

Article

Price Discovery and the Accuracy of Consolidated Data Feeds in the U.S. Equity Markets

Brian F. Tivnan ^{1,2,*}, David Slater ¹, James R. Thompson ¹, Tobin A. Bergen-Hill ¹,
Carl D. Burke ¹, Shaun M. Brady ³, Matthew T. K. Koehler ¹, Matthew T. McMahon ¹,
Brendan F. Tivnan ² and Jason G. Veneman ^{1,*}

¹ The MITRE Corporation, McLean, VA 22102, USA.; dslater@mitre.org (D.S.); jrthompson@mitre.org (J.R.T.); tbergenhill@mitre.org (T.A.B.-H.); cburke@mitre.org (C.D.B.); mkoehler@mitre.org (M.T.K.K.); mcmahon@mitre.org (M.T.M.)

² Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA.; brendantivnan@cvsdvt.org

³ Center for Model-Based Regulation, Davidsonville, MD 21035, USA; sbrady@cmbreg.org

* Corresponding author: btivnan@mitre.org (B.F.T.); veneman@mitre.org (J.G.V.)

Received: 30 September 2018; Accepted: 25 October 2018; Published: 28 October 2018



Abstract: Both the scientific community and the popular press have paid much attention to the speed of the Securities Information Processor—the data feed consolidating all trades and quotes across the US stock market. Rather than the speed of the Securities Information Processor (SIP), we focus here on its accuracy. Relying on Trade and Quote data, we provide various measures of SIP latency relative to high-speed data feeds between exchanges, known as direct feeds. We use first differences to highlight not only the divergence between the direct feeds and the SIP, but also the fundamental inaccuracy of the SIP. We find that as many as 60% or more of trades are reported out of sequence for stocks with high trade volume, therefore skewing simple measures, such as returns. While not yet definitive, this analysis supports our preliminary conclusion that the underlying infrastructure of the SIP is currently unable to keep pace with the trading activity in today's stock market.

Keywords: market microstructure; price discovery; latency

1. Introduction

The scientific community and the popular press have paid much attention to the speed of the Securities Information Processor (SIP)—the data feed consolidating all trade and quote messages across the US stock market—relative to other data services (Tivnan and Tivnan 2016). Here, we address the importance of the SIP.

Elsewhere, we have focused on the importance of the SIP to a key measure of market quality—efficient, price discovery (Tivnan et al. 2017). Here, we focus on extending that analysis to assess the accuracy of the SIP. We use Trade and Quote data to provide various measures of SIP latency relative to direct feeds—the high-speed, data feeds between exchanges. Using first differences, we highlight the fundamental inaccuracy of the SIP. We find that 60% or more of trades are reported out of sequence for stocks with high volume. This disordering of trades skews simple measures, such as returns. While preliminary, this analysis supports our conclusion that the underlying infrastructure of the SIP is currently unable to keep pace with the trading activity in today's U.S. equity markets.

In the sections that follow, we provide an overview of the National Market System and the SIP while adding clarity to the debates surrounding the SIP. We summarize the relevant literature of previous attempts to clarify the role of the SIP within the National Market System, as well as describe our subsequent contribution to this literature. We describe our methods and data in greater detail,

followed by a presentation of our findings. We then conclude the paper with a brief discussion of the implications of our findings.

Overview of the National Market System

The National Market System (colloquially known as the “stock market”) includes all market centers where investors can buy and sell shares of publicly traded companies. To facilitate the efficient exchange of capital and shares in the National Market System (NMS), each market center is required to publish both the Best Bid (i.e., the highest price at which an investor is willing to pay for a single share of a given stock) and the Best Offer (i.e., the lowest price at which an investor is willing to sell a single share of a given stock), as well as how many shares the investor is willing to sell at that price. Across the entirety of the NMS, the highest bid and the lowest offer comprise what is known as the National Best Bid and Offer (NBBO) and the difference between the Best Bid and Best Offer is known as the spread.

The NBBO reflects a distillation of the order flow across all the stock exchanges comprising the NMS. Figure 1 provides a graphical depiction of the three major datacenters of the NMS, all of which are located in northern New Jersey. Table A1 in the Appendix A also depicts the distribution of the families of exchanges across these datacenters. As depicted in Figure 1, the communications infrastructure connects the datacenters by dedicated, high-speed networks known as Direct Feeds depicted in red and black, as well as by the Security Information Processor depicted in pink. The Direct Feeds provide high-speed communications channels (e.g., faster than 50% of the speed of light (Angel 2014)) where all orders flow between the exchanges. The Securities Information Processor, or SIP, consolidates and distills all the order flow from the Direct Feeds to determine and disseminate the NBBO, as well as report all completed trades. We developed and posted an animation to provide a more illustrative depiction of these interactive dynamics between the direct feeds and the SIP¹.

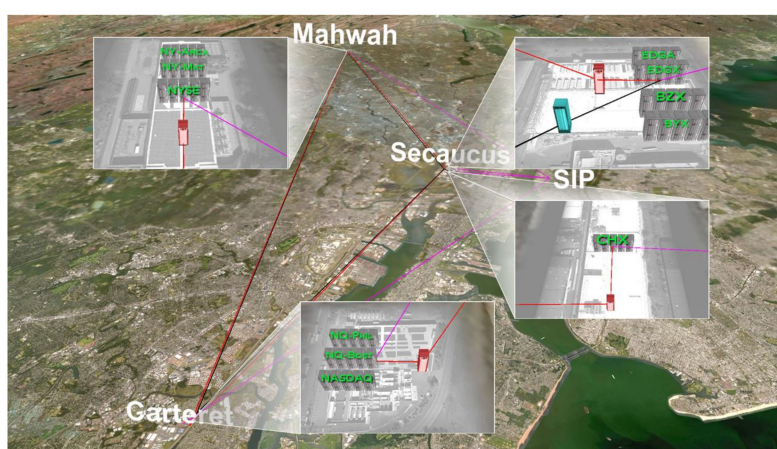


Figure 1. Graphical depiction of the three major market centers in the National Market System (NMS). SIP, Securities Information Processor.

We conclude this overview of the National Market System with a description of the two Trade Reporting Facilities; namely, the NYSE Trade Reporting Facility (NTRF) and the NASDAQ Trade Reporting Facility (QTRF). Both Trade Reporting Facilities consolidate the trade reports from all Alternative Trading Systems (colloquially known as “Dark Pools”). The NTRF consolidates Dark Pool trade reports of NYSE-listed stocks whereas the QTRF consolidates Dark Pool trade reports of NASDAQ-listed stocks.

¹ The MITRE Corporation. “An Example of High Frequency Trader (HFT) Latency Arbitrage.” <https://www.youtube.com/watch?v=1l1jnbBaFok&feature=youtu.be>. Date of information retrieval: 30 September 2018.

Unlike an exchange, which must display best, local quotes, a Dark Pool is a market center where trading occurs without displaying any local quotes. Prevailing regulations still require that price discovery at Dark Pools are driven by the NBBO—same as at each exchange. While not depicted in Figure 1, we include Dark Pools here in our analysis for completeness, since trade reports from Dark Pools contribute to the overall, message traffic consolidated and disseminated by the SIP.

2. Review of Relevant Literature

There is widespread agreement that Mandelbrot (1963) was one of the first to characterize the dynamic behavior of prices in modern markets; later to be extended by Cont (2001). A related study (Johnson et al. 2012, 2013) identified some anomalous behavior in the dynamics of market prices, which deviated from previous expectations (Mandelbrot 1963; Cont 2001). At the time, however, the resolution of our data prevented us from making any definitive determinations about the causal mechanisms of these behaviors.

Initially in the press and later in his book, Scott Patterson was one of the first to draw attention to the speed discrepancy between the SIP and the direct feeds (Patterson 2012a, 2012b), only later to be followed in 2014 by Michael Lewis' (Lewis 2014) popular book, *Flash Boys*. Neither of these authors fully explored the implications of the speed discrepancy, instead relying heavily upon the works of financial industry practitioners.

A few industry practitioners were the first to identify the speed discrepancy between that of the SIP and the direct feeds. Arnuk and Saluzzi (2012) were one of the first to highlight these latency discrepancies. Bodek (2013) detailed the sweeping changes instituted by RegNMS, changes which stimulated the technological arms race and the insatiable demand for greater speeds from the direct feeds. Furthermore, Nanex Research (2014a) highlighted the spoils, which can be extracted in specific instances by those market participants so fortunate as to subscribe to the direct feeds.

Scientific contributions to this debate have been preliminary, possibly due to the limited availability of authoritative data within the scientific community. While authoritative datasets exist, subscriptions to them are extraordinarily expensive and therefore largely available only to specialized, market participants along with regulatory agencies.

Ding et al. (2014) provide one of the first scientific analyses of direct feed and SIP latencies. More recently, Bartlett and McCrary (2017) attempted to revise and generalize earlier findings (Nanex Research 2014a) by quantifying the total profits from subscribing to the direct feeds.

One common theme prevails throughout these analyses: The implication that the SIP accurately depicts market dynamics, albeit subject to some latency. We address this limitation of previous studies by highlighting and quantifying the inaccuracies of the SIP.

3. Methods

The implications of ever increasing market speeds can be a difficult topic to understand even for those immersed in market data daily. The following is an attempt to break down some empirical analyses on granular data from US stock markets that illustrate some of the implications from SIP latencies and the subsequent impact on the accuracy of the SIP. We rely on Trade and Quote data spanning the National Market System. More precisely, we purchased the NxCore dataset from Nanex Research for the following periods: The calendar year 2014, and the second and third quarters of 2015.

Many of the subsequent examples stem from market activity on a representative day (i.e., 11 August 2015), which we chose for three reasons. First, 11 August 2015 was a not a noteworthy day in the U.S. equity markets (i.e., the absence of any market-wide events). Second, this representative day closely precedes a notable day in the U.S. equity markets (i.e., 24 August 2015), which is known colloquially as "Manic Monday." Third, there were no announced modifications to the SIP between 11 August and 24 August. Given these three reasons above, it is realistic to assume the same underlying, market infrastructure for both days. Therefore, 11 August 2015 serves as a useful proxy for a representative day in the U.S. equity markets.

We capture both breadth and depth in our analyses. We achieve breadth in two ways. First, we analyzed the entire population of the 7993 unique tickers (i.e., stocks and exchange-traded funds) that printed quotes and trade reports to the SIP on the representative day. The reader will note that an exchange-traded fund (ETF) is a security similar to a mutual fund, but an ETF trades throughout the day in the same manner as a stock. Second, we also closely analyzed SPY (i.e., the SPDR S&P 500 ETF). It is important to note that the SPY is often used as a proxy of the U.S. equity market. Unlike the S P 500 index, which cannot be directly traded, the SPY represents one of the largest and most liquid, market-wide securities. For the SPY, we analyzed price, spread and market crosses. A special case of the spread occurs when the spread becomes negative. Such instances of a negative spread for the NBBO depict a crossed market, or simply *cross* for short. While a small spread indicates general agreement on price across all market participants, a large spread and crosses indicate little agreement on price. Both large spreads and crosses are indicators of inefficient, price discovery.

We also achieve depth of our analysis in a comprehensive manner. Rather than limiting our analyses to any one stock or relying solely on the large-cap stocks that comprise the Dow 30 as in previous studies (Bartlett and McCrary 2017), we chose a convenience sample of 14 stocks. This heterogeneous set of stocks varies on market capitalization, listed exchange and trading volume. This sample ranges from Apple (i.e., stock ticker: AAPL), one of the largest and most actively traded stocks in the world to Acme United Corporation (i.e., stock ticker: ACU), which is a lightly traded, small-cap stock.

Our analyses include some depictions of common descriptive measures of market performance, namely price and trade volume, as well as the monetary value of that trading activity. We then de-construct this market activity by specific exchange and components of the SIP.

We continue our analyses with a focus on the latencies between the direct feeds and the SIP, beginning with simple counts and frequency distributions. We extend this analysis by comparing first differences to highlight the divergence between the direct feeds and the SIP. We conclude with an analysis of these measures for comparison with known market events.

4. Results

As described above, we begin with empirical evidence confirming our focal date as a representative day in the U.S. equity markets. From there, we provide depictions of common descriptive measures of market performance. We then de-construct this market activity by specific exchange and components of the SIP. We then present our analysis of first differences to highlight the divergence between the direct feeds and the SIP. We conclude with an analysis of these measures for comparison with known, market events.

4.1. Representative Day in the U.S. Equity Markets

For simplicity and clarity in the presentation of our analyses, we chose a representative day in the U.S. equity markets (i.e., 11 August 2015) for its distinction from a noteworthy day (i.e., 24 August 2015) within only a two-week span. Figures 2 and 3 provide the empirical evidence to support this distinction. Both figures plot the price, spread and market crosses for SPY.

Figure 2 depicts the market dynamics for SPY on 24 August 2015. Leading up to and through the opening of the market day, the widening spread and frequent crosses indicate a tremendous uncertainty of the price. Even as spreads narrowed and price stabilized throughout the market day, frequent crosses persisted. These are all indicators of inefficient, price discovery.

Figure 3 depicts the market dynamics for SPY on 11 August 2015. Unlike the dynamics of “Manic Monday,” Figure 3 reflects a “typical market day.” More precisely, spreads remain small throughout the market day and crosses are intermittent yet fleeting while price is largely stable. Moreover, Figure 3 provides no evidence that would lead one to conclude that the underlying, market infrastructure is overwhelmed in any meaningful way on this day.

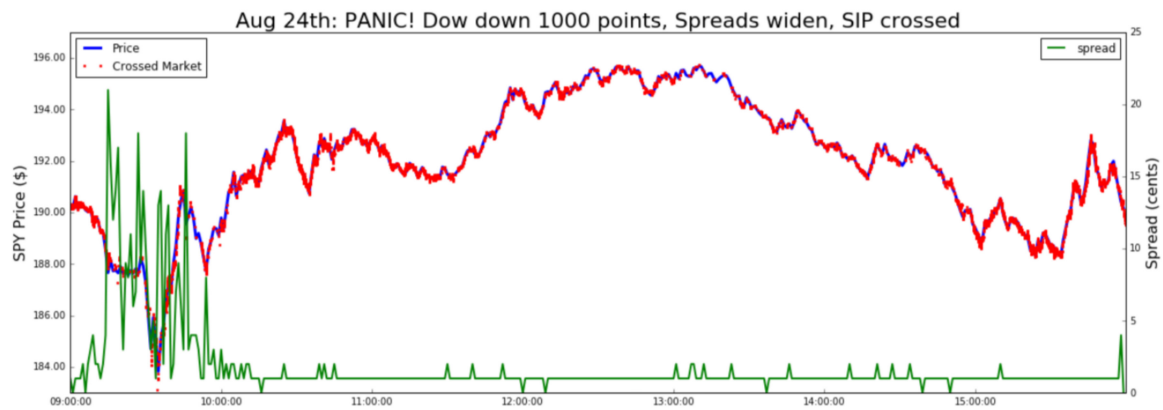


Figure 2. Crosses and Spreads in the SPY on 24 August 2015—“Manic Monday.”



Figure 3. Crosses and Spreads in the SPY on 11 August 2015—“typical market day.”

4.2. Common, Descriptive Measures of Market Activity

We naturally follow with a time-series depicting the price of a share of Apple over the course of a day. Figure 4 shows the price of Apple, including both pre-market (04:00 to 09:30 hours or 0 to 19,800 seconds in the figures below) and after-hours trading (16:00 to 20:00 hours or 43,200 to 57,600 seconds in the figures below). Figure 4 is typical of the only chart most people see when looking at the stock market. With data at the microsecond level, even in this typical view, one starts to see some anomalous behavior with the spikes at around 44,000 seconds. Those spikes are after-hours trades which, along with pre-market, are not usually shown on popular websites (e.g., Yahoo Finance). This begins to illustrate the complexity dynamics of price behaviors across an entire day, not just during the middle of the day.

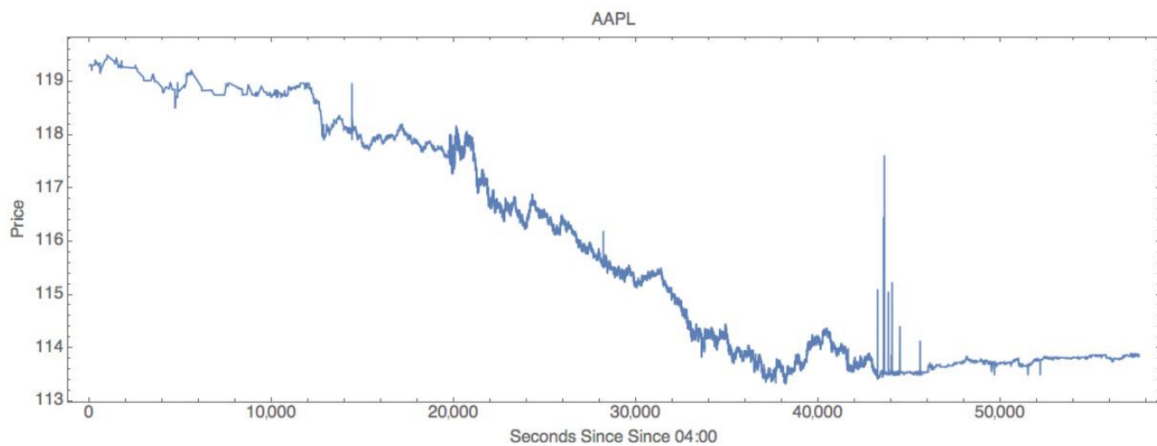


Figure 4. Apple stock price on 11 August 2015 at each second of the day from 04:00 until 20:00.

In Figure 5, we illuminate the interesting dynamics associated with the number of trades per second for a single stock. Here, one will notice the clear difference in activity during pre-market, regular market (09:30–16:00), and after-hours trading (16:00–20:00).

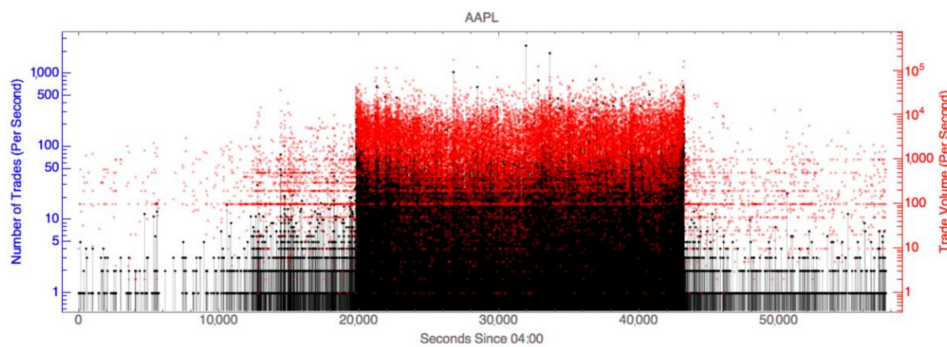


Figure 5. Trades in Apple stock on 11 August 2015 at each second of the day from 04:00 until 20:00.

To get a sense of how much capital is circulating, a look into the number of dollars traded per second in Figure 6 reveals some astounding numbers. Here, one sees that at the highest spike \$40 million in Apple stock is traded in one second.

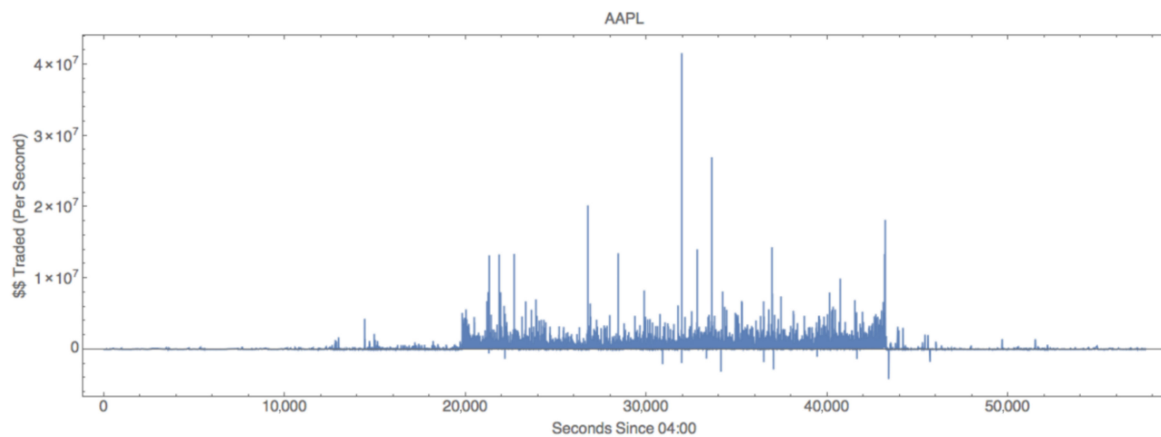


Figure 6. Dollars traded in Apple on 11 August 2015 at each second of the day from 04:00 until 20:00.

To put things in perspective the cumulative traded capital in Figure 7 shows a steady climb to approximately \$10 billion traded in a single day for Apple stock. At such a high rate, the \$40 million spikes, as seen in Figure 6, hardly register as outliers.

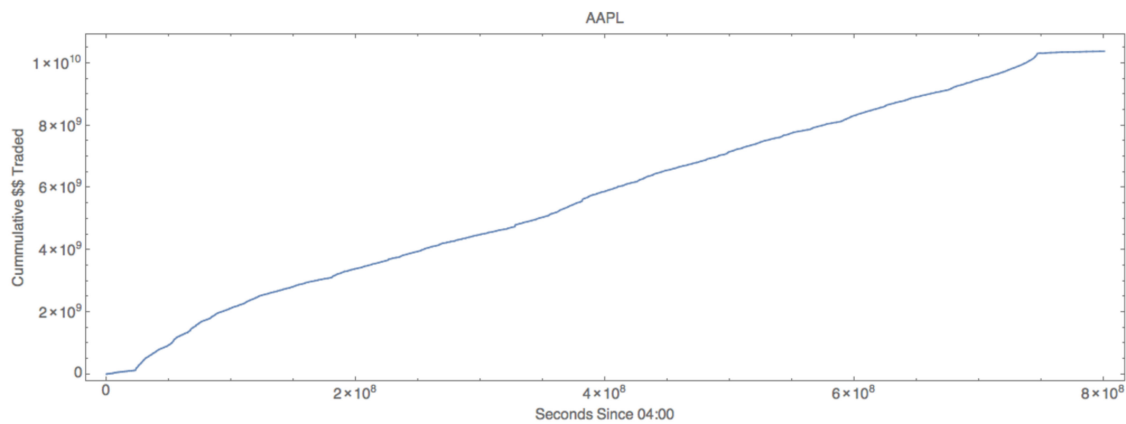


Figure 7. Cumulative dollars traded in Apple stock on 11 August 2015 summed each second from 04:00 until 20:00.

4.3. De-construct Market Activity by Exchange and Components of the SIP

Figure 8 shows how assets are traded at a high volume at the multiple exchanges that comprise the National Market System. The opening and closing hours differ per exchange and are the reason that some lines start late or stop short. The start of regular trading at 09:30 (19,800 seconds since 04:00) is seen in Figure 8 as the sharp increase in volume at all markets at opening time.

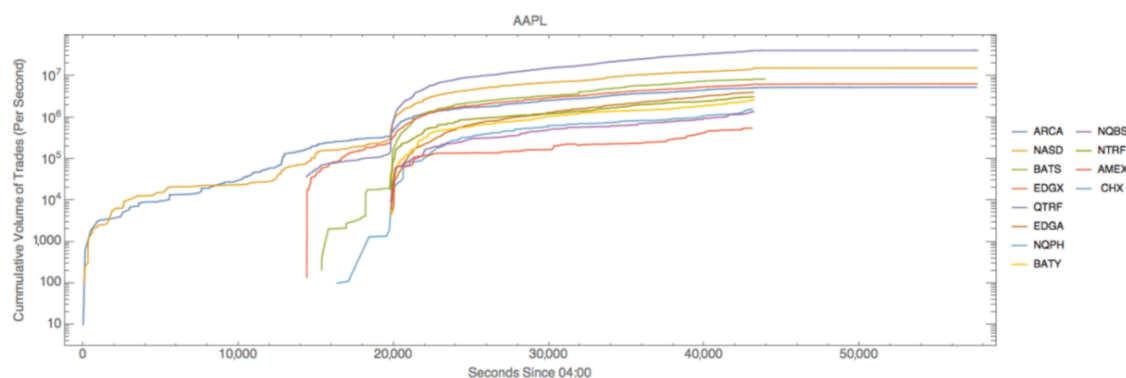


Figure 8. Cumulative trade volume by exchange for Apple stock on 11 August 2015 summed each second from 04:00 until 20:00.

Here, we begin to tie the above financial measures with the concept of bandwidth capacity of the underlying communications networks, it may be more intuitive to think of trade volume and quote lots (a quote lot is typically 100 shares) in terms of messages. A message is defined as an atomic unit of communication that describes a number of shares or lots for a particular asset (e.g., 100 shares of Apple stock). Since there are far more quote messages than there are trade messages (i.e., roughly 10 quote messages for each trade report), the following figures of quote messages per day represent the upper bound of traffic on communication networks.

For simplicity, we have referred to the SIP as a single entity to this point in our analysis. In fact, the SIP is comprised of three, distinct components: SIP A, SIP B and SIP C. The SIPs link “the U.S. markets by processing and consolidating all protected bid/ask quotes and trades from every trading venue into a single, easily consumed data feed (Consolidated Tape Association).” SIP A and B are operated by the Consolidated Tape Association and consolidate market activity for securities listed on New York Stock Exchange (NYSE) on SIP A and securities listed on NYSE ARCA, NYSE MKT, BATS and regional exchanges on SIP B (Consolidated Tape Association). SIP C is operated by NASDAQ for NASDAQ-listed securities (Unlisted Trading Privileges).

In Figure 9, one can again clearly see the opening and closing of the regular market day. Figure 9 shows messages that are recorded at each of the three SIPs. During the market day, there are very few instances that less than 1000 messages per second are printed to the SIPs, while there are times, such as the market open and close, when message traffic exceeds 100,000 messages per second.

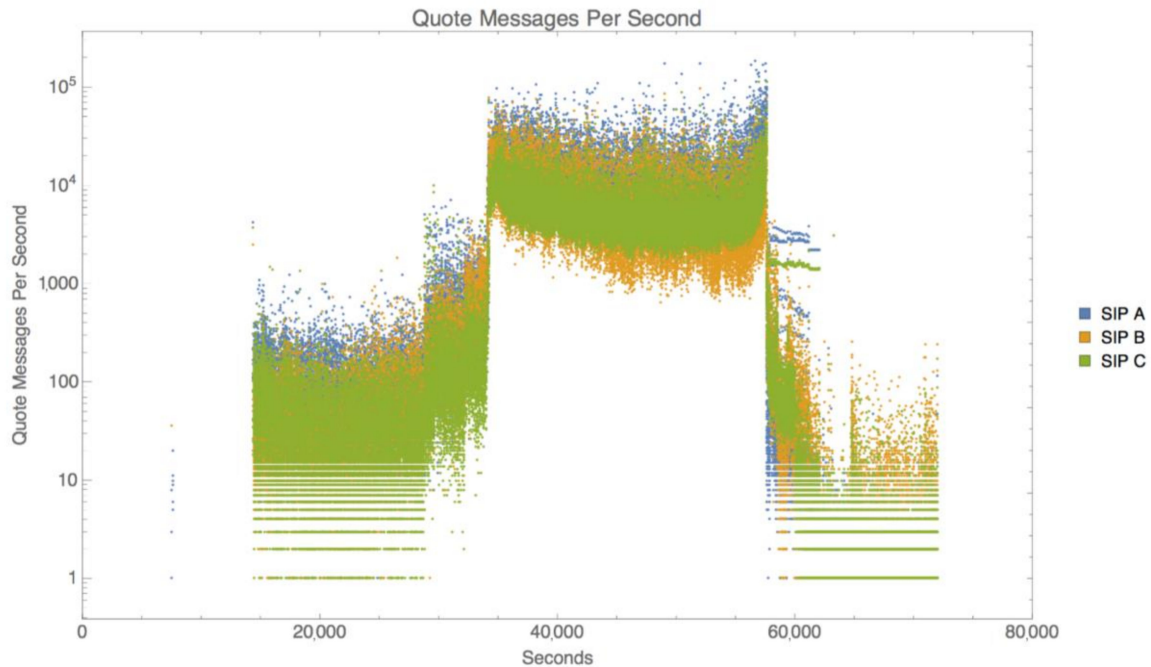


Figure 9. Quote messages by SIP for Apple stock on 11 August 2015 for each second from 04:00 until 20:00.

4.4. Latencies between the Direct Feeds and the SIP

Recall that the SIP exists to unify the market in time via efficient discovery of equity prices. To get a picture of how long it takes for trades to go from being reported at an exchange to being aggregated and recorded by the SIP for all market participants to see, one can look at the SIP latency. Latency is defined here by the timestamp from the SIP minus the timestamp from when the reporting exchange sent the trade or quote information to the SIP. The latency in microseconds for trades at each of the three SIPs for all stocks is shown in Figures 10–12.

From these three figures, one can see that the median, standard deviation and range in latency are similar across each SIP. The median latencies across all SIPs and all exchanges are within 700 μ s of each other. From this, one might infer that trade reporting latencies are similar.

While similarities in latency median, standard deviation and range certainly do exist, there are at least two exceptions to these similarities. The first exception is the Chicago Stock Exchange (CHX). It was first reported elsewhere that CHX latencies are outliers from that of the other exchanges, as well as that CHX has limited share of the trading volume despite often contributing to efficient, price discovery (Nanex Research 2014b). Our findings here confirm CHX latencies as outliers. The second exception relates to the NTRF where Dark Pool trades are consolidated for NYSE-listed stocks.

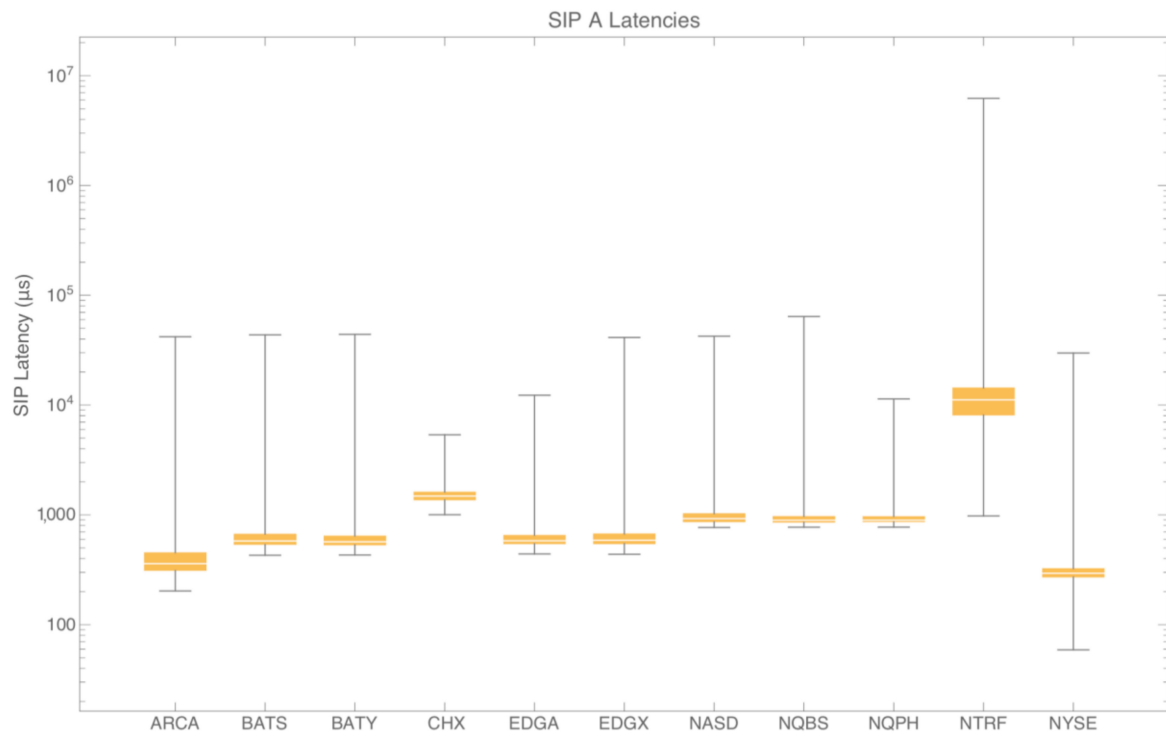


Figure 10. Box plots of latencies for all trades of all stocks traded on 11 August 2015 in microseconds reported by SIP A and separated by exchange.

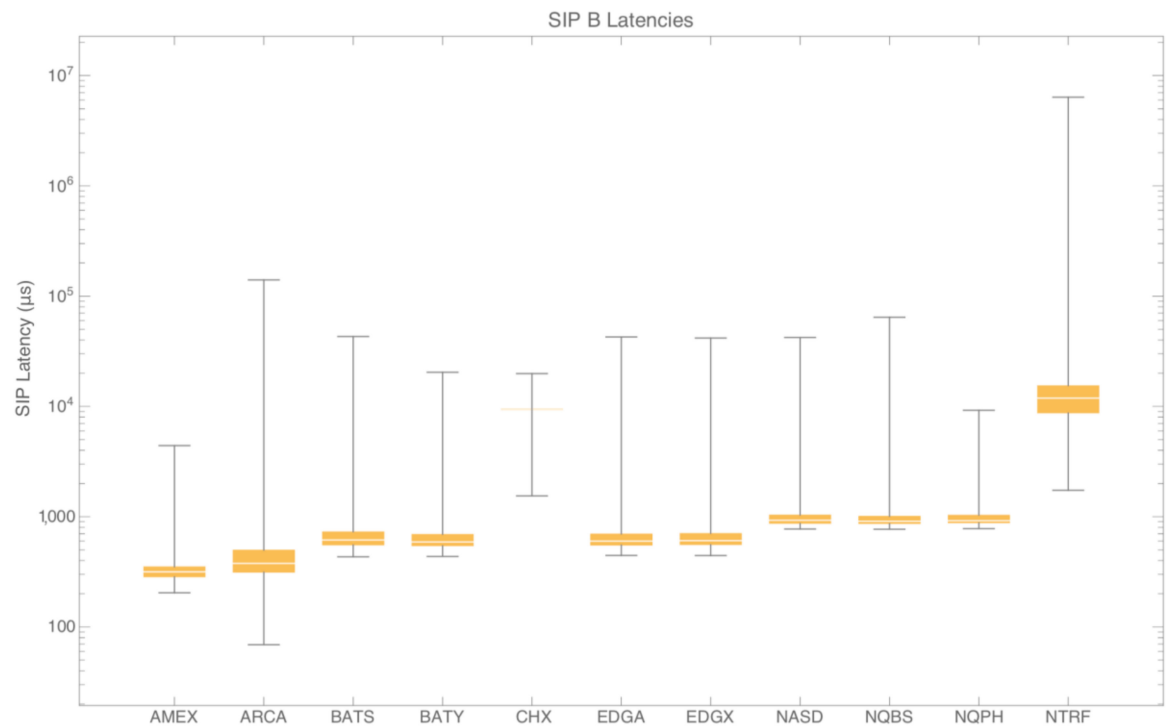


Figure 11. Box plots of latencies for all trades of all stocks traded on 11 August 2015 in microseconds reported by SIP B and separated by exchange.

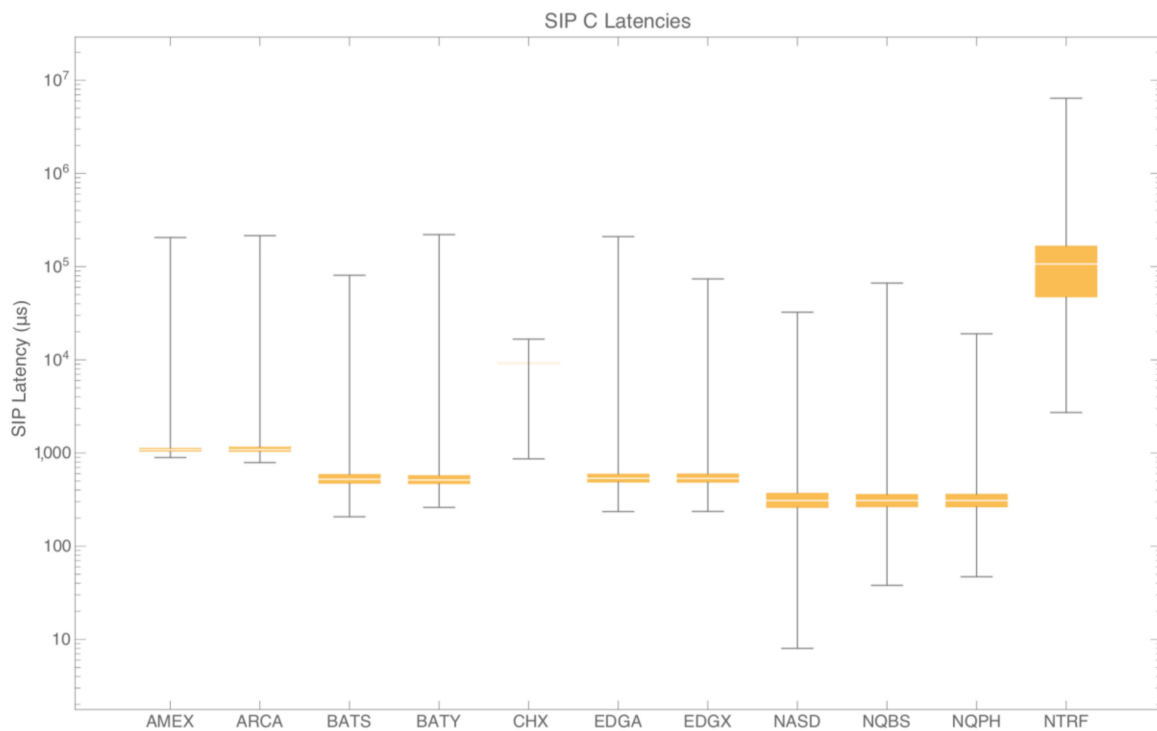


Figure 12. Box plots of latencies for all trades of all stocks traded on 11 August 2015 in microseconds reported by SIP C and separated by exchange.

Looking at latency in more detail, Figure 13 shows the latency in trade reporting for a single stock (i.e., AAPL). Figure 13 presents a far different picture from the SIP latencies and one can see that the latency differs significantly within and between exchanges. The outlier reporting exchange here is QTRF. This clearly shows that the exchanges matter when observing latency.

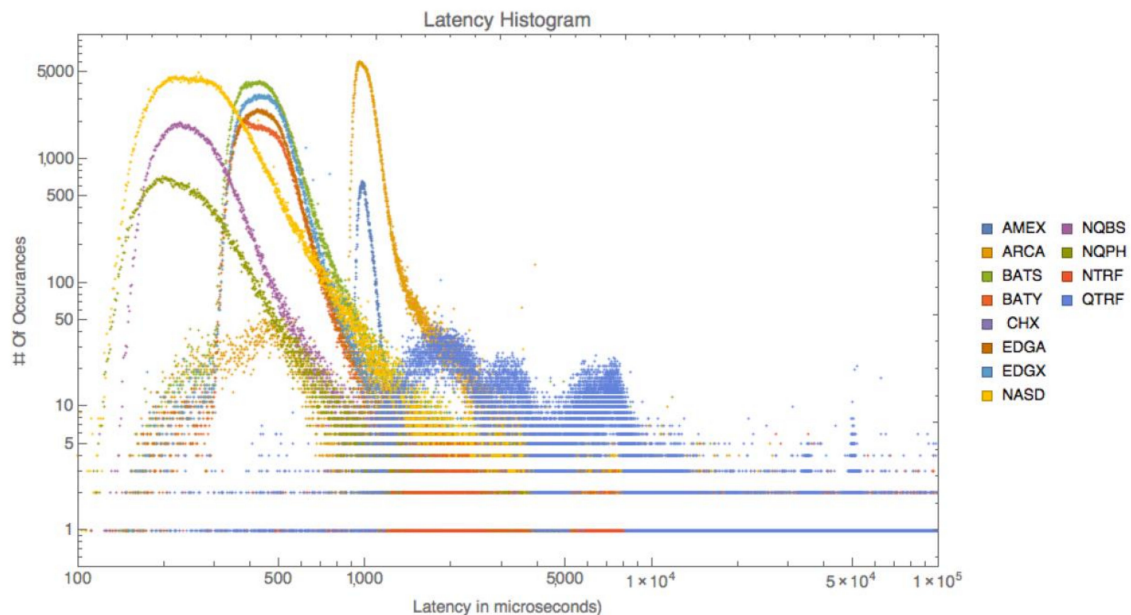


Figure 13. Latency histogram for Apple stock traded on 11 August 2015 by exchange for AAPL (truncated at 10^5 microseconds).

4.5. First Differences Highlight Inaccuracy of the SIP

Figure 14 shows the difference between the best bid and best ask, (i.e., the spread). We generated Figure 14 from the timestamps of the bid and ask quotes as printed to the SIP. This analysis shows that there are a non-trivial number of instances where the spread is negative. Indicative of inefficient price discovery, a negative spread means that one is willing to pay more than a counterparty is willing to sell at that specific time (e.g., buying a soda for \$1.05 when the sticker price is only \$0.99). This is neither desirable for the buyer nor is it allowed by current market rules (Securities and Exchange Commission), so the SIP might be disseminating inaccurate reports. To assess the accuracy of the SIP, one can look at the spread at each exchange in Figure 15 and see that at any exchange the spread never goes negative. However, instances of a negative spread do occur, particularly a negative spread for the NBBO. Such instances of a negative spread for the NBBO depict a *crossed market*, or simply *cross* for short. For completeness, a spread of zero, known as a *locked market*, depicts another market condition precluded by current market rules. All available orders at the best bid should have been executed against all available orders at the best offer.

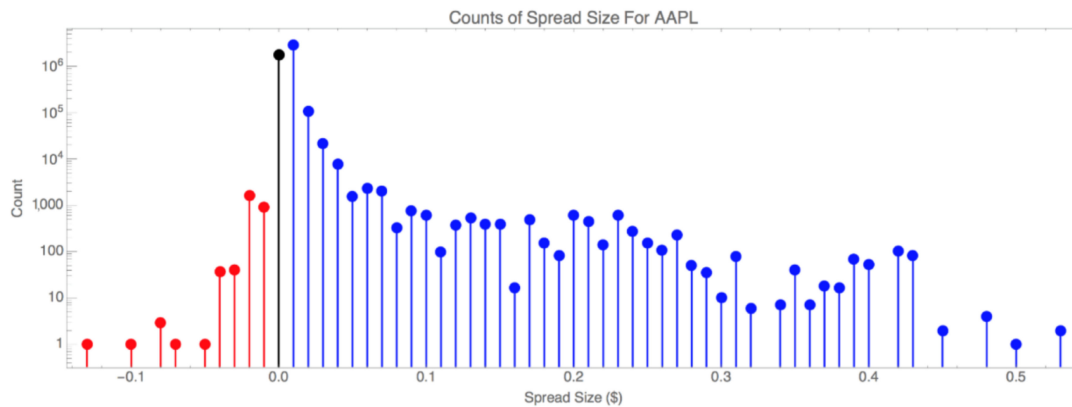


Figure 14. Bid/ask spread in dollars vs. number of occurrences for Apple stock on 11 August 2015.

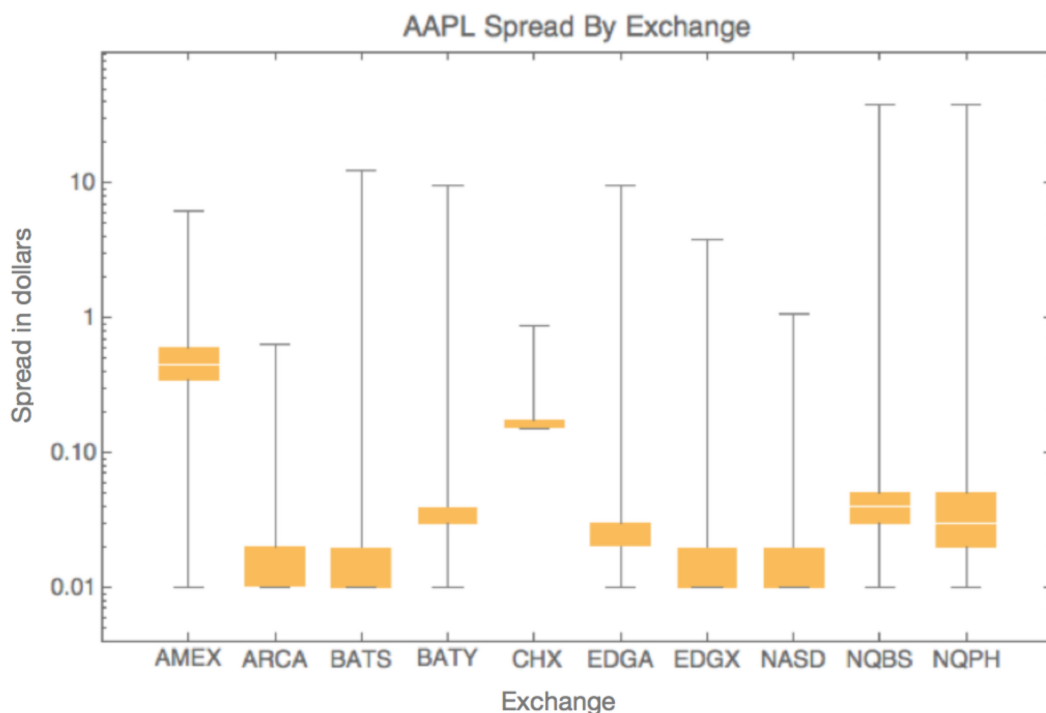


Figure 15. Box plots of spread in dollars for Apple on 11 August 2015 by exchange.

If the SIP appears wrong and the exchanges are never crossed, then this discrepancy may be due to the time it takes for messages to travel from an exchange to the SIP along communications networks. The presence of timing related anomalies indicates the importance of understanding the coupling between the financial markets and communications infrastructure and the implications of how these systems are designed.

These pricing anomalies are present across many stocks (e.g., the component stocks of the Dow 30), but for brevity and space considerations, we limit our treatment here to a convenience sample of 14 stocks listed in Table 1. Some patterns are indicated that give preliminary evidence of a relationship between message traffic and apparent crosses and locks (equal bid and asks which are also disallowed by market rules). If this relationship holds it gives further weight to the importance between finance and communication infrastructure coupling. Figure 16 indicates that as the number of quote messages increases, there is an increasing number of apparent market crossings as seen at the SIP. In Figure 16, the price of the stock is given in the color chart to the right. Stocks that trade for less than one dollar (i.e., penny stocks, blue to blue-green in the figure) operate under different market rules and appear to follow a different relationship between number of messages and crosses. This preliminary result could indicate that market rules also shape the coupling between these infrastructures.

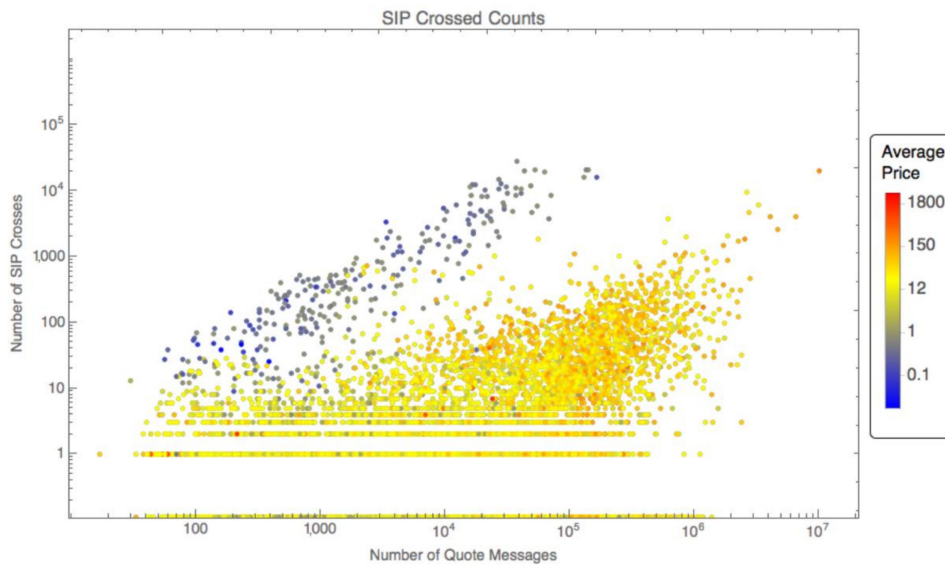


Figure 16. Counts of SIP crosses for all stocks traded on 11 August 2015.

Figure 17 shows the number of times that a stock is apparently locked (i.e., a spread of 0) at the SIP. Figure 17 provides additional evidence that increasing message traffic correlates to more locked stocks.

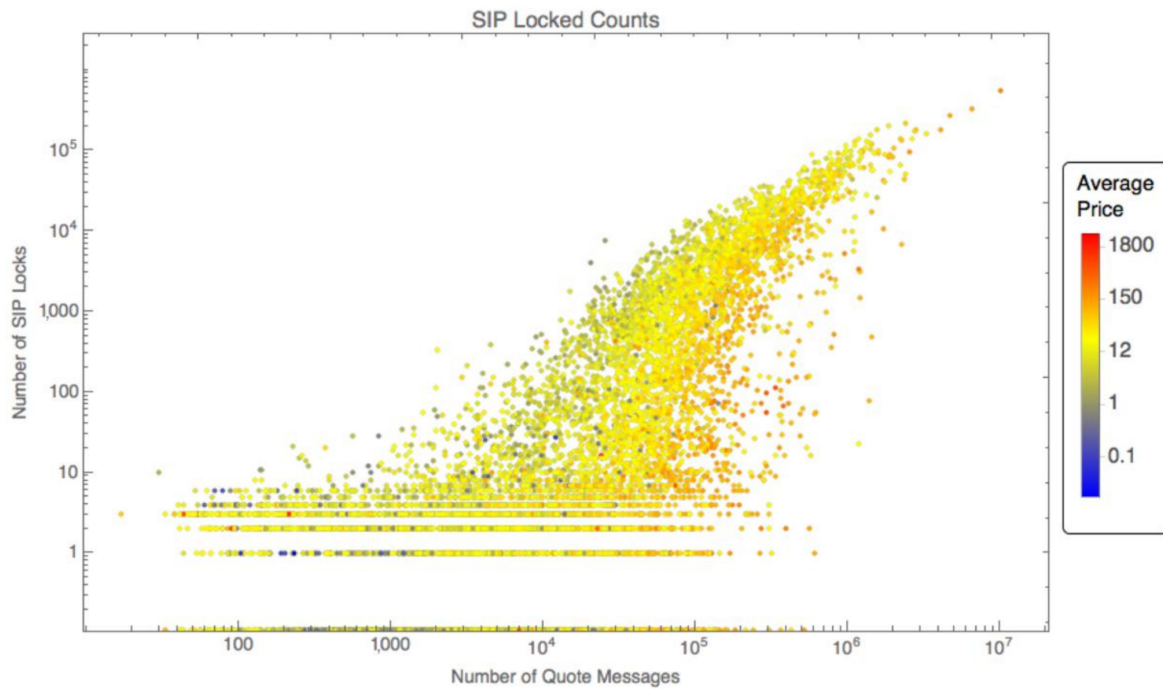


Figure 17. Counts of SIP locks for all stocks traded on 11 August 2015.

The modern financial system relies on efficient and stable communication networks for dissemination of up-to-date information. To get a picture of what can happen in a short time window, one can look at the number of events (trade reports or quote messages) for a stock that occur between the time a trade or quote update is sent to the SIP and the time the SIP reports that update to SIP subscribers—here, we refer to this delay as the latency window. Figure 18 shows that it’s very common to see from hundreds to thousands of events occur within this latency window for Apple stock.

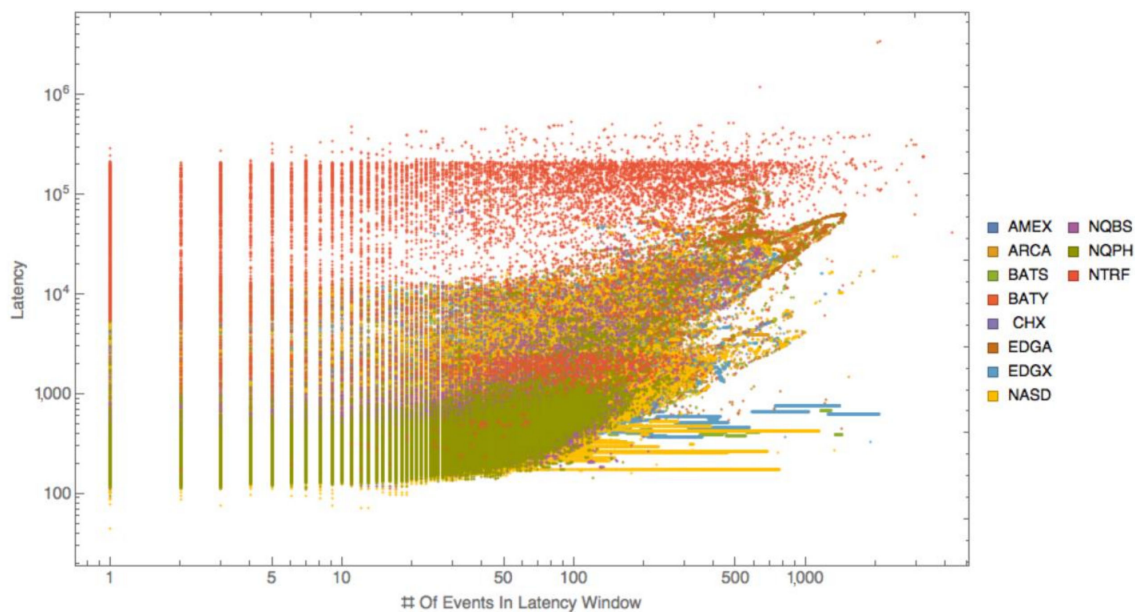


Figure 18. Number of events in latency window for Apple stock on 11 August 2015.

4.6. Crossings Between Order Books Present Arbitrage Opportunities

The appearance of a relationship between the number of apparent market locks and crosses and message (trade report or quote update) volume prompts a deeper look into the number of events

within latency windows. Figures 19 and 20 are aligned vertically on the page to reflect time of day along the x axis. In Figure 19, we depict a lightly traded stock (i.e., Second Sight Medical Products, EYES) from our convenience sample to compare with a heavily traded stock (i.e., Apple, AAPL) in Figure 20. In Figure 19, the reader will notice that EYES has limited trading activity before and after the trading day, whereas Figure 20 depicts significant trading before and after hours in AAPL.

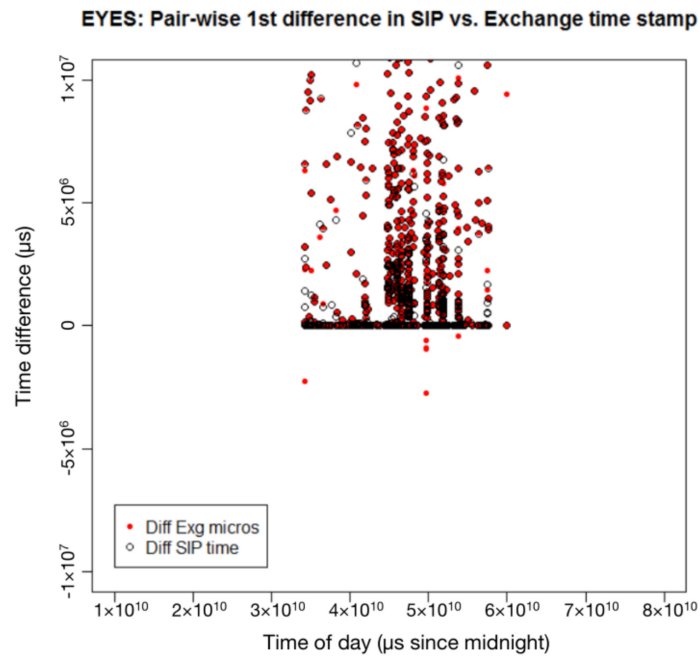


Figure 19. Time difference in SIP vs. exchange time stamp for EYES stock on 11 August 2015.

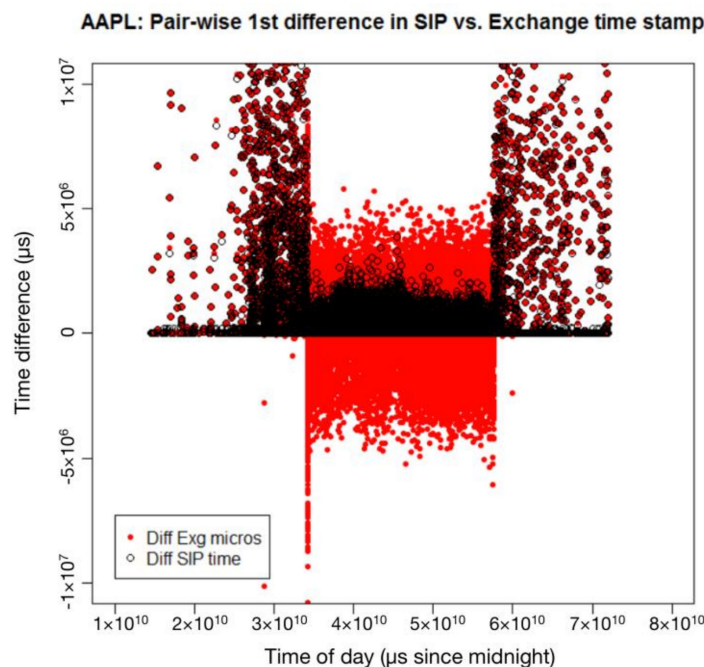


Figure 20. Time difference in SIP vs. exchange time stamp for AAPL stock on 11 August 2015.

In Figure 19, we can see that there are times where the SIP and exchange time do not line up (places where the red and white circles don't line up). For a highly traded stock like AAPL, we see from Figure 20 that there are far more occasions where the difference in time stamps is out of sequence.

A closer analysis of Figures 19 and 20 yields additional insights. Because we sorted by the SIP timestamp using the Precision Time Protocol (Precision Time Protocol), all first differences should be positive. Therefore, all negative differences in the SIP timestamps, as depicted in Figures 19 and 20, indicate that the SIP may often misrepresent the accurate sequence of trades.

We see similar features in a number of stocks across a range of trading volume in Table 1, which shows the total number of trades for a particular asset, and the percentage of those that were reported out of sequence by the SIP. A strong linear trend (slope = 0.66, R-squared = 0.99) exists between the number of trades and the number out of sequence. While additional analyses would be necessary to pinpoint drivers of that trend, such in-depth analyses could prove worthwhile as they might illuminate arbitrage opportunities. These arbitrage opportunities could arise from a crossed market. Recall that a crossed market indicates a negative spread (i.e., a displayed, best bid exceeds a displayed, best offer). If permitted, trading in a crossed market might present opportunities for true arbitrage. As an illustrative example, trading in a crossed market might enable Trader C to purchase an even lot of shares of Apple at \$116.00 per share from Trader A and immediately sell the same even lot of Apple shares for \$116.02 per share to Trader B. The reader will note that in the absence of Trader C in the market, Traders A and B might be natural counterparties at \$116.01 per share of Apple, each experiencing price improvement, which is why trading is halted in crossed markets. To exploit this potential arbitrage opportunity, Trader C must anticipate the crossed market and exploit a latency advantage via a direct-feed connection to the market. While market participants may have the ability to act on these arbitrage opportunities, they may exist for only a very small window of time (e.g., 500 microseconds or less) (Tivnan et al. 2017). These fleeting opportunities for latency arbitrage clearly indicate the ramifications of the tight interdependence between financial markets and the communication networks that connect them.

Table 1. Trades out of sequence for selected stocks on 11 August 2015.

Stock	Percent Out of Sequence	Total Trades	Listing Exchange
AAPL	66.2%	482,578	NASDAQ
BAC	50.1%	97,303	NYSE
XOM	43.0%	86,834	NYSE
GOOG	56.3%	69,085	NASDAQ
DNR	43.3%	51,185	NYSE
IBM	26.4%	29,960	NYSE
SHAK	27.9%	16,349	NYSE
KLIC	30.3%	3304	NASDAQ
GBX	20.6%	2804	NYSE
WBMD	29.9%	2428	NASDAQ
EYES	10.3%	1653	NASDAQ
BRKA	0.3%	304	NYSE
OHGI	3.8%	286	NASDAQ
ACU	0.0%	1	NYSEMKT

5. Conclusions

Above, we uncover clear limitations to the accuracy of the SIP, namely its inability to preserve the correct ordering of trades. Rather than solely relying on a largely homogenous set of stocks (e.g., the components of the Dow 30), we instead opted for a heterogenous set of stocks that varied on market capitalization, listed exchange and trading volume. This heterogeneity yielded convincing evidence that the inaccuracy of the SIP relates to trade volume. Therefore, one can reasonably infer that the physical infrastructure underlying the SIP has finite limitations, which are routinely exceeded for certain stocks.

While the impacts from such SIP inaccuracy could be extensive, we conclude by highlighting two significant implications. First, when the SIP fails to capture the actual order of trades, such inaccuracies will skew stock returns. While a simple measure, returns play a fundamental role in nearly all measures

of risk; therefore, the inaccuracies of the SIP uncovered here could reveal extensive limitations for risk management. Finally, given the preponderance of algorithmic trading in today’s U.S. equity markets, disordered trades and subsequently skewed returns could result in positive feedback driving markets away from efficient, price discovery.

While convincing, our findings on the accuracy of the SIP are not exhaustive, so we therefore encourage researchers to consider additional dimensions of analysis for future research. A longitudinal study might address the accuracy of the SIP over time. Another study might compare the accuracy of the SIP for component stocks of prevailing market indices (e.g., Dow 30, S P 500 and Russell 3000). Lastly, we suggest that a direct extension of this study of the SIP accuracy of trades includes an analysis of updated quotes.

Author Contributions: Conceptualization, B.F.T. (Brian F. Tivnan), S.M.B., M.T.K.K., M.T.M., D.S., J.R.T., B.F.T. (Brendan F. Tivnan) and J.G.V.; Data curation, B.F.T. (Brian F. Tivnan), D.S., T.A.B.-H. and J.G.V.; Formal analysis, B.F.T. (Brian F. Tivnan), D.S., J.R.T. and M.T.K.K.; Funding acquisition, B.F.T. (Brian F. Tivnan) and S.M.B.; Investigation, M.T.K.K.; Methodology, B.F.T. (Brian F. Tivnan), D.S., J.R.T., Carl Burke, M.T.K.K., M.T.M. and J.G.V.; Project administration, B.F.T. (Brian F. Tivnan) and J.G.V.; Resources, B.F.T. (Brian F. Tivnan) and S.M.B.; Software, D.S., J.R.T., T.A.B.-H., Carl Burke, M.T.K.K. and M.T.M.; Supervision, B.F.T. (Brian F. Tivnan) and J.G.V.; Validation, B.F.T. (Brian F. Tivnan); Visualization, D.S., J.R.T., T.A.B.-H., Carl Burke, M.T.K.K. and B.F.T. (Brendan F. Tivnan); Writing—original draft, B.F.T. (Brian F. Tivnan) and J.G.V.; Writing—review & editing, B.F.T. (Brian F. Tivnan), B.F.T. (Brendan F. Tivnan) and J.G.V.

Funding: The researchers are grateful for external funding which supported the initial phase of this study. B.F.T., D.S., J.R.T., T.A.B.-H., C.D.B., S.M.B., M.T.T.K., M.T.M., and J.G.V. were supported by U.S. Department of Homeland Security award #HSHQDC-14-D-00006. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Homeland Security or the U.S. Government.

Acknowledgments: The authors gratefully acknowledge the following: Collaborative contributions from Richard Bookstaber, Michael Foley, Christine Harvey, Eric Hunsader, Neil Johnson and Mark Paddrik; helpful insights from Anshul Anand, Chris Danforth, David Dewhurst, Peter Dodds, Jordan Feidler, Andre Frank, Bill Gibson, Frank Hatheway, John Ring, Chuck Schnitzlein, Colin Van Oort, Tom Wilk and attendees at the 2016 International Congress on Agent Computing. The authors’ affiliation with The MITRE Corporation is provided for identification purposes only and is not intended to convey or imply MITRE’s concurrence with, or support for, the positions, opinions or viewpoints expressed by the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Exchanges and Trade Reporting Facilities.

Exchange Family	Exchange Name	Abbreviations	Geographic Location
BATS	BATS	BATS/BZX	Secaucus, NJ
	BATS-Y	BATY/BYX	Secaucus, NJ
	Direct Edge A	EDGA	Secaucus, NJ
	Direct Edge X	EDGX	Secaucus, NJ
Chicago NASDAQ	Chicago Stock Exchange	CHX	Secaucus, NJ
	NASDAQ	NASD	Carteret, NJ
	NASDAQ-Boston	NQBS	Carteret, NJ
NYSE	NASDAQ-Philadelphia	NQPH	Carteret, NJ
	New York Stock Exchange	NYSE	Mahwah, NJ
	New York Stock Exchange—ARCA	ARCA	Mahwah, NJ
	New York Stock Exchange—Market	AMEX/NY-MKT	Mahwah, NJ
SIP	NYSE Trade Reporting Facility	NTRF	Mahwah, NJ
SIP	NASDAQ Trade Reporting Facility	QTRF	Carteret, NJ

References

- Angel, James J. 2014. When Finance Meets Physics: The Impact of the Speed of Light on Financial Markets and Their Regulation. *Financial Review* 49: 271–81. [[CrossRef](#)]
- Arnuk, Salvatore, and Joseph Saluzzi. 2012. *Broken Markets*. Upper Saddle River, New Jersey: FT Press.

- Bartlett, Robert P., and Justin McCrary. 2017. How Rigged Are Stock Markets?: Evidence from Microsecond Timestamps. Available online: <https://www.nber.org/papers/w22551> (accessed on 30 September 2018).
- Bodek, Haim. 2013. *The Problem of HFT: Collected Writings on High Frequency Trading & Stock Market Structure Reform*. Norwalk: Decimus Capital Markets.
- Consolidated Tape Association. Available online: <https://www.ctaplan.com/index> (accessed on 30 September 2018).
- Cont, Rama. 2001. Empirical Properties of Asset Returns. *Quantitative Finance* 1: 223–36. [CrossRef]
- Ding, Shengwei, John Hanna, and Terrence Hendershott. 2014. How Slow is the NBBO? A Comparison with Direct Exchange Feeds. *Financial Review* 49: 313–32.
- Johnson, Neil, Guannan Zhao, Eric Hunsader, Jing Meng, Amith Ravindar, Spencer Carran, and Brian Tivnan. 2012. Financial Black Swans Driven by Ultrafast, Machine Ecology. Available online: <https://arxiv.org/abs/1202.1448> (accessed on 30 September 2018).
- Johnson, Neil, Guannan Zhao, Eric Hunsader, Hong Qi, Nicholas Johnson, Jing Meng, and Brian Tivnan. 2013. Abrupt Rise of New Machine Ecology Beyond Human Response Time. *Scientific Reports* 3: 2627. [CrossRef] [PubMed]
- Lewis, Michael. 2014. *Flash Boys: A Wall Street Revolt*. New York: Norton & Company.
- Mandelbrot, Benoit. 1963. *Variation of Certain Speculative Prices*. New York: Springer.
- Nanex Research. 2014a. NBBO Misfiring. Available online: <http://www.nanex.net/aqck2/4616.html> (accessed on 30 September 2018).
- Nanex Research. 2014b. Perfect Pilfering. Available online: <http://www.nanex.net/aqck2/4661.html> (accessed on 30 September 2018).
- Patterson, Scott, and Jennifer Strasburg. 2012a. Traders Navigate a Murky New World. *Wall Street Journal*, April 9.
- Patterson, Scott. 2012b. *Dark Pools*. New York: Crown Pub.
- Precision Time Protocol is “a protocol used to synchronize clocks throughout a computer network”. Available online: https://en.wikipedia.org/wiki/Precision_Time_Protocol (accessed on 30 September 2018).
- Securities and Exchange Commission. Regulation NMS. Available online: <https://www.sec.gov/rules/final/34-51808.pdf> (accessed on 30 September 2018).
- Tivnan, Brian F., and Brendan F. Tivnan. 2016. Impacts of Market Liquidity and Heterogeneity in the Investor Decision Cycle on the National Market System. Paper presented at SWARMFEST 2016: 20th Annual Meeting on Agent-Based Modeling & Simulation, Burlington, VT, USA, July 31–August 3.
- Tivnan, Brian F., Matthew T. K. Koehler, David Slater, Jason Veneman, and Brendan F. Tivnan. 2017. Toward a Model of the U.S. Stock Market: How Important is the Securities Information Processor? Paper presented at the 2017 Winter Simulation Conference, Las Vegas, NV, USA, 3–6 December; Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Inc., pp. 1181–92.
- Unlisted Trading Privileges. Available online: <http://www.utpplan.com/overview> (accessed on 30 September 2018).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).