for an example from the data. Table C.5 presents out-of-order results identical to what we show in Table C.4, except we only consider events on the same exchange as where the trade occurred.

Figure C.4: Out-of-Order Events within 500 μs of a Trade: Average Number of Events



Notes. This figure computes the expected number of events (quote updates or trades) that is within 500 microseconds (μs) of a trade according to exchange timestamps but is reported out of order using SIP timestamps for each symbol on June 20, 2019. The left (right) plot shows the expected out-of-order events erroneously reported by the SIP as occurring before (after) the trade. We plot the number of events on the y-axis and the number of trades on the x-axis. We highlight several example symbols that span a range of trading activity: Equifax (EFX), Whirlpool (WHR), Harley Davidson (HOG), Starbucks (SBUX), Nike (NKE), Bank of America (BAC), Apple (AAPL), and the SPDR S&P 500 ETF (SPY). See Table C.4 for specific numbers on the example symbols.
This figure examines the 500 microseconds (μs) before or after a trade (left and right plot, respectively). We compute the probability that there is at least one quote change or trade in the 500 μs before or after a given trade. We compute the probability for each symbol on June 20, 2019. We plot the probability on the y-axis, and plot it against the number of trades on that day on the x-axis. We also plot the probability if all quote changes and trades for a given symbol were distributed uniformly and i.i.d. throughout the trading day in gray. We highlight several example symbols in bold that represent a range of trading activity: Equifax (EFX), Whirlpool (WHR), Harley Davidson (HOG), Starbucks (SBUX), Nike (NKE), Bank of America (BAC), Apple (AAPL), and the SPDR S&P 500 ETF (SPY). See Table C.3 for specific numbers on the example symbols.

From those events, we report the probability that one of them is OoO Before and OoO After, as well as the expected number. While it is not as obvious from the table, the securities with the largest OoO Before numbers are those that are listed on NYSE and report to the NYSE SIP. This is because these errors are usually not due to geographic latency, but rather the way the SIP processes and disseminates messages. As a result, this problem is much more concentrated on NYSE SIP symbols.

We also examine these issues at the exchange-SIP level, similar to Table 2 in the main text. That is, we compute the same out-of-order numbers for all trades on a given exchange-SIP combination but, when counting out-of-order events, we only consider events that are on the same exchange as the trade. Table C.6 reports the same-exchange out-of-order probabilities and counts.

Panel A shows that there are consistently same-exchange out-of-order events. For example,

Table C.4: Out-of-Order Event within 500 μs of a Trade

Panel A: Example Symbols

	Pr(≥1 Event)		E[#Events]	
	OoO Before	OoO After	OoO Before	OoO After
S&P 500 ETF (SPY)	73.54%	39.85%	11.35	6.05
Apple (AAPL)	37.20%	35.18%	2.33	1.84
Bank of America (BAC)	77.97%	34.33%	17.71	9.26
Starbucks (SBUX)	45.16%	33.74%	3.67	2.12
Nike (NKE)	64.17%	31.22%	5.74	3.48
Harley Davidson (HOG)	66.14%	26.88%	7.07	3.13
Whirlpool (WHR)	49.29%	22.70%	1.99	1.36
Equifax (EFX)	42.22%	21.28%	1.87	1.28

Panel B: Distribution (Based on Trade Count)

	Pr(≥1 Event)		E[#Events]	
	OoO Before	OoO After	OoO Before	OoO After
Mean	41.51%	22.75%	3.24	1.88
10th Pctile	20.99%	19.25%	0.71	0.54
25th Pctile	11.44%	15.28%	0.54	0.54
50th Pctile	42.69%	26.44%	3.34	1.81
75th Pctile	45.44%	35.15%	5.93	3.00
90th Pctile	64.56%	26.67%	4.38	1.87

Notes. This table presents statistics for out-of-order event in the 500 microseconds before and after a trade. We label an event (i.e., a quote of trade) as "Out of Order Before," or abbreviated (OoO Before), if the event occurs within 500 microseconds after the trade (not including the nanosecond of the trade) according to exchange time but the SIP reports it before the trade according to SIP timestamps. Analogously, we label an event as "Out of Order After" (or OoO After) if an event occurs in the 500 microseconds before a trade (not including the nanosecond of the trade) according to exchange time, but is reported by the SIP after the trade according to SIP timestamps. We present the probability of at least one out-of-order event before and after, and report the expected number of out-of-order events before and after. We consider all trades and quotes for all symbols on June 20, 2019. Panel A provides the numbers for a set of example symbols. Panel B shows the out-of-order statistics for symbols throughout the distribution of symbols based on trade activity.

a trade on Nasdaq in an NYSE-listed symbol will have at least one Nasdaq event occur after the trade with exchange timestamps but before the trade according to the SIP about 23% of the time. The percentage of OoO Before is similar for NYSE Arca, and slightly smaller for BZX, BYX, and EDGX in Secaucus. Interestingly, trades in Mahwah that report to the Mahwah SIP have larger OoO Before issues than OoO After. This pattern is in contrast with Table 2. So, while stale SIP NBBO issues are more prevalent for NYSE Arca trades in NYSE symbols in general, the subset of same-exchange out-of-order events skews heavily towards OoO Before events, or a SIP with a look-ahead bias.

Table C.5: Out-of-Order Events within 500 μs of a Trade on the *Same Exchange*

Panel A: Example Symbols

	Pr(≥1 Event)		E[#Events]	
	OoO Before	OoO After	OoO Before	OoO After
S&P 500 ETF (SPY)	23.92%	7.80%	1.10	0.36
Apple (AAPL)	3.26%	0.94%	0.10	0.04
Bank of America (BAC)	26.54%	6.86%	1.81	0.63
Starbucks (SBUX)	5.09%	0.77%	0.15	0.04
Nike (NKE)	23.60%	3.08%	0.73	0.11
Harley Davidson (HOG)	30.58%	1.83%	0.93	0.06
Whirlpool (WHR)	6.70%	2.78%	0.12	0.05
Equifax (EFX)	5.21%	3.22%	0.09	0.06

Panel B: Distribution (Based on Trade Count)

	Pr(≥1 Event)		E[#Events]	
	OoO Before	OoO After	OoO Before	OoO After
Mean	8.91%	1.84%	0.25	0.08
10th Pctile	0.27%	0.55%	0.00	0.01
25th Pctile	0.33%	0.23%	0.01	0.00
50th Pctile	3.28%	0.22%	0.09	0.01
75th Pctile	5.46%	0.82%	0.20	0.03
90th Pctile	10.56%	2.70%	0.22	0.05

Notes. This table presents statistics for out-of-order event in the 500 microseconds before and after a trade, but only considers events of the same exchange as the trade. We label an event (i.e., a quote of trade) as "Out of Order Before," or abbreviated (OoO Before), if the event occurs within 500 microseconds after the trade (not including the nanosecond of the trade) according to exchange time but the SIP reports it before the trade according to SIP timestamps. Analogously, we label an event as "Out of Order After" (or OoO After) if an event occurs in the 500 microseconds before a trade (not including the nanosecond of the trade) according to exchange time, but is reported by the SIP after the trade according to SIP timestamps. We present the probability of at least one out-of-order event before and after, and report the expected number of out-of-order events before and after. We consider all trades for all symbols on June 20, 2019. For each trade, we consider the subset of events within 500 microseconds on the same exchange. Panel A provides the numbers for a set of example symbols. Panel B shows the out-of-order statistics for symbols throughout the distribution of symbols based on trade activity.

Table C.6: Out-of-Order Events on the *Same Exchange* by Exchange and SIP

Panel A: NYSE SIP (Mahwah)						
			Pr(≥1 OoO Event)		E[#Events OoO]	
Exchange	Location	# Trades	Before	After	Before	After
All Exchanges		19,557,153	16.95%	4.23%	0.61	0.20
NASDAQ	Carteret	5,286,140	23.44%	8.69%	0.86	0.44
NASDAQ BX	Carteret	704,031	5.77%	2.28%	0.23	0.04
BZX	Secaucus	2,093,681	19.84%	2.98%	0.70	0.08
BYX	Secaucus	1,931,203	15.46%	2.95%	0.46	0.06
EDGX	Secaucus	1,504,301	15.65%	2.76%	0.49	0.07
EDGA	Secaucus	1,233,410	9.49%	1.78%	0.23	0.04
IEX	Secaucus	966,867	0.81%	0.93%	0.05	0.03
NYSE Arca	Mahwah	2,321,615	25.59%	3.24%	1.36	0.24
NYSE	Mahwah	2,792,464	10.33%	2.22%	0.13	0.14
Panel B: Nasdaq SIP (Carteret)						
			Pr(≥1 OoO Event)		E[#Events OoO]	
Exchange	Location	# Trades	Before	After	Before	After
All Exchanges		9,221,786	3.27%	0.79%	0.12	0.03
NASDAQ	Carteret	3,831,340	5.21%	1.23%	0.13	0.06
NASDAQ BX	Carteret	265,614	0.35%	1.03%	0.00	0.05
BZX	Secaucus	1,036,584	2.94%	0.67%	0.12	0.02
BYX	Secaucus	685,681	2.39%	0.64%	0.07	0.02
EDGX	Secaucus	717,208	2.51%	0.77%	0.13	0.02
EDGA	Secaucus	478,122	1.51%	0.46%	0.05	0.01
IEX	Secaucus	426,704	0.49%	0.30%	0.03	0.01
NYSE Arca	Mahwah	1,346,439	1.47%	0.02%	0.17	0.00
NYSE	Mahwah	196,920	1.50%	0.33%	0.17	0.01

Notes. This table presents statistics for out-of-order event in the 500 microseconds before and after a trade. We label an event (i.e., a quote of trade) as "Out of Order Before," or abbreviated (OoO Before), if the event occurs within 500 microseconds after the trade (not including the nanosecond of the trade) according to exchange time but the SIP reports it before the trade according to SIP timestamps. Analogously, we label an event as "Out of Order After" (or OoO After) if an event occurs in the 500 microseconds before a trade (not including the nanosecond of the trade) according to exchange time, but is reported by the SIP after the trade according to SIP timestamps. We consider all trades, but we only consider events from the same exchange as the trade when labeling events as out of order (i.e., we identify out-of-order events on the same exchange). We present the probability of at least one same-exchange out-of-order event before and after, and report the expected number of same-exchange out-of-order events before and after. We consider all trades and quotes for all symbols on June 20, 2019. Panel A shows the statistics over trades on each exchange for NYSE SIP symbols. Panel B does the same but for Nasdaq SIP symbols.

D Latency-Induced Errors

D.1 How an Out-of-Order SIP affects the SIP NBBO

In this appendix, we provide examples of how the two types of out-of-order errors, documented in Section 4.3, affect the SIP NBBO and SIP NBBO midpoint. See Section 5.1 for more details on how these errors affect the SIP NBBO midpoint and why that object is key to understanding how out-of-order errors affect signing and spreads.

Out-of-Order Before → Look-Ahead Bias

When a quote change that occurs after a trade is erroneously reported before the trade on the SIP, it can trigger an NBBO update, which in turn can change the prevailing midpoint used to sign trades and measure spreads.

Consider the following example. There are two exchanges: exchange A has a BBO of 10.00/10.03, and exchange B has a BBO of 10.01/10.04. The NBBO is 10.01/10.03, with a midpoint of 10.02. The actual sequence of events is as follows: there is a trade on A at 10.03, then a change in exchange A's BBO to 10.01/10.05, and a change in exchange B's BBO to 10.02/10.06. After all of this, the new NBBO is 10.02/10.05.

The SIP often records quote changes that occur after a trade as occurring before the trade, even on the same exchange. If both of the quote updates on exchanges A and B, which occurred after the trade, are reported by the SIP before the trade, they will change the prevailing NBBO at the time of the trade. That is, the SIP will report the two quote updates, which lead to an NBBO of 10.02/10.05, then will report the trade at 10.03. Thus, the prevailing NBBO midpoint at the time of the trade is 10.035. Table D.1 summarizes the sequence of events. Panel A shows the actual sequence, and Panel B shows the SIP-reported sequence.

Using Lee-Ready and the NBBO midpoint, we can see that the actual sequence of events signs the trade as buyer-initiated. Panel B shows that using Lee-Ready with the SIP NBBO midpoint will sign this trade as *seller-initiated*. In addition, the effective spread using the sequentially accurate data is 0.02, computed by taking the half spread of 0.01 (trade price of 10.03 minus midpoint of 10.02) and multiplying by 2. Using the SIP midpoint, the trade has an effective spread of 0.01. Note that the SIP-assigned sign is different for the trade, so we compute the spread treating the trade as seller-initiated. If we were to fix the sign as a

Table D.1: Out-of-Order Before: Actual vs. SIP-Reported Sequences

Panel A: Actual Sequence

Order	Event	A's BBO	B's BBO	NBBO	NBBO Mid
0	Initial State	10.00 | 10.03	10.01 | 10.04	10.01 | 10.03	10.020
1	**Trade on A at 10.03**				
2	A Quote Update	10.01 | 10.05	10.01 | 10.04	10.01 | 10.04	10.025
3	B Quote Update	10.01 | 10.05	10.02 | 10.06	10.02 | 10.05	10.035

Signing: Trade at 10.03 > midpoint of 10.02 → Buyer-Initiated
Effective Spread: (Trade at 10.03 − midpoint of 10.02) × 2 → 0.02

Panel B: SIP-Reported Sequence

Order	Event	A's BBO	B's BBO	NBBO	NBBO Mid
0	Initial State	10.00 | 10.03	10.01 | 10.04	10.01 | 10.03	10.020
1	A Quote Update	10.01 | 10.05	10.01 | 10.04	10.01 | 10.04	10.025
2	B Quote Update	10.01 | 10.05	10.02 | 10.06	10.02 | 10.05	10.035
3	**Trade on A at 10.03**				

Signing: Trade at 10.03 < midpoint of 10.035 → Seller-Initiated
Effective Spread: (midpoint of 10.035 − Trade at 10.03) × 2 → 0.01

buy (based on the true direction), the trade would have an even smaller SIP-assigned spread of -0.01.

The key intuition for why this sequentially inaccurate SIP (from OoO Before updates) could lead to signing and spread errors is as follows. When there is a trade, there will likely be a response to that trade. We usually think of these updates as reflecting at least some price impact from the trade. If there is at least some probability that the trade is informed, quotes should be revised in the direction of the trade (Glosten and Milgrom, 1985). This means that a buy is more likely to lead to upward revisions of quotes (similarly, a sell leads to downward revisions). If these quote updates are reported before the trade, some of the price impact from the trade itself is imputed into the prevailing midpoint before the trade. This is the sense in which quote updates that occur after a trade but are reported before it on the SIP introduce a *look-ahead bias.*

How often signs are flipped and spreads are changed is open to measurement. There are factors that may soften the detrimental effects of the out-of-order SIP issue. For example, it is possible that for any given out-of-order event, prices do not change but only quantities. That is, if there is a buyer-initiated trade, the best bid quantity may increase and the best offer quantity may decrease. This is still consistent with Glosten and Milgrom (1985) logic, but results from discrete tick sizes (Hagströmer, 2021). Importantly, the quantity updates do not change the midpoint price. Another possibility is that quote changes that occur after a trade move in the opposite direction (in terms of prices or sizes). For example, a buyer-initiated trade may lead to the NBBO going down. This would be inconsistent with Glosten and Milgrom (1985), but could be a result of updates that were en-route to the exchange before the trade was observed by market participants, or could be a result of other public signals (e.g., activity in other symbols or public news) that, on net, lead to revisions in the opposite direction. An example is when there is public news, and traders have differential hedging demand (some send marketable buy orders and others send more competitive sell limit orders). The prevalence of each of these possibilities is succinctly captured by trade signing errors (shown in Section 5.2) and the direction and size of the spread bias (shown in Section 5.3).

Out-of-Order After → Stale NBBO

Another possibility is that quote changes before a trade are reported after the trade on the SIP (i.e., OoO After). This can occur if the trade occurs after the quote changes, possibly in response to the quote change itself. This means that the prevailing SIP NBBO will be stale in that it has not incorporated information from quote changes on other exchanges.

We provide an example to highlight the effect of a stale NBBO. Just as before, there are two exchanges: exchange A has a BBO of 27.00/27.03, and exchange B has a BBO of 27.01/27.04. The NBBO is 27.01/27.03, with a midpoint of 27.02. The actual sequence of events is as follows: exchange B's BBO changes to 27.02/27.05, and a trade occurs on A at 27.03. Based on the actual sequence of events, the prevailing NBBO just before the trade is 27.02/27.03 and the NBBO midpoint is 27.025. If the quote change on exchange B, which occurred before the trade, is not reported by the SIP till after the trade, SIP users will observe a different (stale) prevailing NBBO of 27.01/27.03 with a midpoint of 27.02. Table D.2 summarizes the sequence of events: Panel A shows the actual sequence and Panel B shows the SIP-reported sequence.

If we were to use Lee-Ready and the NBBO to sign the trade, we would sign the trade as buyer-initiated, regardless of whether we use the actual sequence in Panel A or the SIP sequence in Panel B. In fact, nearly any signing rule will lead to the same conclusion. This highlights the fact that distortions from a stale SIP (i.e., OoO After) are somewhat bounded, whereas distortions from a look-ahead SIP (i.e., OoO Before) are theoretically unbounded.

Table D.2: Stale NBBO: Actual vs. SIP-Reported Sequences

Panel A: Actual Sequence

Order	Event	A's BBO	B's BBO	NBBO	NBBO Mid
0	Initial State	27.00 | 27.03	27.01 | 27.04	27.01 | 27.03	27.020
1	B Quote Update	27.00 | 27.03	27.02 | 27.05	27.02 | 27.03	27.025
2	**Trade on A at 27.03**				

Signing: Trade at 27.03 > midpoint of 27.025 → Buyer-Initiated
Effective Spread: (Trade at 27.03 − midpoint of 27.025) × 2 → 0.01

Panel B: SIP-Reported Sequence

Order	Event	A's BBO	B's BBO	NBBO	NBBO Mid
0	Initial State	27.00 | 27.03	27.01 | 27.04	27.01 | 27.03	27.020
1	**Trade on A at 27.03**				
2	B Quote Update	27.00 | 27.03	27.02 | 27.05	27.02 | 27.03	27.025

Signing: Trade at 27.03 > midpoint of 27.02 → Buyer-Initiated
Effective Spread: (Trade at 27.03 − midpoint of 27.02) × 2 → 0.02

We discuss this point further in the midpoint and measurement section below. Also note that if exchange B's quote update included a new best bid of 27.03 or greater, the actual sequence of updates would lead to a locked or crossed NBBO, which could lead to signing discrepancies.[54]

However, stale quotes can regularly lead to distortions in the effective spread. The SIP-reported sequence does not reflect the fact that the spread narrowed (via Exchange B's higher bid) before the trade. The trade could have been a response to the bid increasing or, more likely, participants observed information that led them to believe prices should be moving up, consistent with some participants increasing bids and others hitting the offer.

Just as in the look-ahead case, the sequence errors could have gone the other way or led to no change in the midpoint. For example, B could have ticked down before a trade at the offer on A, or the change in B could have been just for the best bid and offer size with no change in price. We think the example above, as well as a version with only quantity changes, is the most likely given that events clustered in sub-millisecond intervals are likely to be economically related.

[54] We do not highlight this case because it is less likely because of Reg NMS and the order protection rule. The locked/crossed case is possible in instances where a participant sends an ISO order to trade the best offer of 27.03 on exchange A at roughly the same time it sends an ISO order to be bid at 27.03 or higher on exchange B (see Section 2.2 for more context).

E Additional Signing Results

E.1 Signing Accuracy for NYSE Arca Trades

One simple diagnostic for potential issues with the SIP is when the SIP NBBO is locked or crossed. Reg NMS stipulates that the market centers must prevent a locked or crossed market. Practically, this means that exchanges will not post a displayed limit order if it would lock or cross an existing limit order on another exchange. The challenge is that exchanges all have different and delayed views of what other exchanges' best quotes are (even with direct-feed data and fast connections). The SIP, with even greater latency, exacerbates this issue and will more often show a locked or crossed market.

There is a natural challenge to sign trades using Lee-Ready and the SIP NBBO midpoint when the SIP NBBO is locked or crossed. This market state is difficult to make sense of, and is almost surely a result of latency-induced issues. As a result, we think it can serve as a simple diagnostic for when signing with the SIP is most problematic. As a reminder, the BBO on any given exchange can never be locked or crossed, and thus always provides a reasonable midpoint when signing a trade. For our signing rule, we will use the EX BBO for all round lot trades and all odd lot trades at the prevailing EX BB or BO.

Table E.1 provides the signing accuracy numbers for the LF and SIP signing methodologies. We split the sample into various categories to isolate differences in accuracy when the SIP NBBO is and is not locked/crossed. While trades when the SIP is locked/crossed accounts for around 8.5% of the sample in trade and dollar terms, the trade signing errors are extremely high for the SIP, while our LF approach is nearly unaffected. Note that these trades, which occur in moving markets, are typically tossed from academic studies. The trade signing accuracy, especially for the subsets that are signed with the EX BBO (first four rows), is extremely high for the LF method, approaching 100%. In all subsets, the LF method significantly outperforms the SIP approach, and this is especially pronounced when the SIP is locked or crossed.

E.2 Signing Difference as an Accuracy Proxy

We provide additional support for using the rate of signing differences as a proxy for the difference in the accuracy of our two methods.

Table E.1: NYSE Arca Signing Accuracy: When the SIP is Locked or Crossed

	Trades				Dollars			
Trade Subset	Latency-Free	SIP	Diff.	*% Obs.*	Latency-Free	SIP	Diff.	*% Obs.*
Round Lots	95.5%	90.0%	5.4%	*44.0%*	96.8%	91.2%	5.6%	*75.9%*
Round Lots (SIP Locked/Crossed)	99.3%	78.5%	20.9%	*5.5%*	99.4%	76.7%	22.7%	*8.0%*
Odd Lots (At EX BBO)	98.3%	95.8%	2.6%	*25.8%*	99.4%	96.4%	3.1%	*8.8%*
Odd Lots (At EX BBO + SIP Locked/Crossed)	99.8%	82.6%	17.2%	*2.4%*	99.8%	76.7%	23.1%	*0.5%*
Odd Lots (Inside EX BBO)	89.8%	74.9%	15.0%	*21.7%*	87.4%	75.0%	12.4%	*6.8%*
Odd Lots (Inside EX BBO + SIP Locked/Crossed)	94.3%	86.2%	8.1%	*0.6%*	90.5%	83.2%	7.3%	*0.1%*

Notes. This table presents the signing accuracy for our Latency-Free signing methodology ("Latency-Free") and the Lee-Ready methodology using the SIP NBBO midpoint ("SIP"). We present the overall accuracy in terms of trade observations, shares, and dollar trade volume. We split the sample of trades by round lot, odd lots at the EX BBO, and odd lots inside the EX BBO. Within each of these three groups, we split based on trades that occur when the prevailing SIP NBBO was locked or crossed (SIP NBB $\geq$ SIP NBO). We report the size of the subset in terms of percentage of trade and dollar observations.

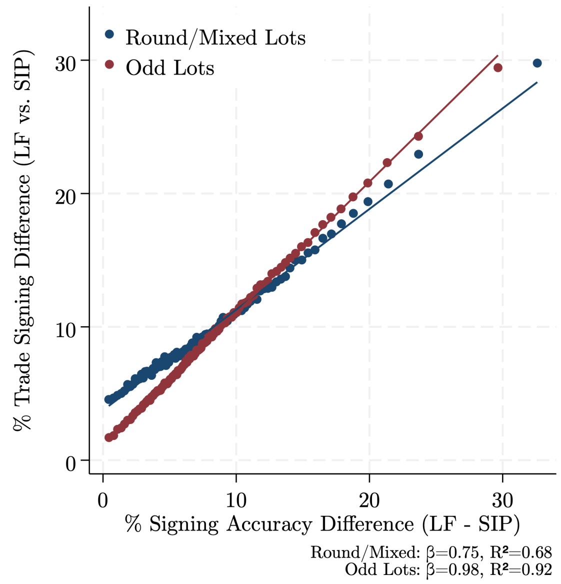
First, we repeat Figure 8 from the main text, but separately report the results for round/mixed lots and odd lot trades. We separate trades on this dimension because we know that a significant fraction of odd lot trades occur inside the EX BBO, and we use the stale LF NBBO to sign these trades. Figure E.1 provides a binscatter plot of the percent of trades signed differently between the two methods (y-axis) against the signing accuracy difference (x-axis). The unit of observation is at the symbol-day level, and we split the sample into round lots and odd lots. We weight each observation by the number of Arca trades to better map to aggregated trade accuracy numbers. We also report the univariate regression coefficients and R^2s from a similarly weighted regression for each sample.[55] The plots show that there is a clear positive relation between trade signing differences and accuracy differences. That is, the more accurate the LF method is relative to the SIP, the greater the signing difference between the two methods. The relationship holds for both round and odd lots, though the round lot result is slightly flatter than a one-for-one relationship and the odd lots show almost an exact one-for-one relationship. The flatness of the slope for round lots is driven mostly by errors when there are no accuracy differences – even when a symbol-date has little to no accuracy difference, there are still some trade signing differences between the LF and SIP rules.

To further validate our proxy as capturing the signing accuracies of the SIP method, we estimate the following weighted regression:

$$\text{SignDiffPct}_{i,d} = \alpha + \text{Accuracy}_{i,d} + \varepsilon_{i,d}, \tag{E.1}$$

[55] Just as in the main text, we only consider cases where the accuracy difference is positive. See the footnote in the main text for our reasoning.

Figure E.1: Signing Difference vs. Signing Accuracy Difference: NYSE Arca



Notes. This figure shows two binscatter plots of two signing variables, where the unit of observation is at the symbol-day level. The first variable is the difference in true signing accuracy between the Latency-Free (LF) and SIP signing methodologies, which is plotted on the x-axis. The second variable is the percent of trades that are assigned different signs between the two methodologies. We compute these two variables separately for round/mixed lots and odd lots, and for each symbol-day. Each plot gives a weight to each symbol-day in proportion to the number of trades in that symbol-day. We show the slope and R^2 from an OLS regression for round lots and odd lots for each plot in the bottom right.

where $\text{SignDiffPct}_{i,d}$ is the percentage of trades signed differently between the LF and SIP methods for symbol i on day d and weights are given based on the number of Arca trades. For $\text{Accuracy}_{i,d}$, we use three measures: (1) the accuracy difference between LF and SIP methods (just as in Figure 8), (2) the percentage signing accuracy level of the LF method alone, and (3) the percentage signing accuracy level of the Lee-Ready SIP method alone. We also estimate the weighted regression separately for round and odd lots.

Table E.2 reports the regression estimates. The first three columns report the results for all trades. Column (1) shows that accuracy is strongly related to the simple trade difference measure and has a high R^2, repeating the right panel of Figure 8. Columns (2) and (3) show that most of the variation comes from variation in SIP accuracy rather than our LF method's accuracy – the SIP's slope is 12 times as large in magnitude and the R^2 is more than 150 times as large. We see a similar pattern when we break the sample into round lots, in columns (4) to (6), and odd lots in columns (7) to (9). If anything, the pattern is more stark for odd lots – most if not all of the variation in trade signing differences seems to come from variation in SIP accuracy.

Table E.2: Signing Difference as a Proxy for Signing Accuracy: NYSE Arca

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Accuracy Diff (pp)	0.85			0.75			0.98		
	(143.47)			(97.71)			(237.62)		
LF Accuracy (%)		-0.03			-12.40			5.92	
		(-2.36)			(-7.72)			(3.97)	
SIP Accuracy (%)			-0.38			-31.37			-56.56
			(-23.67)			(-23.59)			(-24.71)
# Obs.	84,277	84,277	84,277	68,083	68,083	68,083	76,100	76,100	76,100
R^2	0.802	0.002	0.376	0.677	0.028	0.351	0.918	0.002	0.505
Sample	All	All	All	Round Lots	Round Lots	Round Lots	Odd Lots	Odd Lots	Odd Lots

Notes. The table provides regression estimates of signing measures at the symbol-day level. The dependent variable for all regressions is the percent of trades that are assigned different signs between the two methodologies. Accuracy Difference is the difference in true signing accuracy between the Latency-Free (LF) and SIP signing methodologies, measured as a percentage of trades. LF Accuracy and SIP Accuracy are the percent of trades signed accurately by the LF and SIP methods, respectively. We estimate the regressions separately for all trades within a symbol-day (columns 1-3), for round and mixed lots trades (columns 4-6), and odd lots (columns 7-9). We require at least 100 trades in a symbol-day to be included (e.g., column 4 requires 100 round/mixed lot trades in a symbol-day to be included). T-statistics are shown in parenthesis, and computed with standard errors clustered at the symbol and day level.

These regressions suggest that as SIP accuracy worsens, it increases the gap in trade signing differences between the two methodologies. That indicates that as the trade signing differences increase between the LF and SIP approaches, this is driven by a decrease in accuracy by the SIP approach.

E.3 Signing Difference Results

In Section 5.2.2, we present results on signing differences between the LF and SIP approaches. We also provide difference rates for all trades within an exchange-SIP pair in Table 5. In Table E.3, we reproduce the table but in dollar volume terms rather than trade terms. That is, we ask what fraction of dollars are mis-signed by exchange-SIP pair. We do this separately for round lot and odd lot trades.

The dollar volume results are very similar to the trade results in the main text. If anything, Table E.3 shows that the distribution of dollar volume is more skewed towards the major, maker-taker exchanges like Nasdaq and NYSE Arca. This suggests that larger dollar trades are less likely to occur on "inverted" exchanges like BYX and EDGA, which have a smaller fraction of dollar volume compared to trade observations, and more likely to occur on Nasdaq and Arca. Otherwise, the signing difference numbers are largely similar.

The signing difference results by exchange and SIP also combine all trade observations including round and odd lot trades. In the main text, we describe in detail why we create

Table E.3: LF vs. SIP Signing Differences by Exchange and SIP (Dollar Volume)

		NYSE SIP (Mahwah)		Nasdaq SIP (Carteret)	
Exchange	Location	Sign Diff.	*% of Obs.*	Sign Diff.	*% of Obs.*
All Exchanges		13.71%		9.88%	
NASDAQ	Carteret	20.40%	*25.00%*	6.97%	*52.08%*
NASDAQ BX	Carteret	10.36%	*3.09%*	13.76%	*2.44%*
BZX	Secaucus	15.26%	*10.93%*	12.82%	*9.72%*
BYX	Secaucus	6.01%	*6.22%*	9.59%	*3.89%*
EDGX	Secaucus	13.17%	*6.97%*	11.64%	*6.68%*
EDGA	Secaucus	7.30%	*4.01%*	9.46%	*2.87%*
IEX	Secaucus	30.51%	*5.01%*	29.82%	*4.80%*
NYSE Arca	Mahwah	8.44%	*17.30%*	7.52%	*11.36%*
NYSE	Mahwah	7.59%	*17.10%*	11.07%	*2.63%*

Notes. This table reports the percentage of dollar volume where our Latency-Free (LF) signing methodology differs from the Lee-Ready methodology using the SIP NBBO midpoint. We report the numbers separately for symbols that report to the NYSE SIP in Mahwah and the Nasdaq SIP in Carteret. The top row includes all trades for each SIP. In the other rows, we report the signing difference percentages and the percentage of dollar volume on each exchange to provide context for which exchanges have the most activity. We only report figures for the top 9 exchanges, which account for more than 95% of dollar volume. We show the percentage of volume in italics to differentiate from the signing difference percentages. Also for context, we can report that our data has around 1.75 times as much dollar volume for NYSE SIP symbols compared to Nasdaq SIP symbols.

the LF signing method, which uses the EX BBO for round lots and odd lots that trade at the EX BBO or worse, and then the delayed LF NBBO to sign odd lots inside the EX BBO. See the main text for more details. With the signing rule in mind, it makes sense to separate out these three categories of trades because they are signed differently and because they vary in ways that also may be related to signing differences.

We report trade signing differences by exchange and SIP in Table E.4 but further separate the results by all round lot trades, odd lot trades at the EX BBO or wider ("Odd Lot (At)") and odd lot trades inside the EX BBO ("Odd Lot (In)"). We also report the percentage of observations within each lot category across exchanges. Panel A reports the numbers for all NYSE SIP symbols, and Panel B shows Nasdaq SIP symbols. The table is mostly consistent with the results in Table 5 from the main text, with a few interesting nuances. In general, odd lots at the EX BBO have smaller signing differences than round lots and odd lots inside the EX BBO. This pattern suggests that these trades are easy for the SIP to get right. One interpretation is that odd lot trades at the EX BBO are not very informative, and thus do not have a flurry of activity that follows (that the SIP could get out of order) or there is a flurry of activity but it does not move the SIP NBBO midpoint at all or enough to change sign. This pattern is true for all exchanges on both SIPs, and especially true on "inverted" taker-maker exchanges like BX, BYX, and EDGA. This may point further point to the idea that odd lot trades at the EX BBO on taker-maker exchanges are especially uninformative.

E.4 Signing and Characteristics

We consider several symbol-level characteristics and relate them to trade-signing differences. Because characteristics are at the symbol-day level, we also aggregate trade signing differences at the symbol-day level. We require a symbol-day to have at least 100 trades across all exchanges to be included in the sample. We compute signing differences for each symbol-day as the percent of dollar volume where the sign assigned by our LF signing methodology and the Lee-Ready SIP methodology differ. Weighting by dollar volume within each symbol-day naturally incorporates round and odd lots by blending them based on their contribution to overall dollar volume. We use signing differences as the dependent variable in the following regression on a characteristic $x_{i,d}$:

$$\text{SignDiffPct}_{i,d} = \alpha + \beta \cdot x_{i,d} + \varepsilon_{i,d}. \tag{E.2}$$

We use several characteristics for $x_{i,t}$, including log closing share price, log intraday dollar volume, log time-weighted percent quoted spread, log annualized intraday volatility, log absolute intraday return, and the 1-minute variance ratio. The volatility and return measures are in percentage terms before taking logs, and the quoted spread is in basis points before taking logs. We also winsorize spreads and volatility at the 1st and 99th percentiles before taking logs.

We estimate the univariate regression in Equation E.2 and present the results in a series of binscatter plots in Figure E.2. The variables that are visually and statistically strongly related to signing differences are the quoted spread, intraday volatility, and dollar volume. While the overall signing results from Sections 5.2.1 and 5.2.2 suggest that the SIP signing errors are unavoidable, Figure E.2 shows that there is significant dispersion in just how severe the signing errors might be. In particular, symbols with a high share price, high dollar volume, and a low quoted spread are where signing errors are likely to be greatest. These are also the types of characteristics that describe ETFs and prominent stocks that might be used in event studies.

We combine all variables in a multivariate regression and also include indicator variables for whether the symbol is an ETF and whether the symbol is Nasdaq-listed (i.e., a Nasdaq SIP symbol). The estimates are presented in Table E.5. The most notable differences when

Table E.4: Latency-Free vs. SIP Signing Differences: By Exchange, SIP, and Lot Type

Panel A: NYSE SIP (Mahwah)

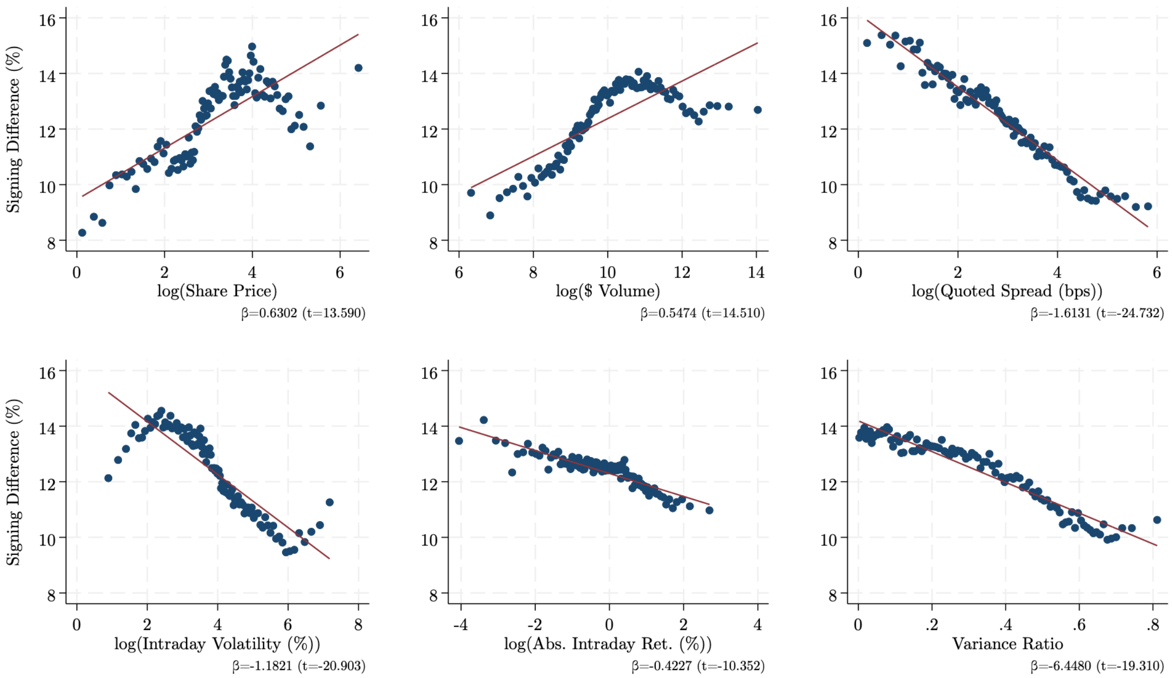
		Signing Difference			% Of Obs.		
Exchange	Location	Round Lots	Odd Lots (At)	Odd Lots (In)	Round Lots	Odd Lots (At)	Odd Lots (In)
All Exchanges		15.09%	8.53%	14.83%			
NASDAQ	Carteret	23.27%	16.60%	25.19%	23.53%	32.88%	28.62%
NASDAQ BX	Carteret	10.30%	2.27%	7.05%	3.96%	3.29%	3.85%
BZX	Secaucus	17.63%	9.72%	17.64%	11.61%	7.99%	9.93%
BYX	Secaucus	5.60%	0.97%	4.64%	9.46%	10.26%	10.31%
EDGX	Secaucus	14.20%	6.22%	17.35%	7.36%	7.57%	8.90%
EDGA	Secaucus	7.12%	1.08%	5.50%	5.85%	7.45%	7.50%
IEX	Secaucus	30.59%	13.59%	14.16%	6.71%	1.89%	3.08%
NYSE Arca	Mahwah	10.17%	5.45%	8.19%	11.74%	11.85%	9.76%
NYSE	Mahwah	9.70%	3.59%	10.08%	14.81%	13.79%	15.38%

Panel B: Nasdaq SIP (Carteret)

		Signing Difference			% Of Obs.		
Exchange	Location	Round Lots	Odd Lots (At)	Odd Lots (In)	Round Lots	Odd Lots (At)	Odd Lots (In)
All Exchanges		11.52%	4.20%	12.31%			
NASDAQ	Carteret	8.93%	5.01%	8.45%	37.87%	50.19%	40.70%
NASDAQ BX	Carteret	13.74%	1.68%	3.65%	3.34%	2.43%	3.68%
BZX	Secaucus	15.05%	7.24%	17.21%	12.09%	7.98%	10.41%
BYX	Secaucus	8.32%	1.25%	5.61%	7.06%	7.11%	8.77%
EDGX	Secaucus	12.84%	4.55%	16.64%	7.77%	6.54%	8.78%
EDGA	Secaucus	8.06%	1.18%	6.32%	5.17%	6.03%	6.10%
IEX	Secaucus	32.55%	11.77%	12.32%	6.81%	1.60%	2.92%
NYSE Arca	Mahwah	7.47%	2.03%	23.93%	12.87%	14.49%	14.32%
NYSE	Mahwah	12.94%	3.22%	27.29%	2.89%	1.24%	1.73%

Notes. This table reports the percentage of trades where our Latency-Free (LF) signing methodology differs from the Lee-Ready methodology using the SIP NBBO midpoint. Panel A presents the numbers for NYSE SIP symbols that report trades to Mahwah. Panel B does the same except for Nasdaq SIP symbols that report to Carteret. We further separate the trade difference percentage by the exchange the trade took place on and whether it was a round/mixed lot, odd lot (< 100 shares) at the EX BBO ("Odd Lots (At)"), or odd lot inside the EX BBO ("Odd Lots (In)"). We also report the fraction of observations by each lot type on each exchange to provide context for which exchanges have the most activity. We only report figures for the top 9 exchanges, which account for more than 95% of trades.

Figure E.2: Signing Differences vs. Characteristics



Notes. This figure presents six binscatter plots of trade signing differences vs. several symbol characteristics. The signing difference, plotted on the y-axis, is the percentage of dollar volume where our Latency-Free (LF) and the SIP signing methodologies assign different signs for each symbol-day. Each of the six plots correspond to the following characteristics: log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. The unit of observation is at the symbol-day level. We also show the line from an OLS regression, and report the beta and t-stat at the bottom right of each plot. Standard errors are clustered at the symbol-day level.

including all characteristics are intraday volatility, which flips sign when including the other variables, and the coefficients on the indicators, which are economically large and statistically significant. The signing differences are more than 2 pp greater for ETFs, whereas Nasdaq-listed symbols are more than 3 pp smaller than for NYSE SIP symbols. In the multivariate regression, the statistically strongest effects come from the quoted spread, intraday volatility, and the variance ratio.[56]

We can also report that symbol fixed effects soak up much of the variation captured by these variables. However, even with symbol and date fixed effects, the quoted spread and intraday volatility remain strongly statistically significant and have a similar magnitude. In addition, dollar volume increases in significance and magnitude, suggesting that dollar volume can

[56]Intraday volatility happens to be highly correlated with the quoted spread. However, estimating the multivariate regression without the quoted spread still leads to a positive coefficient for intraday volatility.

Table E.5: Signing Differences on Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Price	0.926						0.751
	(22.39)						(6.22)
Log $ Volume		0.704					-0.306
		(20.19)					(-2.97)
Log Quoted Spread			-1.321				-2.496
			(-20.76)				(-18.36)
Log Intraday Volatility				-0.957			2.584
				(-17.25)			(19.89)
Log Abs. Intraday Ret.					-0.416		0.130
					(-9.55)		(3.51)
Variance Ratio						-5.521	-4.304
						(-16.57)	(-10.56)
𝕀[is ETF]							1.999
							(9.40)
𝕀[is UTP Symbol]							-3.795
							(-14.93)
Intercept	9.414	5.597	16.099	16.033	12.245	14.134	12.794
	(47.38)	(18.49)	(45.97)	(44.40)	(59.83)	(51.69)	(14.30)
# Observations	95,956	95,956	95,956	95,956	95,956	95,956	95,956
R^2	0.034	0.03	0.071	0.044	0.008	0.039	0.177

Notes. This table reports coefficients and t-statistics from regressing trade signing differences on characteristics at the symbol-day level. The signing difference is the dependent variable and is the percentage of dollar volume where our Latency-Free and the SIP signing methodologies assign different signs for each symbol-day. The independent variables are log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. For column 7, we include all characteristics, and also include an indicator that is 1 if the symbol is an ETF and if the symbol is Nasdaq-listed (reports to the Nasdaq SIP operated by Nasdaq). The unit of observation is at the symbol-day level. Standard errors are clustered at the symbol and day level.

help understand the variation in signing differences within a symbol across different trading days.

We also note that these regressions use signing difference percentages as the dependent variable. We can repeat the exercise using the true accuracy differences between the two signing methods, but we must limit the sample to only NYSE Arca trades. Table E.6 in the appendix reports the regression results. The patterns are very similar and help further validate our trade difference measure as a proxy for accuracy differences.

E.5 Order Imbalance

A key input into computing imbalances is signing trades. We examine trade imbalances in our sample using both the LF and SIP methods. Since imbalances are usually associated with

Table E.6: Signing Accuracy Differences on Characteristics: NYSE Arca

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log Price	1.118						1.757
	(26.27)						(17.90)
Log $ Volume		0.709					-1.053
		(23.59)					(-12.48)
Log Quoted Spread			-1.267				-2.154
			(-23.91)				(-19.11)
Log Intraday Volatility				-0.922			2.167
				(-18.60)			(16.67)
Log Abs. Intraday Ret.					-0.418		0.238
					(-7.64)		(6.38)
Variance Ratio						-5.310	-4.960
						(-21.44)	(-13.28)
$\mathbb{I}$[is ETF]							3.324
							(12.78)
$\mathbb{I}$[is UTP Symbol]							-1.839
							(-9.16)
Intercept	1.671	-1.594	8.797	8.749	5.107	6.916	9.479
	(11.18)	(-5.49)	(37.71)	(33.56)	(44.23)	(45.07)	(13.34)
# Observations	94,295	94,295	94,295	94,295	94,295	94,295	94,295
R^2	0.038	0.024	0.05	0.032	0.006	0.027	0.118

Notes. This table reports coefficients and t-statistics from regressing trade signing accuracy differences on characteristics at the symbol-day level. The accuracy difference is the dependent variable and is the percentage of dollar volume of our Latency-Free signing accuracy and the SIP signing accuracy for each symbol-day. We only consider NYSE Arca trades, where we have true trade directions. The independent variables are log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. For column 7, we include all characteristics, and also include an indicator that is 1 if the symbol is an ETF and if the symbol is Nasdaq-listed (reports to the Nasdaq SIP). The unit of observation is at the symbol-day level. Standard errors are clustered at the symbol-day level.

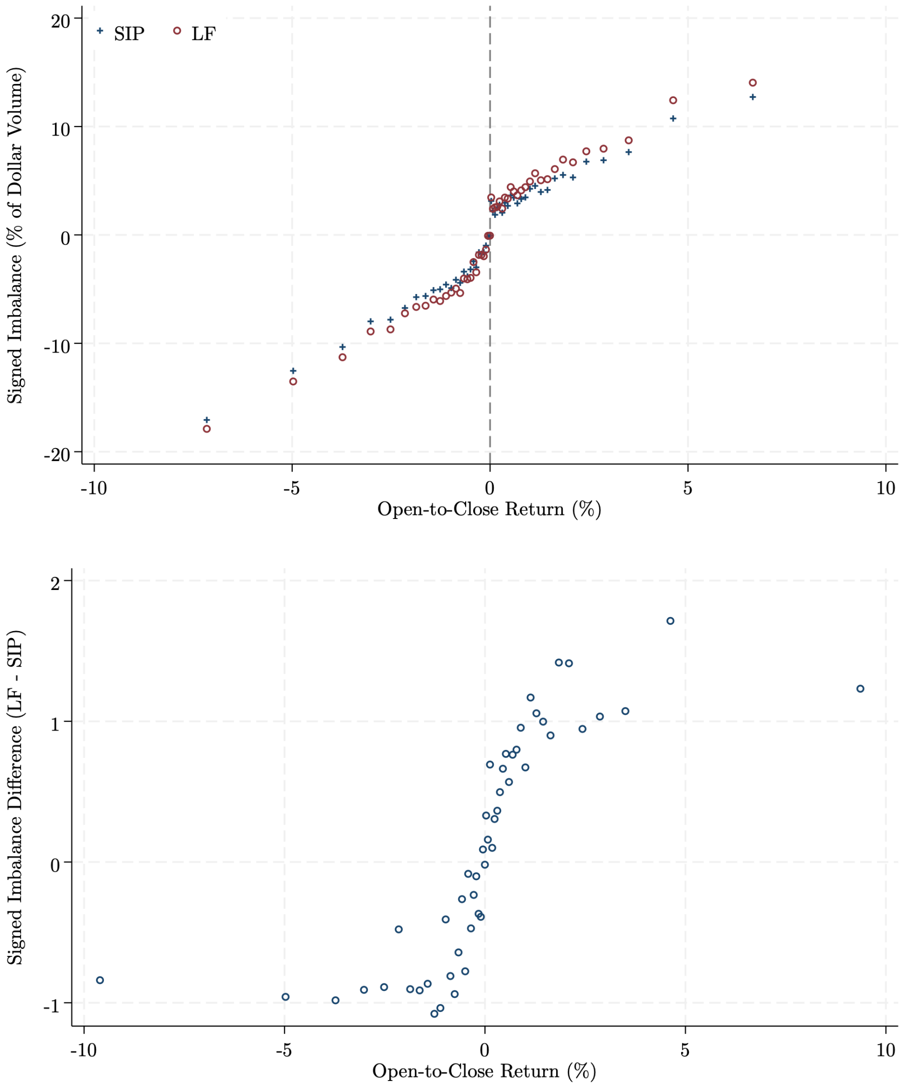
returns – more extreme returns materialize in part due to more extreme trade imbalances – we examine trade imbalance, computed separately for trades, signed using each methodology, against intraday returns.

We study the association between daily intraday (open-to-close) returns and dollar-volume-weighted order imbalances, where the imbalance is computed using trades signed with the LF and SIP methods. We do this for each symbol-day with at least 100 trades, and we winsorize the imbalance measures and intraday returns at the 1st and 99th percentiles.

The top plot of Figure E.3 shows the imbalance measures against open-to-close return. The LF imbalance measure has more extreme imbalance values relative to the SIP imbalance as returns increase in magnitude. A somewhat more muted imbalance is consistent with the signing and out-of-order results (and the spread results in the next section). That is, the SIP signing methodology tends to sign trades as sells when the trades are actually a buy followed by price adjustments shortly after – the price adjustments after the trade may be incorporated into the SIP ex-ante, which causes the SIP NBBO midpoint to increase in the direction of the trade and possibly even flip the sign of the trade.

The bottom plot of Figure E.3 shows the difference between the LF imbalance and the SIP imbalance. The plot show more clearly that the gap in imbalance increases as returns get further from zero. The range of imbalance differences is from -1% to almost 2%, and positive returns have a more steadily increasing gap than negative returns, where the gap is an imbalance difference of about 1 pp from -1% returns and below.

Figure E.3: Latency-Free and SIP Imbalance vs. Open-to-Close Return



Notes. The figure shows two binscatter plots on signing imbalance. The first shows the imbalance measures (buyer minus seller initiated dollar volume) based on trade signs assigned with the SIP methodology and our Latency-Free (LF) methodology on the y-axis. We plot the intraday return on the x-axis. The unit of observation is at the symbol-day level. The lower plot shows the LF imbalance measure minus the SIP imbalance measure on the y-axis, which highlights the differences between the two measures.

F Additional Effective Spread and Price Impact Results

F.1 Spreads by Geographic and SIP Latency

We provide additional tables to support and extend the analysis in Section 5.3. Specifically, we reproduce Table 8 from the main text, which shows how often the LF effective spread (ES) is greater than or less than the SIP ES, except in dollar volume terms instead of at the trade observation level. Similarly, we reproduce Table 9, which reports the estimated magnitude of the LF and SIP ES, weighted by dollar volume instead of equal-weighted across trades.

Table F.1 provides the results on what fraction of dollar volume have an LF effective spread greater than or less than the SIP effective spread. This table is identical to Table 8, except it reports the numbers based on dollar weights instead of by trade observations. The main takeaway is that the LF ES is slightly more likely to be greater than the SIP ES when measured in dollar terms. This applies to both NYSE and Nasdaq SIP symbols, though there is a slightly bigger percentage point increase for Nasdaq. The increase in LF > SIP in dollar terms relative to trade terms is distributed across exchanges. We also note that there is a slight increase of LF < SIP percentages in dollar terms relative to trade terms. The takeaway from Table F.1 is that larger dollar trades tend to be associated with more clustering of events around the trade, which leads to greater out-of-order issues and a more severe look-ahead bias in the SIP NBBO. This story is consistent with Easley and O'hara (1987) – trade size is positively correlated with information content.

We also reproduce Table 9 from the main text but weighted by dollar volume. Table F.2 shows the LF and SIP ES estimates in basis points for all trades in NYSE and Nasdaq SIP symbols, as well as for all trades in each exchange-SIP combination. The numbers are universally smaller, consistent with the well known fact that large-cap stocks that have more dollar volume have smaller spreads. This is also why NYSE SIP symbols have a smaller dollar-weighted ES than Nasdaq symbols. Other than that, the numbers are very similar to Table 9 in the main text. NYSE SIP symbols have a larger bias, though Nasdaq SIP symbols have a very slightly smaller bias.

Table F.1: Latency-Free vs. SIP: Effective Spread Comparison by Exchange in Dollars

		NYSE SIP (Mahwah)		Nasdaq SIP (Carteret)	
Exchange	Location	LF > SIP	LF < SIP	LF > SIP	LF < SIP
All Exchanges		23.11%	4.93%	20.55%	5.25%
NASDAQ	Carteret	44.73%	2.00%	21.72%	7.18%
NASDAQ BX	Carteret	15.38%	2.59%	7.02%	4.95%
BZX	Secaucus	28.29%	2.85%	28.21%	2.88%
BYX	Secaucus	7.55%	1.61%	11.34%	2.39%
EDGX	Secaucus	23.59%	2.52%	24.96%	2.91%
EDGA	Secaucus	10.47%	1.94%	13.49%	2.35%
IEX	Secaucus	5.54%	2.50%	4.28%	3.23%
NYSE Arca	Mahwah	14.26%	8.94%	23.57%	2.96%
NYSE	Mahwah	13.48%	10.69%	18.24%	3.25%

Notes. This table presents the percentage of trades where the effective spread (ES) measure differs when using the latency-free (LF) NBBO midpoint vs. the SIP NBBO midpoint. We sign every trade using our Latency-Free signing methodology. We separately report the percentage of trades where the LF ES is strictly less than and strictly greater than the SIP ES. We compute the percentages for all trades on each SIP ("All Exchanges"), but also compute the percentages for each exchange-SIP combination (where the trade occurred and which SIP the symbol reports to). For the results broken out by exchange, we use the nine highest-volume exchanges, which account for more than 95% of trades.

F.2 Price Impact Results

In the spreads analysis in Section 5.3, we focus only on effective spreads. Here, we expand the analysis to price impact at three different horizons (500 ms, 1 min, 5 mins).

Conceptually, the price impact measures should be highly correlated with the effective spread measures because both are anchored by some notion of a prevailing midpoint. Our approach, using the LF NBBO midpoint (LF), and the approach predominantly used in the literature, using the SIP NBBO midpoint (SIP), both use the same computation for spreads – the log difference between the trade price and the midpoint. The only difference is the choice of midpoint. Similarly, price impact is computed as the signed log difference between the midpoint price at some point in the future and the prevailing midpoint at the time of the trade. Because the midpoint in the future is at a somewhat random point in time throughout the day and most points in time have no activity (as documented in Section 4.2), the LF and SIP midpoints in the future are almost always identical. This holds even if we look 500 milliseconds into the future, a relatively short period of time based on human perception but long relative to the clustering of activity. For reference, we note that most 500 *microsecond*

Table F.2: Latency-Free vs. SIP: Effective Spread by Exchange in Dollars

		NYSE SIP (Mahwah)			Nasdaq SIP (Carteret)		
Exchange	Location	LF ES	SIP ES	% Diff	LF ES	SIP ES	% Diff
All Exchanges		3.14	2.55	-18.95%	5.86	5.27	-10.01%
NASDAQ	Carteret	2.94	1.51	-48.55%	7.43	6.78	-8.77%
NASDAQ BX	Carteret	3.08	2.75	-10.56%	3.80	3.66	-3.56%
BZX	Secaucus	2.60	1.88	-27.51%	3.76	2.98	-20.82%
BYX	Secaucus	3.57	3.38	-5.35%	5.10	4.70	-7.82%
EDGX	Secaucus	2.85	2.16	-24.17%	4.66	3.81	-18.29%
EDGA	Secaucus	3.41	3.20	-6.14%	4.78	4.41	-7.87%
IEX	Secaucus	0.99	0.89	-10.03%	1.31	1.28	-2.08%
NYSE Arca	Mahwah	3.39	3.18	-6.10%	4.69	4.12	-11.99%
NYSE	Mahwah	3.82	3.63	-4.97%	2.86	2.47	-13.46%

Notes. This table presents the effective spread, measured in basis points. We compute the spread using the latency-free (LF) NBBO midpoint and the SIP NBBO midpoint, then report the percentage difference the SIP measure is from the LF measure (e.g., SIP ES / LF ES - 1). We sign every trade using our Latency-Free signing methodology. We compute the average effective spread for all trades on each SIP ("All Exchanges"), but also report the averages for each exchange-SIP combination (where the trade occurred and which SIP the symbol reports to). For the results broken out by exchange, we use the nine highest-volume exchanges, which account for more than 95% of trades.

intervals, a period of time 1000 times smaller than 500 *milliseconds*, have no activity. That means looking at some point in time after a trade is very unlikely to have another trade and has nothing happening more than 99% of the time for the typical stock.

If the future midpoint used to compute price impact is effectively fixed, then all variation comes from the prevailing midpoint and the fraction of the effective spread that is attributable to price impact/adverse selection (see the decomposition in Section 5.3.1 for details). To the extent the price impact results are different from the effective spread results amounts almost entirely to differences in the fraction of the effective spread that is for adverse selection. In addition, we do not analyze the realized spread because it is the remainder of the effective spread minus the adverse selection/price impact component – the ES and PI results together can be used to infer the realized spread results.

Table F.3 shows the results for price impact at 500 ms, 1 min, and 5 mins. We provide the results in trade and dollar terms, and decompose the sample three ways, just as in Table 7. The overall results, especially at the 500 microsecond level, are similar to the estimates for the effective spread. The difference between the SIP and LF price impact, what we interpret as bias, goes down as the price impact horizon increases. Part of this is due to the fact that the level of price impact goes up with horizon – at 1 minute and 5 minutes, the price impact is greater than the effective spread seen in Table 7. This indicates that more than the entire effective spread is lost to adverse selection by market makers over these horizons.[57] So, while the gap between the SIP and LF price impact measures in basis points is roughly constant, the level of price impact goes up with horizon, which makes the difference smaller. However, the mechanism is the same – the look-ahead bias in the SIP moves the SIP NBBO midpoint towards the trade price *and* the future price-impact-induced midpoint, which biases both effective spreads and price impact in the same way (just with different normalization).

We also present price impact estimates for the LF and SIP by exchange and SIP. Table F.4 presents the results, which correspond to the effective spread results in Table 9. The results are very similar, which is expected given the logic provided above. The most notable differences are from the bias in smaller taker-maker exchanges like BYX and EDGA. There, the price impact bias is greater than the effective spread bias from the main text. This is somewhat mechanical â these trades have less overall price impact, so while the bias is the same in basis points, when scaled relative to the amount of price impact, the bias is larger as a percentage (essentially normalizing by a smaller price impact number relative to a larger effective spread number). Otherwise, the patterns are very similar – the bias is largest on exchange-to-SIP combinations where latency is greatest.

F.3 Spreads, Price Impact, and Characteristics

Just as in Section E.4, our findings will have an even greater potential impact on future studies if spread biases are related to other variables used in research. We consider the same symbol-level characteristics as in Section E.4 and relate them to LF vs. SIP spread differences. Because characteristics are at the symbol-day level, we must aggregate ES and PI measures at the symbol-day level. We require a symbol-day to have at least 100 trades across all exchanges to be included in the sample. We compute spread differences by taking the dollar-weighted ES over all trades in a symbol-day, then taking the log difference between the SIP ES and the LF ES. We apply the same procedure for PI. We weight by dollar volume, just as in Section E.4, because it essentially weights the contribution of each trade (and odd lot trades) by the size within a symbol-day (but does not compare across symbols since we aggregate up to the symbol-day). We call these measures ESDiffPct and PIDiffPct.

[57] This does not necessarily mean that market makers experience losses. It could be that the worst trades tend to come from investors submitting limit orders that are disproportionately adversely selected and/or rebates may allow market makers to have positive trading profits because the net-of-fee price is more attractive to market makers than gross prices may suggest.

Table F.3: Latency-Free vs. SIP: Price Impact Magnitudes

Panel A: Price Impact (500 ms)

	Trade Weighted			Dollar Weighted		
Sample	LF	SIP	% Diff	LF	SIP	% Diff
All Trades	5.305	4.380	-17.44%	3.439	2.901	-15.65%
Decomposition 1:						
Round Lots	6.214	5.069	-18.42%	3.536	2.989	-15.48%
Odd Lots (At EX BBO)	5.016	4.044	-19.38%	3.480	2.693	-22.61%
Odd Lots (Inside EX BBO)	3.347	3.003	-10.29%	2.461	2.232	-9.33%
Decomposition 2:						
NYSE SIP	4.431	3.530	-20.34%	3.072	2.482	-19.20%
Nasdaq SIP	6.968	5.997	-13.93%	4.091	3.645	-10.92%
Decomposition 3:						
SIP Not Locked/Crossed	5.046	4.309	-14.61%	3.231	2.905	-10.08%
SIP Locked/Crossed	8.908	5.374	-39.67%	6.133	2.843	-53.64%

Panel B: Price Impact (1 min)

	Trade Weighted			Dollar Weighted		
Sample	LF	SIP	% Diff	LF	SIP	% Diff
All Trades	6.875	5.982	-12.99%	4.397	3.905	-11.18%
Decomposition 1:						
Round Lots	7.853	6.772	-13.76%	4.491	3.999	-10.95%
Odd Lots (At EX BBO)	6.651	5.668	-14.77%	4.423	3.641	-17.69%
Odd Lots (Inside EX BBO)	4.687	4.338	-7.46%	3.463	3.232	-6.67%
Decomposition 2:						
NYSE SIP	5.663	4.799	-15.26%	3.713	3.126	-15.82%
Nasdaq SIP	9.181	8.233	-10.32%	5.612	5.291	-5.72%
Decomposition 3:						
SIP Not Locked/Crossed	6.700	5.996	-10.50%	4.239	3.973	-6.27%
SIP Locked/Crossed	9.310	5.790	-37.81%	6.439	3.024	-53.03%

Panel C: Price Impact (5 mins)

	Trade Weighted			Dollar Weighted		
Sample	LF	SIP	% Diff	LF	SIP	% Diff
All Trades	6.811	6.094	-10.53%	4.393	4.022	-8.43%
Decomposition 1:						
Round Lots	7.734	6.915	-10.59%	4.517	4.144	-8.24%
Odd Lots (At EX BBO)	6.608	5.740	-13.14%	4.176	3.564	-14.66%
Odd Lots (Inside EX BBO)	4.737	4.408	-6.95%	3.389	3.251	-4.08%
Decomposition 2:						
NYSE SIP	5.529	4.911	-11.18%	3.432	3.158	-7.97%
Nasdaq SIP	9.249	8.342	-9.80%	6.102	5.559	-8.90%
Decomposition 3:						
SIP Not Locked/Crossed	6.652	6.126	-7.91%	4.273	4.120	-3.58%
SIP Locked/Crossed	9.024	5.649	-37.40%	5.944	2.759	-53.58%

Notes. This table presents the price impact at 500 milliseconds (Panel A), 1 minute (Panel B), and 5 minutes (Panel C), measured in basis points. We compute each measure using the latency-free (LF) NBBO midpoint and the SIP NBBO midpoint. We aggregate across all trades in our sample, and report the trade-weighted and dollar-weighted average, then report the percentage difference the SIP measure is from the LF measure (e.g., SIP PI / LF PI - 1). We sign every trade using our Latency-Free signing methodology. We further decompose these percentages into three subsets: round and mixed lot vs. odd lot trades at the EX BBO vs. odd lot trades inside the EX BBO, NYSE SIP vs. Nasdaq SIP symbols, and when the SIP is or is not locked or crossed.

Table F.4: Latency-Free vs. SIP: Price Impact by Exchange

Panel A: Price Impact (500 ms)

		NYSE SIP (Mahwah)			Nasdaq SIP (Carteret)		
Exchange	Location	LF PI	SIP PI	% Diff	LF PI	SIP PI	% Diff
All Exchanges		4.43	3.53	-20.34%	6.97	6.00	-13.94%
NASDAQ	Carteret	5.38	3.23	-40.02%	8.54	7.44	-12.88%
NASDAQ BX	Carteret	2.35	1.95	-16.75%	3.63	3.58	-1.42%
BZX	Secaucus	5.03	3.83	-23.90%	7.04	5.55	-21.16%
BYX	Secaucus	2.32	2.10	-9.49%	3.89	3.45	-11.49%
EDGX	Secaucus	5.34	4.29	-19.67%	7.84	6.49	-17.24%
EDGA	Secaucus	2.05	1.80	-11.94%	3.32	2.85	-14.34%
IEX	Secaucus	1.17	1.00	-14.71%	1.34	1.31	-2.02%
NYSE Arca	Mahwah	5.93	5.74	-3.19%	8.26	7.09	-14.15%
NYSE	Mahwah	4.73	4.66	-1.49%	5.14	4.40	-14.37%

Panel B: Price Impact (1 min)

		NYSE SIP (Mahwah)			Nasdaq SIP (Carteret)		
Exchange	Location	LF PI	SIP PI	% Diff	LF PI	SIP PI	% Diff
All Exchanges		5.66	4.80	-15.25%	9.18	8.23	-10.33%
NASDAQ	Carteret	6.21	4.10	-33.98%	10.98	9.93	-9.57%
NASDAQ BX	Carteret	4.27	3.72	-12.92%	5.84	5.70	-2.46%
BZX	Secaucus	5.61	4.48	-20.02%	8.52	6.99	-17.91%
BYX	Secaucus	4.58	4.35	-5.08%	6.26	5.66	-9.51%
EDGX	Secaucus	6.74	5.73	-15.04%	10.55	9.29	-11.96%
EDGA	Secaucus	3.46	3.06	-11.51%	4.84	4.34	-10.36%
IEX	Secaucus	1.61	1.51	-6.28%	1.81	1.77	-1.94%
NYSE Arca	Mahwah	7.07	6.95	-1.69%	11.04	9.99	-9.49%
NYSE	Mahwah	6.15	6.19	0.75%	6.29	5.54	-12.00%

Panel C: Price Impact (5 mins)

		NYSE SIP (Mahwah)			Nasdaq SIP (Carteret)		
Exchange	Location	LF PI	SIP PI	% Diff	LF PI	SIP PI	% Diff
All Exchanges		5.53	4.91	-11.17%	9.24	8.34	-9.78%
NASDAQ	Carteret	5.89	4.04	-31.42%	10.93	9.99	-8.60%
NASDAQ BX	Carteret	4.55	4.00	-12.11%	6.30	6.12	-2.84%
BZX	Secaucus	5.30	4.51	-14.81%	8.61	7.07	-17.89%
BYX	Secaucus	4.95	4.68	-5.48%	7.26	6.26	-13.72%
EDGX	Secaucus	6.44	5.82	-9.67%	10.50	9.36	-10.86%
EDGA	Secaucus	3.75	3.16	-15.82%	4.94	4.51	-8.74%
IEX	Secaucus	1.44	1.52	5.59%	2.05	1.70	-17.10%
NYSE Arca	Mahwah	6.93	7.05	1.65%	10.80	9.99	-7.51%
NYSE	Mahwah	5.90	6.47	9.68%	6.21	5.39	-13.13%

Notes. This table presents the price impact at 500 microseconds (Panel A), 1 minute (Panel B), and 5 minutes (Panel C), measured in basis points. We compute each measure using the latency-free (LF) NBBO midpoint and the SIP NBBO midpoint, then report the percentage difference the SIP measure is from the LF measure (e.g., SIP PI / LF PI - 1). We sign every trade using our Latency-Free signing methodology. We compute the average price impact for all trades on each SIP ("All Exchanges"), but also report the averages for each exchange-SIP combination (where the trade occurred and which SIP the symbol reports to). For the results broken out by exchange, we use the nine highest-volume exchanges, which account for more than 95% of trades.

While our LF ES and PI measures are much more stable than the SIP counterparts, there are still some extreme observations, since LF ES and LF PI can be extremely small for outlier symbol-days, particularly for symbol-days with very little trading activity. As a result, we winsorize each measure at the 2.5 and 97.5 percentiles.[58] We use ES and PI differences as the dependent variable in estimating

$$\text{DiffPct}_{i,d} = \alpha + \beta \cdot x_{i,d} + \varepsilon_{i,d}, \tag{F.1}$$

where $x_{i,d}$ is a symbol-level characteristic. We use the same characteristics as the tests in Section E.4, where we examine signing differences and characteristics: log closing share price, log intraday dollar volume, log of the winsorized percent quoted spread, log of the winsorized annualized intraday volatility, log absolute intraday return, and the 1-minute variance ratio.

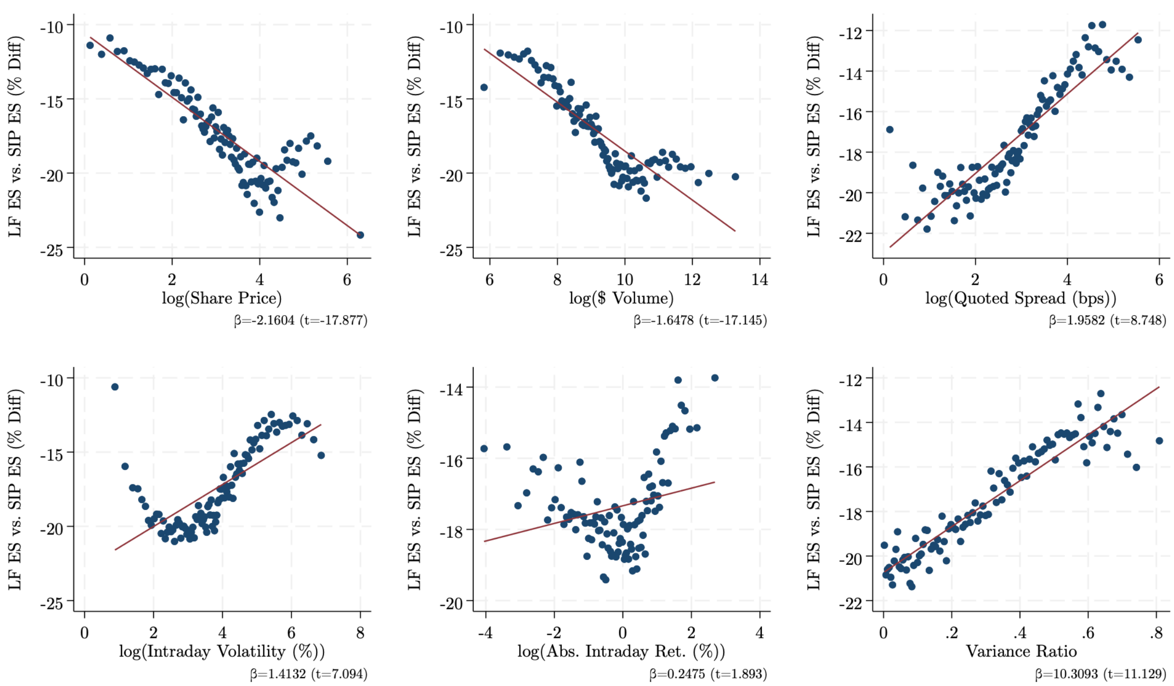
We first present binscatter plots, which show the data and the linear relationship between spread differences and characteristics. Figure F.1 presents the plots. The correlation between ES differences and these six characteristics mirrors the signing-characteristics plots in Figure E.2. For example, as the share price and dollar volume increase, the SIP ES is even smaller relative to the LF ES *and* signing differences are larger. These patterns further emphasize that these characteristics seem to be jointly related to trade signing and spread differences. For example, lower intraday volatility, lower variance ratios, lower quoted spreads, higher share prices, and higher dollar volume are all associated with greater trade signing differences and even smaller SIP effective spreads.

We also repeat the exercise for our price impact measure and present the binscatter plots in Figure F.2. Interestingly, the patterns flip for intraday volatility, quoted spreads, and variance ratios. All three of these characteristics had a positive relationship with the ES difference – for example, a smaller variance ratio is associated with a relatively smaller SIP ES. However, these characteristics are negatively related to the PI difference – a smaller PI difference is related to a *larger* variance ratio.

Understanding the change in patterns is somewhat nuanced. It is easy to understand why the relationship between ES differences and characteristics is mirrored by trade signing differences and characteristics: both are highly contingent on the look-ahead bias in the SIP NBBO (the trade price is "fixed"). The measure of SIP PI is slightly different in that it depends not only on the SIP NBBO, but also on future prices. If there is a correlation

[58] The results are not very sensitive to the choice of winsorization cutoffs, as long as there is some winsorization (e.g., the results hold for even 0.5 and 99.5). We use these cutoffs because they lead to what we think are more realistic magnitudes for the outliers.

Figure F.1: Effective Spread Differences vs. Characteristics

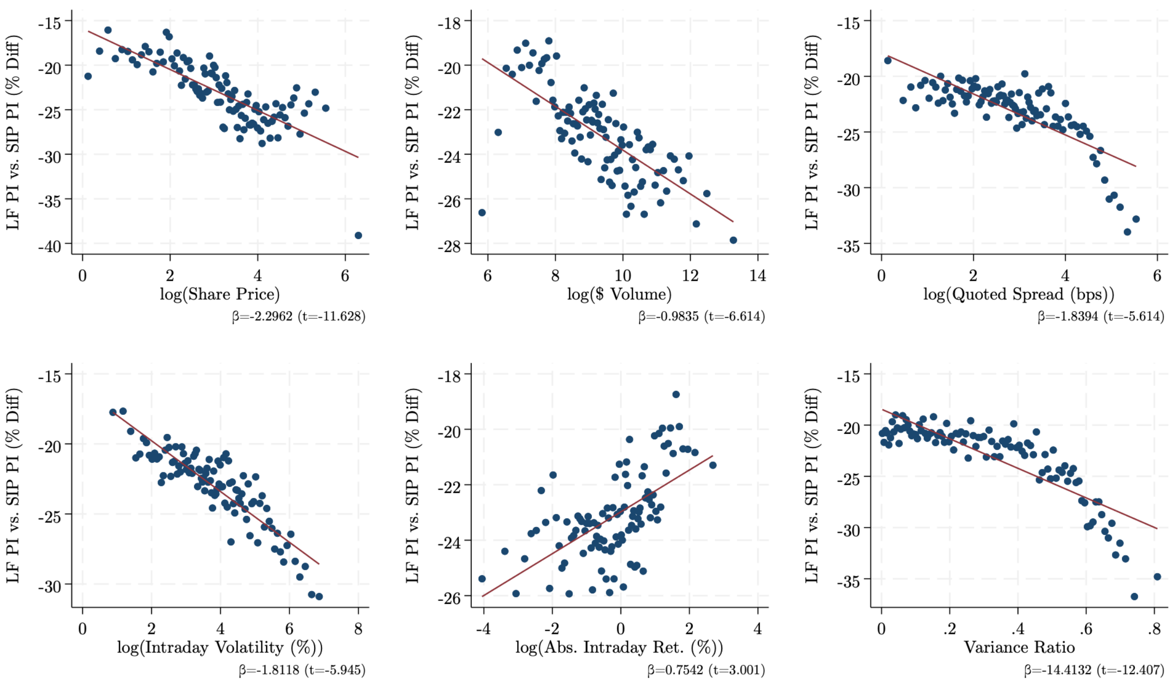


Notes. This figure presents six binscatter plots of the difference in the effective spread (ES) using two different NBBO benchmarks vs. several symbol characteristics. The ES difference, plotted on the y-axis, is the log difference in ES computed with the latency-free (LF) NBBO minus the SIP NBBO. The effective spread is computed using the LF and SIP NBBOs for each symbol day. Because this object can be unstable with few observations, we require at least 100 trades per symbol-day to be included. Each of the six plots correspond to the following characteristics: log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. The unit of observation is at the symbol-day level. We also show the line from an OLS regression, and report the beta and t-stat at the bottom right of each plot. Standard errors are clustered at the symbol-day level.

between price impact and how much the SIP NBBO midpoint erroneously moves ex-ante, it can lead to different results for the PI patterns in Figure F.2 than the ES patterns in Figure F.1.

We present the regression results from estimating Equation F.1 for ES and PI differences. We estimate univariate regressions, then estimate a multivariate regression with all characteristics. The multivariate regressions also include indicators for whether the symbol is an ETF or whether the symbol is a Nasdaq SIP symbol. As highlighted before, some characteristics have the same correlation with ES and PI (e.g., price and dollar volume). Others flip signs when added to the multivariate regression, and some characteristics have different signs for ES and PI.

Figure F.2: Price Impact Differences vs. Characteristics



Notes. This figure presents six binscatter plots of the difference in the 0.5 second Price Impact (PI) using two different NBBO benchmarks vs. several symbol characteristics. The PI difference, plotted on the y-axis, is the log difference in PI computed with the latency-free (LF) NBBO minus the SIP NBBO. The price impact is computed using the LF and SIP NBBOs for each symbol day. Because this object can be unstable with few observations, we require at least 100 trades per symbol-day to be included. Each of the six plots correspond to the following characteristics: log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. The unit of observation is at the symbol-day level. We also show the line from an OLS regression, and report the beta and t-stat at the bottom right of each plot. Standard errors are clustered at the symbol-day level.

Table F.5: Effective Spread and Price Impact Differences on Characteristics

	LF vs. SIP Effective Spread							LF vs. SIP Price Impact (500 ms)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Log Price	-2.160						-3.026	-2.296						-7.597
	(-17.88)						(-10.37)	(-11.63)						(-18.32)
Log $ Volume		-1.648					1.179		-0.983					2.428
		(-17.14)					(4.73)		(-6.61)					(7.60)
Log Quoted Spread			1.958				3.442			-1.839				-2.533
			(8.75)				(9.29)			(-5.61)				(-5.46)
Log Intraday Volatility				1.413			-3.317				-1.812			-2.192
				(7.09)			(-11.39)				(-5.94)			(-5.60)
Log Abs. Intraday Ret.					0.247		-0.509					0.754		0.461
					(1.89)		(-5.67)					(3.00)		(3.44)
Variance Ratio						10.309	6.953						-14.413	-7.772
						(11.13)	(7.36)						(-12.41)	(-7.65)
$\mathbb{I}$[is ETF]							4.163							-4.087
							(8.30)							(-6.21)
$\mathbb{I}$[is UTP Symbol]							5.668							3.560
							(6.20)							(2.68)
Intercept	-10.572	-2.042	-22.961	-22.84	-17.336	-20.742	-21.193	-15.895	-13.98	-17.907	-16.152	-22.977	-18.453	-4.452
	(-22.42)	(-2.38)	(-25.41)	(-22.39)	(-43.86)	(-33.45)	(-8.71)	(-22.84)	(-10.28)	(-18.40)	(-13.51)	(-52.56)	(-32.33)	(-1.71)
# Observations	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584	95,584
R^2	0.018	0.015	0.016	0.010	0.000	0.013	0.051	0.007	0.002	0.005	0.006	0.001	0.009	0.040

Notes. This table reports coefficients and t-statistics from regressing effective spread (ES) and 0.5 second price impact (PI) differences on characteristics at the symbol-day level. The ES and PI differences are the dependent variables and are the log differences in ES and PI computed with the latency-free (LF) NBBO minus the SIP NBBO. The ES and PI are computed using the LF and SIP NBBOs for each symbol day. Because this object can be unstable with few observations, we require at least 100 trades per symbol-day to be included. The independent variables are log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. Columns 1 to 7 show the ES results, and columns 8 to 14 show the PI results. For column 7 and 14, we include all characteristics, and also include an indicator that is 1 if the symbol is an ETF and if the symbol is Nasdaq-listed (reports to the Nasdaq SIP operated by Nasdaq). The unit of observation is at the symbol-day level. Standard errors are clustered at the symbol-day level.

G Diagnostic Information for Signing Accuracy and Spread Bias

In this appendix, we outline our signing methodology, describe where we expect it to be accurate (and inaccurate), and identify what types of studies are most likely affected (and unaffected) by using the de-facto standard SIP NBBO Lee-Ready signing methodology.

G.1 Proposed Signing Methodology

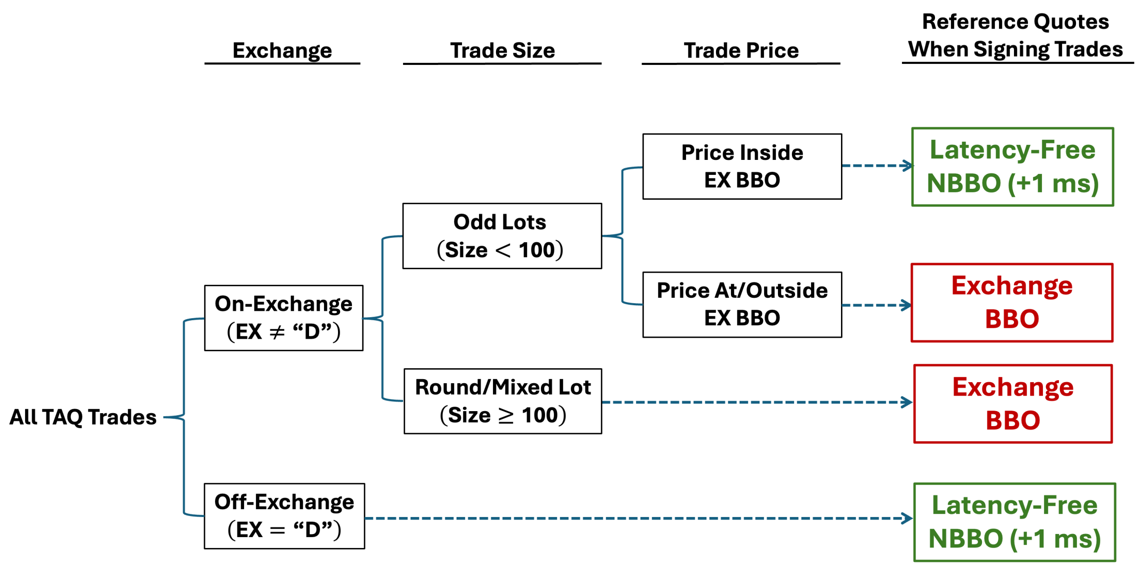
For clarity and accessibility, we present the full signing methodology proposed in Section 5.2.1 in Figure G.1. The figure presents a chart on how to sign trades based on three observables in the TAQ data: exchange, size, and price.

As discussed in Appendix B.1, one source of uncertainty in the proposed signing methodology is that we do not have the true sign of executions against hidden limit orders in the book. Since we cannot directly observe these types of trades, we do not know how our signing methodology performs with these trades. We also describe these trades in Appendix B.1, and they appear to be somewhat different than the sample of Arca trades we do have, including that there are, unsurprisingly, round lot trades inside the EX BBO.

As discussed in Appendix B.1, we think our signing methodology is likely to perform best for the trades we do not have in our sample. The reason why we have a different methodology for odd lot trades inside the EX BBO and off-exchange trades is that they may be more likely to trade when the market is moving and there are stale orders that get picked off. We believe that hidden round lot trades inside the EX BBO or NBBO are unlikely to be stale in this manner, and thus recommend the EX BBO to sign all round lot trades. However, we are open to simplifying the rule to allow for all exchange trades to be signed using the EX BBO if the trade price is at our outside the EX BBO, otherwise, use the delayed LF NBBO. The logic behind such a methodology is to make a clear dividing line for trades we can definitively sign accurately (EX BBO trades) and those where the signing is ambiguous.

While we cannot test this alternative methodology for many trades that will likely fall inside the EX BBO (the hidden order executions), we have tested the methodology for our Arca direct-feed sample that we can infer the true trade direction. We find that this alternative is marginally worse than the methodology we propose in Figure G.1. Again, we raise the alternative here because it is understandable that some researchers may want to take a stand

Figure G.1: Signing Methodology



Notes. This figure presents our proposed signing methodology using a flow chart. We start with all TAQ trades, then follow the chart based on observables on TAQ: the exchange the trade took place on (using the TAQ exchange code), the trade size, and the trade price. Based on the chart, each trade is signed using Lee-Ready with a reference price of either the midpoint of the exchange BBO or the midpoint of the latency-free NBBO delayed by 1 millisecond.

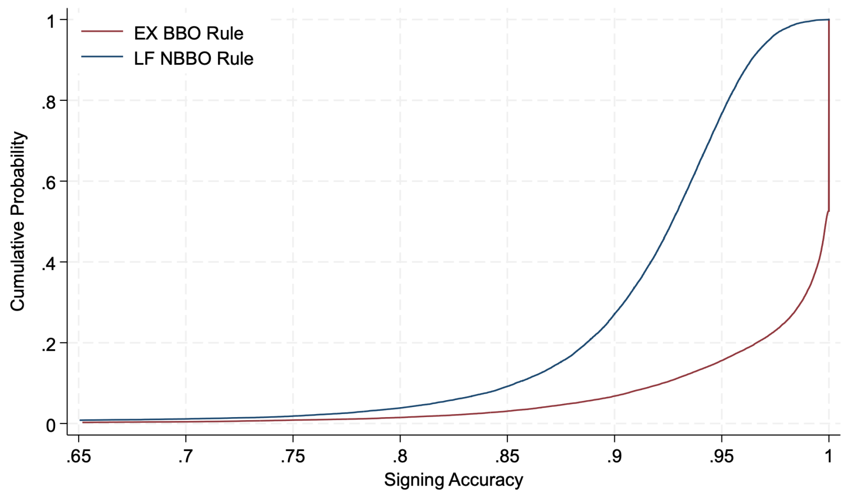
on how to sign hidden order executions and/or off-exchange trades. This is because these hidden trades may be more similar to the odd lot trades inside the EX BBO, which are signed more accurately with the LF NBBO than the EX BBO. We raise this alternative methodology to acknowledge that it also has similar logical backing to our main proposed methodology.[59]

G.2 Where Does the Latency-Free Signing Methodology Perform Well?

In this appendix, we describe where our signing methodology for Arca exchange trades from our direct-feed data, introduced in Section 5.2, is highly accurate and in what situations it is less accurate.

[59]As a reminder, Battalio et al. (2026) are able to sign some of these hidden order executions, specifically from retail order wholesaler(s) using a proprietary dataset. They find that the Latency-Free signing methodology works about as well as direct-feed data that was provided by the wholesaler(s) themselves.

Figure G.2: CDF of Signing Accuracy by Signing Rule Type



Notes. This figure presents two CDFs for signing accuracy of NYSE Arca trades. We split the sample into all trades that use the EX BBO rule and the LF NBBO rule, then compute signing accuracy within these two samples at the symbol-day level. We then separately plot the CDF of signing accuracy at the symbol-day for the EX BBO and LF NBBO trades. We require at least 100 trades within a symbol-day-rule to be included.

To get an overall sense of accuracy, we provide a CDF of signing accuracy for NYSE Arca trades, where the unit of observation is at the symbol-day-rule. By rule, we mean the set of trades that use the EX BBO rule (i.e., round lots and all odd lot trades at or outside the prevailing EX BBO) and all other trades that use the delayed LF NBBO rule (i.e., odd lots at a trade price inside the EX BBO). We use all symbol-days with at least 100 trades within each rule, and we use NYSE Arca trades in the direct-feed dataset since we can perfectly measure accuracy. Note that since the direct-feed data do not allow us to sign hidden order executions, as described in Appendix B.1, we omit these trades from our sample.

Figure G.2 provides the CDF for accuracy separately for all trades at the symbol-day level using the EX BBO and trades using the delayed LF NBBO. The figure shows that a significant fraction of symbol-days – nearly 50% – have an EX BBO accuracy of 100%. The errors for EX BBO trades come from price slide trades, described in Section 5.2, and other hidden order executions. Perhaps unsurprisingly, the LF NBBO trades are significantly worse, with around 20% of symbol-days with an accuracy of around 95% or more, and 75% of symbols-days with an accuracy of 90% or more. There is inherently more guesswork in these trades because any trade that happens at prices different from the best prevailing quotes are usually matched with (odd lot) limit orders we cannot observe. As mentioned in the main text, these trades will all be signed using the EX BBO once odd lot limit orders are displayed and disseminated on the SIP starting in May 2026.

We also provide regressions of signing accuracy at the symbol-day-rule level on characteristics to get a sense of where our signing methodology is most likely to have errors. This approach is distinct from the characteristics analysis in Appendix E.4 in that we are assessing overall accuracy here instead of comparing our signing methodology with the Lee-Ready SIP methodology (i.e., accuracy instead of accuracy differences). This is the sense in which this appendix provides a guide for where our signing methodology is likely to produce more or less accurate inference.

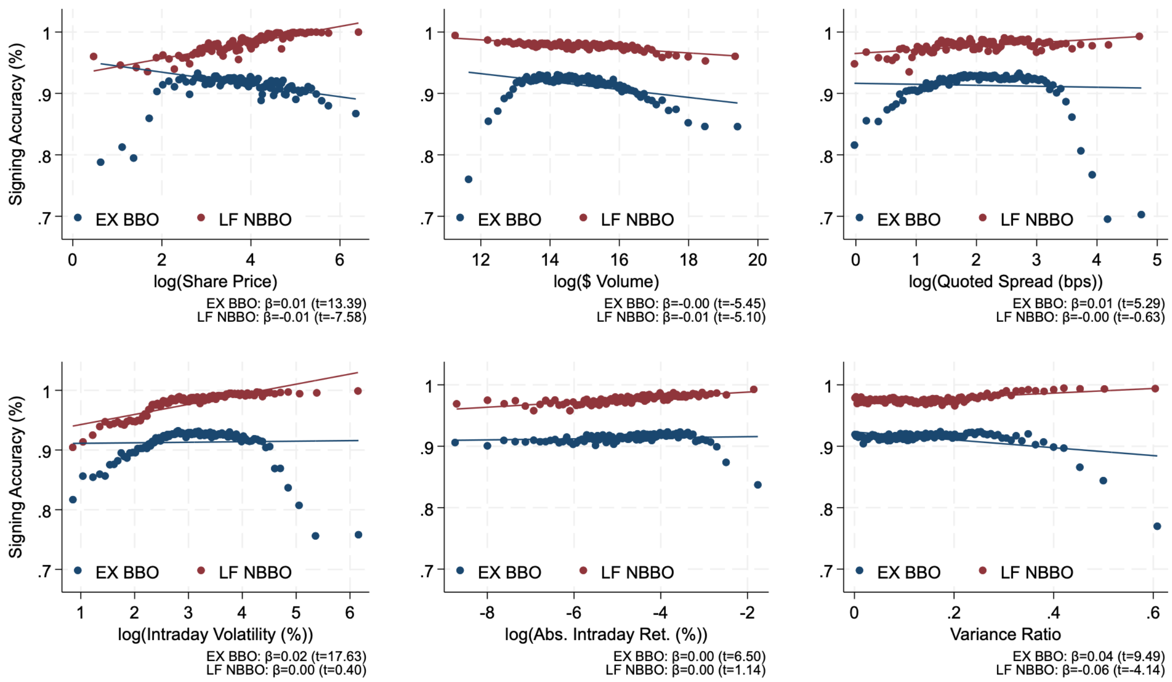
Figure G.3 provides binscatter plots to show how signing accuracy varies with characteristics. In every plot, the EX BBO method reliably outperforms the LF NBBO method, suggesting that symbols that have disproportionately more odd lot trades inside the EX BBO are likely to have more signing errors. There are some statistically significant patterns for the EX BBO rule, but they have small beta estimates. For example, price, volume, quoted spread, and absolute return all have statistically significant slopes for the EX BBO method, but the slopes are economically small at 0.01 or less. Thus, we would predict that a symbol with a share price of \$10 would have a 2.3 pp reduction in signing accuracy for EX BBO-signed trades compared to a symbol with a share price of \$100 ($0.01 \times (\log(100) - \log(10)) = 0.023$). Examining the non-linearities and outliers, the plots also reveal that symbols with extremely low volatility, extremely low quoted spreads, and extremely low share price tend to be unusually inaccurate among trades using the EX BBO rule. Overall, the very worst accuracy comes from trades using the LF NBBO rule that have extremely low share price ($<$ \$5), extremely low volume ($<$ \$200K), extremely low or high quoted spreads ($<$ 1 bps or $>$35 bps), extremely high intraday volatility, or extremely high variance ratios (all of these characteristics are undoubtedly correlated). In these extreme cases, the accuracy is still quite low and can range from 70% to 85%.

G.3 What Types of Analyses are Affected?

Our paper identifies the root cause of the look-ahead bias in the TAQ data when using the SIP NBBO and SIP timestamps. The look-ahead bias creates trade signing errors and a bias in the effective spread. In this appendix, we discuss when these issues are most likely to be problematic for researchers, policy makers, and other consumers of the TAQ data. We focus on the types of analyses that use trade signing as a necessary procedure to conduct economic analysis. Some types of analysis may fall into more than one of the categories below.

Since this section likely affects hundreds or even thousands of studies, we do not cite any

Figure G.3: NYSE Arca Signing Accuracy vs. Characteristics



Notes. This figure presents six binscatter plots of NYSE Arca signing accuracy vs. several symbol characteristics. We separate the sample into trades signed with the EX BBO rule and the delayed LF NBBO rule. The signing accuracy is plotted on the y-axis, and is computed for each symbol-day with at least 100 NYSE Arca trades for a given signing rule. Each of the six plots correspond to the following characteristics, also computed at the symbol-day: log of closing share price, log of dollar trading volume, log of the time-weighted quoted spread in basis points, log of the annualized intraday 1-minute volatility in percent, log of the absolute percent intraday return, and the variance ratio. We also show the line from an OLS regression for each rule, and report the beta and t-stat at the bottom right of each plot. Standard errors are clustered at the symbol-day level.

specific papers but rather provide a characterization of how we think our results will likely affect a particular *type* of study. We leave it to authors and researchers to assess their specific use case via the guide below. Of course, these critiques apply more specifically to studies using the TAQ, but they can also apply to studies using other types of aggregated trade and quote data.

Affected Studies

1. **Estimate Effective Spreads.** Of all TAQ variables constructed, one of the most widely used measures throughout finance is the effective spread. As we document in detail in Section 5.3, the effective spread is biased downwards because of the look-ahead bias in the SIP that we document in Section 4.3. A plain use of the effective spread is

likely to understate transaction costs. However, there are subtleties – as we describe in Appendix E.4, the bias is correlated with microstructure variables like dollar volume, intraday volatility, share price, etc. This means studies that use the effective spread as a control or outcome variable in a cross-sectional or panel setting will likely have a more complex set of biases to address. That is, studies will have to address both the average negative bias and the bias that varies across symbols and time. This is potentially problematic when the variable of interest is the effective spread itself or when it is used as a control.

2. **Estimate Price Impact or Realized Spreads.** Effective spreads are often decomposed into two components: price impact/adverse selection and the realized spread. Each component is typically measured as

$$\mathrm{RS}_{i,t} = 2 \times \mathrm{Sign}_{i,t} \times \log\left(\frac{p_{i,t}}{m_{i,t+\ell}}\right) \tag{G.1}$$

$$\mathrm{PI}_{i,t} = 2 \times \mathrm{Sign}_{i,t} \times \log\left(\frac{m_{i,t+\ell}}{m_{i,t}}\right), \tag{G.2}$$

for symbol i for trade at time t, where p and m represent the trade price and midpoint, respectively, and ℓ is some interval of time to examine the future midpoint (e.g., 1 minute after the trade). This equation makes clear that the bias in the effective spread shows up almost completely in the price impact component of the effective spread. The reason is that the key object that creates the bias is $m_{i,t}$, which contains information from after the trade. That is, the same mechanism that creates a bias in the effective spread also creates a bias in price impact. The realized spread is less affected because it uses two objects that have little to no bias – the trade price and the midpoint from some point in the future. The second reference price is worth clarification – as long as the interval of time is longer than a 1 ms (most use at least a second), the future midpoint price is likely to land in an empty interval (Section 4.2 shows that more than 99% of 500 μs intervals for nearly all symbols have nothing happening). That is, the midpoint in the future is far enough in the future that it avoids the cluster of events near the trade itself and lands in an empty interval where latency will not have an affect on the prevailing quotes.

This means that the realized spread is likely measured relatively accurately, but price impact measures are biased. That means all of the critiques for the effective spread also apply to price impact. In addition, studies that use the adverse selection contribution to the effective spread (average price impact divided by average effective spread) are likely to understate the adverse selection contribution (and, similarly, overstate the realized spread contribution).

3. **Contributions to Price Discovery.** Many papers in the microstructure literature test how trades and quotes contribute to price discovery. These papers use data at the trade and time level. As we will discuss below, studies at the trade level are specifically affected by the SIP look-ahead bias. But in addition to the unit of observation, many of these studies look specifically at the sequence of trades and quote changes and interpret how past events contribute to the permanent component of price changes in the future. That is, sequencing is at the heart of the analysis of many of these studies. While the types of analysis is wide-ranging, the look-ahead bias we document suggests that many of these studies will understate the contribution of trades to price discovery because some of the permanent price change induced by a trade will be erroneously counted as occurring before a trade. Just as above, since the bias is related to symbol characteristics, there are many more complex subtleties that will depend on the specific setting of the analysis. Time-weighted analyses, that use regular calendar time intervals, are less prone to be affected. For example, price discovery in the cross-section of symbols sometimes uses calendar time returns and aggregates activity within that time to make for better apples-to-apples comparisons across symbols. As discussed further below in the "Unaffected Studies" section, time-weighted analyses are likely to put more weight on periods where there is no activity.

4. **Informed Trading Measures.** Many measures of informed trading or adverse selection rely on buy-sell imbalances. As described in Appendix E.5, imbalance measures are likely attenuated towards zero, especially with trades that have more sudden price impact after the trade. Since the SIP will erroneously report events after the trade as occurring before the trade, the prevailing midpoint will move towards the trade price and increase the chance that the trade will flip sign. As an example, this means that on a day with more buyer-initiated trades, more of the buy trades will be signed incorrectly, resulting in the imbalance measure showing less buying than what actually occurred on the day. This is likely to feed into informed trading measures as attenuation bias – informed trading is likely more extreme than the measures suggest.

5. **Imbalance Measures.** While imbalance measures are sometimes used as an input into informed trading measures or adverse selection costs (see above), imbalance is also used as a control variable in the analysis of short-selling, etc. Since a SIP imbalance measure is subject to attenuation bias, when included in a regression as a control variable, it will not fully strip out the effects of imbalances, biasing the coefficient on

the variable of interest. In addition, imbalance measures are sometimes used to proxy for dealer inventory, both as an empirical measure and as an input to models. These will be similarly attenuated.

6. **Event Studies.** One of the most common ways to use TAQ is to study trading activity leading up to and after an event. The types of events studies covers an extremely wide range: earnings announcements, FOMC and other macroeconomic news announcements, index reconstitutions, derivatives expiration dates, stock splits and reverse splits, IPOs and SEOs, merger and acquisition announcements, extreme market days (e.g., the "Flash Crash"), days with a change in securities regulation, ex-dividend dates, analyst and credit rating changes, and days with insider trading. Some of these studies use TAQ for intraday prices at regular time intervals, which are likely unaffected by our findings (see "Unaffected Studies" below). However, many examine the price informativeness of trades, effective spread and price impact dynamics, lead-lag relationship across securities, order imbalance patterns, and indicators of abnormal trading, all of which are likely to be affected by the look-ahead bias in the SIP. In addition, some of these events are used as exogenous shocks in, e.g., a diff-in-diff setting. However, these shocks are likely to affect symbol characteristics (e.g., share price or trading volume), which are related to the error rate for signing (Appendix E.4) and the size of the bias in spreads and price impact (Appendix F.3). This means that the look-ahead bias present in the SIP may affect not only the actual measures used in the analysis, but also the *changes* in measures conditional on the event itself.

7. **Trade Time and Trade Weights.** Studies are often conducted either implicitly or explicitly in trade time (also referred to as business time or event time). These studies either have the unit of observation at the trade level (which implicitly means the analysis uses trade time) or explicitly weight by trade observations. These studies are highly susceptible to the look-ahead bias we document. The reason is exactly because of the clustering and out-of-order issues we document in Sections 4.2 and 4.3. In essence, these studies focus specifically on periods of time where the look-ahead bias is likely to manifest. This applies to analyses that use many of the variables already discussed (spreads, price impact, imbalances), but any microstructure variable where the sequence of quotes and or trades is relevant. This is in contrast to calendar time or time weighted studies, which we discuss further below.

8. **Non-U.S. Equity Market Studies.** Our study focuses only on the U.S. securities market. However, modern technology has allowed for the proliferation of venues, such that fragmentation is now increasingly common. This means that trading venues in

different geographical locations may offer trading in the same contracts or securities. This means that products like the SIP are becoming more prevalent. One example is the EuroCTP, which operates much like the SIP and is scheduled to be launched in Q3 2026. As markets become more technologically advanced, we think the look-ahead bias raised in this paper is likely to become more prevalent. However, with more advanced technology and better availability of data, we hope that solutions, like the ones we propose in Sections 5.2 and 5.3, can also be implemented. The first step in providing solutions is to make as much data from venues available on consolidated feeds, including, at the very least, venue-assigned timestamps.

Unaffected Studies

1. **Calendar Time Measures.** Many microstructure variables involve using calendar time. For example, the time-weighted quoted spread or measures of intraday volatility with 1-minute or 5-minute returns are commonly used measures constructed from TAQ data. These measures are only marginally affected by the look-ahead bias we document. The reason is that the prices and quotes used to construct these measures are very likely to be drawn from periods where nothing is happening – as documented in Section 4.2, in nearly all symbols, 99% of 500 μs intervals have literally no activity. Thus, that for any random point in time, there is likely nothing going on and there is nothing for the SIP to get sequentially out of order.

2. **Direct-Feed or Exchange Order Book Data.** These types of data come directly from an exchange and are extremely precise. Studies using this type of data are unlikely to be affected. However, if these data are used in conjunction with SIP data, they may still be subject to look-ahead bias concerns based on how and the extent to which the SIP data is used.

3. **Non-Fragmented Markets.** When examining symbols that only trade in non-fragmented markets, it is likely that the analysis is not subject to the look-ahead bias we document. We say “likely” because it is possible that the dissemination of trade and quote data, even from a single venue, may not be a perfect representation of order book data. in that case, the way in which data is disseminated, in particular the possible choice to disseminate quotes before trades, may still create a look-ahead bias.

4. **Studies Using Exchange Timestamps.** Exchange timestamps (i.e., participant timestamps) have been available via TAQ since 2015. There have been other studies

that use these exchange timestamps in lieu of SIP timestamps, some of which are mentioned in the related literature in the main text (Section 2.1). Since these timestamps are always sequentially accurate with respect to activity on a given exchange, these trades and quotes will not be prone to a look-ahead bias. It is possible to use the timestamps to generate a different type of bias, which is to use information that traders will not have had (e.g., our latency-free NBBO does this). However, our purposes are to create a benchmark for spread measurement. For signing, the prevailing EX BBO can always be observed before a trade, and the 1 ms delayed LF NBBO can (almost always) be observed by the time of the trade.[60]

[60]In principle, there can be transmission latencies that are longer than 1 ms, though they are rare and usually involve receiving data from exchanges that are not very active.