EFA Eastern Finance Association

The Financial Review 49 (2014) 313–332

The Financial Review

How Slow Is the NBBO? A Comparison with Direct Exchange Feeds

Shengwei Ding

Wells Fargo Securities

John Hanna

Terrence Hendershott*

University of California

Abstract

This paper provides evidence on the benefits of faster proprietary data feeds from stock exchanges over the regulated "public" consolidated data feeds. We measure and compare the National Best Bid and Offer (NBBO) prices in each data feed at the same data center. Price dislocations between the NBBOs occur several times a second in very active stocks and typically last one to two milliseconds. The short duration of dislocations makes their costs small for investors who trade infrequently, while the frequency of the dislocations makes them costly for frequent traders. Higher security price and days with high trading volume and volatility are associated with dislocations.

Keywords: market data, transparency, high-frequency trading

JEL Classifications: G10, G14

**Corresponding author*: Hass School of Business, 545 Student Services Bldg #1900, University of California, Berkeley, Berkeley, CA 94720; Phone: (510) 643-0619; Fax: (510) 643-1412; E-mail: hender@haas.berkeley.edu.

Redline provided technological support for the project. We thank Gideon Saar, two anonymous referees, and the guest editor, Michael Goldstein, for helpful comments.

1. Introduction

Financial markets have evolved from manual, human-based, single-venue floor trading to ultra-fast, low-latency, multi-venue, fully automated electronic trading. Regulators have continuously updated rules to cope with the change. In the United States, the Securities and Exchange Commission (SEC)'s regulatory objectives include maintaining fair, orderly, and efficient markets (O'Hara and Macey, 1999). By examining the differences between publicly provided market data and data sold directly from the exchanges we provide empirical evidence pertinent to assessing the transparency and fairness of the U.S. equity markets.[1] Our results characterize the amount of latency, the frequency and magnitude of price differences due to latency, and the potential costs to investors arising from latency. We study trading in Apple for one day to illustrate the details of latency in the data. We then examine a set of 24 securities for 16 days in May 2012. We find that using public information imposes small costs for investors trading infrequently and not trading at times when price dislocations between the public/regulated and direct exchange data feeds are more likely. In contrast, active traders are at a substantial disadvantage if they use the public data.

Broadly speaking, there are two trading systems in the United States: registered exchanges and alternative trading systems. The registered exchanges are required to provide the best bids and offers to be included in the consolidated quotation system (CQS) and are also required to file any rule changes with the SEC. The alternative trading systems include electronic communication networks and dark pools which do not provide best quotes to CQS, but are required to match trades within a National Best Bid and Offer (NBBO). In this study, we deal only with the quotation system based on registered exchanges.

Trading occurs on 13 U.S. equity exchanges during our sample period (see O'Hara and Ye, 2011, for evidence on trading across exchanges and batstrading.com for more recent data). With many exchanges trading stocks simultaneously, how it can be ensured that the submitted order is executed at the best bid and offer price across all exchanges? This concern prompted the SEC to establish Regulation National Market System (Reg NMS) in 2007 to protect fair access to the best price for investors, particularly retail investors. Based on Reg NMS, exchanges are required to provide the quotes to the primary exchanges such as NYSE and NASDAQ. The Security Information Processors, known as SIPs for NYSE and NASDAQ, gather the data

[1] We use public data to refer to market data provided under Section 11A of the Exchange Act. What we refer to as proprietary data typically includes more detailed data, for example, limit orders not at the best price, and is not consolidated before distribution. Both data feeds are available to any subscriber, but the proprietary data are significantly more expensive.

from all exchanges and publish their respective NBBO.[2,3] Stock brokers are required by Reg NMS to execute the retail customer trades at the NBBO or better.

Not all market participants have equal access to trade and quote information. Both physical proximity to the exchange and the technology of the trading system contribute to the latency. In addition, gathering and processing data takes time and also causes delay. The NBBO from the NASDAQ SIP may not be the fastest NBBO investors can obtain from the market. The delay is significant to the extent that investors cannot get the optimal price if they have a large amount to be traded. Also, there are delays in trade execution that cause the shown best price to be no longer available at the moment an order reaches the market. Thus, there is uncertainty whether NBBO prices can translate into trade prices. To mitigate problems such as the NBBO requirement, the SEC allows trading via intermarket sweep orders (ISO) and dark pools. ISO is a trade execution method in which an investor sends orders to multiple exchanges for immediate execution, disregarding whether such a price is the best nationwide.

The potential of deriving the NBBO more quickly opens opportunities for companies to directly subscribe to different exchanges, allowing them to calculate a faster NBBO compared to the SIP NBBO. This study provides possibly the first public evidence that access to exchanges and fast calculation of the NBBO could generate profitable opportunities.

Different market participants have different levels of interest in quantifying latency costs. For traditional fund managers whose trading frequency is days or even longer, it is debatable whether they should directly pay attention to latency costs. For institutional investors who commonly adopt algorithmic trading strategies, such as volume-weighted average price (VWAP) or time-weighted average price,[4] their reliance on third-party algorithmic trading software often makes them aware of the latency cost but not to the extent that they monitor it closely. For these investors latency is relevant to the execution of their trades, but not to their asset allocation and portfolio choices. High-frequency traders (HFTs) decide which stocks to buy and sell continuously in real time, so the latest and most accurate information is crucial to them. To the extent that HFTs have an informational advantage over less

[2] The August 22, 2013 failure of the NASDAQ SIP and subsequent halt of trading highlights the central role the SIP plays. This failure stemmed from insufficient capacity at the SIP and raised awareness of the SIP's importance. For details, see http://www.nasdaq.com/press-release/nasdaq-omx-provides-updates-on-events-of-august-22-2013-20130829-00686

[3] For stocks listed on the NYSE, the SIP that provides the NBBO is the CQS and the equivalent NASDAQ system for its listed securities is called the Unlisted Trading Privileges Quote Data Feed. In this study, we obtain the data through the NASDAQ SIP, which provides the NBBO for all the NASDAQ listed stocks. The cost of a co-located server using only SIP data is approximately $7,000 per month. A server also using the direct exchange feed costs roughly three times that amount. The incremental costs are split close to equally between the cost of the direct exchange data and higher network bandwidth requirements.

[4] Hu (2009) provides an analysis of VWAP's role in measuring transaction costs.

well-informed investors, all traders and/or their brokers must be aware of latency issues.

Having access to less up-to-date information complicates trading in a number of ways. Price and execution become less certain because orders at particular prices may change between the time of the last update and the time an order reaches the market. The longer the latency, the larger the uncertainty. This uncertainty can be compounded when trading is fragmented across many markets. Traders with access to more recent prices can also devise various strategies to profit from slower investors. These strategies can range from picking off stale orders in public markets to taking advantage of any stale prices utilized by dark pools.

Latency in the public data reduces transparency to those investors viewing the public data as opposed to the direct data feeds.[5] This difference in transparency is a source of unfairness across investors. Foucault, Roell and Sandas (2003) investigate how investors with slower data are more likely to be picked off by investors constantly monitoring market conditions.[6] Ready (1999) and Stoll and Schenzler (2006) empirically examine how slower traders' orders provide a free trading option for those traders with lower latency. Easley, Hendershott and Ramadorai (2014) study an upgrade in the NYSE's trading system which reduced the latency of off-floor traders. They find that this reduction in off-floor traders' latency as compared to that of on-floor traders improves liquidity and raises stock prices.

The speed at which investors receive new information is a form of differential information across investors. A number of theoretical models explore different aspects of informational asymmetry related to the trading process. Slower market data are a simple informational asymmetry. Hirshleifer, Subrahmanyam and Titman (1994) and Foucault, Hombert and Rosu (2013) explicitly model the strategy of a trader receiving information just ahead of other investors. Easley, O'Hara and Yang (2012) show that when exchanges provide differential access to trade information liquidity is reduced, volatility is increased, and prices are lowered. Cespa and Foucault (2014) show that reductions in insiders' access to post-trade information relative to outsiders may also increase prices by reducing the risk to outsiders.

Empirically measuring informational differences across investors is difficult as investors' information set typically is not observable. O'Hara, Yao and Ye (2013)

[5] Data processing and transmission lead to latency in market data regardless of whether trading is fragmented or centralized. Huang (2002), Barclay, Hendershott and McCormick (2003), and Goldstein, Shkilko, Van Ness and Van Ness (2008) study competition among equity markets. Stoll (2001) posits that investors and brokers can virtually integrate markets together through technology. Latency in market data exists even when securities are traded on a single exchange; see "High-Speed Traders Exploit Loophole," *Wall Street Journal*, May 1, 2013, for a discussion of how this occurs on the Chicago Mercantile Exchange. Data latency is one source of quoted prices on different exchanges being "locked" or "crossed" (Shkilko, Van Ness and Van Ness, 2008).

[6] Moallemi and Saglam (2013) model another cost of latency for investors trying to capture the bid–ask spread by using limit orders. Gai, Yao and Ye (2013) examine how congestion due to order arrivals at NASDAQ can increase the latency in market data.

study the importance of trades smaller than 100 shares in the price discovery process. These trades are not reported in the public data but are reported in markets' proprietary data feeds. Our results complement O'Hara, Yao and Ye (2013) by quantifying another advantage for investors with access to proprietary data feeds: lower latency in observing quotes.

2. Data collection

We collected data on a system provided by Redline Trading Solutions located at BATS' data center.[7] The dedicated server used in this study was connected to various exchanges directly (BATS, Direct Edge A/X, NASDAQ's TotalView feed) as well as to the NASDAQ SIP. We focus on NASDAQ listed stocks so we use the NASDAQ component of the SIP. Ideally, the server should connect with all 13 exchanges to obtain the lowest latency updates and consolidate quotes to generate its own NBBO. If all data feeds exist, the system can generate an NBBO in the same way as the SIP with lower latency. Data from the exchanges and the SIP are given time stamps at the server used in this study so discrepancies in the clocks at the different exchanges and the SIP do not affect our measurement of latency.

For this study, the server is directly connected with the following exchanges: two BATS exchanges, BYX and BZX; two Direct Edge exchanges, EDGA and EDGX; and NASDAQ. In order to construct an NBBO, including data from all exchanges, we combine the direct exchange feeds with the other exchange components from the SIP.

The synthetic NBBO is constructed with the following two rules:

- Use direct data feeds BATS, Direct Edge (EDGE), and NASDAQ and the SIP top of book BBO for other exchanges to build our NBBO
- If BATS, EDGE, or NASDAQ has a new price better than current SIP price, update our NBBO with the new price; if BATS, EDGE, or NASDAQ is alone at the SIP NBBO and has a new price worse than current SIP price, update our NBBO with the new price

The NBBO generated by this approach is not perfect. For exchanges in which the information comes through the SIP, there is no benefit at all. Also, if some updates arrive with short delay (direct feed) and other updates arrive with longer delay (SIP), the amount of time each update stays on top of the book deviates from the real value.

[7] The commercial product used to construct an NBBO is called the InRush™ Accelerator ticker plant. Redline is a company specializing in providing low-latency market data and order-execution systems to its customers. The system used an IBM server with 24 CPUs collocated at Savvis' NJ2 center, where BATS hosts its BZX and BYX exchanges. Savvis provides outsourced Internet infrastructure services and low-latency connectivity to major financial exchanges. The server is co-located with the BATS exchange, which results in the lowest possible latency for BATS. Additional details of the data collection are available in an appendix available online: http://faculty.haas.berkeley.edu/hender/NBBO_Appendix.pdf

Figure 1

AAPL bid/ask price from NASDAQ SIP NBBO on May 9, 2012

However, if more than 90% of the top of book updates are from exchanges with direct feeds, this system generates a synthetic NBBO quite accurately with low latency.

3. Data latency

To illustrate the magnitude of the latency between the public (SIP NBBO) and proprietary (synthetic NBBO) data we initially focus on Apple (AAPL), on a single day, May 9, 2012. Later we expand our analysis to examine 24 securities in the month of May 2012. Figure 1 shows the bid and ask prices for AAPL on May 9, 2012. AAPL's price rose between 1% and 2% from the opening price of \$563.70 to the closing price of \$569.18. The average bid–ask spread is \$0.1621, which is roughly three basis points of AAPL's stock price, consistent with AAPL being highly liquid.

As is apparent from the difficulty in distinguishing the two lines in Figure 1, examining how the bid and ask prices differ between the SIP NBBO and the synthetic NBBO requires much higher time resolution than a daily graph can provide. To illustrate this effect, Figure 2 plots the bid and ask prices for the SIP NBBO from 9:31 a.m. to 9:32 a.m. with dots marking when the synthetic bid and ask prices differ from the SIP. Differences appear multiple times each second while clustering around price changes. This suggests the natural intuition that a security's volatility plays an important role in the value of more up-to-date price information.

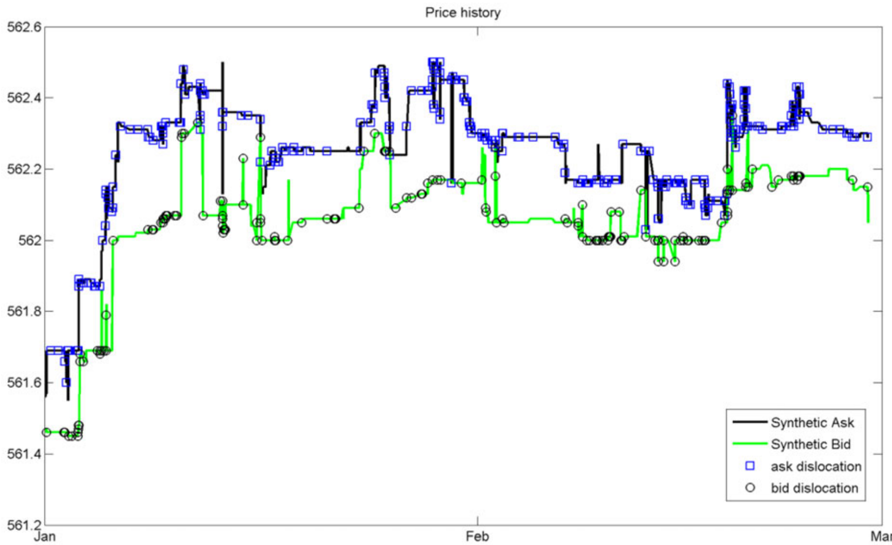

Figure 2

AAPL NBBO from 9:31 to 9:32 a.m. on May 9, 2012

Dots are when the synthetic NBBO is updated before the SIP NBBO

Figure 2 demonstrates that dislocations between the SIP NBBO and the synthetic NBBO occur frequently, possibly several times a second. Figure 2 does not provide information about how long those price dislocations last. Quote changes occur when a limit order improves the best price or the depth at the best price is cancelled or executed against. The changes occur first in the synthetic NBBO and then subsequently in the SIP NBBO. The latency between the two data sources can be quantified by calculating the amount of time between the time stamps of the NBBOs:

$$Latency = Timestamp_{SIP} - Timestamp_{Synthetic}.$$

Figure 3 provides a histogram of the distribution of latency for AAPL on May 9 by market centers. The average latency on BATS is larger than those on EDGE and NASDAQ because the server is located just next to the BATS data center. Thus, the updates directly from BATS arrive immediately on the direct data feeds. The feeds from NASDAQ and EDGE arrive at the BATS data center with some delay due to the distances between the data centers, reducing their latency relative to the SIP. The amount of time it takes for information to be routed between market centers and the SIP determines latency in an absolute sense, but the latency perceived by market participants depends on their perspective; that is, from

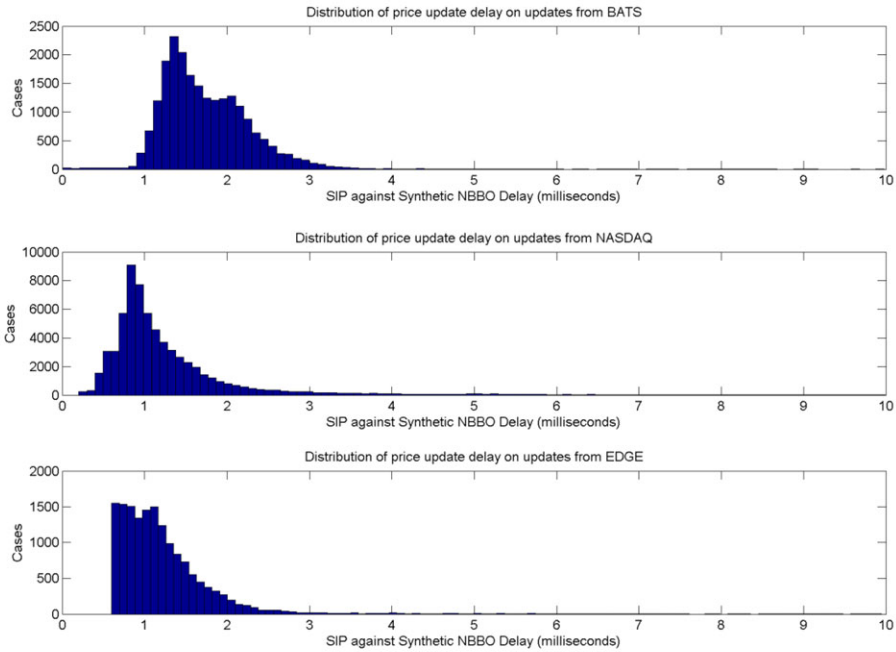

Figure 3

Histogram of latency for AAPL on May 9, 2012, for BATS, NASDAQ and EDGE

which data center latency is measured. Measuring latency is challenging because it depends on location (Wissner-Gross and Freer, 2010). Thus, our exact latency numbers depend on our precise measurement approach and the location of our server.

Averaged across all exchanges latency is about 1.5 milliseconds.[8] As a comparison, the average time it takes to execute a market order is less than one-fifth as large, roughly 300 microseconds. Therefore, brokers waiting for the NBBO information to decide what price and which exchange to route market orders to can face disadvantages. The SEC's 2010 concept release on equity market structure wrote that latency at the SIPs themselves was about five milliseconds. Our latency measure for BATS, which incorporates both latency at the SIP and between the BATS and the SIP, is less than half that size, suggesting that latency has fallen over time.

[8] Figure 4 excludes EDGE observations with a latency less than 600 microseconds. Some of these were due to an issue known at the time, and later corrected, in the EDGE feeds.

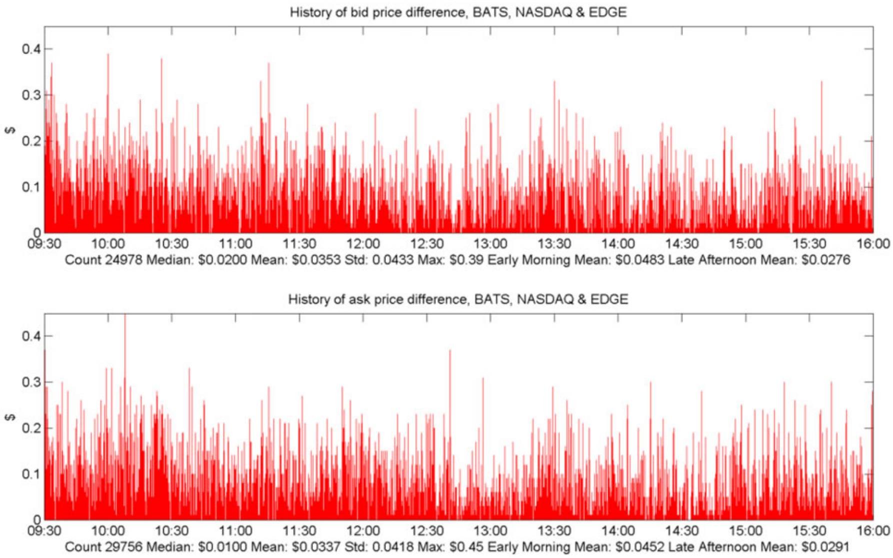

Figure 4

Price dislocations for AAPL on May 9, 2012

4. Price dislocations

The latency in Figure 3 demonstrates the delays in price information that investors receive in the SIP NBBO. The magnitude of latency-generated price dislocations is particularly meaningful to investors. Figure 4 shows the price dislocations for AAPL throughout the trading day on May 9. Price dislocations occur at the bid almost 25 thousand times in the day and at the ask nearly 30 thousand times. There are 23,400 seconds during the 9:30 a.m. to 4:00 p.m. trading day, so price dislocations occur more than twice per second on average.

Figure 4 reports the median price dislocation as being the tick size of \$0.01. However, many price dislocations are greater than \$0.10, making the mean price dislocation 3.4 cents, more than three times greater than the median. If an investor routes orders based on the stale SIP NBBO then the investor can lose this amount on each share. Figure 3 shows that these dislocations are short-lived at only several milliseconds. Therefore, while dislocations are costly and frequent, their impact on infrequently trading investors can be quite small as prices are dislocated less than 1% of the time.

Figure 5 examines how price dislocations occur across exchanges at the ask (the bid looks similar). Dislocations occur most often on NASDAQ and are slightly smaller there. NASDAQ has the largest market share in APPL, so it is not surprising that NASDAQ is where the differences appear most often.

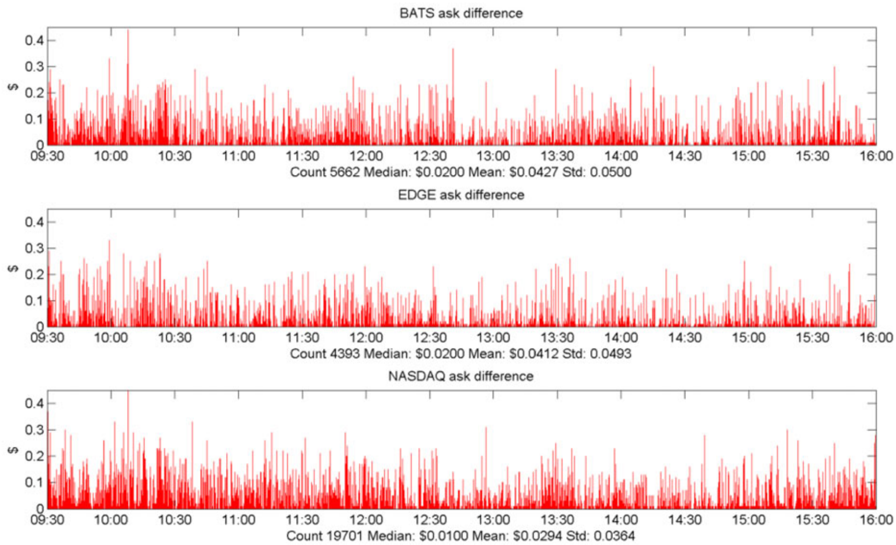

Figure 5

Price dislocations by exchange for AAPL on May 9, 2012

5. The cost of latency

Up to this point we have illustrated the frequency and magnitude of the price dislocations between the SIP NBBO and the synthetic NBBO using direct data feeds from the exchanges. These price dislocations can impact investors in a number of ways depending on their trading strategies. The simplest example would be an investor routing a market order to the exchange with the best price at the SIP NBBO. In this case the frequency and durations of price dislocations provide an estimate of how often the investor's order could go to the wrong exchange. Figure 4 shows that there are 54,734 price dislocations for AAPL on May 9 during the 6.5-hour trading day. This corresponds to 2.34 dislocations per second on average. Estimating that dislocations last as long as the latency shown in Figure 3 of approximately 1.5 milliseconds implies that for 3.51 milliseconds of each second the SIP NBBO and synthetic NBBO differ. This could result in a buy or sell market order going to the wrong market roughly half that often: 0.175% of the time. Figure 4 shows that the average price dislocation is $0.034. Simply multiplying this times the percentage of the time a dislocation occurs yields an expected price dislocation of $0.006 per 100 shares for a market order entered randomly throughout the day. Multiplying this dollar amount by AAPL's May 9 trading volume of 17,167,989 shares yields $942, representing 0.001 of a basis point of dollar volume traded. This suggests that investors randomly routing market orders are unlikely to face meaningful costs due to data latency. However, if investors are more likely to trade when dislocations are occurring, then latency is more costly.

Although price dislocations have small effects on infrequently trading investors, investors that are continuously in the market can be substantially disadvantaged. One example involves dark pools that use the NBBO as a reference price at which orders are matched. If the NBBO is based on the SIP and not the NBBO constructed from exchanges, then the dislocations illustrated in Figure 4 are incorporated in the trading prices in the dark pool. If an HFT monitors the proprietary and SIP NBBOs the trader can enter a buy order when the synthetic NBBO is above the SIP NBBO. If the trader initiating the trade in the dark pool at the SIP NBBO can exit the position at the midpoint of the synthetic NBBO, a profit of half the price dislocation is realized. That profit comes at the expense of the investor who had an order resting in the dark pool.

To illustrate the above logic we provide a simple example. Assume BATS updates AAPL's bid price from $530 to $531, and the ask price remains at $532. This changes the mid-price from $531 to $531.5. In the first 1.5 milliseconds, slower traders are not aware of the price change. If some such traders have placed an order to trade at mid-price in a dark pool, then faster traders can buy the stock at $531 in the dark pool when the synthetic NBBO gets updated. After 1.5 milliseconds, traders can sell it for $531.5 in the dark pool. In this case the trade gains 50% of the price dislocation. Dark pools represent roughly 11% of trading volume, corresponding to 1,888,478 shares of AAPL on May 9. If half of the average dislocation of 3.4 cents is captured on this volume then fast traders would make a profit of $32,510 in a single stock on a single day. The profit figure represents an upper bound on the profits of this type of strategy because it assumes all dark pool trades occur during price dislocations on dark pools using the SIP NBBO for prices. While AAPL is one of the highest-volume stocks, the dollar figure illustrates the possible magnitude of profits and costs stemming from latency for traders continuously in the market. How much of such hypothetical profits could be captured in practice and whether data latency enables other types of profitable strategies require further study.

The above calculations illustrate that latency costs can be very low for infrequently trading investors, but that latency in data can be quite costly for very active investors. There are many other possible costs of latency. For example, we have focused solely on price dislocations for marketable orders. Adjusting limit orders with slow data can result in worse queue position, which reduces the likelihood of the order being filled. Having slower data also reduces the accuracy of information on the quantities at the best prices, which complicates filling larger orders.

6. Price dislocations over time

May 9 is a single day so we next examine its representativeness by studying AAPL on other days in May 2012. In addition to each day's number of dislocations, Figure 6 plots AAPL's intraday volatility based on the percentage difference

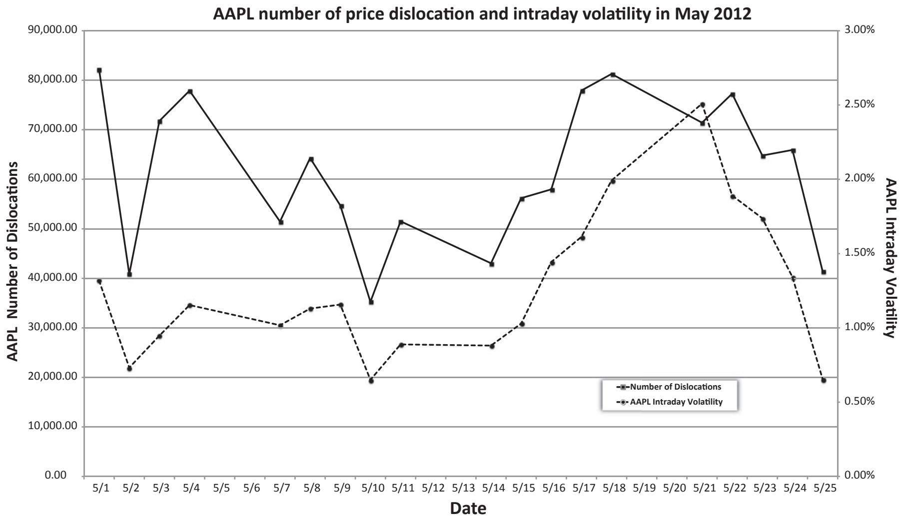

Figure 6

Price dislocations and intraday volatility for AAPL in May 2012

between each day's high and low prices. The largest number of dislocations is 81,279 on May 18. That day AAPL had its second highest intraday volatility in the sample period of 2% and its highest daily trading volume of over 26 million shares (not shown). The three lowest numbers of dislocations occur on May 2, 10, and 25, 40,486, 35,264, and 41,467, respectively, which have the lowest intraday volatilities, 0.73%, 0.65%, and 0.66%, and trading volumes, 15, 12, and 12 million shares, respectively.

The number of dislocations and volatility clearly move together in Figure 6 and the correlation between the two series is 0.71. This relation is not surprising as higher volatility implies more price changes and dislocations occur when the bid and ask prices change. Trading activity also impacts the frequency of bid and ask changes. Trading volume and stock price volatility generally are highly correlated and in the AAPL sample the daily correlation is 0.86. Hence the correlation between the number of dislocations and trading volume also is high.

Thus, the daily analysis for AAPL shows that dislocations are greater when volatility and trading volume are higher. If these same relations hold within each trading day, then the potential costs of latency calculated in this paper represent a lower bound. For example, the costs calculated in Section 5 take daily averages of the number of dislocations, the average size of the dislocations, and trading volume. If these are positively correlated as in Figure 6, then by Jensen's inequality the average of the number of dislocations times the size of the dislocation times the trading volume is larger than the average of the number of dislocations times the average of the size of the dislocation times the average of the trading volume.[9]

7. Price dislocations across securities

While AAPL is a very important stock for investors, its high price makes the tick size of one cent small as a percentage of share price. Its finer pricing grid and rapid trading activity both could lead to price discrepancies occurring quite often in AAPL. To examine the differences across various types of securities we turn to data for 24 securities. The securities were selected to represent a broad cross-section of characteristics including share price, market capitalization, and trading volume. Market capitalization, share price, volatility, trading volume, and number of trades are taken from CRSP.[10] Volatility is measured as the percentage difference between the day's highest and lowest price. The number of securities is limited to ensure that the latency measures do not arise from any congestion on the server collecting the data.[11] The sample period is the 16 trading days from May 4 through May 25, 2012. This sample period is a few days shorter than the sample for AAPL shown in Figure 6 because the first few days of May were used for testing and data were not collected for all securities.

Table 1 provides information on how these securities vary in terms of characteristics and trading activity. Each variable is measured daily and Table 1 reports the average across days in the sample. Dislocations are measured by their number in thousands, their total value (number times size) in thousands of dollars, and their average size in percentage of share price.[12] A Herfindahl index is calculated daily using trading volume at each of the exchange codes in the NYSE's Trade and Quote database. While these are the best publicly available data, they underestimate the true fragmentation of trading as the trade report facilities aggregate trading from a number of different trading venues. This aggregation possibly causes the variation in the Herfindahl to be relatively small.

[9] Calculating a precise estimate of realized cost of latency is very difficult because synchronizing trades and price dislocations across exchanges is possible only if the latency in the different data feeds is known. In addition, some market centers, for example, dark pools, are only required to report their trades within 30 seconds.

[10] Two exchange traded funds, PowerShares QQQ and ProShares Ultrapro Short QQQ, are included. Market capitalization is reported in CRSP for these, but its meaning is less well defined for ETFs. The results are not sensitive to removing these two securities.

[11] Queuing models demonstrate that the server's utilization increases latency in the server, as opposed to latency in the data feeds themselves, nonlinearly. By keeping the number of securities small, the server's average utilization was kept very low to avoid congestion on the server contributing to the latency measures.

[12] For May 10, 2012, data on Arena Pharmaceuticals (ARNA) are missing from CRSP. For EDS on May 15 and CLNT on May 16 there are zero price dislocations.

Table 1

Sample descriptive statistics

The daily sample extends from May 4 to 25, 2012 for 24 securities. Market capitalization is in billions of dollars. Volatility is the percentage difference between the day's highest and lowest prices. Trading volume is in millions of dollars. Number of trades is in thousands. Dislocations are measured by their number in thousands, their total value (average number times average size) in thousands of dollars, and their average size in percentage of share price.

Ticker symbol	Security name	Market capitalization ($B)	Share price ($)	Volatility (%)	Trading volume ($M)	Number of trades (000)	Herfindahl index	Avg no. of dislocations (000)	Min no. of dislocations (000)	Max no. of dislocations (000)	Value of dislocations ($000)	Average dislocations (%)
AAPL	AAPL	522.68	558.98	1.32	10,741.29	138.04	0.23	60.78	35.26	81.28	1.90	0.01
ALXA	Alexza Pharmaceuticals	0.05	0.41	5.96	0.77	1.62	0.29	0.50	0.11	1.76	0.00	0.57
AMZN	Amazon.com	99.48	220.80	1.34	959.68	34.49	0.23	19.74	12.40	27.36	0.52	0.01
ARNA	Arena Pharmaceuticals	0.98	5.23	4.29	122.02	51.72	0.31	1.54	0.00	11.03	0.02	0.21
BRCD	Brocade Communications	2.27	4.96	2.13	33.13	18.33	0.20	0.05	0.03	0.08	0.00	0.21
CLNT	Cleantech Solutions	0.01	3.69	7.51	1.10	0.87	0.29	1.44	0.00	9.76	0.02	0.44
DNDN	Dendreon	1.32	8.57	3.42	60.34	29.77	0.23	0.82	0.19	2.88	0.01	0.13
EDS	Exceed	0.06	2.04	3.65	0.12	0.18	0.35	0.27	0.00	0.97	0.00	0.78
FSLR	First Solar	1.33	15.28	3.82	96.18	29.46	0.29	6.77	2.94	17.58	0.07	0.07
GPRO	Gen Probe	3.68	81.13	0.19	90.89	6.32	0.19	0.89	0.12	2.29	0.01	0.02
HOLX	Hologic	4.54	17.17	1.36	85.98	22.64	0.26	0.93	0.47	1.86	0.01	0.06
INTC	Intel	134.19	26.67	0.95	1,057.12	125.74	0.20	0.44	0.23	0.69	0.00	0.04
MDRX	Allscripts Healthcare	2.08	10.91	1.52	68.96	28.10	0.23	0.42	0.22	0.57	0.00	0.09
MNKD	Mannkind	0.31	1.84	2.91	2.99	4.34	0.30	0.03	0.01	0.06	0.00	0.58
PPHM	Peregrine Pharmaceuticals	0.05	0.50	4.66	0.41	1.20	0.31	1.01	0.20	4.69	0.00	0.54
QGEN	Qiagen	3.95	16.76	0.89	20.08	7.72	0.21	0.72	0.31	1.76	0.01	0.06
QQQ	Powershares QQQ Trust	30.85	63.17	0.85	4,014.79	120.99	0.18	1.48	0.87	2.20	0.01	0.02
RMBS	Rambus	0.49	4.46	2.04	3.50	4.03	0.27	0.17	0.09	0.26	0.00	0.23
SIRI	SIRIUS XM Radio	7.71	2.03	2.54	167.32	47.31	0.28	0.03	0.00	0.21	0.00	0.49
SOHU	Sohu.com	1.71	44.96	1.90	29.86	4.81	0.23	4.47	1.90	7.35	0.08	0.04
SQQQ	Proshares Trust	0.14	40.14	2.45	102.18	8.26	0.21	16.71	1.74	35.04	0.18	0.04
VVUS	Vivus	2.36	23.64	2.30	74.66	18.70	0.21	3.54	1.76	7.91	0.04	0.04
YRCW	YRC Worldwide	0.04	5.80	3.79	1.12	0.66	0.36	1.23	0.62	2.79	0.02	0.23
ZNGA	Zynga	1.68	7.65	4.58	169.74	53.87	0.28	1.41	0.49	3.63	0.01	0.13

AAPL is the largest, highest-price, and most actively traded security in our sample. AAPL has three times more dislocations than the next highest security, Amazon.com, but the average size of these dislocations is only one basis point. The smallest firm is Cleantech Solutions, with a market capitalization of approximately \$10 million and about one million dollars per day in trading volume. Alexza is the lowest price stock at \$0.41. These small, low-priced securities have only 5 and 50 dislocations per day, but these dislocations are large at 21 and 57 basis points. The differences between the minimum and maximum number of daily dislocations demonstrate that the substantial time variation in the number of dislocations shown for AAPL in Figure 6 is also present in other securities. The dislocations statistics suggest that the illustrative numbers for the costs of latency for AAPL in Section 5 do not directly generalize to other securities. Because AAPL is a very active high-priced stock the percentage size of dislocations is much smaller, but dislocations occur much more often than in other securities.

The descriptive statistics in Table 1 suggest a number of interesting possible relations among security characteristics and dislocations. To examine these more systematically, Table 2 provides pairwise correlations among the variables for the 384 security-day observations. The reciprocal (inverse) of price typically is used because the tick size is fixed at one cent for securities priced above one dollar and to mitigate the impact of high-priced securities. The correlations among the security characteristics are not surprising: larger stocks are higher priced with lower volatility and higher trading activity. Trading volume is negatively correlated with volatility because of the negative cross-sectional relation between them. As suggested by Table 1 the number and value of dislocations are highly correlated at 0.95. These two variables are negatively correlated with the average percentage dislocations –0.28 and –0.24, respectively. Security characteristics that are positively correlated with the number of dislocations generally are negatively correlated with the percentage average size of these dislocations.

Table 2 calculates the correlations among the variables in Table 1, in which observations across firms are pooled. This pooling mixes together cross-sectional and time series correlations. The correlation between volatility and the number of dislocations in Table 2 has the opposite sign as shown for AAPL in Figure 6. The graphical correlation is purely time series whereas the pooled correlation is both cross-sectional and time series.

To estimate the pairwise time series relations between the variables and the dislocation measures, Table 3 estimates univariate regressions with securities fixed effects for each of the three dislocation measures on the security characteristics. Because trading volume and the number of trades have a 0.73 correlation we will focus only on trading volume, as the results for number of trades are similar. In addition, because there are no stock splits in our sample period and share price is incorporated in inverse price, market capitalization is not included. Therefore, the four security characteristics and three measures of dislocation lead to 12 separate regressions. Each coefficient in Table 3 is from estimation of a regression of the

Table 2

Correlations

The daily sample extends from May 4 to 25, 2012 for 24 securities. Volatility is the percentage difference between the day's highest and lowest prices. Trading volume is in dollars. Dislocations are measured by their number, their total value (average number times average size) in dollars, and their average size in percent.

	Market capitalization	Share price	Volatility	Trading volume	Number of trades	Herfindahl index	No. of price dislocations	Value of dislocations	Average dislocation
Market cap	1								
Share price	0.94	1							
Volatility	−0.20	−0.24	1						
Trading volume	0.90	0.88	−0.19	1					
Trades	0.64	0.91	−0.16	0.73	1				
Herfindahl	−0.16	−0.20	0.37	−0.18	−0.31	1			
No. of dislocations	0.88	0.92	−0.12	0.85	0.49	−0.14	1		
Value of dislocations	0.92	0.95	−0.14	0.90	0.52	−0.11	0.97	1	
Avg. dislocation	−0.25	−0.32	0.45	−0.25	−0.35	0.55	−0.28	−0.24	1

Table 3

Univariate regressions with security fixed effects

The daily sample extends from May 4 to 25, 2012 for 24 securities. Regressions are conducted on each measure of price dislocation for each independent variable, so the table reports coefficients for 12 regressions. Each regression includes security fixed effects. Volatility is the percentage difference between the day's highest and lowest prices. Trading volume is in shares. Dislocations are measured by their number in thousands, their total value (average number times average size) in thousands of dollars, and their average size in percent. Logarithms are taken of share price, share trading volume, and the number of and value of dislocations. Statistical significance is calculated controlling for heteroskedasticity. Standard errors are in parentheses.

	Log(no. of price dislocations)	Log(value of dislocations)	Average size of dislocations
Log(price)	1.16**	1.22**	−0.11**
	(0.20)	(0.20)	(0.02)
Volatility	14.68**	16.97**	0.95*
	(3.03)	(3.12)	(0.38)
Log(trading volume)	0.61**	0.66**	0.01
	(0.08)	(0.08)	(0.01)
Herfindahl	−3.67**	−3.95**	0.21
	(1.06)	(1.13)	(0.20)

*/** denote significance at 0.95/0.99 level.

column dislocation measure on the row security characteristic and security fixed effects. Hence, Table 3 corresponds to time series only correlations between the variables. Statistical significance is calculated controlling for heteroskedasticity. The regressions are of the form:

$$dislocation_{i,t} = \alpha_i + \beta x_{i,t} + \varepsilon_{i,t},$$

where $dislocation_{i,t}$ is the dislocation measure for security i on day t, α_i is the fixed security effect, and $x_{i,t}$ is the characteristic for security i on day t. Logarithms are taken of price, share trading volume, and the number of and value of dislocations to account for the substantial cross-sectional heteroskedasticity in those variables. Share trading volume is used to separate the effects of share price effects and trading volume.

Table 3 shows that the time series correlations between the number of price dislocations and volatility, trading volume, and higher price all are positive, as are the correlations with the total value of those dislocations. The channels by which price and volume could lead to more dislocations are straightforward as all of these lead to more frequent limit order book updates. A higher trading concentration reducing dislocations could arise from liquidity providers revising their quotes more frequently as the number of markets increases. In the time series, volume and volatility are positively correlated (Tauchen and Pitts, 1983) and in Table 3 higher volatility is

Table 4

Panel regressions

The daily sample extends from May 4 to 25, 2012 for 24 securities. Regressions are conducted on each measure of price dislocation. Regressions are performed with and without security fixed effects. Volatility is the percentage difference between the day's highest and lowest prices. Trading volume is in shares. Dislocations are measured by their number in thousands, their total value (average number times average size) in thousands of dollars, and their average size in percent. Statistical significance is calculated controlling for heteroskedasticity. Standard errors are in parentheses.

	Log(no. of price dislocations)		Log(value of dislocations)		Average size of dislocations	
Log(price)	1.15**	1.11**	1.46**	1.16**	−0.09**	−0.11**
	(0.06)	(0.21)	(0.05)	(0.21)	(0.01)	(0.02)
Volatility	37.11**	6.32*	39.78**	8.45**	0.80*	1.22*
	(5.81)	(2.88)	(5.30)	(2.97)	(0.36)	(0.51)
Log(trading volume)	−0.16**	0.46**	−0.20**	0.47**	−0.02**	−0.02
	(0.04)	(0.09)	(0.04)	(0.09)	(0.00)	(0.02)
Herfindahl	2.48	−1.37	4.58**	−1.50	0.61**	0.09
	(1.39)	(1.06)	(1.21)	(1.07)	(0.19)	(0.18)
Constant	−2.22**		−7.54**		0.48**	
	(0.76)		(0.71)		(0.09)	
Fixed effects	N	Y	N	Y	N	Y
Observations	381	381	381	381	381	381
R-squared	0.63	0.92	0.78	0.94	0.63	0.85

*/** denote significance at 0.95/0.99 level.

associated with more frequent and larger dislocations. This result does not hold in the cross-section because volatility and trading volume are negatively correlated as can be seen in Table 1. Finally, there is a significant common component to both volatility and trading volume across assets (Wang, 2002), which leads to a common component in dislocations across securities.

The correlations in Tables 2 and 3 are useful for assessing the relations between dislocations due to latency and security characteristics. Table 2 shows nontrivial correlations among the security characteristics which make understanding their marginal impact more difficult as the collinearity makes statistical inference less precise. For example, market capitalization and trading volume have a correlation of 0.90 in Table 2.

Table 4 conducts panel regressions with and without security fixed effects of the dislocation measures regressed on the security characteristics. The regressions are of the form:

$$dislocation_{i,t} = \alpha + \beta' X_{i,t} + \varepsilon_{i,t},$$

where $dislocation_{i,t}$ is the dislocation measure for security i on day t, α is the intercept, and $X_{i,t}$ is the vector of characteristics for security i on day t; in the regressions with fixed effects α is replaced by α_i. As in Table 3, logarithms are taken of price, share trading volume, and the number of and value of dislocations.

As in the univariate regressions in Table 3, the coefficients on volatility are positive and statistically significant for all dislocation measures. The more prices change, the more often dislocations occur and the larger those dislocations are. Volatility can also lead to liquidity providers wanting to adjust their quotes more often, which can be another source of dislocations.

Share price matters because the tick size is constant at one cent for all securities. A larger share price means that the tick size is smaller as a percentage of share price. Hence, smaller percentage price changes translate into discrete prices ticking up or down more often in smaller percentage terms in higher priced securities. When prices move, opportunities for dislocations occur. Therefore, higher prices lead to more, but smaller, dislocations due to the tick size effect.

Trading volume could be associated with more or fewer dislocations. Trading volume leading to more changes in the quotes can lead to more dislocations. For example, if all depth at the best bid price is taken out by a marketable order, then a dislocation can occur. In addition, dislocations could lead to higher trading volume as high-frequency trading firms attempt to capitalize on these opportunities. On the other hand, higher trading volume could be associated with higher depth and liquidity and, therefore, more stable quotes. More stable quotes would lead to fewer and smaller dislocations. The positive coefficient on trading volume in the specifications with firm fixed effects is consistent with trading volume being positively associated with dislocations in the time series. The cross-sectional association of trading volume with dislocations is negative as seen in the specifications without fixed effects. After controlling for other security characteristics, the Herfindahl trading concentration index is not robustly associated with dislocations.

8. Conclusion

In this study we compare the NBBO from the public/regulated SIP and the NBBO from proprietary data feeds from the exchanges. Price dislocations between the NBBOs occur several times a second in AAPL and typically last one to two milliseconds. The brevity of dislocations mitigates costs for investors trading infrequently. However, the frequency of the dislocations makes them costly for frequent traders. Examination of 24 securities over 16 trading days indicates that higher security price, trading volume, and volatility are associated with dislocations.

How well does current market data regulation meet its goals? The answer depends on many factors including: (i) how much could further regulatory intervention possibly reduce the small costs for infrequent traders? (ii) how much regulation is needed to protect frequent traders from possible market power by exchanges in the pricing of their proprietary data? (iii) how inefficient is it for all frequent traders to purchase the data from exchanges and then consolidate it? (iv) how effective are the incentives for technological innovation by the more heavily regulated public data providers?

References

Barclay, M., T. Hendershott, and T. McCormick, 2003. Competition among trading venues: Information and trading on electronic communication networks, *Journal of Finance* 58, 2637–2665.

Cespa, G. and T. Foucault, 2014. Sale of price information by exchanges: Does it promote price discovery? *Management Science* 60, 148–165.

Easley, D., M. O'Hara, and L. Yang, 2012. Differential access to price information in financial markets. *Working paper*, Cornell University.

Easley, D., T. Hendershott, and T. Ramadorai, 2014. Levelling the trading field, *Journal of Financial Markets* 17, 65–93.

Foucault, T., J. Hombert, and I. Rosu, 2013. News trading and speed. *Working paper*, HEC Paris.

Foucault, T., A. Roell, and P. Sandas, 2003. Market making with costly monitoring: An analysis of the SOES controversy, *Review of Financial Studies* 16, 345–384.

Gai, J., C. Yao, and M. Ye, 2013. The externalities of high frequency trading. *Working paper*, University of Illinois.

Goldstein, M., A. Shkilko, B. Van Ness, and R. Van Ness, 2008. Competition in the market for NASDAQ-listed securities, *Journal of Financial Markets* 11, 113–143.

Hirshleifer A., Subrahmanyam, and S. Titman, 1994. Security analysis and trading patterns when some investors receive information before others, *Journal of Finance* 49, 1665–1698.

Hu, G., 2009. Measures of implicit trading costs and buy-sell asymmetry, *Journal of Financial Markets* 12, 418–437.

Huang, R., 2002. The quality of ECN and NASDAQ market maker quotes, *Journal of Finance* 57, 1285–1319.

Moallemi, C. and M. Saglam, 2013. The cost of latency, *Operations Research* 61, 1070–1086.

O'Hara, M. and J. Macey, 1999. Regulating exchanges and alternative trading systems: A law and economics perspective, *Journal of Legal Studies* 28, 17–54.

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

O'Hara, M., C. Yao, and M. Ye, 2013. What's not there: The odd-lot bias in TAQ data, *Journal of Finance*, forthcoming.

Ready, M., 1999. The specialist's discretion: Stopped orders and price improvement, *Review of Financial Studies* 12, 1075–1112.

Shkilko, A., B. Van Ness, and R. Van Ness, 2008. Locked and crossed markets on NASDAQ and the NYSE, *Journal of Financial Markets* 11, 308–338.

Stoll, H., 2001. Market fragmentation, *Financial Analyst Journal* 57, 16–20.

Stoll, H.R. and C. Schenzler, 2006. Trades outside the quotes: Reporting delay, trading option or trade size? *Journal of Financial Economics* 79, 615–653.

Tauchen, G.E. and M. Pitts, 1983. The price variability-volume relationship on speculative markets, *Econometrica* 51, 485–505.

Wang, J., 2002. Trading volume and asset prices, *Annals of Economics and Financial* 3, 299–359.

Wissner-Gross, A.D. and C.E. Freer, 2010. Relativistic statistical arbitrage, *Physical Review E* 82, 056104.