

In the blink of an eye: exchange-to-SIP latency and trade classification accuracy *

Craig Holden[†], Matthew Pierson[‡] and Jun Wu[§]

March 2025

Abstract

We develop two new methods to match trades and quotes with latency adjustments in the era of fast trading. The Relative Best Bid and Offer (RBBO) method reconstructs exchanges top-of-the-book, accounting for exchange-to-exchange direct feeds, data center co-location, and message latency. A simpler latency timestamp-adjusted method modifies Holden and Jacobsen (2014) with minimal code changes. Testing 650 million TAQ trades with order book data, both methods improve Lee and Ready (1991) classification accuracy from 86% to 92%, reshaping estimates of liquidity, order imbalance, and informed trading. These latency adjustments prevent false positives and ensure robust inferences in market microstructure research.

JEL Classification: C15, G12, G20

Keywords: Latency, Trade Classification, Exchange Network, Direct Feed

*The authors want to thank NYSE for providing data and support for this paper. The authors thank Robert Battalio, Dion Bongaerts, Allen Carrion, Bidisha Chakrabarty, Thomas Ernst, Björn Hagströmer, Pankaj Jain, Robert Jennings, Wendy Ju, Christos Koutsoyannis, Heinrich Lutjens, Phil Mackintosh, Albert Menkveld, Pamela Moulton, Rabih Moussawi, Luis Palacios, Eugenio Piazza, Steven Poser, Dale Rosenthal, Mehmet Saglam, Ion Lucas Saru, John Shim, Sunil Wahal, Chaojun Wang, Mao Ye, Anthony Zawadzki, Kathryn Zhao, Doris Zhou, and participants at the 2020 Indiana University Bloomington Seminar, the Microstructure Exchange, the 2023 FMA Chicago Annual Meeting, the 2023 SFA Annual Meeting, and the 2024 SQA/CQA Blackrock Trading Seminar for helpful comments.

[†]The late Professor Craig Holden is at Kelley School of Business, Indiana University Bloomington, [Professor Craig Holden's memorial website](#).

[‡]Matthew Pierson is at WRDS, the Wharton School, University of Pennsylvania, mpiers@wharton.upenn.edu.

[§]Jun Wu is at WRDS, the Wharton School, University of Pennsylvania, jun5@wharton.upenn.edu, corresponding author.

1 Introduction

Since the passage of Regulation National Market System (known as “Reg NMS”) in 2005, equity markets have become substantially faster, while competition between exchanges has led to greater geographical and market concentration (SEC (2020a)). With technological infrastructure improvements leading to processing times approaching the speed of light, the geography of exchange data centers has become the first-order determinant of equity market speed (NYSE (2018)). The U.S. equity markets have become focused on three cities in New Jersey – Mahwah, Carteret, and Secaucus – where, in 2020, 96% of all lit venue orders were executed. Datacenter geography has led to further practical changes in the local calculation of the National Best Bid and Offer (NBBO) by each exchange, with multiple “correct” NBBOs existing simultaneously at different exchanges. Industry leading market makers have adapted to this race by consuming exchanges’ direct feeds at co-located data centers in each of the three cities. A best execution algorithm then computes the Relative Best Bid and Offer (RBBO) for each venue and decides where to route the orders to maximize fill rate, minimize implementation shortfall, and comply with Reg NMS.¹

Yet the academic literature lags behind. Both current and past work evaluating financial markets relies on the Security Information Processor (SIP) consolidated data feed (i.e., NYSE’s TAQ data) and accurate trade initiation assignment. These studies have identified two methodological choices – trade assignment algorithms and the relationship between reported trade times and reported reference quote times in TAQ (Bessembinder (2003)). Prior literature has primarily focused on the former (e.g., Lee and Ready (1991), Lee and Radhakrishna (2000), Finucane (2000), Ellis, Michaely, and O’Hara (2000), Chakrabarty, Li, Nguyen, and Van Ness (2007), Easley, Lopez de Prado, and O’Hara (2016), and Duarte, Hu, and Young (2020)), but despite the monumental shifts in the speed of equity trading, conventional trade assignment methods continue to follow the methodology of Bessembinder (2003), which matches trades with instantaneously available quotes, or assigns arbitrary lags of 1 or 5 seconds. Holden and Jacobsen (2014) established the empirical practice to match trades with the prevailing previous

¹From discussions with major market makers, every microsecond in latency counts to ensure their best execution algorithm performance. In 2020, Virtu sued NYSE for the “last mile” of millimeter wave tower poles on the roof of NYSE’s datacenter. See Virtu’s SEC [comment letter](#) for more detail.

millisecond quotes in Daily TAQ data.² This practice is imperfect and researchers occasionally encounter abnormal cases where trades and quotes do not line up, causing biases in trade assignment and empirical measures.³

Our primary contribution is to uncover the cause of this timestamp mismatch and show that researchers do not accurately measure liquidity, order imbalance, and other measures in today’s U.S. equity markets. Neither the NBBO itself, as recorded by either of the two Securities Information Processors (SIPs), nor any of the conventional, arbitrary time adjustments made when matching trades with reference quotes, provide accurate trade initiation, assignment, or order flow inferences. We develop two new trade-to-quote matching methods, the Relative Best Bid and Offer (RBBO) method and the Latency Timestamp-adjusted (LTA) method, that adjust for message latency between exchanges and construct locally different NBBOs (which we refer to as RBBOs). The RBBO method incorporates both latency timestamp-adjustments as well as RBBOs, and signs an average of about 3.8 million trades, 555 million shares, and \$24.28 billion in trade value more accurately each day than conventional SIP-time-based trade-to-quote matching methods. The new assignment of these trades leads to an improvement in matching accuracy of 5.88 p.p., improving future and potentially altering past research inferences.

Utilizing latency adjustments between data centers and the RBBO for each exchange, we use both the LTA and RBBO methods to evaluate the accuracy of several trade assignment algorithms using 19 months of NYSE and NYSE ARCA exchange order book data (i.e, Direct Feed). These data indicate whether the trade is buyer or seller-initiated and indicate the exact timestamp, to the nanosecond, of the trade record. Evaluating the trade assignment accuracy of the RBBO method compared to conventional methods against order book data boosts trade assignment accuracy by 5.25 p.p. (LR: 5.88%, EMO: 4.84%, CLNV: 4.92%).⁴

However, the RBBO method is computationally expensive. Because computing resources could be constrained for some researchers, we developed a computing inexpensive second method, the Latency Timestamp-adjusted (LTA) method. LTA implements timestamp iterations to adjust for latency, and delivers an equivalent trade signing accuracy to RBBO. It only adds two

²Rosenthal (2012) also propose estimators for quotes prevailing at trade time.

³We dissect the shortcomings of the conventional methods in Appendix A.2.

⁴Referring to the buy/sell trade classification methods widely used in the literature, “LR” denotes Lee and Ready (1991); “EMO” denotes Ellis, Michaely, and O’Hara (2000); “CLNV” denotes Chakrabarty, et al (2007).

lines of code to Holden and Jacobsen (2014)’s SAS code and requires no additional computing time. The LTA method finishes spread computation for all stocks on a given day in 30-40 minutes in SAS. Compared to the RBBO method which takes about 3 hours to compute, LTA saves roughly 50 days of computing time on WRDS’ servers for a regular user to compute liquidity measures for all symbols in a 19-month sample.⁵ Using LTA improves trade classification accuracy by an average of 4.73 p.p (LR: 5.30%, EMO: 4.36%, CLNV: 4.43%). This accounts for roughly 90% of the improvement between conventional methods and the RBBO method, with effectively no loss in computation time.

RBBOs have been computed to implement our suggested methodology, and the resulting more accurate trade assignments, and its upgrade to all WRDS Intraday Indicator Dataset (IID) variables, will be available on WRDS. Further, we provide code that implements the Latency Timestamp-adjusted (LTA) method in Appendix A.5. In Table VI, we compare the computation time, coding complexity, and trade classification accuracy of the two proposed methods and benchmark them against existing methods in the literature. Latency adjustments can be applied to Daily TAQ since August 2015, when the participant exchange timestamp became available. We use the LTA set of results as our main result because of how convenient it is to implement the latency adjustment in a widely-circulated and frequently-used SAS code by Holden and Jacobsen (2014). Since LTA is our second-best approach, we, conservatively, use this as our primary method in the paper.⁶

Improved trade classification accuracy is economically significant in several ways. First, trade classification accuracy improves the measurement of liquidity, informed trading, and other measures. Constructing common measures across several literatures, we find improved trade classification accuracy introduces broad and significant changes in the estimates of absolute percentage order imbalance (28.9 b.p., or 2.55% larger), probability of informed trading (PIN, 34.1 b.p., 1.93% larger), effective spread (1.9 b.p., 5.23% larger), realized spread (2.4 b.p., 14.15% smaller), and price impact (4.3 b.p., 26.19% larger) in our sample. The LTA method’s improvements to these prevalent measures are unrelated to midpoint bias adjustment by Hagströmer

⁵WRDS has researched alternative programming languages to consume TAQ such as Python and PostgreSQL. As of today, SAS is the most memory-efficient programming language when handling historical TAQ data. Certain data steps and sorting function in Python/PostgreSQL could take ten times longer than SAS (with SAS views).

⁶All tests are repeated using the RBBO method and their robustness are included in the online Appendix.

(2021). The impact of these measures result in an absolute average difference of 10% from their current, conventional calculations. While extrapolating from these differences is difficult, studies using these measures may find substantially altered economic significance, statistical significance, or both.

To provide further evidence of the economic importance of our methods, we test their impact in a plausible research design. Conventional SIP-time-based trade classification methods would yield false positive results in event studies by examining the exogenous shock of a transition to faster technology in NYSE’s servers on several common liquidity measures. The NYSE exchanges individually transitioned to a system called “Pillar,” which NYSE estimates is up to 95% faster than its previous technological infrastructure (NYSE (2019)). We confirm this, finding that the average quote (trade) message latency at NYSE decreased from 74.32 (90.90) microseconds to 24.37 (32.72) microseconds, with a significant reduction of 49.95 (58.17) microseconds.

Properly accounting for latency should exclude the mechanical effects of this technological change on equity markets, leaving on economic effects, if any. We examine the average stock’s effective spread, realized spread, and price impact changes around the transition of the New York Stock Exchange to the Pillar system on July 13, 2020. Using conventional SIP-time-based trade classification methods, this transition had a significant effect on market liquidity. Realized spreads have a pre-Pillar mean of 6.7 b.p., and a post-Pillar mean of 6.3 b.p, with a significant reduction of 0.4 b.p., and price impact has a pre-Pillar mean of 12 b.p., and a post-Pillar mean of 12.6 b.p., with a significant increase of 0.6 b.p.. The pre- and post-migration difference is greater when limiting to NYSE trades only. Both the RBBO method and the LTA method remove latency in the computation of liquidity measures and, therefore, exclude the effects caused by this shock to market measures, and show a precisely estimated null effect. With the LTA method, we find no significant difference in effective spreads, realized spreads, or price impact across the entire market or in NYSE-only stocks.⁷ This finding aligns with the fact that an exchanges technological upgrade in latency reduction does not systematically alter the economics of market-making over a short period. It does, however, mechanically alter the construction methods of liquidity measures that rely on 5-minute windows, which are sensitive

⁷In untabulated results, the RBBO method shows the same insignificant pre- and post-Pillar event study results as LTA.

to an initial market-wide response (Hasbrouck and Saar (2013)).

Trade classification accuracy reduces the appearance of edge cases within equity market data. Trades linked to locked and crossed spreads, which imply negative or zero economic returns to market making, are substantially reduced. And trades linked to outside NBBO quotes, which appear to violate the best execution mandate of Reg NMS, are more robust and properly counted. Using conventional methods of trade classification, we measure a baseline of trades matched to locked, crossed, and outside NBBO quotes, respectively. We find that 7.79% of these “abnormal” quotes matched trades using conventional methods. Using the LTA method, we find 1.42%, 0.045%, and 3.89% of quotes matched to locked, crossed, and outside NBBO quotes, respectively, for a total of 5.02% of total abnormal quotes; a significant reduction of 2.77%.

Finally, to understand why our methods achieve 92% accuracy and how far future work could improve our methods, we decompose incorrectly signed trades and analyze them further in Appendix A.4. Using the Lee and Ready (1991) algorithm and the LTA method, approximately 1.71% of total trades trigger the tick test, which assigns trade direction based on price changes from the prior trade. Of these 1.71% trades assigned using the tick test, 1% are correctly assigned, while the remaining 0.71% are incorrect.⁸ Of the remaining 98.29% of trades classified using bid-ask midpoint inference, 8.1% are incorrectly assigned (with 90.19% of the total correctly assigned without the tick test). Of these 8.1%, 1.49% are trade-through exempt, 2.50% are \$0.01 away from triggering the tick test (and would be correctly assigned if the tick test was triggered), and 1.57% have a hidden order executed within one second prior, which may lead to improper matching due to hidden order priority. While all trade assignment algorithms we evaluate use the tick test at some point, future work that develops new trade assignment algorithms would improve trade classification accuracy by limiting the usage of the tick test, dealing with trade through exempt orders, improving tick test triggers, and developing a better method for classifying trades around hidden order executions. Improvements on all fronts could lead to a potential trade classification accuracy of 97.46%. While we have investigated inaccurately assigned trades thoroughly, the remaining 2.54% defy a simple explanation, which could

⁸The tick test, by construction, could sign one of the four tick movement scenarios incorrectly (Aitken and Frino (1996)).

be related to the lower latency of the direct feed, weather, or other factors (see Appendix A.3).

Our two new proposed methods (LTA and RBBO) have substantially improved trade assignment accuracy and economically significant changes to many other common research methods. These changes may impact past and future research inferences and suggest that new work using any of these methods uses either of our methods to assign trade direction. Our trade classification methods provide a way to handle the technological, seasonal, and other changes routinely altering message travel time between exchanges and their preferred message routing connections.

The rest of the paper is organized as follows: Section 2 briefly reviews the relevant literature, Section 3 provides institutional detail and outlines our methodology, Section 4 presents our data in detail, Section 5 discusses our main results, and Section 6 concludes.

2 Literature

This study intersects with a wide number of studies in financial research, as it affects nearly every empirical measure that relies on accurate intra-day equity market prices. To demonstrate the breadth of this discussion, like Bessembinder (2003), we can summarize the task of assigning trade direction from trade and quote data by breaking it apart into two tasks: (a) the use of an algorithm for trade direction assignment and (b) the matching of the prevailing best quote available to a market participant. This paper improves trade direction assignment by addressing the latter, while a concentrated literature has sprung up in addressing the former, including Lee and Ready (1991), Ellis, Michaely, and O’Hara (2000), Chakrabarty, et al. (2007), and Easley, Lopez de Prado, and O’Hara (2016).

A further literature has arisen analyzing the accuracy of these trade classification algorithms by matching trade and quote data with limit order book data, which identifies the initiating side of a trade. Among these are, most notably, Ellis, Michaely, and O’Hara (2000), Lee and Radhakrishna (2000), Finucane (2000), Odders-White (2000), Theissen (2000), Werner (2003), Peterson (2003), Chakrabarty, et al. (2007), Asquith, Oman, and Safaya (2010), Blais and Protter (2012), Chakrabarty, Moulton, and Shkilko (2012), and Chakrabarty, Pascual, and Shkilko (2015). Prior accurate classification rates vary greatly across these studies, with an average

accurate classification rate of 77.53%, ranging from 52% to 93%. While we know of no other work addressing trade and quote message latency and its effects on trade classification accuracy, besides a brief discussion by Bessembinder (2003), the microstructure literature has begun to investigate exchange latency along several other dimensions. Notably, Ding, Hanna, and Hendershott (2014) use proprietary data on direct feeds from exchanges to market participants, while Ernst, Sokobin, and Spatt (2021), and Bartlett and McCrary (2019) use TAQ’s new participant time variable, which we use to calculate message latencies between the sending exchange and the requisite SIP. However, these studies do not explicitly use this measure to study exchange latency.

Fundamentally, several separate fields of study and their commonly used measures are influenced by trade classification accuracy. First, the short-selling literature often uses order imbalance measures as a proxy for the number of short sellers in the market. Differential buy and sell trade assignment directly affects the construction of this measure. Among this literature, several studies that use order imbalance as their measure include Chordia, Roll, and Subrahmanyam (2002), Sarkar and Schwartz (2009), Boehmer, Jones, and Zhang (2013), and Boehmer and Wu (2013).

Second, the study of high-frequency trading (HFT) often relies on trade classification accuracy for evaluating HFTs impact on liquidity and other metrics. Among these include Conrad, Wahal, and Xiang (2015), van Kervel (2015), Clark-Joseph, Ye, and Zi (2017), Brogaard, Hendershott, and Riordan (2017), Chordia, Green, and Kottimukkalur (2018), and Yao and Ye (2018).

Third, many informed trading studies utilize price impact measures. Several of these consist of, for example, Duarte and Young (2009), Collin-Dufresne and Fos (2015), Bernile, Hu, Tang (2016), Lou and Shu (2017), Back, Crotty, and Li (2018), Brogaard, Hendershott, and Riordan (2019), Duarte, Hu, and Young (2020), and Brogaard and Pan (2022).

Finally, liquidity and trading cost measures like effective spread and realized spread are evaluated relative to a trades direction. These measures are extensively used to evaluate market liquidity, including by Hasbrouck (2009), Fang, Noe, and Tice (2009), Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010), Fang, Tian, Tice (2014), Collin-Dufresne and

Fos (2015), Bessembinder, Carrion, and Venkataraman (2016), Lou and Shu (2017), Barardehi, Bernhardt, and Davies (2019), Weller (2019), Comerton-Forde, Grégoire, Zhong (2019), Chung, Lee, and Rösch (2020), Bessembinder, Hao, and Zheng (2020), Conrad and Wahal (2020), Albuquerque, Song, and Yao (2020), and Hagströmer (2021). Altogether, a 2014 review of the empirical liquidity literature by Holden, Jacobsen, and Subrahmanyam (2014) references 143 studies. This serves as a starting point of our review of the literature, as the number of studies impacted by or using these measures is only briefly evaluated here.

3 Institutional Detail and Methodology

3.1 Geographic Locations of the Exchanges

Of first-order concern, geographic distance determines message latency, which determines what the top of each exchanges book displays. Accounting for message latency ensures that no trade is measured as executed against a quote that had either (1) not yet arrived on that exchange’s book (caused by contemporaneous matching) or (2) become buried deeper in the book by a better bid or offer (caused by too great a lag). Reconstructing the top of each exchanges book at any given moment is necessary to match quotes and trades accurately.

Figure 1 maps the 14 U.S. exchanges and UTP/CTA SIP data center locations. The map shows that the exchanges and SIPs are co-located in three cities in New Jersey – Mahwah, Carteret, and Secaucus. The straight-line geographic distance between Mahwah and Carteret is 35 miles, Mahwah and Secaucus is 21 miles, and Carteret and Secaucus is 17 miles. Trades in one city cannot be matched with an away quote without adjusting for travel latency. In this sense, it is inappropriate to match TAQ trades contemporaneously or by lagging for an arbitrary amount of time to infer trade side classification. When the exchange is co-located with the SIP, the geographic distance is approximately zero, and the geographic latency is minimal. Away exchanges have varying distance, further compounded by technological factors, like gateway and processing latency, that add to geographic latency.

Empirical geographic latency is approaching these theoretical minimums. Table I reports the median quote and trade latency between the exchanges and corresponding SIPs on a repre-

sentative day in our sample. On the CTA Network, CBOE exchanges have latencies of around 400 microseconds (μs), and are co-located in Secaucus. Nasdaq exchanges, with a latency of around 540 μs , are co-located in Carteret. It is noticeable that NYSE-owned exchanges have a latency of around 100 μs with the CTA SIP. These exchanges are all located in Mahwah and co-located with the CTA SIP. Thus, the 100 μs observed latency accounts for each exchanges gateway latency and SIP processing time. Because there is no SIP located in Secaucus, the latency between Carteret and Secaucus is unobservable in the CTA network in Daily TAQ.

Latency from the UTP SIP is presented in Columns 3 and 4 in Table I. As Nasdaq exchanges and the UTP SIP are co-located in Carteret, they have the lowest latency (about 18-24 μs). CBOE exchanges, located in Secaucus, have an average latency of 200 μs when communicating with the UTP SIP in Carteret. NYSE exchanges have a latency of around 370 μs . Both the UTP and CTA SIPs communicate between Mahwah and Carteret. Yet, comparing Columns 1 and 2 with Columns 3 and 4, CTA communication (540 μs) is substantially slower than UTP communication (370 μs). Notably, the 100 μs latency between NYSE exchanges and the CTA SIP is not approximately equivalent to the 100 μs difference of UTP and CTA communication between Mahwah and Carteret. The three separable elements in latency that we can identify between the UTP SIP and the CTA SIP – geographic latency, gateway latency, and processing latency – appear to be distinct.

3.2 Relative Best Bid Offers (RBBO) and Local Consolidation

Reg NMS mandates a National Best Bid and Offer (NBBO), which sets the best available quotation across all exchanges in U.S. equity markets. However, the NBBO, displayed by either the UTP or CTA SIP feeds (collectively called the SIP), has used slower technology in its connectivity and processing hardware for some time (CTA (2020)). Further, UTP servers, administered by NASDAQ, and CTA servers, administered by NYSE, are housed within each exchange groups data centers in separate cities (UTP (2021), CTA (2019)). NBBO qualifying quotes that are broadcast to the SIP in a distant city, selected as the new NBBO, and broadcast back to the sending city or all other cities add substantial message latency and are at greater risk of being picked off or becoming stale before they arrive at away or other local exchanges.

The best quote available that fulfills all the conditions of the NBBO on each exchange at any moment in time is not necessarily the current prevailing NBBO as displayed by the SIP, yet it is the NBBO *as the exchange perceives it* that exchanges execute their matching engines against an exchange-specific BBO, which we call the Relative Best Bid and Offer (RBBO).⁹ The message latency between exchanges causes time-based market segmentation, fracturing the National Market System into 14 local exchange books.¹⁰ We estimate that the official NBBO differs from exchanges constructed RBBO 15% of the time. This error rate leads to an average of \$32 billion of trades signed differently each day.

With latency causing RBBOs, this issue is exacerbated by nearly no exchange communicating with other exchanges via the SIP as an intermediary. Exchanges connect directly to each other, communicating via what are called Direct Feeds. Direct Feeds use much faster technology than the SIP, and constantly broadcast their best available NBBO qualifying quotes, using these quotes to match trades without confirmation from the SIP.¹¹

Over the years, exchanges have switched from using SIP data feeds as their primary connection to other exchanges to using direct feeds. Figure 2 illustrates this point, with the use of direct feeds rising from 50% of all connections between exchanges to 85% by the end of our sample.¹² Direct feeds allow exchanges to receive quote updates much earlier than the SIP feed in multiple ways. First, rather than “waiting” on the time it takes an order to route from one away exchange to the requisite SIP and then to a different exchange, the two exchanges communicate directly. This substantially decreases geographic latency. Second, direct feeds offer superior technology to the SIPs, utilizing millimeter wave towers and other advances not integrated into the SIP infrastructure, reducing the technological drivers of latency.

With NBBO qualifying orders arriving to exchanges several hundred microseconds before

⁹Per correspondence with NYSE: “Each exchange calculates its own NBBO based on the data feed it is processing at the time, again as defined in Rule 7.37(e) Use of Data Feeds”.

¹⁰From its most recent Market Data Proposal SEC (2020b), the SEC itself is aware of this, noting “currently market participants may already observe multiple NBBO quotes. Therefore, the Commission preliminarily believes that the decentralized consolidation model would result in no meaningful difference in practice with respect to the existence of multiple NBBOs”.

¹¹The SIP feed is sent via fiber optics cables, while the direct feed is transmitted via much faster millimeter waves or laser beams. See technical specifications in Phil Mackintosh’s [article](#).

¹²The pervasiveness of direct feeds is not just well known, but well accepted by market participants. For example, Mehmet Kinak, Head of Trading for T.Rowe Price at the time, noted in an SEC Roundtable that, “If a broker is routing using SIP data, they are not routing my flow” (SEC (2018)).

they would have if they had been routed through the SIP, exchanges often have substantially different NBBO quotes based on both what the SIP transmits as the current NBBO as well as quotes received from other exchanges. Exchanges following best execution, as mandated by Reg NMS, leads exchanges to execute trades according to their local best prices rather than what the SIP may display.¹³ Despite the intention of the National Best Bid and Offer, exchanges must execute against the best prices in the market as they view them and that can differ from how the SIP aggregates them. With direct feeds substantially altering this “view” by speeding up the arrival time of messages, exchanges are executing trades routinely against a locally consolidated RBBO that differs from the current NBBO. Our own estimate finds that city-based RBBO calculations differ from the SIP-generated NBBO 15% of the time. The operational structure of U.S. equity markets is that exchanges each have their RBBO, constructed from direct feeds from other exchanges.¹⁴

Yet, as of today, no data consolidator aggregates direct feeds across all exchanges and makes such data available to academics.¹⁵ The only accessible data are Daily TAQ trades and quotes generated by the SIP.¹⁶ Therefore, we are constrained to construct a methodology that recreates these RBBOs rather than identifying latency and RBBO for each exchange by order matching across different order book datasets. We discuss these methods in the following section.

3.3 Methodology

3.3.1 Measuring Latency

Originally, trade or quote time (`time_m`) in Daily TAQ refers to the time when the Security Information Processor (SIP) completes processing a message. We denote this “time_m” timestamp in Daily TAQ as the “SIP time”. In August 2015, Daily TAQ introduced a new timestamp called the “Participant Exchange Time (`part_time`)”, which refers to the time when the trade or quote

¹³Certain orders, such as trade-through exempt orders, are legally able to be executed outside the prevailing NBBO.

¹⁴Many exchanges do not use the “RBBO” nomenclature, but instead refer to it as an exchange-specific NBBO.

¹⁵Many practitioners have no choice but to consolidate the direct feeds in-house themselves.

¹⁶Researchers can purchase order book data from exchanges, but it is costly. We find a rough estimate of over \$150,000 per month for the data subscription and platform access fees to access the raw order book data files from the 14 U.S. exchanges at the time of our sample. This estimate does not include data processing or storage costs, as well.

message leaves the exchange. In Daily TAQ, the Participant Exchange time is always earlier than the SIP time. We describe the relationship and sequence between SIP time and the Participant Exchange Time as following: first, a given quote or trade message is processed by an exchange’s matching engine. The message leaves the exchange at Participant Exchange Time and starts to travel to the SIP. Depending on the geographic location, it would take “Exchange-to-SIP Travel Time” for the quote/trade to arrive at the SIP. Once it passes through the SIP gateway, the SIP processes the quote/trade for the Consolidated Tape. The SIP completes processing the quote/trade message at the SIP time.

An exchange is affiliated with either the UTP network (and therefore reports to the UTP SIP) or the CTA network (reporting to the CTA SIP). The SIP may process quotes and trades at different speeds depending on the traffic at the gateway and the processing load of the SIP. The SIP Processing Time commonly differs from the same exchange-to-SIP communication for quotes and trades. SIP Gateway Latency also varies within the day depending on traffic and may change when NYSE/Nasdaq upgrades their SIP hardware.¹⁷ Similar to the SIP, exchanges have their own gateway latency, which is included in the Participant Exchange Time.¹⁸

As a special case, the Investors Exchange (IEX) uses a 38-mile coil of wire (The Speedbump) to slow down orders going to IEX. This speedbump delays the brokers orders traveling to IEX but does not affect the SIP-to-exchange and exchange-to-exchange communications.

After understanding the differences in the two timestamps in Daily TAQ, we introduce the notion of latency. Extrapolating Equation 1, we define latency as:

$$\begin{aligned} \text{Latency} &= \text{SIP Time} - \text{Participant Exchange Time} \\ &= \text{Exchange-to-SIP Travel Time} + \text{SIP Gateway Latency} + \text{SIP Processing Time}, \end{aligned} \quad (1)$$

where SIP time is the time when SIP records the trade or quote message and the Participant Exchange time is when the trade or quote message leave the exchange and start to route to other exchanges (via direct feed) or to the SIP.

We propose two methods that adjust the trade-to-quote matching process for i) message latency and ii) RBBOs. The RBBO method integrates both features, while the Latency Timestamp-

¹⁷We discuss possible determinants of exchange-to-SIP latency in Appendix A.3.

¹⁸To simplify, we do not difference the inbound gateway latency and the outbound gateway latency. Some exchange such as Nasdaq offers their gateway latency data to academics. See [Nasdaq Trading Reports](#) on “Missed Opportunities and Latency Gateway” for more information.

adjusted (LTA) method only incorporates latency adjustments.

3.3.2 The Relative Best Bid and Offer (RBBO) Method

Adjusting for latency suggests we should match trades at the trade’s Participant Exchange Time with quotes that would have enough time from their sending exchange to arrive by that point. However, we are not able to observe the latency of direct feeds between exchanges, so we proxy for this latency using the SIP that is co-located with that exchange, or in the case of exchanges in Secaucus where there is no SIP, the one-way latency between a SIP city (Mahwah or Carteret) and Secaucus. In short, we lay out this approach as follows:

$$\begin{aligned}
 \text{Trade @ ExecutingExchange Time} \rightarrow \\
 \text{matched with, prevailing BBO Quote @} \\
 [\text{QuotingExchange Time} + \text{QuotingExchange's co} \\
 \text{-- located SIP-to-ExecutingExchange Latency}]
 \end{aligned} \tag{2}$$

This creates a measure of latency across exchanges that is updated based on contemporaneous weather, network, and technological conditions. This offers the best estimate for latency between data centers and across exchanges. However, it has some downsides. Although this latency adjustment method modifies timestamps for the latency between SIPs and exchanges, it does not represent the order flow for exchanges communicating through direct feeds. The speeds of direct feeds are unknown outside the market participants who subscribe to them. However, it is important to note that the speed of light binds these speeds. For a beam of light to transit between any pair of the three cities, it would take 188 μs between Mahwah and Carteret, 133 μs between Mahwah and Secaucus, and 91 μs between Carteret and Secaucus. This is 50%, 54%, and 72% less than the median latencies observed in Table I. While substantially different, the latencies we observe in our estimates are likely within the same order of magnitude as those of direct feeds. Finally, despite being unable to observe direct feed latency, we find that our latency estimates substantially improve trade classification accuracy over arbitrary adjustments and leave improved latency estimates for future work.

The second aspect of this method creates a city-based Relative BBO (RBBO) reflecting exchanges sending and receiving quote updates between each other using direct feeds. Recreating a city-based RBBO is important, as exchanges often route market orders to the exchange with

the national best prices. Using Rule 605 exchange reports, we find that 36.68% of executed market orders (share volume inferred) at NYSE and ARCA are routed out to other exchanges.¹⁹ Further, we use a city-based RBBO measure, assuming gateway latency within data centers is approximately equivalent for all exchanges within the center. That is, a message arriving to Mahwah arrives at all the exchanges in Mahwah at approximately the same time, given that all exchanges there are co-located within the same NYSE data center.

With this in mind, our RBBO construction is as follows. Given three co-located cities for the major US exchanges: Mahwah exchanges and the CTA SIP, denoted as [Mahwah]; Secaucus exchanges, denoted as [Secaucus]; and Carteret exchanges and the UTP SIP, denoted as [Carteret]. To create each city’s RBBO, we sort quotes using the “Quote@Exchange Time” for every exchange co-located in the same city to compute the city’s best bids and asks at “Quote@City Time”. Next, we generate a latency reference for each city-to-city pair and compute the travel latency as the time it takes for the most recent quote message to travel from one city to another. This means computing $\text{Travel Latency} = (\text{UTP/CTA}) \text{ SIP Time} - \text{Participant Exchange Time}$ for each quote traveling between Mahwah-Carteret, Carteret-Secaucus, and Secaucus-Mahwah for UTP and CTA separately.²⁰ After that, we adjust the timestamp of the quotes from away cities using this latency reference. We denote the latency-adjusted time of the away quotes arriving at the local (standing) city at “Quote@City Latency-adjusted Time,” which reflects the actual time a quote from away cities could arrive at the local exchange in the standing city. We sort the away quotes at “Quote@City Latency-adjusted Time” along with the local quotes at “Quote@City Time” and compile all the quotes into a Relative BBO quote. According to the NBBO rules, we match the trades at “Trade@City Time” with the prevailing RBBO quotes. Finally, we consolidate all trade-to-quote matches across the three cities. For example, the trades executed at the Mahwah exchanges are matched using the trades’ “Trade@Mahwah Time” with the RBBO quotes that are composed of (i) Carteret’s and Secaucus’ away quotes at “Quote@Mahwah Latency-adjusted Time” and (ii) Mahwah’s quotes at Quote@Mahwah Time”. This approach is referred to as the “RBBO” method throughout the paper.

¹⁹Li, Ye, and Zheng (2023) Table III shows similar level of market order routing activity at 33.64%.

²⁰In total, we will have six latency references between each pair of the three cities times two (one for UTP and another for CTA). We have to have one for each UTP/CTA network because the SIP gateway latency and the SIP processing time are distinct between the two SIPs.

3.3.3 The Latency Timestamp-adjusted (LTA) Method

While the RBBO method is computationally expensive and requires extra coding efforts, we suggest a second method that is a lot more convenient. This method takes advantage of knowing the trade took place at a given exchange ex-post in Daily TAQ Trades and brings the Trade@ExecutingExchange time early. It also avoids constructing an RBBO for each city. Subtracting the “NBBO Quotes SIP-to-TradeExecutingExchange Travel Time” from both the trade and the quote timestamps in Equation 2, we have:

$$\begin{aligned} &Trade @ [TradeEXExchange Time - Quote's SIP-to-TradeEXExchange Travel Time] \\ &\rightarrow \text{matched with, prevailing NBBO Quote @ Quote SIP Time.} \end{aligned} \quad (3)$$

As the NBBO Quote’s TradeExecutingExchange-to-SIP Travel Time is not directly observable in Daily TAQ, we proxy this time using the time it takes the trade message to report to the SIP:

$$\begin{aligned} &Proxy : \\ &Quote's SIP-to-TradeEXExchange Travel Time \\ &\approx Trade's TradeEXExchange-to-SIP Travel Time \\ &= Trade's SIP Time - Trade's Participant Exchange Time \\ &= Trade Latency. \end{aligned} \quad (4)$$

The trade message latency (Trade@SIP Time - Trade@ExecutingExchange Time), which is trade’s SIP Time less the trade Participant Exchange Time, have both timestamps directly observable in Daily TAQ trades.²¹ In short, this computing-saving latency adjustment method proposes to re-write Equation 3 using Equation 4 as:

$$\begin{aligned} &Trade @ [TradeEXExchange Time - Quote's SIP-to-TradeEXExchange Travel Time] \\ &= [Trade's Participant Exchange Time - Trade Latency] \\ &\rightarrow \text{matched with, prevailing NBBO Quote @ Quote SIP Time.} \end{aligned} \quad (5)$$

The latency timestamp-adjusted method does not require intensive coding and can be easily applied by adding two lines in the SAS code provided by Holden and Jacobsen (2014). Appendix A.5 provides the two-liner SAS code that applies the computation in Equation 5 – the first line computes the latency for the trade message, and the second line overwrites the “time m” variable as trades participant exchange time less trade latency. Table VI shows the coding complexity and the average time it takes to run this code on all trades and symbols for one day in Daily TAQ. We denote this approach as the “LTA” method in the remaining sections of the paper.

²¹This proxy rests on the assumption that the “NBBO message TradeExecutingExchange-to-SIP travel time” is equivalent to the “Trade message TradeExecutingExchange-to-SIP travel time”.

4 Data

We first construct a sample of trades where trade direction is already known to assess the impact of our methods on trade classification accuracy. Limit order book data, used in previous studies such as Ellis, Michaely, and O’Hara (2000), Odders-White (2000), and Chakrabarty, et al. (2007), among many others, provides this source. We create a sample of limit orders and TAQ trades by matching order executions in the NYSE and NYSE ARCA Integrated Feeds from December 2017 to June 2019 (19 months) with trades in Daily TAQ. From these two order books (the Integrated Feed), we retrieve the actual side of the order from order submission messages. We match order book execution messages to trades in Daily TAQ based on date, symbol, trade price, trade size (shares), and the trade time (to the exact nanosecond).

We restrain our testing sample to regular, unconditional trades in Daily TAQ at NYSE and NYSE ARCA to avoid complications with partial executions of conditional orders such as Intermarket Sweeps. Appendix A.1 demonstrates how an order execution between a market order and a sitting limit order in order book data matched to a trade record in Daily TAQ.²² We match about 77% of regular, unconditional trades on NYSE and ARCA, ending with 654,170,045 trades matched using TAQ and the limit order books of these two exchanges.²³ We use this sample to test trade classification accuracy using our RBBO and LTA methods. We also gather data on exchange order routing activities. We collected Rule 605 monthly reports from NYSE and NYSE ARCA for this same sample period.²⁴

In further tests, we apply our two trade assignment methods to a more extended sample, beginning on the first date when the Participant Exchange Time becomes available (August 2015) till September 2019 (49 months). We denote this sample as the “Full Sample”, with our trade classification testing sample denoted as the “Order Book Sample”. Like the Order Book

²²We filter for market order to limit order executions and exclude resting limit order to limit order executions for simplicity. We observe that over 97.25% of the order executions at NYSE and over 99.99% of the order executions at NYSE ARCA are market-to-limit-order executions.

²³We only consider unique matched trades in the order book and TAQ trades that have the same symbol, price, share, and the exact nanosecond execution time. Further, we do not match trades against orders that arrive at the same nanosecond as another order, or if duplicate orders appear in the data. We do not have any more data constraints.

²⁴For the Rule 605 data, we filter for market orders and marketable limit orders, and we compute the percentage of routed market order shares over total executed shares.

sample, we include all trades and quotes for the entire cross-section of stocks.²⁵ We utilize this sample to study the improvement in trade-to-quote matching quality with respect to abnormal quotes, common measures derived from TAQ trade assignment, and our NYSE Pillar migration test.

5 Empirical Results

5.1 Evaluating Trade Classification Algorithms

We begin our analyses by evaluating the accuracy of each of our two new trade classification methods. While both of our methods alter the timing of matching trades and quotes, we remain agnostic on choice of assignment algorithm. Therefore, we test our methods using three of the most commonly used trade classification algorithms - Lee and Ready(1991) (LR), Ellis, Michaely, and O’Hara (2000) (EMO), and Chakrabarty, et al. (2007) (CLNV). To properly evaluate the accuracy of our methods, we also need a source of known, accurate trade origination. We follow prior studies evaluating trade classification accuracy and use order book data as this source. We match limit order book data from NYSE and ARCA, which correspond to trade executions on those exchanges at the exact nanosecond (see [Data](#) for detail). This allows us to verify trade initiation direction, and evaluate how well these common trade assignment algorithms do before and after implementing our methods. Table II describes these results.

Our RBBO method significantly outperforms SIP time trade classification methods. We find an improvement of trade classification accuracy to 92.63% (LR), 92.44%, (EMO), and 92.42% (CLNV), for a statistically significant improvement of 5.88%, 4.94%, and 4.92%, respectively from conventional methods. This compares to conventional assignment accuracy baselines of 86.75% (LR), 87.50%, (EMO), and 87.50% (CLNV). Using our LTA method, we find an improved trade classification accuracy of 92.05% (LR), 91.86%, (EMO), and 91.93% (CLNV), respectively. The majority of accuracy improvement offered by these methods, then, is due to adjusting for latency when matching trades and quotes. Construction of the RBBO is robust, as these two methods are statistically significantly different in every case, but the largest improvement offered

²⁵Similar to the Order Book sample, we do not have any more data constraints.

by the RBBO method is due to the latency adjusted portion, which is constructed into the LTA method as well.

Not only do these methods together offer substantial improvement over current methods on average, but they improve accuracy in all cases. Figure 3 demonstrates the time series of daily trade classification accuracy in our sample. Our latency adjusted methods outperforms the conventional SIP time method and Bartlett and McCrary (2019)’s Direct NBBO method on each day throughout the Order Book Sample.²⁶ The “Direct NBBO” method suggested by Bartlett and McCrary (2019) improves trade classification accuracy by 1.21%, but, because Bartlett and McCrary (2019) assume latency is zero between exchanges and omit the time cost of exchange-to-exchange communications, this improvement is not comparable to the LTA method or the RBBO method. Figure 3 also reports the monthly market order execution activities. On average, 36.68% of the executed market orders at NYSE and NYSE ARCA are routed out to other exchanges. The conventional method and Direct NBBO accuracy drops when exchange order routing activity increases.

Our trade assignment methods may vary over the price, size, or volume of stocks, affecting the implementation of common research methods. We present these results in Table III. We find that trade classification accuracy improvement is greatest in the most liquid stocks, while accuracy monotonically decreases across trade classification methods with trading volume. Comparing differences in trade classification accuracy between conventional methods and the Latency Timestamp-adjusted (LTA) method, the highest daily volume quintile of stocks has an accuracy improvement of 5.32%, while the lowest quintile has an accuracy improvement of 4.12%. Differencing high vs low quintiles leads to a difference of 1.19%. We infer from these results that our methods have the greatest impact among stocks that are traded or quoted frequently enough for NBBO eligible quotes to become stale before making the roundtrip to the relevant SIP. This may particularly impact prior and future research that is dependent or sensitive to trading volume.

Our latency adjustment methods also improve the rate at which we identify locked or crossed spreads, or outside the NBBO/RBBO trade execution. Table V reports these results. The

²⁶When there is no technical issue affecting the participant time recording on the consolidated trade tape.

Latency Timestamp-adjusted (LTA) method shows an improvement of -4.18% in locked spreads, -0.05% in crossed spreads, and 1.28% trades executed outside the NBBO. Across all three of these measures, our methods improve the percentage of trades matched to abnormal quotes by -2.77% compared to using conventional methods.

While the RBBO method is superior to the Timestamp-adjusted (LTA) method, the former is more computationally expensive and difficult to code. Table VI shows that the LTA method adds two lines to the existing code by Holden and Jacobsen (2014) and takes about 30-40 minutes to compute for one day’s TAQ trades and quotes.²⁷ The RBBO method would take about 3 hours to compute all trades for a given day and takes a lot more effort to code.

Creating an RBBO, rather than relying on the inaccurate NBBO, and adjusting for latency between exchanges is important in accurately assigning trade direction in equity markets. These two improvements to current trade assignment methods substantially improve the accuracy of these methods by about 6%. This improvement is strengthened among high volume stocks, making liquidity or volume-related inferences particularly sensitive to errors in classification accuracy. Finally, these methods owe much of their accuracy improvement to adjusting for latency, and the computing time of the LTA method is exactly identical to current research methods. The construction of RBBOs, while computationally expensive, is significant and accounts for an additional 58 bps of improvement in trade classification accuracy. Due to the limited marginal benefit of RBBO and the convenience of LTA, we use the LTA set of results as our main results in the following analysis.

5.2 Trade Classification Accuracy and Common Measures

An improvement of trade classification accuracy is only necessarily useful on its own if it has implications for research inferences. Our two new trade classification methods affect several measures used in finance research. Using our proposed methods, we show that order imbalance, effective spreads, realized spreads, and price impact are all economically and statistically different from conventional methods.²⁸

²⁷We give a range of computing time because it could change as the data file sizes of TAQ trades and quotes vary from one day to another.

²⁸We also note here that, by implying alterations in research inferences, past research may both negatively or positively impacted. That is, the statistical significance of results may lose precision or persist, the point estimates

Comparing common liquidity measures, we examine percentage daily effective spreads, realized spreads, and price impact. These liquidity measures are computed as:

$$\text{Percent Effective Spread}_k = \frac{2D_k(P_k - M_k)}{M_k}, \quad (6.1)$$

$$\text{Percent Realized Spread}_k = \frac{2D_k(P_k - M_k^{+5})}{M_k}, \quad (6.2)$$

$$\text{Percent Price Impact}_k = \frac{2D_k(M_k^{+5} - M_k)}{M_k}, \quad (6.3)$$

where D_k is an indicator variable that takes value of +1 if the trade k is a buy and -1 if it is a sell; P_k is the trade price of trade k ; M_k is the midpoint of the NBBO quotes assigned to trade k using the interpolated time technique; M_k^{+5} is the bid-ask mid-point five minutes after the midpoint M_k . We simple-average the percent effective spread, realized spread, and price impact across all market-hour trades for a given stock-day to calculate the stock daily simple averaged effective spread, realized spread, and price impact.

We follow the classic Easley, Kiefer, O'Hara, and Paperman (1996) model (equation 7.1 and 7.2) and compute annual PIN for 2018 as:

$$PIN = \frac{a \times u}{a \times u + \varepsilon_b + \varepsilon_s}, \quad (7.1)$$

where $\{a, d, u, \varepsilon_b, \varepsilon_s\}$ are the set of parameter vectors in the likelihood function

$$L(\vartheta|B, S) = (1-a)e^{-\varepsilon_b} \frac{\varepsilon_b^B}{B!} e^{-\varepsilon_s} \frac{\varepsilon_s^S}{S!} + ade^{-(u+\varepsilon_b)} \frac{(u+\varepsilon_b)^B}{B!} e^{-\varepsilon_s} \frac{\varepsilon_s^S}{S!} + a(1-d)e^{-\varepsilon_b} \frac{\varepsilon_b^B}{B!} e^{-(u+\varepsilon_s)} \frac{(u+\varepsilon_s)^S}{S!}, \quad (7.2)$$

where B and S are the daily number of buys and sells. a denotes the probability that an information event occurs on any day. If an information event occurs, it can be bad news with the probability of d or good news with the probability of $1 - d$. Informed trades who know the quality of the information and submit orders on any day at the arrival rate of u .

We present these results in Table IV. In Panel A, we estimate baseline daily percentage

may grow larger, smaller, or change sign, and/or the economic magnitude of these results may be smaller or larger. that is, net effects are difficult to discern. In any of these cases, past research may not only be adversely impacted – prior studies may become even more important in their position within the literature.

effective spreads of 0.363%, compared to LTA method spreads of 0.382% (a significant difference of 0.019% larger). With effective spreads used as both a measure of liquidity and trading costs, these substantially larger effective spreads imply that prior research has been using substantially smaller trading costs than many be realized in equity markets.

Daily percentage realized spreads are significantly smaller. Conventional methods find realized spreads of 0.205%, compared to LTA estimates of 0.176% (a significant difference of 0.024% smaller). We find an opposite relationship with daily percentage price impact, with conventional method estimates of 0.168%, LTA estimates of 0.212% (a significant difference of 0.043% larger). This relationship is mechanical, as realized spreads and price impact are a decomposition of effective spreads.

The effect of larger price impact dominates this difference. Hasbrouck and Saar (2013) find that the majority of the probability of a trade, using a hazard model, comes immediately after a quote arrives on the book. This makes intuitive sense as a new quote arrives, resting orders within an exchanges book that match this quote are successfully matched, and a trade is executed. In the context of latency, the measure of price impact captures the time period immediately following the execution of a trade and the assignment of a new Best Bid and Offer on an exchanges book. A time window that is too early from this trade omits the RBBO/NBBO response to this trade and estimates significantly lower price impact, while the correct time window accurately captures the immediate response to the arrival of a new quote documented by Hasbrouck and Saar (2013) that accounts for a substantial portion of the price impact of a trade.

Next, maximum-likelihood estimated informativeness models (e.g. the classic PIN by Easley, Kiefer, O'Hara, and Paperman (1996) and its variations) infer probabilities of informed trading based on order imbalances, and these models rely on the accuracy of trade assignment. We also test these models in Table IV Panel B. Conventional SIP time methods produce estimates of the Probability of Informed Trading of 17.692% for common stocks in 2018, while our LTA method produces estimates of 18.033%, a significant difference of 0.341% larger.

The implications for substantially larger price impact or informed trading than previous studies have estimated are challenging to parse. The high-frequency trading literature has

often found larger price impact and market volatility in the presence of algorithmic traders (e.g. Brogaard, Hendershott, and Riordan (2019)). Furthermore, this measure is often seen as a proxy for the information content of a trade. Studies of informed trading have created measures of information priced in the cross-section, studied leaks before macro events, or found it otherwise difficult to measure (e.g. Easley and O’Hara (2004), Duarte and Young (2009), Bernile, Hu, Tang (2016), and Duarte, Hu, and Young (2020)). It is unclear how a substantially larger unconditional estimate of price impact affects these results across these applications and others. Overall, an unconditionally larger price impact than previously understood within equity markets has potentially broad-based effects.

Finally, absolute order imbalance (i.e. absolute of buys less sells) is a commonly used short selling measure that depends on accurate trade assignment. These measures are often used to measure the appearance of short sellers in daily stock volume (e.g. Chordia, Roll, and Subrahmanyam (2002)), or as a control for its association with liquidity, information, and market sided frictions. We present results on order imbalance in Table IV Panel B. Using conventional methods of trade classification absolute order imbalance throughout our sample averages 11.472%, while averaging 11.765% for the LTA method, a significant difference of 0.289% larger. With altered buy and sell estimates, our methods influence the treatment propensity of a given observation in these studies. While we cannot use these baseline averages to make any claims about the results of previous studies, these altered means do imply that there are potential changes to either the results within them, or changes in the economic magnitude of the results of this work.

Combined with other common measures used by financial researchers, the impacts may be disparate and wide-reaching. Altogether, these common measures differ substantially from estimates using current trade and quote matching methods. With an improvement in trade and quote matching, we offer not just an improvement in methodology, but a change in estimates for a wide range of finance topics covering liquidity, trading costs, short selling, and informed trading measures.

5.3 NYSE PILLAR Transition

Next, we show that our methods do not simply have implications, but specifically alter research inferences. To do this, we use a plausible research design. Using a technological shock to latency at the New York Stock Exchange (NYSE), we find that conventional methods report false positive findings in common liquidity measures, while our latency timestamp-adjusted method shows no significant change to market liquidity.

The NYSE Pillar system is a new trading technology platform developed by NYSE to replace its legacy trading systems. For the NYSE exchange, this event took place on July 13, 2020. NYSE judges it to be up to 95% faster (NYSE (2019)). We observe the reduction in latency in the event of NYSEs Pillar migration and report the pre- and post-event latency in Table VI. This technological shock to latency results in latency dropping from 74 μ s to 24 μ s for quotes at NYSE and from 90 μ s to 32 μ s for trades. A 10-day window implies a slow transition period, but Figure 5 shows that this reduction in latency was immediate and transitioned over a single day. Pre and post-event latencies for quotes and trades were also consistent and stable, with little volatility.

We use a $[-5, +5]$ trading day window to study market liquidity before and after this transition event. We examine the average stocks effective spread, realized spread, and price impact changes around the transition of the NYSE Pillar system on July 13, 2020. Using conventional methods, we find that this transition significantly affected market liquidity in a $[-5, +5]$ trading day window. As shown in Table VIII, the average realized spread has a pre-Pillar mean of 0.067%, and a post-Pillar mean of 0.063%, with a significant reduction of 0.004%. The average price impact has a pre-Pillar mean of 0.120%, and a post-Pillar mean of 0.125%, with a significant increase of 0.006%. The before-migration and after-migration difference is greater when limiting to NYSE trades only.

However, after implementing our latency timestamp-adjusted (LTA) method, we find no significant change in the realized spread and price impact across all stocks post-Pillar. The decrease in latency mechanically creates a larger window, both by the $\sim 50 \mu$ s added to the initial trade and the greater frequency of trades that may occur within the same time window. Because of this, greater price movements may enlarge the price impact of a trade, similarly leading to a

mechanical reduction in the realized spread since, holding effective spread constant, the increase of one will lead to the decrease of another. Adjusting for latency keeps this window constant relative to the speed of financial markets. This finding implies that an exchanges technological upgrade does not necessarily cause economic impacts or function as an exogenous shock in equity markets. Put differently, latency reduction in equity markets does not systematically alter the economics of market making, which conventional methods would find due to mechanical effects in common liquidity measures – a false positive inference.

Studies that use this or shocks to latency difference-in-difference designs may find significant effects where there are none when using conventional trade assignment methods. This arises from mechanically inflated price impact and deflated realized spread estimates. This may affect a host of topics including the impact of HFT, profits to market making, trading costs, and short-term market efficiency. While the scope of this placebo research design is limited, it is possible that other designs and measures are influenced by our trade assignment measures. Failure to account for RBBOs and latency impacts research inferences.

5.4 The Remaining 8%

Despite our methods’ improvements, about 8% of matched trades remain incorrectly classified. We examine the root causes of this in Appendix A.4. Of the remaining 8.81% of trades incorrectly classified using the Latency Adjusted method, 0.71% of this total are due to incorrect tick test matching, 1.49% are trade through exempt, 2.5% are within \$0.01 of triggering a successful tick test result, 1.57% have a hidden order executed within the surrounding second, with 2.54% remaining inaccurately classified for unknown reasons. The reasons for many of these issues are simple.

First, the tick test has four fundamental outcomes, one of which is incorrectly classified. Therefore, assuming equal probability across all four outcomes implies 75% accuracy, while the greater the proportion of wrong tick assignment in the sample, the lower the accuracy of the test.²⁹ In our sample, 1.71% of trades trigger the tick test, with 0.71% incorrectly classified, leading to a tick test accuracy of 58%.

²⁹Naturally, all 4 outcomes do not necessarily occur with equal probability empirically. However, there is some evidence that this is often the case. See Aitken and Frino (1996).

Secondly, 1.49% of incorrectly classified trades stem from Trade Through Exempt (TTE) orders. These orders are exempt from being executed at the NBBO and may be executed against any limit orders resting below the best bid or above the best offer. Third, 1.57% of incorrectly classified trades are executed within a second of a hidden order on the same exchange. While this time band is quite large, it may be possible that the presence of a hidden order within an exchanges matching engines queue may alter the timing of what best quote was available when the displayed trade was executed, particularly in the context of a hidden orders differing execution priority compared to a traditional market or limit order.

Finally, 2.5% of inaccurately assigned trades are \$0.01 away from triggering the tick test, which, if triggered, would make them accurately signed. This is simply a limitation of the Lee and Ready (1991) algorithm, though notably, neither EMO nor CLNV have rules that would render these trades accurately signed.

In summary, there are several problems left to solve for more accurate trade and quote matching. First, a trade classification algorithm or trade and quote matching method that effectively deals with hidden and trade through exempt orders would improve accuracy by 3.06%. Second, any method that optimizes the use of the tick test, by expanding its usage trigger by a tick, but also minimizing its inaccuracy could lead to a theoretical improvement of 3.21%. Studies like Hagströmer (2021) offer some advancement in research methods that hinge on these arbitrary assignments. Finally, for the remainder of orders that are inaccurately classified for unknown reasons, our measures accurately estimate an individual order’s latency across exchanges, data centers, and cities. However, direct feeds, which are known to use faster technology than the SIP messages we use to estimate message latency, gateway latencies, unobserved by our estimates, weather, and other factors may all affect the latency of a given order between exchanges and when it is executed. We hope future work on this topic may resolve these issues and improve trade classification accuracy further.

6 Conclusion

We provide two alternatives to traditional methods of trade-to-quote matching in Daily TAQ. The first method adjusts for quote message latency and reconstructs an exchanges perspective of the NBBO at that moment, a method which we call the Relative Best Bid and Offer (RBBO). Our second method, the Latency Timestamp-adjusted (LTA) method, which is convenient to apply, adjusts for quote message latency, matching trades to quotes that an exchange would have present on its book at the moment a trade was executed, as displayed by the SIP. We find substantial improvements in trade classification accuracy by testing these methods. This improvement in accuracy could affect research inferences by impacting several common measures of liquidity, short selling, trading costs, and informed trading. We confirm this by exploring how a technological shock to the speed of equity markets would, under conventional methods, appear to cause a significant shock to the spreads and price impact. Our methods shed further light on the effects of latency and geography in today’s high-speed equity markets. These issues manifest in the latency between exchanges, direct feeds between exchanges compared to SIP routing and its effects on this latency, and the exchange-specific construction of RBBOs in response to these latency issues.

References

- Aitken, M., and A. Frino, 1996, The determinants of market bid ask spreads on the Australian stock exchange: Cross-sectional analysis, *Accounting & Finance* 36:51-63.
- Albuquerque, R., S. Song, and C. Yao, 2020, The price effects of liquidity shocks: A study of the SECs tick size experiment. *Journal of Financial Economics* 138:700-24.
- Asquith, P., R. Oman, and C. Safaya, 2010, Short sales and trade classification algorithms, *Journal of Financial Markets* 13:157-73.
- Back, K., K. Crotty, and T. Li, 2018, Identifying information asymmetry in securities markets, *The Review of Financial Studies* 31:2277-325.
- Barardehi, Y. H., D. Bernhardt, and R. J. Davies, 2019, Trade-time measures of liquidity, *The Review of Financial Studies* 32:126-79.
- Bartlett, R. P., and J. McCrary, 2019, How rigged are stock markets? evidence from microsecond timestamps, *Journal of Financial Markets* 45:37-60.
- Bernile, G., J. Hu, and Y. Tang, 2016, Can information be locked up? informed trading ahead of macro-news announcements, *Journal of Financial Economics* 121:496-520.
- Bessembinder, H., 2003, Issues in assessing trade execution costs, *Journal of Financial Markets* 6:233-57.
- Bessembinder, H., L. T. Allen Carrion, and K. Venkataraman, 2016, Liquidity, resiliency and market quality around predictable trades: Theory and evidence, *Journal of Financial Economics* 121:142-66.
- Bessembinder, H., J. Hao, and K. Zheng, 2020, Liquidity provision contracts and market quality: Evidence from the New York stock exchange, *The Review of Financial Studies* 33:44-74.
- Blais, M., and P. Protter, 2012, Signing trades and an evaluation of the lee-ready algorithm, *Annals of Finance* 8:1-13.
- Boehmer, E., C. M. Jones, and X. Zhang, 2013, Shackling short sellers: The 2008 shorting ban, *The Review of Financial Studies* 26:1363-400.
- Boehmer, E., and J. J. Wu, 2013, Short selling and the price discovery process, *The Review of Financial Studies* 26:287-322.
- Brogaard, J., T. Hendershott, and R. Riordan, 2017, High frequency trading and the 2008

- short-sale ban, *Journal of Financial Economics* 124:22-42.
- Brogaard, J., T. Hendershott, and R. Riordan, 2019, Price discovery without trading: Evidence from limit orders, *The Journal of Finance* 74:1621-58.
- Brogaard, J., and J. Pan, 2022, Dark Pool Trading and Information Acquisition, *The Review of Financial Studies* 35:2625-2666.
- Chakrabarty, B., B. Li, V. Nguyen, and R. A. V. Ness, 2007, Trade classification algorithms for electronic communications network trades, *Journal of Banking & Finance* 31:3806-21.
- Chakrabarty, B., P. C. Moulton, and A. Shkilko, 2012, Short sales, long sales, and the leeready trade classification algorithm revisited, *Journal of Financial Markets* 15:467-91.
- Chakrabarty, B., R. Pascual, and A. Shkilko, 2015, Evaluating trade classification algorithms: Bulk volume classification versus the tick rule and the lee-ready algorithm, *Journal of Financial Markets* 25:52-79.
- Chordia, T., T. C. Green, and B. Kottimukkalur, 2018, Rent seeking by low-latency traders: Evidence from trading on macroeconomic announcements, *The Review of Financial Studies* 31:4650-87.
- Chordia, T., R. Roll, and A. Subrahmanyam, 2002, Order imbalance, liquidity, and market returns, *Journal of Financial Economics* 65:111-130.
- Chung, K. H., A. J. Lee, and D. Rösch, 2020, Tick size, liquidity for small and large orders, and price informativeness: Evidence from the tick size pilot program, *Journal of Financial Economics* 136:879-99.
- Clark-Joseph, A. D., M. Ye, and C. Zi, 2017, Designated market makers still matter: Evidence from two natural experiments, *Journal of Financial Economics* 126:652-67.
- Collin-Dufresne, P., and V. Fos, 2015, Do prices reveal the presence of informed trading?, *The Journal of Finance* 70:1555-82.
- Comerton-Forde, C., V. Grégoire, and Z. Zhong, 2019, Inverted fee structures, tick size, and market quality, *Journal of Financial Economics* 134:141-64.
- Comerton-Forde, C., T. Hendershott, C. M. Jones, P. C. Moulton, and M. S. Seasholes, 2010, Time variation in liquidity: The role of market-maker inventories and revenues, *Journal of*

- Finance* 65:295-331.
- Conrad, J., and S. Wahal, 2020, The term structure of liquidity provision, *Journal of Financial Economics* 136:239-59.
- Conrad, J., S. Wahal, and J. Xiang, 2015, High-frequency quoting, trading, and the efficiency of prices, *Journal of Financial Economics* 116:271-91.
- CTA, 2019, NMS network custom FAQs, [CTA website](#).
- CTA, 2020, CTA pillar technology migration, [CTA website](#).
- Ding, S., J. Hanna, and T. Hendershott, 2014, How slow is the nbbo? a comparison with direct exchange feeds, *Financial Review* 49:313-32.
- Duarte, J., E. Hu, and L. Young, 2020, A comparison of some structural models of private information arrival, *Journal of Financial Economics* 135:795-815.
- Duarte, J., and L. Young, 2009, Why is pin priced?, *Journal of Financial Economics* 91:119-38.
- Easley, D., and M. O'Haram 2004, Information and the cost of capital, *Journal of Finance* 59:1553-83.
- Easley, D., M. L. de Prado, and M. OHara, 2016, Discerning information from trade data, *Journal of Financial Economics* 120:269-85.
- Easley, D., N. M. Kiefer, M. OHara, and J. B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *The Journal of Finance* 51:1405-36.
- Ellis, K., R. Michaely, and M. OHara, 2000, The accuracy of trade classification rules: Evidence from nasdaq, *Journal of Financial and Quantitative Analysis* 35:529-51.
- Ernst, T., J. Sokobin, and C. Spatt, 2021, The value of off-exchange data, Working paper.
- Fang, V. W., T. H. Noe, and S. Tice, 2009, Stock market liquidity and firm value, *Journal of Financial Economics* 94:150-69.
- Fang, V. W., X. Tian, and S. Tice, 2014, Does stock liquidity enhance or impede firm innovation?, *The Journal of Finance* 69:2085-125.
- Finucane, T. J, 2000, A direct test of methods for inferring trade direction from intra-day data, *Journal of Financial and Quantitative Analysis* 35:553-76.
- Golovachev, Y., A. Etinger, G. Pinhasi, and Y. Pinhasi, 2019, The effect of weather conditions on millimeter wave propagation, *International Journal of Circuits, System, and Signal*

- Processing* 13:690-5.
- Hagströmer, B, 2021, Bias in the effective bid-ask spread, *Journal of Financial Economics* 142:314-37.
- Hasbrouck, J, 2009, Trading costs and returns for u.s. equities: Estimating effective costs from daily data, *The Journal of Finance* 64:1445-77.
- Hasbrouck, J, and G. Saar, 2013, Low-latency trading, *Journal of Financial Markets* 16:646-79.
- Holden, C. W., and S. Jacobsen, 2014, Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions, *Journal of Finance* 69:1747-85.
- Holden, C. W., S. Jacobsen, and A. Subrahmanyam, 2014, The empirical analysis of liquidity, *Foundations and Trends in Finance* 8:263-365.
- Lee, C. M., and B. Radhakrishna, 2000, Inferring investor behavior: Evidence from torq data, *Journal of Financial Markets* 3:83-111.
- Lee, C. M., and M. J. Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 46:733-46.
- Li, S., Ye, M., and Zheng, M., 2023, Refusing the best price?, *Journal of Financial Economics* 147:317-337
- Lou, X., and T. Shu, 2017, Price impact or trading volume: Why is the amihud (2002) measure priced?, *The Review of Financial Studies* 30:4481-520.
- NYSE, 2018, Letter from Michael Blaugrund, head of transactions, NYSE, [SEC website](#).
- NYSE, 2019, NYSE Pillar migration – adding efficiency to the marketplace, [NYSE website](#).
- Odders-White, E. R, 2000, On the occurrence and consequences of inaccurate trade classification, *Journal of Financial Markets* 3:259-86.
- OHara, M., C. Yao, and M. Ye, 2014, Whats not there: Odd lots and market data, *The Journal of Finance* 69:2199-236.
- Peterson, M., and E. Sirri, 2003, Evaluation of the biases in execution cost estimation using trade and quote data, *Journal of Financial Markets* 6:259-80.
- Rosenthal, D. W., 2012, Modeling trade direction, *Journal of Financial Econometrics*, 10:390-415
- Sarkar, A., and R. T. A. Schwartz, 2009, Market sidedness: Insights into motives for trade

- initiation, *The Journal of Finance* 64:375-423.
- SEC, 2018, Roundtable on market data products, market access services, and their associated fees, [SEC website](#).
- SEC, 2020a, Market data infrastructure, release no. [34-88216](#).
- SEC, 2020b, Notice of proposed order directing the exchanges and the financial industry regulatory authority to submit a new national market system plan regarding consolidated equity market data, release no. [34-87906](#).
- Shkilko, A., and K. Sokolov, 2020, Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs, *The Journal of Finance* 75:2899-927.
- Theissen, E, 2001, A test of the accuracy of the lee/ready trade classification algorithm, *Journal of International Financial Markets, Institutions and Money* 11:147-65.
- UTP, 2021, Utp plan administration data policies, [UTP website](#).
- van Kervel, V, 2015, Competition for order flow with fast and slow traders, *The Review of Financial Studies* 28:2094-127.
- Weller, B. M, 2019, Measuring tail risks at high frequency, *The Review of Financial Studies* 32:3571-616.
- Werner, I. M, 2003, NYSE order flow, spreads, and information, *Journal of Financial Markets* 6:309-35.
- Yao, C., and M. Ye, 2018, Why trading speed matters: A tale of queue rationing under price controls, *The Review of Financial Studies* 31:2157-83.
- Yeung, W. F., and A. R. Johnston, 1978, Effect of temperature on optical fiber transmission, *Applied Optics* 17:3703-5.

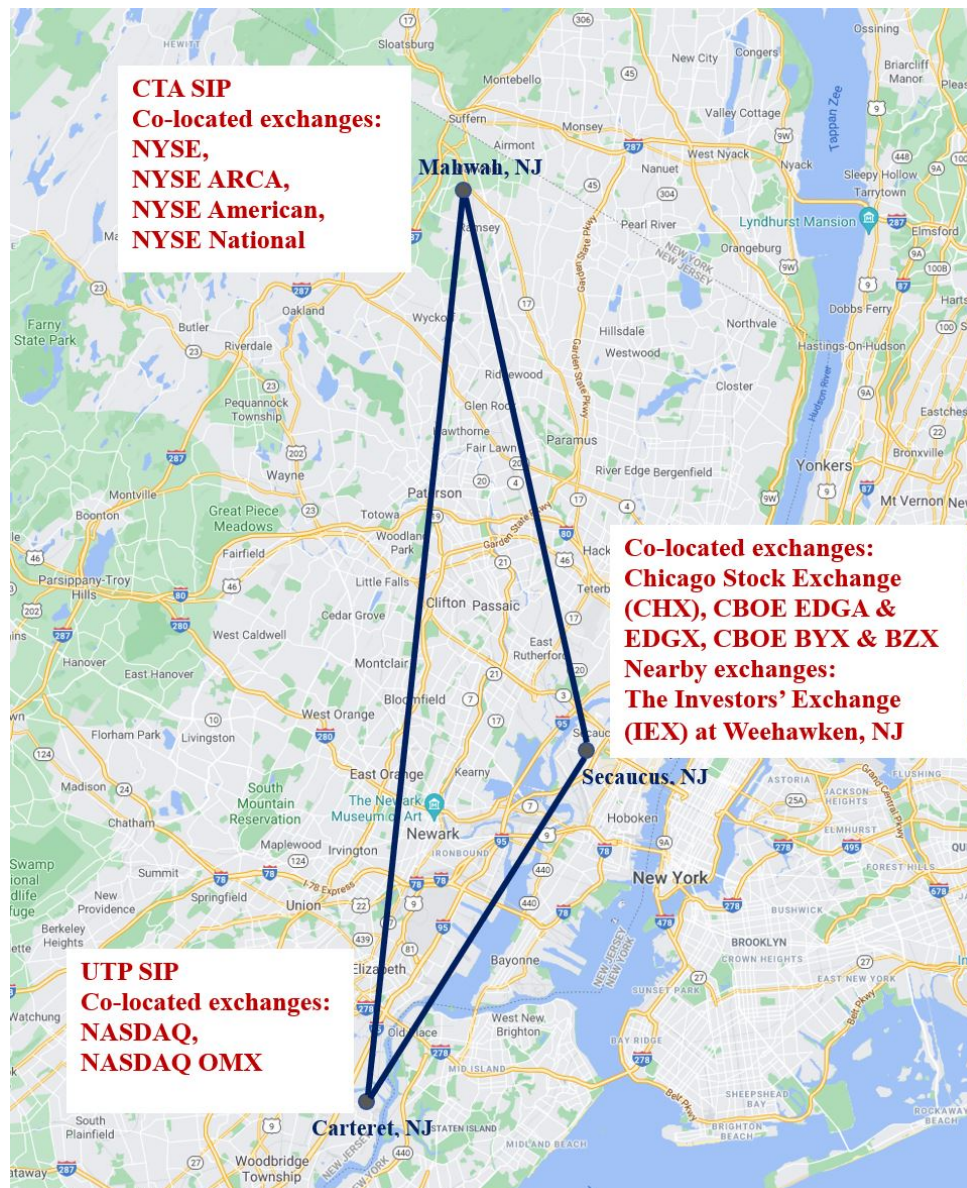


Figure 1. SIPs and Exchange Data Center Locations. This map shows the data center co-locations of the major exchanges in the United States. The three co-located cities are: Mahwah, Carteret, Secaucus in New Jersey. The Investors' Exchange (IEX) is located in Weehawken, NJ, nearby Secaucus, NJ (for simplicity they are marked as one).

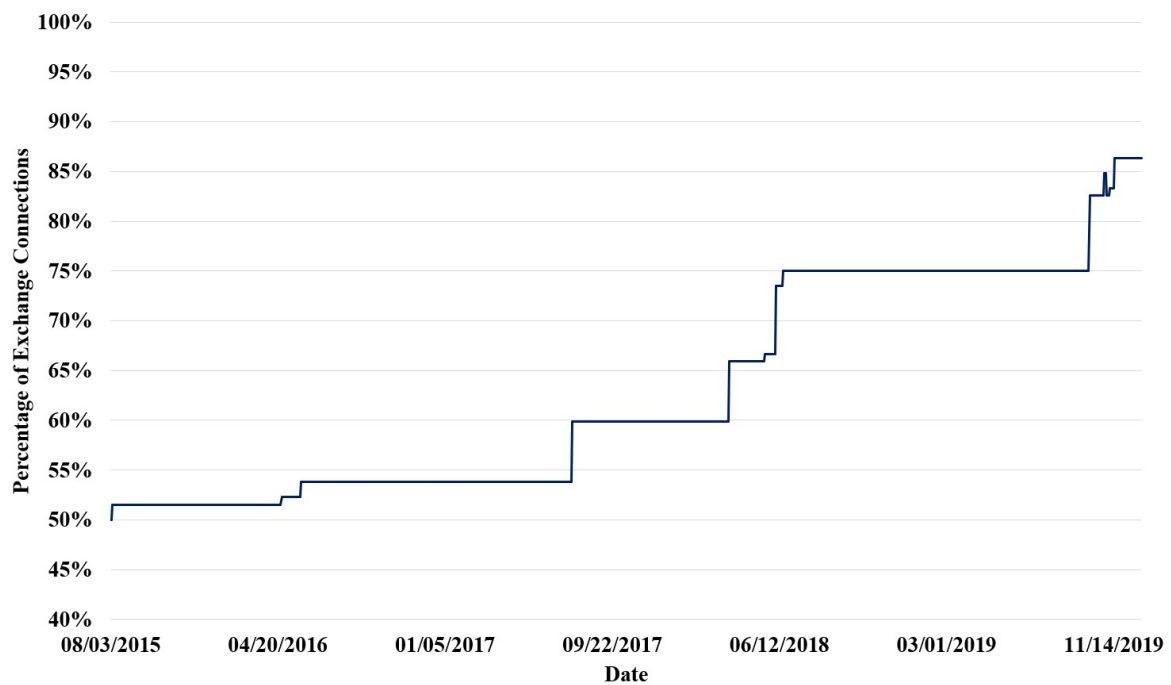


Figure 2. Percent of Direct Feeds as Primary Exchange Connections. This chart plots the percentage of all exchange-to-exchange connection pairs between 10 lit venues from 2015 August to 2019 December (IEX is excluded).

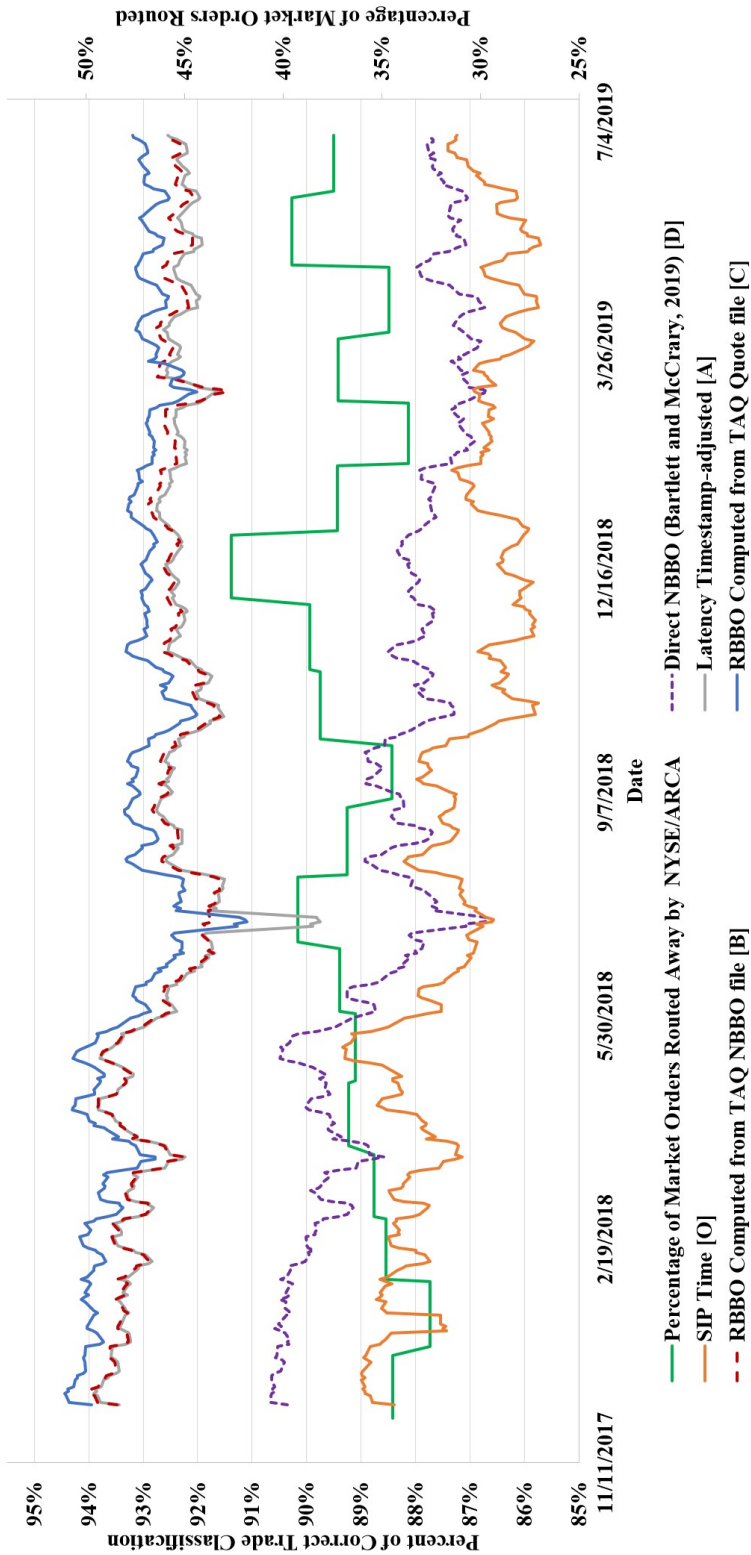


Figure 3. Lee Ready Trade Classification Accuracy and Latency Adjustment. This figure plots the 5-day moving average trade classification accuracy of sampled trades from 2017 December to 2019 June. Over 650 million regular trades from TAQ are unique matched with NYSE and NYSE ARCA order book based on trade price, traded shares, stock symbol, and nanosecond timestamp. Trade direction, buy or sell, is inferred from the order book data where the order submission message records the direction of the trade. The percent of correct trade classification is calculated as the number of trades with the correct trade assignment over the total number of matched trades. We calculate this accuracy ratio for every day in the sample. We plot five series of trade classification accuracy in this chart – [A] the latency Timestamp-adjusted method; [B] RBBO computed using the TAQ complete NBBO file; and [C] RBBO computed from the TAQ quote file, [D] Bartlett and McCrary (2019)’s direct NBBO method, and [O] the conventional SIP time method. The labelling of the five series in the plot corresponds to those in Table VI. We plot the monthly percentage of market orders (share volume inferred) routed away by NYSE and NYSE ARCA (the secondary vertical axis on the right-hand-side) using the exchange Rule 605 reports.

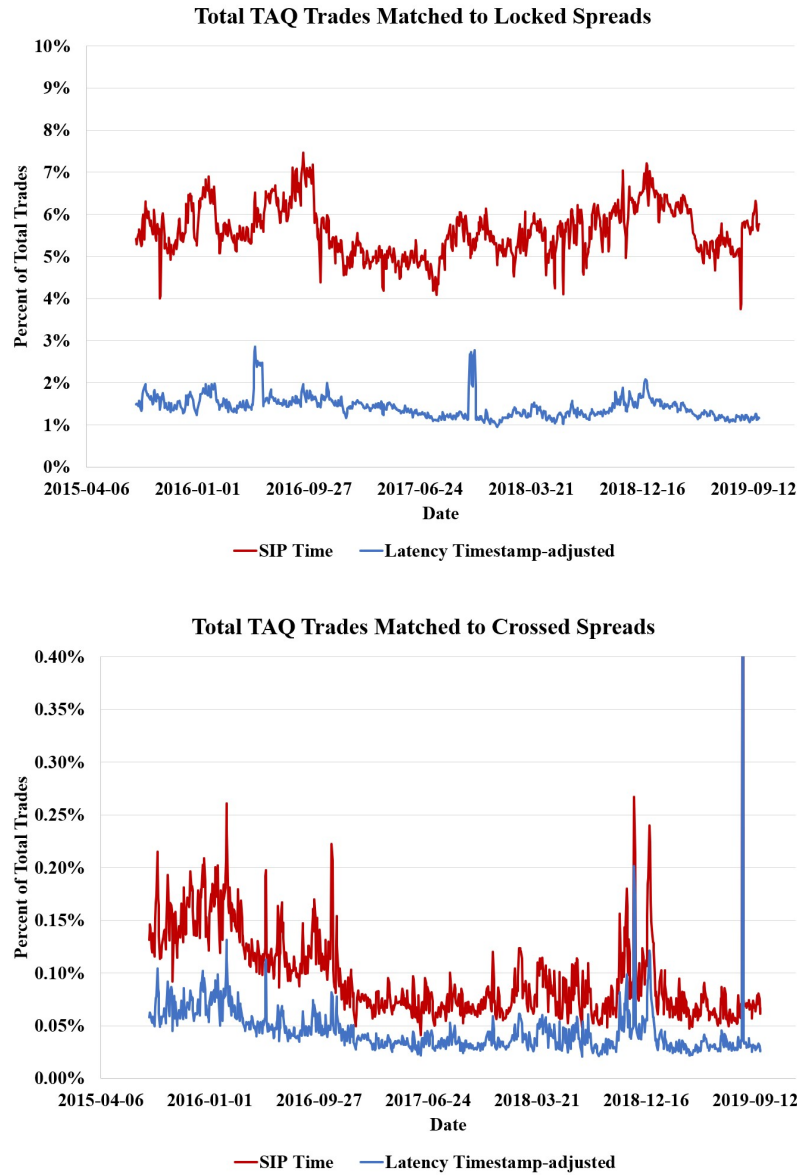


Figure 4. Percent of Trades Matched to Locked Spreads and Crossed Spreads. This chart plots the percentage of all TAQ trades that are matched to its prevailing NBBO quote and the quote is crossed ($NBB > NBO$) or locked ($NBB = NBO$). We plot two series in each chart for comparison from 2015 June to 2019 August. The SIP time series matches trades with their prevailing NBBOs based on trades' and quotes' SIP time. The latency timestamp-adjusted series matches the trades with their prevailing NBBOs and adjusts for the latency between the SIP and the exchange. The upper chart plots the comparison of trades matched to quotes that are locked, and the lower chart plots the comparison of the trades matched to crossed spreads. The spike on 8/12/2019 with 1.4% crossed spread is caused by a network component failure at Mahwah which affected the Consolidated Tape System (CTS). On this spike day, the trade and quote records have gaps during 3:17pm-3:45pm.

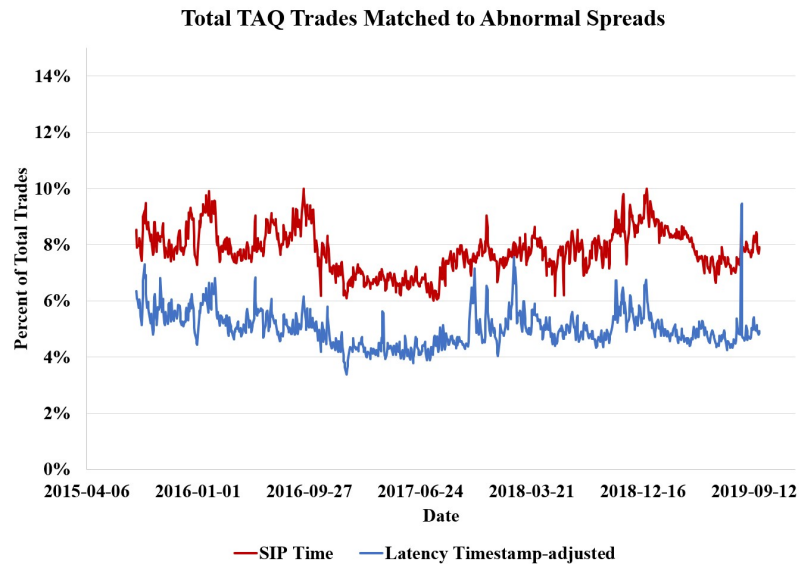


Figure 4 (cont). Percent of Trades Matched to Abnormal Quotes. This chart plots the percentage of all TAQ trades that are matched to its prevailing NBBO quotes and the quotes are abnormal (being locked, crossed, or trade price is outside NBBO) from 2015 June to 2019 August. We plot two series – the SIP time series matches trades with their prevailing NBBOs based on SIP time; the latency timestamp-adjusted series matches the trades with their prevailing NBBOs and adjusts for the latency between the SIP and the exchange.

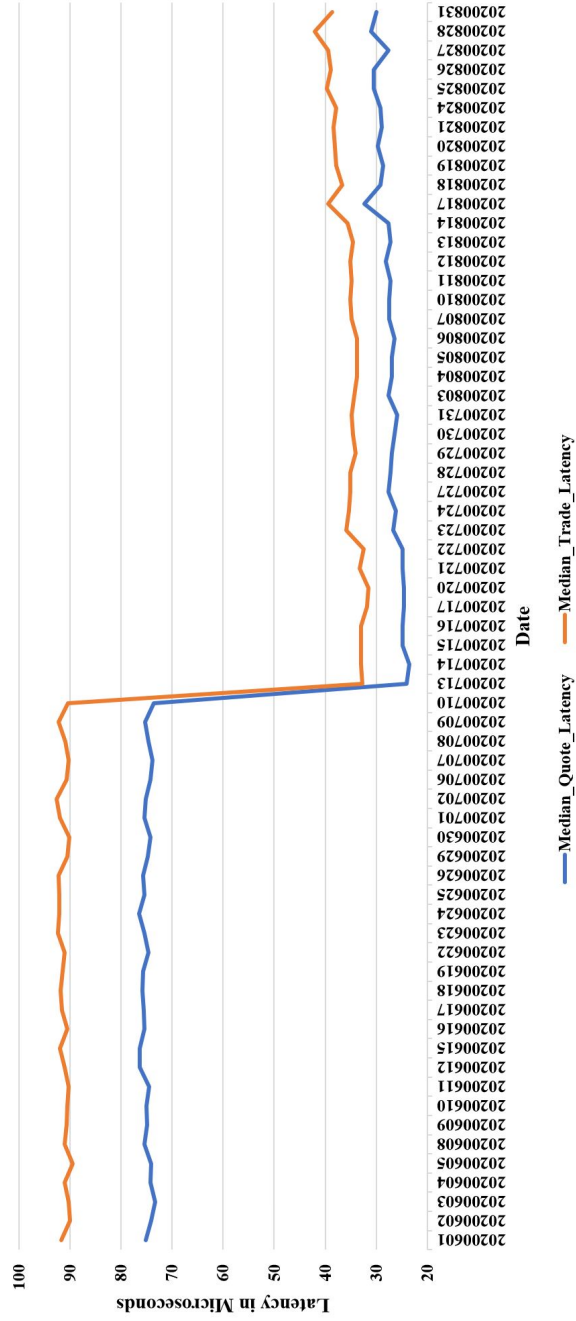


Figure 5. NYSE Trade and Quote Message Latency before and after NYSE Pillar Transition. This figure plots the median trade and quote message latency around the NYSE pillar migration on July 13th, 2020. As observed in the TAQ data, latency dropped significantly after the NYSE Pillar migration. According to NYSE (2019), the migration reduced up to a 90% in latency and reduced cost. Our observations of latency reduction in TAQ matches with the “Binary Pillar gateway” latency reduction reported in NYSE (2019).

Table I
Observed SIP-to-Exchange Trade and Quote Message Latency

We report the latency (SIP time less Participant time) for all trades and quotes per UTP/CTA network for each exchange on a representative trading day (June 3rd, 2019). For each exchange, we report the median quote/trade message latency in microseconds. We also include the city where the exchange's data center is located. For CTA/UTP SIP co-locations on a map, see Figure 1.

Exchange Name (TAQ)	Trade and Quote Message Latency on June 3 rd , 2019					
	CTA Quote Msg	CTA Trade Msg	UTP Quote Msg	UTP Trade Msg	Exchange Data Center	
	Median Latency in Microseconds (μ s)	Median Latency in Microseconds (μ s)	Median Latency in Microseconds (μ s)	Median Latency in Microseconds (μ s)		
NYSE National, Inc (NYSE National)	102.345	148.603	367.240	370.868		Mahwah
New York Stock Exchange, LLC (NYSE)	105.149	153.512	369.059	372.009		Mahwah
NYSE American, LLC (NYSE American)	105.491	158.650	372.404	375.610		Mahwah
NYSE Arca, Inc. (NYSE Arca)	110.999	160.198	373.455	377.399		Mahwah
Chicago Stock Exchange, Inc. (CHX)	10,254.062	1,064.678	8,025.529	570.807		Secaucus
Cboe BYX Exchange, Inc (Cboe BYX)	401.249	456.071	192.712	209.868		Secaucus
Cboe BZX Exchange, Inc (Cboe BZX)	403.041	464.326	193.306	212.089		Secaucus
Cboe EDGA Exchange (Cboe EDGA)	406.564	468.415	200.341	219.669		Secaucus
Cboe EDGX Exchange (Cboe EDGX)	408.735	472.421	200.846	221.421		Secaucus
The Investors Exchange, LLC (IEX)	449.464	491.171	215.936	236.147		Weehawken
NASDAQ OMX BX, Inc.	536.811	578.221	16.666	19.387		Carteret
NASDAQ Stock Market, LLC (NASDAQ)	539.873	586.388	17.042	21.591		Carteret
NASDAQ OMX PSX, Inc.	540.073	593.972	18.056	24.054		Carteret

Table II
Latency Adjustment Methods and Trade Classification Accuracy

We present our two latency adjustment methods and their trade classification accuracy and compare them with the conventional SIP time method [A]. The SIP time [A] matches trades to their prevailing NBBO based on the SIP time. The RBBO method [B] constructs a city-based relative BBO (RBBO) top-of-the-book for each of the three exchange data center co-located cities (Mahwah, Carteret, and Secaucus) using the TAQ quote file. The trades are matched to the prevailing RBBO quotes with latency adjustment between the local city and the away cities. The Latency Timestamp-adjusted method [Panel C] adjusts the TAQ trade and NBBO quote timestamps by the latency that travels from the SIP to the exchange. We report the percentage of matched trades that are signed correctly by the RBBO method [B] and the latency timestamp-adjusted method [C]. The TAQ-to-order-book matched trades in the sample are constructed following the trade matching outlined in Appendix A.1. We test three prevalent buy-sell trade classification algorithms – “Lee Ready” using Lee Ready (1991), “EMO” using Ellis, Michaely, and O’Hara (2000), and “CLNV” using Chakrabarty, et al. (2007). The same trade classification algorithm in [B] and [C] only differs in how trades are matched to the prevailing quote to infer the side of the trade. Our samples include all stocks all trades and does not limit on stock prices or minimum daily trading volume. We report the differences and the significance of the difference for each stock-day for methods [B] and [C] benchmarking with SIP [A] comparatively.

		Trade Classification Accuracy		
		Lee Ready	EMO	CLNV
		Accuracy	Accuracy	Accuracy
		Rate	Rate	Rate
Conventional SIP Time	[A]	86.75%	87.50%	87.50%
Relative Best Bid Offer (RBBO)	[B]	92.63%	92.44%	92.42%
<i>Difference</i>	[B-A]	5.88%*** (164.73)	4.94%*** (147.36)	4.92%*** (144.42)
Latency Timestamp-adjusted (LTA)	[C]	92.05%	91.86%	91.93%
<i>Difference</i>	[C-A]	5.30%*** (125.50)	4.36%*** (121.09)	4.43%*** (120.64)

t-statistics in parentheses, *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$

Table IV
Common Liquidity and Informativeness Measures

We compare the common liquidity and informativeness measures using the latency timestamp-adjusted (LTA) and the conventional SIP time for our order book sample. We filter for 2,700 common stocks (CRSP SHRCD in ('10','11') and listing exchange at NYSE, Nasdaq, and Amex) in our order book sample. We use Lee and Ready (1991) to sign the trades, and calculate simple-average daily percentage effective spread, realized spread, and price impact following Equations 6.1, 6.2, and 6.3, and present the comparison results in Panel A. We calculate the absolute order imbalances (abs(buys-sells)/(buys+sells)) and the probability of informed trading (using Equations 7.1 and 7.2). For the PIN measure, we also check Hessian matrix and drop possible corner solutions that violate the second order condition. We also include a set of liquidity and informativeness measures computed using imbalance-weighted midpoint [IW] following Hagströmer (2021), and benchmark against the regular midpoint [Reg]. All measures are averaged at stock-level. We report the differences and the significance of the difference across the timestamp dimension (SIP time vs LTA) as well as the midpoint modifications (Imbalance-weighted midpoint vs regular midpoint). .

		Panel A Liquidity Measures (in %)							
		[Reg] Regular Midpoint				[IW] Imbalance-weighted Midpoint			
		Effective Spread	Realized Spread	Price Impact		Effective Spread	Realized Spread	Price Impact	
SIP Time	[A]	0.363	0.205	0.168		0.394	0.226	0.173	
LTA	[B]	0.382	0.176	0.212		0.418	0.194	0.230	
<i>Difference</i>	<i>[B]-[A]</i>	0.019*** (58.19)	-0.024*** (-39.93)	0.043*** (52.01)		0.015*** (43.96)	-0.030*** (-39.81)	0.045*** (55.83)	
<i>t</i> -statistics in parentheses, ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10									
		Panel B Order Imbalance and PIN (in %)							
		[Reg] Regular Midpoint		[IW] Imbalance-weighted Midpoint		Difference [IW]-[Reg]			
		Absolute Order Imbalance	PIN	Absolute Order Imbalance	PIN	Absolute Order Imbalance	Annual PIN	Absolute Order Imbalance	Annual PIN
SIP Time	[A]	11.472	17.692	12.975	18.433	1.703*** (73.80)	0.631*** (3.98)	1.703*** (73.80)	0.631*** (3.98)
LTA	[B]	11.765	18.033	13.027	18.696	1.252*** (75.25)	0.403*** (2.50)	1.252*** (75.25)	0.403*** (2.50)
<i>Difference</i>	<i>[B]-[A]</i>	0.289*** (35.52)	0.341** (2.15)	0.0452*** (5.52)	0.263* (1.73)				
<i>t</i> -statistics in parentheses, ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10									

Table V
Latency Adjustment and Abnormal Quotes

This table reports the percentage of all TAQ trades that are matched to its prevailing NBBO quote and the quote is crossed (NBB>NBO) or locked (NBB=NBO) from 2015 June to 2019 August. We present two series for comparison. The SIP time [A] matches trades with prevailing NBBO based on trades' and quotes' SIP time. The latency timestamp-adjusted [B] matches trades with their prevailing NBBOs and adjusts for the latency between the SIP and the exchange. This table reports the percentage of trades matched to quotes that are locked, crossed, or the trade price is outside the matched NBBO quote, along with the sum of all three cases as total abnormal quotes. We report the differences and the significance of the difference for each stock-day between methods [A] and [B] comparatively.

		Trades Matched with Abnormal Quotes			
		Percent of Trades Matched to Locked Quotes	Percent of Trades Matched to Crossed Quotes	Percent of Trades Matched to Outside NBBO Quotes	Percent of Trades Matched to Locked, Crossed, or Outside NBBO Quotes
Conventional SIP Time Latency Timestamp-adjusted (LTA) <i>Difference</i>	[A]	5.608%	0.097%	2.606%	7.792%
	[B]	1.424%	0.045%	3.890%	5.024%
	[B-A]	-4.184%*** (-238.86)	-0.052%*** (-72.21)	1.283%*** (125.03)	-2.767%*** (-146.07)

t-statistics in parentheses, ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10

Table VI
Computation Time, Coding Complexity, and Classification Accuracy

We report the computation time cost, the trade classification accuracy improvement, and the coding complexity of the proposed latency adjustment methods. The LTA method [A] adjusts the TAQ trade and NBBO quote timestamps by the latency that travels from the SIP to the exchange. The RBBO method [C] constructs a city-based relative BBO (RBBO) top-of-the-book for each of the three exchange data center co-located cities (Mahwah, Carteret, and Secaucus) using the TAQ quote file. The trades are matched to the prevailing city RBBO quotes with latency adjustment between the local city and the away cities. We include the benchmark method [O] that runs on the Holden and Jacobsen (2014) SAS code, which matches the trade to the previous millisecond complete NBBO quote, as well as [D] Bartlett and McCrory (2019)'s Direct NBBO method, which matches trades to its prevailing NBBO quotes based on participant exchange time. We report the average SAS code run time of the proposed methods for all stocks on a given day to matches trades to quotes with the associated method, sign the trades with Lee Rerady, and compute the daily liquidity measures. We also include the description of coding complexity of the proposed methods. For the LTA method [A], we provide the two-liner addition to Holden and Jacobsen (2014)'s SAS code in Appendix A.4.

		Average Computing Time for One-day All Stock's Liquidity Measures	Computation Cost Comparison		Coding Complexity
			Lee Ready Trade Classification Accuracy	Accuracy Improvement Compared to [O]	
Latency Timestamp-adjusted (LTA)	[A]	30-40 minutes	0.9205	5.30%***	Add two lines to Holden and Jacobsen (2014)'s SAS code
Near-perfect City-based Relative BBO (RBBO) Computed from TAQ Quote File	[C]	3 hours	0.9263	5.88%***	Generate RBBO file from Daily TAQ Quote file based on datacenter locations
Bartlett and McCrory (2019)'s Direct NBBO Method	[D]	40-50 minutes	0.8795	1.21%***	Match trades to their prevailing NBBO quotes by participant exchange time
Benchmark: Conventional SIP Time Matching Method	[O]	30-40 minutes	0.8675	-	Holden and Jacobsen (2014)'s SAS code

***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10

Table VII
NYSE-to-SIP Latency Before and After Pillar Migration

We report the average quote and trade message latency in a [-5 day, +5 day] event window around NYSE's most recent Pillar upgrade on July 13th, 2020. We report the differences and the significance of the difference for each stock before the Pillar migration [A] and after the Pillar migration [B] comparatively.

		[-5,+5] Day Event Study of NYSE Pillar Migration on July 13 th , 2020	
		NYSE Avg. Quote Message Latency (in microseconds μ s)	NYSE Avg. Trade Message Latency (in microseconds μ s)
Before NYSE Pillar Migration	[A]	74.320	90.892
After NYSE Pillar Migration	[B]	24.371	32.717
<i>Difference</i>	[B-A]	-49.949*** (-113.65)	-58.175*** (-127.88)

t-statistics in parentheses, ***p \leq 0.01, **p \leq 0.05, *p \leq 0.10

Table VIII
Liquidity Measures Around NYSE Pillar Migration

We examine the liquidity measures in a [-5 day, +5 day] trading window around NYSE's most recent Pillar upgrade on July 13th, 2020. We report the average percentage effective spread, percentage realized spread, and percentage price impact before and after the event using both the SIP time trade-to-quote matching method and the latency timestamp-adjusted method. For a stock to be included in this event study, the stock must be trading with no less than 1,000 trades for at least one day in the 5-day pre-event window, as well as at least one day in the 5-day post-event window. For each stock, we compute the 5-day average of the liquidity measure and compare their values before and after event. We report the significance of the liquidity measures, as well as their differences before and after the Pillar migration event for about 4,800 stocks that meet the trading activity filter. We present the event study result for all TAQ trades in panel A, and the result for NYSE (EX='N') trades in panel B.

Panel A All TAQ Trades				
		Percent Effective Spread (%)	Percent Realized Spread (%)	Percent Price Impact (%)
SIP Time				
Before NYSE Pillar Migration	[A]	0.186	0.067	0.120
After NYSE Pillar Migration	[B]	0.188	0.063	0.125
<i>Difference</i>	[B-A]	0.002 (1.41)	-0.004** (-2.01)	0.006*** (2.63)
Latency Timestamp-adjusted (LTA)				
Before NYSE Pillar Migration	[C]	0.205	0.044	0.161
After NYSE Pillar Migration	[D]	0.205	0.042	0.163
<i>Difference</i>	[D-C]	-0.000 (-0.19)	-0.002 (-0.84)	0.001 (0.62)
<i>t</i> -statistics in parentheses, ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10				
Panel B NYSE Trades Only				
		Percent Effective Spread (%)	Percent Realized Spread (%)	Percent Price Impact (%)
SIP Time				
Before NYSE Pillar Migration	[A]	0.223	0.060	0.163
After NYSE Pillar Migration	[B]	0.223	0.046	0.177
<i>Difference</i>	[B-A]	-0.000 (-0.03)	-0.014** (-2.02)	0.014** (2.07)
Latency Timestamp-adjusted (LTA)				
Before NYSE Pillar Migration	[C]	0.263	0.010	0.253
After NYSE Pillar Migration	[D]	0.262	0.009	0.253
<i>Difference</i>	[D-C]	-0.001 (-0.41)	-0.001 (-0.12)	-0.000 (-0.06)
<i>t</i> -statistics in parentheses, ***p ≤ 0.01, **p ≤ 0.05, *p ≤ 0.10				

Appendix A

A.1 How to Match TAQ Trades with Order Book Data

We match NYSE and ARCA order book order execution messages with TAQ trades based on stock symbol, date, trade price, size (trade volume) of the trade, and execution time (in order book) with the exact nanosecond participant time (in Daily TAQ). See an example below:

NYSE ARCA order book, Apple. Inc (symbol=“AAPL”), March 29th, 2019, at 10:03:10.804101376, a market order (order ID=“00016044086562792313”) executed against a sitting limit order (order ID=“00016044086562791062”). In the order book order execution messages (MsgType=103), we could find two records with the same tradeID=“649014”:

MsgType	Symbol	SourceTime	OrderID	Price	Volume	TradeID	...
...							
103	AAPL	10:03:10.804101376	00016044086562791062	189.7000	79	649014	
...							
103	AAPL	10:03:10.804101376	00016044086562792313	189.7000	79	649014	
...							

In Daily TAQ trades, based on the same date, symbol, trade price, size, and exact nanosecond timestamp, we could find the following observation in TAQ trades:

Date	Symbol	SIP Time (Time_m)	Participant Exchange Time (Part_time)	Price	Volume	TradeID	...
...							
20190329	AAPL	10:03:10.804481417	10:03:10.804101376	189.7	79	2991	
...							

The participant time in Daily TAQ trades matches the exact nanosecond SourceTime in order book trade executions.³⁰

The last step, we go to the order submission messages (MsgType=100) and find the direction of the order to decide if the trade is buy-initiated or sell-initiated. In this example, the order book order submission message indicate the market order (OrderID=“...2313”) is a market sell and the sitting limit order (OrderID=“...1062”) is a limit buy.

³⁰There are 26 AAPL trades during the whole second 10:03:10 am to 10:03:11 am on March 29th, 2019. Only one trades with the price of 189.70 and volume of 79 shares. It is the only candidate trade that would match with

MsgType	Symbol	SourceTime	OrderID	Price	Volume	Side	...
...							
100	AAPL	10:03:10.220877312	00016044086562791062	189.7000	79	B	
...							
100	AAPL	10:03:10.799669504	00016044086562792313	189.7000	79	S	
...							

We suggest future research use this approach to match order book executions with TAQ trades. You could also match any broker/dealer or wholesaler proprietary data with Daily TAQ trades based on symbol, date, price, volume, and time. Please be aware that requiring the execution time (in the exchange’s order book) to be the exact nanosecond as the participant time (in Daily TAQ trades) is a very strict filter. According to the [Daily TAQ manual \(V4.0\)](#), page 9, you could allow up to 100 microseconds difference when matching based on timestamps.³¹ This approach work best for matching regular unconditional trades in TAQ on a given exchange. To match conditional trades such as intermarket sweep orders, please consider order routing across exchanges and the latencies (exchange gateway latency and travel latency). To match all trade types, you may want to relax the matching timestamp constraint and allow a matching timestamp difference buffer up to several milliseconds.

In today’s financial markets, Direct Feed is not only used as the primary feed in exchange-to-exchange communications, but also has been widely adopted as the main feed used by industry leading market makers, wholesalers, dark pools, and electronic trading platforms when communicating with exchanges. When the market players connect to an exchange via direct feed, the exchange sends the order execution messages to both the SIP and the market players at the same time. If the market player’s terminal is co-located with the exchange or at a close-by location, they could receive order execution confirmation before the SIP receives and prints the trades to the consolidated trade tape. Therefore, future research might find matching order execution time in the proprietary data from market makers or wholesalers with the trade’s participant exchange time yields better matching outcomes than the trade’s SIP time in Daily TAQ.

the order book order execution, showing it’s the exact same trade.

³¹See the following timestamp note in the Daily TAQ manual: “If from an Exchange: Timestamp 1 denotes the Exchange Matching Engine Publication timestamp for a transaction. Exchanges use a clock sync methodology ensuring that timestamps are accurate within tolerances of 100 microseconds or less. Exchanges provide the timestamp in terms of nanoseconds since Epoch.”

A.2 Why Is the Conventional SIP Time Method Flawed?

We discuss why the existing method of matching trades and quotes using SIP time is inappropriate. In a hypothetical scenario, we assume there are two exchanges – a quote originating exchange [A] and a trade originating exchange [B], and one SIP. The quote originating exchange [A] receives a new limit order whose price is superior to the quote originating exchange’s best price. The exchange’s matching engine then enters this quote into its book. Exchange [A] sends out two messages to other exchanges and the SIP – one message, via the direct feed, contains this quote and sends to all other exchanges that are connected via direct feed, including Exchange [B]; the other message leaves the exchange and network edge travels to the SIP at Quote@Exchange time. After $\text{QuoteExchange-to-SIP}$ travel time, the quote arrives at the SIP. The SIP receives the new quote and begins to process it for (1) the National Best Bid and Offer (NBBO), and (2) the Limit-UP Limit-Down (LULD). The SIP completes processing the quote at Quote@SIP time and the SIP broadcasts the updated National Best Bid Offer (NBBO) to all participant exchanges via the SIP feed. Thus, we re-write Equation 1 for this quote’s Quote@SIP time as:

$$\begin{aligned} \text{Quote@SIP Time} = & \text{Quote@Exchange Time} + \text{QuoteExchange-to-SIP Travel Time} \\ & + \text{SIP Quote Gateway Latency} + \text{SIP Quote Processing Time.} \end{aligned} \quad (8)$$

With the SIP broadcast, an NBBO update message is sent to all participant exchanges via the SIP feed, including the quote originating exchange [A] and the trade originating exchange [B]. If exchange [B] is using the SIP feed, it will receive the new NBBO and enters into its book. The trade originating exchange [B] matching engine considers a market order to be routed away, as the most recent NBBO infers exchange [A] has the national best prices at that moment.³² Exchange [B]’s matching engine sends the market order in-route to exchange [A] and the order is executed at the best prices at [A]. Exchange [A], where the quote originated, now becomes a trade executing exchange, and its matching engine matches the routed-to-here market order with the sitting limit order that has the best price. Once the trade is executed successfully, exchange [A] sends the trade message to the SIP at Trade@EXExchange time (denote “EX” for “Executing”). Therefore, with observable Trade@EXExchange time (which is the trades’ Participant Exchange

³²For simplicity, we use the term “market order” throughout this paper. We refer “market order” as marketable limit order, and orders that are immediately executable relative to the NBBO.

Time in Daily TAQ trades), researchers should match this Trade@EXExchange time with the NBBO at the Quote@SIP time plus the quote’s SIP-to-TradeExecutingExchange travel time. We cannot observe the time when the NBBO quote travels from the SIP and arrives at the trade executing exchange, but we can proxy using the trade’s trade executing exchange to SIP travel time when the exchange reports the trade to the SIP. As for the trade originating exchange [B], the routing of the orders may not always happen (depending on which exchange has the best price and what feed exchange [B] uses as primary), and even it happens, we do not have a timestamp in TAQ to observe.³³

Continuing on the hypothetical scenario, the Trade@SIP time is when the SIP receives the trade message from the trade executing exchange and keeps a record of the trade on the Consolidated Trade Tape. To show this, one can re-write Equation 1 and view the trade’s SIP time as:

$$\begin{aligned} \text{Trade@SIP Time} = & \text{Trade@EXExchange Time} + \text{TradeEXExchange-to-SIP Travel Time} \\ & + \text{SIP Trade Gateway Latency} + \text{SIP Trade Processing Time.} \end{aligned} \quad (9)$$

Finally, along with the trade records in the consolidated trade tape print, the SIP records the trade executing exchange [A] as the variable “EX” of the trade in Daily TAQ. In this hypothetical scenario, we assume the exchange’s matching engine gateway latency is negligible. If the market order fails to execute after routing from exchange [B] to exchange [A], for example, when the best prices in exchange [A] is no longer available the moment [A] receives the routed order, exchange [A] might have sent a new quote to the SIP. And Exchange [B] will re-process that order based on the state of its book at that time.³⁴

If the communications between the two exchanges [A] and [B] are established via the direct feed, it saves the amount of time it takes the quote originating exchange [A] sending the quote

³³Using the exchange 605 data, we find 36.68% of the market orders (share volume inferred) at NYSE and NYSE ARCA are routed away to other exchanges. Exchanges do not have to report to SIP when orders are routed out to other exchanges and one cannot observe the trade originating exchange of an routed-away order in TAQ.

³⁴In Daily TAQ, cancelled trades are not related to this scenario. No cancellation message or trade print will be reported and recorded by the SIP. Order routing is not directly observable in Daily TAQ.

to the SIP and the time it takes the SIP to process and broadcast to all participating exchanges. With direct feeds, the quote originating exchange [A] sends the quote message to all exchanges via direct feed. The market order at exchange [B] is routed to exchange [A] for execution immediately. Therefore, this hypothetical scenario also explains why the direct feed is faster than the SIP feed, even without any technological difference, and the motivation for exchanges to prefer direct feeds over the SIP feed. We conclude that the current SIP time matching method that matches trade using Trade@SIP time with prevailing NBBO using Quote@SIP time would result in the trade being late by (taking Equation 9 subtract Equation 8):

$$\begin{aligned}
& \textit{QuoteExchange-to-SIP Travel Time} + \textit{TradeEXExchange-to-SIP Travel Time} \\
& + \textit{SIP Trade Gateway Latency} - \textit{SIP Quote Gateway Latency} \quad (10) \\
& + \textit{SIP Trade Processing Time} - \textit{SIP Quote Processing Time}.
\end{aligned}$$

As shown in Equation 10, as long as the sum of the time elements is non-zero, the existing trade-to-quote matching using the SIP time method omits such exchange-to-SIP latency factors and could cause trades misaligned with the NBBO quotes that they likely executed against. In 2010, the sum of these time elements was as large as 10 milliseconds. And as a result, the SIP time method could produce inaccurate trade inference.

A.3 Weather, Temperature, Traffic, Transmission Method, and Latency

As briefly discussed in the paper, we find latency is not constant during the day. In fact, it varies and could be affected by the weather above new jersey, the temperature of the day, and the traffic load of the quote and trade messages. As we shown in latency decomposition, the geographic locations of the exchanges and the SIPs are the key determinant of the exchange-to-SIP latency and set the lower bound of latency by the speed of light. With some preliminary summary statistics, we find the three factors listed above are also likely affecting the exchange-to-SIP latency. We encourage future research to discover all possible determinates of latency and conclude more rigorously with hypothesis testing.

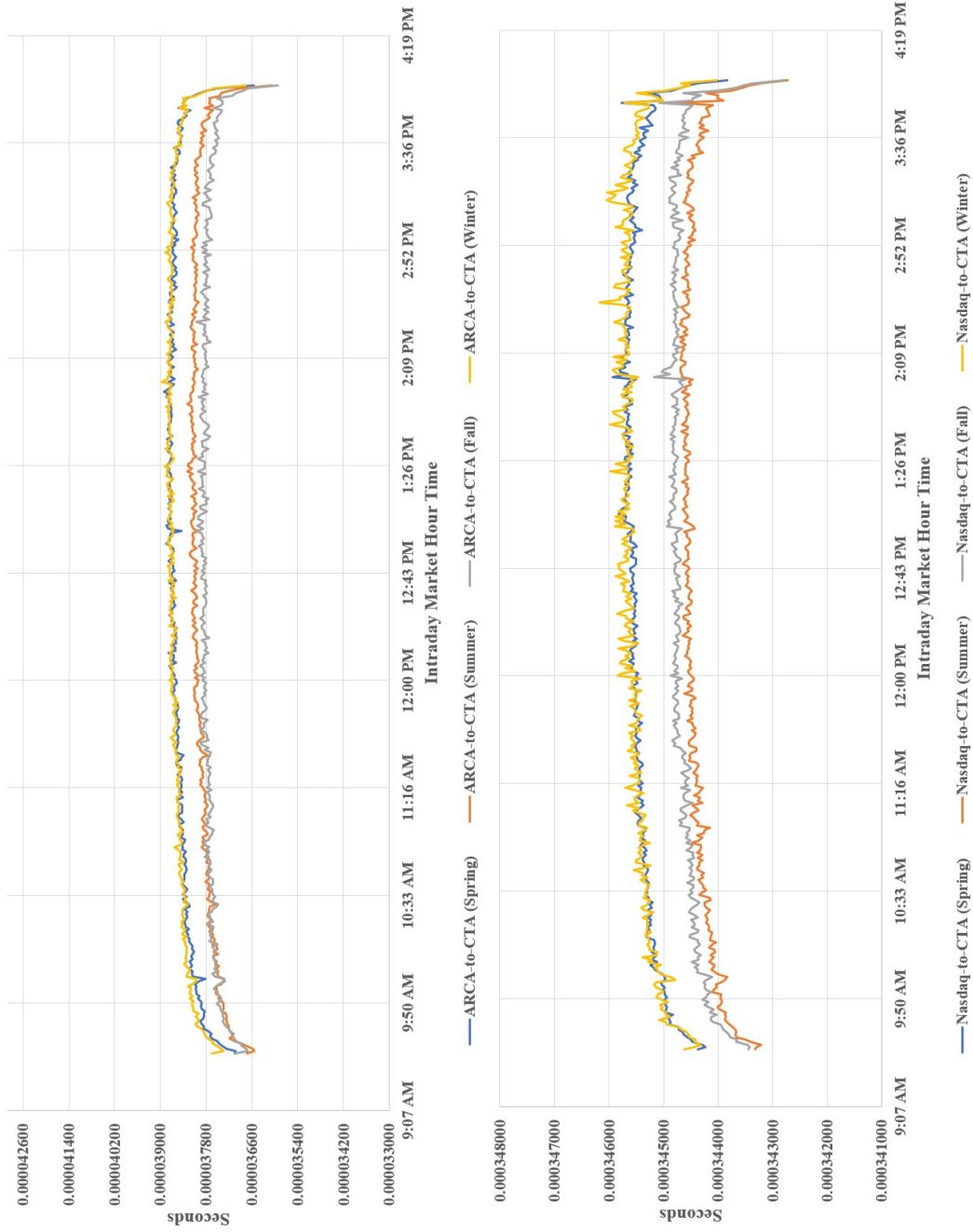
The physical transmission methods are distinct between the direct feed and the SIP feed. In TAQ, as we observe, both the CTA and the UTP network communicates with their SIPs through optical fiber cables. On the other hand, the direct feed participants utilize the wireless Metro Millimeter Wave (MMW). According to Nasdaq, “the MMW offering utilizes millimeter wave technology to deliver ultra-low latent market data to customers in the Nasdaq Data Center as well as many of the major New Jersey metro trading hubs. Utilizing innovative wireless network technology, MMW delivers critical data 40% to 50% faster than any other fiber-based network.”³⁵ The communication from Carteret exchanges to Mahwah CTA SIP is on average 539 microseconds. The same city-to-city communication, the Mahwah exchanges to the Carteret UTP SIP takes much less time at 369 microseconds. As both the UTP and the CTA SIP use the same transmission media, the smaller latency of the UTP may indicate the faster gateway latency at Carteret and the lower processing latency at the UTP SIP. Physics and optical research show that temperature may affect the transmission speed of optical cable fiber, while weather may slow down communications via the MMW towers. To see the seasonality in latency, we plot the average daily latency between NYSE ARCA and the CTA SIP, as well as the latency between NYSE ARCA and the UTP SIP by four seasons. The seasons are assigned as December to February - Winter, March to May - Spring, June to August - Summer, and September to November - Fall. We report the average latency by season in 2021 (post-Pillar). We also plot the seasonal latency in graphs.

³⁵See [Nasdaq UTP MMW specification](#).

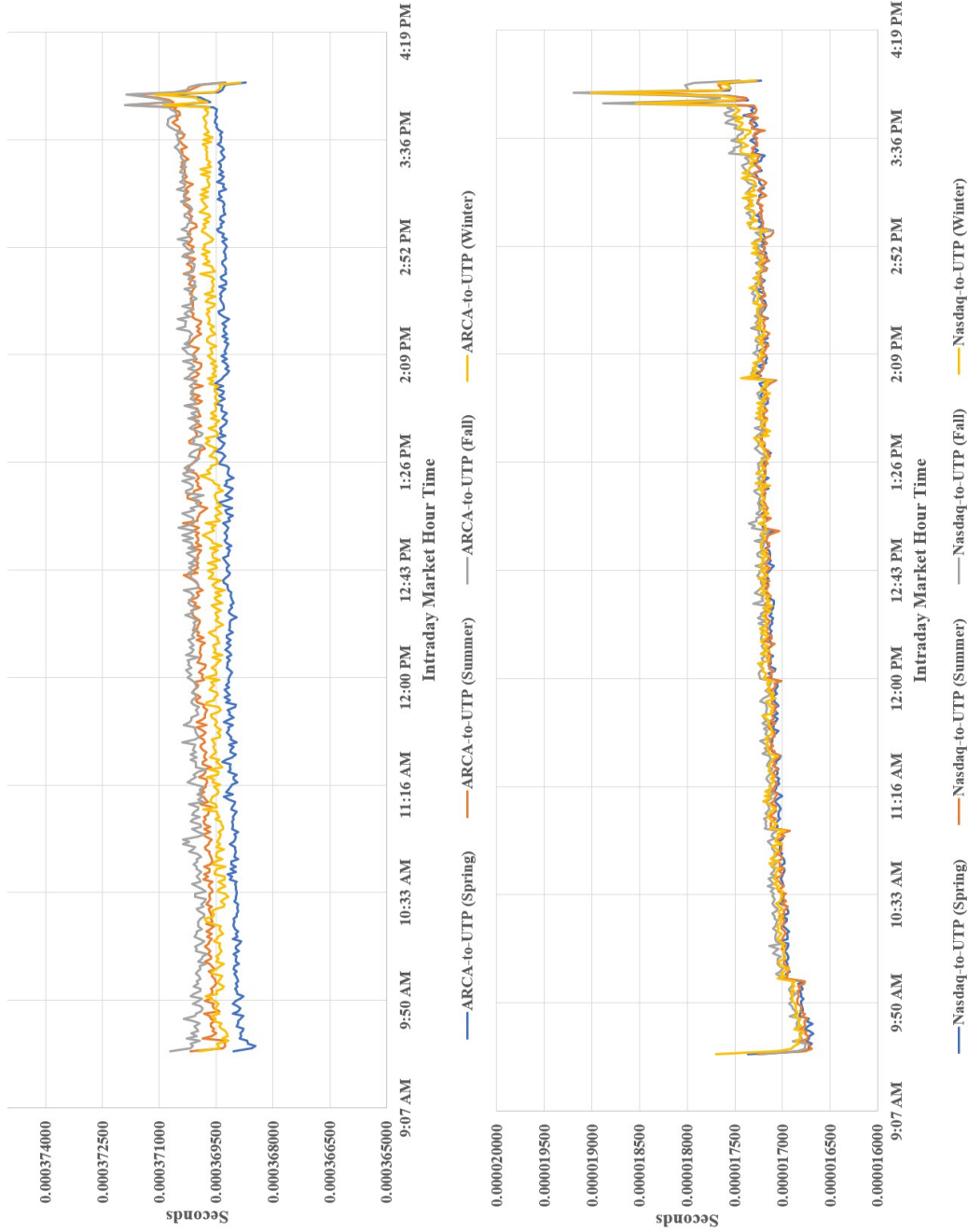
Since the NYSE ARCA matching engine is co-located with the CTA SIP, the ARCA-to-CTA SIP travel latency is minimal, remaining as the optical cable fibers that connect the servers room-to-room. The latency between Nasdaq and the CTA SIP is much greater, as the Nasdaq matching engine is in Carteret and the CTA SIP is located in Mahwah. The Nasdaq-to-CTA SIP travel latency is the main component in the latency of Nasdaq-to-CTA communications. The CTA plot in Appendix Figure A.3(a) implies latency is smaller in warmer temperatures (summer and fall) and larger in colder temperatures (winter and spring). This is in line with the findings of physics studies that examine the optical fiber cable performance in various temperature settings. Yeung and Johnston (1978), the earliest study on that topic, find that “plastic clad fused silica fibers are subject to transmission losses at lower temperatures due to changes in the optical index of the cladding polymer.” As shown in the UTP plot in Appendix Figure A.3(b), the seasonal latency variation is noticeable for the ARCA-to-UTP transmissions. For exchanges consuming the direct feeds, their communications could be vulnerable to weather conditions. “Electromagnetic wave propagation in the atmosphere is affected by the composition of the air and by meteorological conditions like fog and rain” (see Golovachev (2019)). With historical data on weather above New Jersey, researcher could examine how weather effect the latency for exchanges consuming direct feed. For example, Shkilko and Sokolov (2020) use extreme weather conditions above New Jersey as shocks to microwave networks used by distant fast traders. We also notice latency spikes at the opening and closing auctions. As a lot of trades are executed during auctions, we conjecture message volume traffic is another determinant of the exchange-to-SIP latency.

In short, we explained why we consider weather, temperature, and traffic could affect the exchange-to-SIP (SIP feed) and exchange-to-exchange communications (direct feed), we connect to the distinct physical nature of the CTA network and UTP network transmission methods.

	ARCA-to-CTA latency (in ms)		Nasdaq-to-CTA latency (in ms)		ARCA-to-UTP latency (in ms)		Nasdaq-to-UTP latency (in ms)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Spring	38.505	0.305	345.440	0.277	369.190	0.217	17.124	0.178
Summer	37.916	0.305	344.399	0.275	369.972	0.294	17.131	0.174
Fall	37.718	0.230	344.633	0.280	370.157	0.242	17.219	0.216
Winter	38.562	0.267	345.517	0.298	369.568	0.191	17.184	0.189



Appendix Figure A.3(a). ARCA-to-CTA and Nasdaq-to-CTA Latency by Seasons. The upper graph plots the average daily intraday minute-by-minute latency of ARCA-to-CTA SIP communications by season in 2021. The lower graph plots the Nasdaq-to-CTA latency.



Appendix Figure A.3(b). ARCA-to-UTP and Nasdaq-to-UTP Latency by Seasons. The upper graph plots the average daily intraday minute-by-minute latency of ARCA-to-UTP SIP communications by season in 2021. The lower graph plots the Nasdaq-to-UTP latency.

A.4 What Causes the Remaining 8% Inaccurate Classifications?

We use one day’s trades that are matched between the limit order book (LOB) order executions and the Daily TAQ trades to analyse why the trade classification accuracy stops at 92% accurate. We decompose the remaining 8% inaccurate classifications and examine the trade conditions and trade singing mechanisms for what could have gone wrong. The following is based on 1,768,622 matched LOB-TAQ trades on June 3rd, 2019 for NYSE ARCA. They are 91.19% accurately signed using Lee and Ready (1991) with latency adjustment, and 8.81% are inaccurate.

Tick Test Used? (Locked or crossed spread)	No 98.29% of all trades	Incorrect 8.10% of all trades	1.49% are trade thru exempt (TTE)	Reason 2: Legit TTE (about 31.28% of all trades are TTE)
			2.50% is \$0.01 away to trigger tick test (TT) and will be correct with TT	Reason 3: Lee and Ready (1991) has a specific threshold to trigger TT
			1.57% are hidden orders executed within 1 second	Reason 4: Hidden order executions have priority over regular orders
	Yes 1.71% of all trades	Correct 90.19% of all trades	2.54% unknown (2.20% at NBBO and incorrect; 0.34% inside NBBO and incorrect)	
			Reason 1: Tick test could be inaccurate, 1-in-4 tick movement scenarios.	Aitken and Frino (1996)
		Incorrect 0.71% of all trades		
		Correct 1% of all trades		

A.5 Two-liner Latency Adjustment in the SAS Code

Please refer to Holden and Jacobsen (2014) for their SAS code (version 2018-03-16) regarding how to generate complete NBBO and merge trades with prevailing complete NBBO quotes. We suggest adding two lines to their code to adjust for SIP-to-Exchange latency below:

```
... ..  
(line #270)  
/* STEP 6: CLEAN DAILY TRADES DATA - DELETE ABNORMAL TRADES */  
data trade2;  
set DailyTrade;  
where Tr_Corr eq '00' and price gt 0;  
LATEN=TIME_M - PART_TIME; *line 1, create a latency variable;  
TIME_M=PART_TIME - LATEN; *line 2, replace SIP time with latency adjustment;  
drop LATEN Tr_Corr Tr_Source TR_RF Part_Time RRN TRF_Time Sym_Suffix Tr_SCond Tr_StopInd;  
run;  
... ..
```

Note: this two-liner adjustment is the exact “Latency Timesamp-adjusted (LTA) Method” in Tables II, III, IV, V, VI and VIII.

A.6 Order Book Trial Data on WRDS

WRDS provides a sample of TAQ Integrated Feed. You can access [samples](#) of NYSE and NYSE ARCA order book data on wrds-cloud at: /wrds/samples/sasdata/nyse/order_book. You can use the order book data [manual](#) for variable definitions.

You may refer to the sample code in Appendix A.5 that applies the latency timestamp-adjusted method proposed in this paper.

Please kindly note that the two-day sample of Integrated Feed trials (20211004 and 20211005) are outside the sample period used in this paper, but the latency adjustment’s improvement to Lee Ready trade classification accuracy holds robust for these two days (4.36% and 4.11%, respectively).