How rigged are stock markets?
Evidence from microsecond timestamps

Robert P. Bartlett, III*
University of California, Berkeley

Justin McCrary**
University of California, Berkeley
Columbia University
NBER

Abstract

Using new data from the two U.S. securities information processors (SIPs) between August 6, 2015 and June 30, 2016, we examine claims that high-frequency trading (HFT) firms use direct feeds to exploit traders who rely on SIP prices. Across $3.7 trillion of trades, the SIPs report quote updates from exchanges 1,128 microseconds after they occur. However, the SIP-reported National Best Bid and Offer (NBBO) matches the NBBO calculated without reporting latencies in 97% of all SIP-priced trades. Liquidity-taking orders gain on average $0.0002/share when priced at the SIP-reported NBBO rather than the instantaneous NBBO, but aggregate gross profits are just $14.4 million. These findings indicate that direct feed arbitrage is not a meaningful source of HFT profits, nor can it explain the arms race for trading speed.

JEL Classification: G10, G15, G18, G23, G28, K22
Keywords: latency arbitrage, high-frequency trading; SIP; market structure

***Statement of Financial Disclosure and Conflict of Interest*: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.**

* rbartlett@berkeley.edu, 890 Simon Hall, UC Berkeley, Berkeley CA 94720. Tel: 510-542-6646. Corresponding author.
** jrm54@columbia.edu, 521 Jerome Greene Hall, Columbia University, New York, NY 10027. Tel: 212-854-0278.

"Some have suggested that exchanges that use the SIP data to calculate the NBBO provide unfair opportunities to sophisticated traders to engage in risk-free latency arbitrage."

Senate testimony of Joseph Ratterman, Chief Executive Officer of BATS Global Markets, June 14, 2014.

1. Introduction

Concerns over the different speeds at which market participants access information and the resulting potential for adverse selection in equity markets have occupied center stage in recent years. In particular, the emergence of low-latency trading strategies that can exploit sub-second information asymmetries has led not just to economic research, but also to extensive regulatory scrutiny, litigation, and the approval in 2016 of the Investors Exchange (IEX) as a new stock exchange. Describing high-frequency trading (HFT) as "one of the greatest threats to public confidence in the markets," New York attorney general Eric Schneiderman in 2014 launched a series of high profile lawsuits against U.S. dark pools, exchanges, and HFT firms. Regulators from the U.S. Federal Bureau of Investigation,[1] to the Commodity Futures Trading Commission,[2] to the Securities and Exchange Commission (SEC) have all brought pressure to bear on HFT.[3]

Within this debate, an especially important flashpoint has emerged regarding the differing speeds at which traders can access and process data emanating from the approximately one dozen U.S. stock exchanges. For instance, the controversial use by IEX and several other exchanges of so-called "speed bumps"— intentional delays between the time an order is entered

[1] Scott Patterson and Michael Rothfeld, "FBI Investigates High-Speed Trading," *Wall Street Journal*, March 31, 2014. Available at http://www.wsj.com/articles/SB10001424052702304886904579473874181722310, last accessed May 27, 2019.
[2] Douwe Miedema, "U.S. Futures Regulator CFTC Probing Speed Traders," *Reuters Business News*, April 3, 2014. Available at http://www.reuters.com/article/us-hedgefunds-speed-trading-cftc-idUSBREA321QU20140403, last accessed May 27, 2019.
[3] John McCrank, "Exclusive: SEC Targets 10 Firms in High Frequency Trading Probe—SEC Document," *Reuters Business News*, September 12, 2018. Available at http://www.reuters.com/article/us-sec-investigation-highfrequencytradin-idUSKBN0FM2TW20140717, last accessed May 27, 2019.

on an exchange and the time it is executed or posted—is rooted in a desire to level the playing field between fast traders having preferential access to exchanges' quotation data and other traders on an exchange. Similar concerns about fast traders' preferential access to exchange quotation data motivated the SEC's widely-followed investigation in 2016 of the market-making firm Citadel Securities (Levinson, 2016).

In general, these concerns arise from the institutional fact that trading rules generally require brokers and trading venues to fill market orders at (or better than) the National Best Bid and Offer (NBBO) available across exchanges. Additionally, many venues—particularly non-exchange venues—expressly price transactions by "pegging" them to the NBBO. Market participants can determine the NBBO by looking to its publication by the two centralized securities information processors (SIPs) to which all exchanges are required to report updates to their best bids and offers; however, exchanges are also permitted to provide their quote updates directly to subscribers using superior data feeds.[4] If exchanges provide fast traders with the ability to calculate the NBBO microseconds before other traders relying on the SIPs or other slower data feeds, exchanges are effectively allowing fast traders to foresee changes to the NBBO on which other traders will be transacting, potentially enhancing slower traders' adverse selection costs. Because fast traders would exploit the speed advantage of buying the fastest quote data from exchanges rather than relying on slower data feeds from the SIPs, we refer to this trading behavior as "direct feed arbitrage."

Until recently, understanding the extent to which traders engage in direct feed arbitrage has been hampered by the absence of detailed information concerning the informational advantage of fast traders who obtain data from exchanges' proprietary data feeds rather than from the SIPs. In

[4] Jones (2018) provides an overview of the market data products offered by exchanges.

the meantime, concerns that a principal source of HFT rents comes from exploiting these informational advantages has shaped the broader debate concerning the welfare consequences of the arms race for trading speed. By paying for faster access to exchange trading data, do HFT firms obtain rents in the form of risk-free arbitrage? Or are these concerns just a distraction from understanding the primary sources of HFT profits, whether benign, such as conventional market-making (Menkveld, 2013), or not (such as quote-stuffing (e.g., Egginton et al., 2016))?

In this paper, we use new timestamp data provided by the two U.S. SIPs to conduct the first market-wide analysis of the latency with which the SIPs process quote and trade data, and we present evidence regarding the economic significance of direct feed arbitrage. These data are the result of a regulatory change obligating U.S. exchanges and broker-dealers to report to the appropriate SIP the precise time (measured in microseconds) at which a trading venue either updated a quotation or executed a trade. Moreover, amendments to the SIP operating procedures at this time required the two SIPs to record in microseconds the precise time at which each SIP processed a trade or quotation update submitted by an exchange or broker-dealer. Comparing these two timestamps thus permits an analysis of the SIP processing latency for all trades and quote updates across the entire market. For ease of computation, we focus on all trades involving the Dow Jones 30 from August 6, 2015 through June 30, 2016—approximately the first eleven months of these new reporting requirements.[5]

We find that the mean time gap between the time a quote update is recorded by an exchange matching-engine and the time it is processed by a SIP is just 1,128 microseconds during our sample period. The mean latency for processing trades, however, is approximately 20 times higher, clocking in at 24,255 microseconds. Due to these quote reporting latencies, we show that

[5] We also report extensions of selected key findings to half and three-quarters of the full U.S. listed equities market (by trading volume). These results are qualitatively similar to our results using the Dow Jones 30 alone.

the NBBO reported by the SIP lagged the "true" NBBO on average 10,133 times per day across the Dow Jones 30 stocks.

In addition to describing these new data, we use them to explore the economic significance of direct feed arbitrage. We focus on the costs of trading at stale SIP prices for liquidity takers and for liquidity providers. Somewhat surprisingly, both classes of traders are commonly alleged to be injured by direct feed arbitrage, often at the hand of the other. For instance, the SEC's 2016 investigation into the retail market-making firm Citadel Securities focused on the allegation that market makers filling marketable orders at (or within) the SIP-generated NBBO often did so at stale prices to the disadvantage of retail investors using marketable orders.[6] At the same time, the premise behind the "speed bumps" at IEX and other exchanges is that liquidity providers need protection from HFT firms who use market orders to "pick off" resting limit orders that have been pegged to stale NBBO prices.[7]

While the first strategy has not been examined in the academic literature, the latter strategy is consistent with prevailing models of HFT that examine how the presence of fast traders can raise adverse selection costs for dealers and slower traders using limit orders.[8] At the same time, however, Hoffmann (2014) demonstrates that the very risk of being adversely selected produces

[6] For instance, suppose a direct feed showed the NBBO changing from \$10.00 x \$10.01 to \$9.99 x \$10.00, while the SIP's NBBO remained at \$10.00 x \$10.01. A broker might fill buy orders by selling to them at \$10.01 (the stale NBO reflected in the SIP NBBO) rather than at \$10.00 (the NBO shown in its direct feed). We discuss the Citadel case in more detail in subsection 5.1.

[7] As an illustration of this behavior, consider the following example given in Fox et al. (2015). In it, an institutional investor posts to a dark venue a midpoint buy order for a security when the NBBO is \$161.11 x \$161.15 so that an incoming market order to sell would result in this order being filled at \$161.13. However, if the exchange holding the best ask subsequently decreases its displayed quote from \$161.15 to \$161.12 while the midpoint order rests in the dark pool, a fast trader can detect the new NBBO before the dark venue, providing it a momentary opportunity to send an immediate-or-cancel sell order to the dark venue that will execute at the stale midpoint of \$161.13. Upon receiving confirmation, the fast trader can cover the resulting short position by sending a marketable buy order to an exchange to execute at the new national best bid of \$161.12, producing a penny of risk-free profit. In the meantime, the institutional investor—rather than buying at \$161.115, the actual midpoint—buys at \$161.13.

[8] See, for example, Hendershott and Moulton (2011), Jovanovic and Menkveld (2012), Hoffmann (2014), Brogaard et al. (2015), Budish et al. (2015), Foucault et al. (2016), or Menkveld and Zoican (2017).

strong incentives for liquidity providers to invest in speed to avoid quoting at stale prices. Brogaard et al. (2015) show empirically that market-makers are especially inclined to invest in faster technology. Together, the results in these studies suggest that liquidity providers will trade at venues that rapidly update their estimation of the NBBO, which will limit the opportunity to trade against liquidity providers at stale SIP prices.

To estimate empirically how much traders lose by trading at stale SIP prices, we examine how liquidity takers and liquidity providers fared by trading at prices matching the SIP-generated NBBO rather than the NBBO calculated in a world without any reporting latencies. In general, we ask the following: If every trade occurring at a price equal to the SIP-generated NBBO reflects a trader being subject to adverse selection because of direct feed arbitrage, what are the maximum trading losses to liquidity takers? And what are the maximum trading losses to liquidity providers? To answer these questions, we start by showing how to use the new timestamps reported to the SIPs to reconstruct for each trade in our sample the NBBO that prevailed on the SIP (the SIP NBBO) at the microsecond in which the trade occurred, along with the NBBO that was theoretically possible if there was no latency at all in transmitting quote updates (the Direct NBBO). Reconstruction of this "direct feed" NBBO is made possible by the fact that for each quote update from an exchange, the new timestamp data includes the time at which a quote update was released by the exchange's matching engine and therefore available for distribution over an exchange's proprietary data feed.

With these measures, we estimate over our sample period the gross profits gained and lost on each trade that matched the SIP NBBO rather than the Direct NBBO. Importantly, a trade price that matches a stale SIP price can arise either because a trading venue used the SIP NBBO to price a trade or because a trading venue used direct data feeds but was too slow to update its calculation of the new NBBO before the trade occurred. Regardless of why a trade price matches the stale SIP NBBO, these trades reveal an opportunity for a fast trader to profit from the ability to calculate the new NBBO faster than others in the market. Because zero latency and zero transaction costs are assumed in the Direct NBBO, it is important to note that our methodology provides an outer bound of the overall profitability to fast traders from trading with others at stale SIP prices. Moreover, because trade prices matching the SIP NBBO can arise from venues that actually rely on SIP data, as well as venues that use direct data feeds (but process the data slowly), our approach permits insight into the profitability of direct feed arbitrage strategies despite the increasing use of direct data feeds by many venues.

Overall, our results suggest that SIP reporting latencies generate little scope for exploiting direct feed arbitrage, regardless of whether it is targeted at liquidity takers or at liquidity providers. With respect to liquidity takers, on a size-weighted basis, liquidity-taking trades in our sample that match either the SIP national best bid (NBB) or the SIP national best offer (NBO) would have actually gained on average $0.0002 per share by having their trades priced at the SIP NBBO rather than the Direct NBBO. This number is small in magnitude because, on a size-weighted basis, approximately 97% of SIP-priced trades within our sample occur at a time when the SIP NBBO and Direct NBBO are the same. This finding highlights the low probability that the choice of NBBO benchmark even matters for liquidity-taking trades at the best ask or best offer. Moreover, we find that when the SIP NBBO and Direct NBBO differ, liquidity-taking traders systematically benefit by having their trades priced at the SIP NBBO. We attribute this result to the fact that the NBBO will often increase (decrease) in response to serial buy (sell) orders so that late-arriving buy (sell) orders benefit from the stale quotes that have yet to reflect the new trading interest.

Because there are two sides to every trade, our finding regarding the benefits to liquidity takers of trading at SIP NBBO prices naturally raises the possibility that liquidity providers who trade at stale SIP NBBO prices are being “picked off” by fast traders to earn risk-free profits. To examine whether this is the case, we exploit the fact that such an arbitrage play would require a pair of trades and would thus generate a data residue. However, we find little evidence that these trades are the result of fast traders using market orders to pick off stale limit orders priced at the SIP NBBO to earn risk-free profits by hitting a contemporaneous order on the opposite side of the market. Specifically, our results show that on a size-weighted basis, less than 4% of these liquidity-taking trades could be part of such a strategy, resulting in risk-free profits of approximately $264,000 over our sample period. Permitting the second, off-setting trade to be a passive order executed over the ensuing minute (thus potentially exposing the trader to market risk), we estimate net profits from picking off stale SIP quotes to be approximately $8.5 million over our sample period.

Equally important, while our sample of SIP-priced trades amounts to nearly $3.7 trillion of transaction value over our sample period, we estimate that a liquidity taker capable of picking off every stale quote at the SIP NBBO, where doing so was advantageous to the liquidity taker, would have earned just $14.4 million in gross profits over our sample period before accounting for the costs of the second-leg transaction. By comparison, the trading spreads for these stocks are usually near a penny, so the total trading spreads available to liquidity providers for these $3.7 trillion of trades were roughly $37 billion. Consequently, an HFT firm focused on simply earning the spread on these trades would be competing for gross profits that were well over 2,500 times as great as the gross profits available from an active strategy focused on picking off stale quotes at the SIP NBBO. This latter finding underscores how, at least in the present market, HFT strategies other than direct feed arbitrage are considerably more likely to account for the high speed arms race.

We also estimate the aggregate profitability of these direct feed arbitrage strategies for the entire U.S. equities trading market. Overall, we estimate the annual trading value of SIP-priced trades to be over $40 trillion; however, our estimate of the maximum available profits liquidity providers could earn on these trades from direct feed arbitrage is less than $30 million per year. We similarly estimate the maximum amount of annual gross profits available from picking off stale quotes priced at the SIP NBBO to be approximately $156 to $214 million for exchange trades and $77 to $83 million for non-exchange trades before accounting for any second-leg trades or other trading costs. By comparison, 2016 revenue for Virtu Financial, a single HFT firm subject to SEC reporting obligations, was nearly $700 million, suggesting the profitability of HFT is to be found outside these direct feed arbitrage strategies.

This paper is most closely related to two recent studies of latency arbitrage. Wah and Wellman (2013) estimate the prevalence of latency arbitrage opportunities created by market fragmentation when two or more exchanges create a crossed market (i.e., when the best bid on one exchange creates a NBB that is greater than the NBO). However, their analysis focuses on latency arbitrage strategies designed to exploit crossed markets, while we focus on strategies designed to exploit quote reporting latencies. More relevant to our empirical analysis of direct feed arbitrage is Ding et al. (2014), who study the latency between NBBO updates provided by the publicly-available SIP and NBBO updates calculated using proprietary data feeds for a trader based at the BATS exchange in Secaucus, New Jersey. For such a trader, they find that price dislocations between the two observed NBBOs average 3.4 cents and last on average 1.5 milliseconds. Using a single trading day for Apple, Inc., they use these estimates to conclude that a fast trader could theoretically earn up to $32,000 over the course of the trading day by trading against stale orders in dark pools based on the volume of off-exchange trades. This estimate, however, assumes each off-exchange trade is made during a period of price dislocation. Our data, in contrast, permits analysis of how many trades are actually made during a period of price dislocation across both exchange and non-exchange venues, enabling a precise estimate of the probability that a trade is adversely affected by direct feed arbitrage. Our data also permits an estimate of the trading gains and losses traders experience by having their trades priced at the SIP NBBO. Consequently, our results suggest that such fast traders are not likely to be as highly compensated as the analysis in Ding et al. (2014) suggests.

Finally, while our results establish that there is little scope in today's equity markets for direct feed arbitrage, we caution that these results should not be over-interpreted. In particular, our results do not rule out other types of latency arbitrage that might be prevalent, including other "sniping" strategies such as those studied by Budish et al. (2015).[9] In addition, our results do not rule out the possibility that direct feed arbitrage might have been prevalent in the quite recent past (e.g., 2014), for the simple reason that our data are not available until mid-2015. Nonetheless, our results do clarify that a popular narrative regarding direct feed arbitrage would appear to be scarcely relevant to markets in 2015-2016, and they provide the first broad-based

[9] Budish et al. (2015) study the arbitrage profit opportunities that can arise when the correlation structure between multiple assets breaks down at sub-second time intervals. For instance, they illustrate how prices of the S&P 500 SPDR ETF (traded on the NYSE) often lag those of the S&P 500 e-mini futures (traded on the Chicago Mercantile Exchange) due to the fact that price discovery for the S&P 500 tends to occur in the futures market in Chicago, resulting in stale prices in New York as information travels the roughly 700 miles from Chicago to New York. The authors estimate the potential profitability of a "sniping" strategy that picks off stale quotes in New York by simultaneously trading in Chicago. In this regard, direct feed arbitrage can be viewed as but one example of a broader class of sniping strategies (albeit one in which the lag in prices arises from a venue's choice to utilize the SIP to price transactions, rather than from simply the geographic distance between exchanges). Accordingly, our results speak only to this one sniping strategy that is focused on direct feed arbitrage. For instance, they do not speak to a strategy that seeks to engage in multiple sniping strategies that focus on a variety of correlated assets where the returns might scale with the number of assets targeted.

evidence on the extent of quote, trade, and NBBO latency using the SIPs' new microsecond timestamps.[10]

The remainder of this paper is organized as follows. In Section 2, we provide institutional details regarding the rules governing the dissemination of trade and quote data and the advantage that direct feed data gives fast traders. In Section 3, we summarize the new microsecond timestamps and sample selection choices. We present our empirical estimates of trade and quote reporting latencies in Section 4. In Section 5, we examine the economic consequence to liquidity takers and liquidity providers of having trades in the Dow Jones 30 priced at the SIP NBBO rather than the Direct NBBO. In Section 6, we extend this analysis to entire U.S. equities market. We conclude in Section 7.

2. Institutional background

There are three U.S. national market plans governing the dissemination of quote and trade data for National Market System (NMS) equity securities. These three plans are required by Rule 603 of Regulation National Market System (Reg. NMS) and reflect the historical structure of U.S. equity markets. For exchange trades in NYSE-listed securities (Tape A securities) and securities listed on regional exchanges and their successors (Tape B securities), the Consolidated Trade Association (CTA) Plan requires all exchanges and all broker-dealers supervised by the Financial Industry Regulatory Authority (FINRA) to report last sale information to the Securities Industry Automation Corporation (SIAC), a subsidiary of the NYSE, which acts as the central

[10] A related issue arising from the availability of direct data feeds concerns the fact that any divergence between the SIP NBBO and the NBBO derived from direct data feeds can create the possibility for conflicting trade execution measures depending on which NBBO a venue chooses to use as its pricing benchmark. For instance, in connection with government investigations into the use of direct feed data by Citadel's market-making division, Levinson (2016) reports concerns that Citadel's trade execution statistics were based on the slower SIP data rather than the NBBO available from direct data feeds. In an Internet Appendix, we provide an extensive analysis of how the choice of NBBO benchmark affects a venue's trade execution statistics. Calculating effective spreads using the Direct NBBO rather than the SIP NBBO changes effective spreads by less than 1.3 percentage points for exchange trades and less than a half percentage point for all non-exchange trades.

SIP for any transaction in Tape A and Tape B securities. The Consolidated Quotation (CQ) Plan similarly obligates all exchanges and all broker-dealers supervised by FINRA to report to the SIAC any change in the best bid or best offer (including changes to the number of shares) currently available on each trading venue for Tape A and Tape B securities, which the SIAC uses to calculate the NBBO for these securities.[11] For transactions in NASDAQ-listed securities (Tape C securities), the Unlisted Trading Privileges (UTP) Plan governs reporting obligations for both trades and quotes. Under this plan, exchanges and FINRA members must report last sale information for all exchange trades and all quote updates in any Tape C securities to NASDAQ, which operates as the SIP for transactions in these securities. We refer to the SIP managed by the SAIC as the "NYSE SIP" and the SIP managed by NASDAQ as the "NASDAQ SIP."

While the trade reporting plans initially focused on exchange-based trades, the SEC has required since March 2007 that all off-exchange transactions be reported to a formal FINRA-managed Trade Reporting Facility (TRF) (O'Hara and Ye, 2011). FINRA manages two facilities operated separately by the NYSE and NASDAQ. In combination with FINRA member's trade reporting obligations under the CTA and UTP Plans, this SEC reporting requirement for FINRA members means that off-exchange trades made through a broker-dealer internalizer or in a dark pool are now effectively segregated and reported to the appropriate SIP as having been executed at a FINRA TRF.

In addition to sending market data to the SIPs for consolidation, exchanges and FINRA TRFs are also permitted to sell access to the same transaction data directly to customers through proprietary data feeds. Importantly, the SEC has interpreted Rule 603 to require only that

[11] FINRA operates an Alternative Display Facility (the FINRA ADF) through which non-exchange venues (such as an electronic communications network (ECN)) might choose to disseminate quotations from their subscribers. During our sample period, no venue disseminated any quotations through the FINRA ADF.

exchanges *transmit* data to the SIPs no later than they transmit data through their proprietary data feeds.[12] This implies that traders subscribing to a direct feed avoid the inevitable latency arising from the SIPs' obligation to consolidate and process transaction information before disseminating it.

To establish the magnitude of this delay, Table 1 provides processing times for trade and quote information disclosed by both SIPs from 2014 through the second quarter of 2016. For Tape A and Tape B securities, the time between receipt of a transaction report by the NYSE SIP and its subsequent dissemination of that report averaged 410 microseconds for trades and 450 microseconds for quote updates. Processing times for Tape C securities were slightly higher at 700 microseconds and 750 microseconds, respectively.[13]

[Insert Table 1]

In addition to allowing exchanges to sell their direct feed data, the SEC also allows exchanges to sell co-location services. These services allow customers to place their computer servers in close physical proximity to the exchanges' matching engines to minimize the transit time of the exchanges' market data. For Tape A and Tape B securities, co-location accordingly allows a trader to avoid the additional latency a transaction report experiences when traveling from a market center to the NYSE SIP in the NYSE's Mahwah, New Jersey datacenter (the same datacenter housing the NYSE's matching engine); for Tape C securities, it avoids the latency a report experiences when traveling to the NASDAQ SIP's processing platform in Carteret, New Jersey (the same datacenter housing NASDAQ's matching engine).

[12] See In re NYSE LLC, Exchange Act Release No. 34-67857, at 2 (Sept. 14, 2012).

[13] The secular decline in processing-related latencies shown in Table 1 reflect several initiatives by both SIPs to improve processing speeds.

In light of widespread concerns about the advantages these direct feeds provide fast traders, SEC Chair Mary Jo White requested that the SIPs "incorporate a time stamp in their data feeds to facilitate greater transparency on the issue of data latency" (White, 2015). We use these new timestamps in the analyses below.

3. Data and sample selection

We obtain all trade and quote reports published by the two SIPs for the common stock of firms listed within the Dow Jones 30 as of August 1, 2015. These reports are made available through the NYSE Trade and Quote Daily Files. We focus initially on the Dow Jones 30 in light of popular claims that HFT firms are "overwhelmingly interested in heavily traded" securities (Lewis, 2014: p. 115). Our sample period commences with the full implementation of the new microsecond timestamps on August 6, 2015 (the first full day on which exchanges complied with the new reporting requirements) and ends on June 30, 2016.[14] We focus on quotes and trades occurring during normal trading hours, so we subset the data to include quotes and trades occurring after 9:30:05 and before 15:59:55. As noted in Holden and Jacobsen (2014), the NBBO file of the Daily TAQ file is incomplete; therefore, we manually calculate the NBBO for each security for each microsecond during our sample period using quote updates from the daily TAQ data and the standard Hasbrouck algorithm. In so doing, we restrict our analysis to those quotations that are eligible to establish an exchanges' best offer or best bid (i.e., quotation updates having a condition of A, B, H, O, R, W, or Y). Finally, for our latency analysis in

[14] The implementation date for Tape C securities was July 27, 2015 and August 3, 2015 for Tape A and Tape B securities. However, the BATS Y exchange did not fully commence using the new timestamps until August 6, 2015. Prior to these amendments, SIP messages only carried timestamps marked in milliseconds that indicated when the processing of the messages by a SIP was completed, but not the time a venue processed a trade or a quote. The 2015 timestamp modifications also required clock synchronization among exchanges to ensure that timestamps are accurate within tolerances of 100 microseconds or less. See UTP Vendor Alert #2015 – 7: New Timestamp Definitions for July 2015 Release, available at https://www.nasdaqtrader.com/TraderNews.aspx?id=UTP2015-07.

Section 4, we exclude quote or trade records with missing venue timestamps or with venue timestamps that are subsequent to the SIP timestamp.[15] Imposing these conditions results in a core sample of 481,588,512 trades and 7,260,418,102 quote updates.

We use these data to construct two versions of the NBBO that prevailed at the time of each trade in our sample. In the first version, we calculate the NBBO using the timestamp showing the time (in microseconds) at which a SIP disseminated a quote update. This version reflects the NBBO that was available from the SIP at the moment of each trade; therefore, we designate it as the "SIP NBBO." In the second version, we calculate an alternative NBBO using the new "participant timestamp," which shows the time (in microseconds) at which an exchange-matching engine reported processing a quote update. This alternative version reflects the NBBO at the moment of each trade in a world with no processing or transmission latencies. Because it is derived directly from exchange data, we designate it the "Direct NBBO."

Finally, we further use the participant timestamps to match each trade to the SIP NBBO and Direct NBBO that prevailed at the time the trade was executed. We do so by assigning to each trade a SIP NBBO and a Direct NBBO based on the microsecond at which the trade was executed using the trade's participant timestamp. This approach differs from traditional approaches that assign the SIP NBBO to trades using only the trade's SIP timestamp, which was previously the only timestamp the SIP provided for a transaction. However, a trade's SIP timestamp may not reflect the SIP NBBO that prevailed at the time a venue actually executed the trade due to the transit and processing-related delays associated with the SIP's processing of trade reports. For similar reasons, relying on the SIP timestamp of a trade does not permit

[15] This sample selection rule excludes 64,845,020 quote updates (0.9% of all quotes), only 9 of which are due to missing venue timestamps, and 3,499,562 trade records (0.7% of all trades), none of which are due to missing venue timestamps. Our analyses in Section 5 include in the sample all quote and trade records with venue timestamps that are subsequent to the SIP timestamp, which are excluded in our latency analysis presented in Section 4.

insight into the Direct NBBO that prevailed at the moment a venue executes a trade. Relying on the participant timestamp for trades thus permits a unique insight into how a broker or venue perceived the SIP NBBO and Direct NBBO at the time they were seeking to price transactions, rather than the time at which the SIP processes the trade report.

We additionally classify trades as having been buy- or sell-side initiated based on the SIP NBBO assigned to the trade. In particular, we classify trades priced above the midpoint of the SIP NBBO as buy orders, and we classify trades priced below the midpoint as sell orders.[16] For all trades, we retain the SIP-generated timestamp on a trade report to permit analysis of trade reporting latencies.

Note that the Direct NBBO is a construct rather than a direct measure. No trader has access to the Direct NBBO due to the physical distance between exchange matching engines. Nonetheless, the Direct NBBO provides an in-the-limit representation of the advantages of having access to exchanges' fastest direct feeds. In other words, to the extent that the need to receive and process quotes over direct data feeds diminishes the speed advantage of subscribing to these feeds over the SIP, our use of the Direct NBBO can be viewed as the maximum latency advantage a trader could expect by using direct feeds to construct the NBBO.[17] We note that

[16] For analyses of SIP-priced trades, buy orders are accordingly priced at the SIP NBO, and sell orders are priced at the SIP NBB. We use the SIP-NBBO assigned to a trade as our research question focuses on whether there is harm to traders on venues that price transactions at a potentially stale SIP NBBO.

[17] Even with the need to receive and process quotes over direct data feeds, it is important to note that subscribers to direct data feeds have a structural advantage over SIP subscribers regardless of how fast the SIPs process the NBBO. This advantage arises from the fact that all Tape A and Tape B quote data must be managed, aggregated and disseminated by the NYSE SIP in Mahwah, while all Tape C information is processed by NASDAQ's SIP in Carteret. Thus, assuming just 50 microseconds of processing time by the NYSE SIP, a quote update made in NASDAQ that changes the NBBO for a Tape A security would not be known to a broker relying on the SIP NBBO who is located in Secaucus for 510 microseconds (290 microseconds from Carteret to Mahwah in fiber, 50 microseconds to process, and another 170 microseconds to send the data by fiber to the broker in Secaucus). Were the same broker to use direct feeds to construct the NBBO and assuming it also needed 50 microseconds to process the NBBO upon receipt of a quote update, the broker would see the NBBO change after just 188 microseconds from the quote update being made on NASDAQ (138 microseconds from Carteret to Secaucus and 50 microseconds to process).

Ding et al. (2014), while focused only on a subset of exchanges, take advantage of direct measures.

4. SIP reporting latencies and dislocations of the SIP NBBO and Direct NBBO

We first examine how the reporting latency of the SIP NBBO can result in dislocations between the SIP NBBO and the Direct NBBO. We define reporting latency as the difference between the timestamp of a transaction reported by a SIP (the SIP timestamp) and the participant timestamp, which is the time an exchange matching-engine or broker-dealer records a transaction as having occurred:

$$Latency = Timestamp_{SIP} - Timestamp_{Participant}$$

All timestamps are marked in microseconds; therefore, our measure of latency is in microseconds. We note, however, that the microsecond timestamps for trades by non-exchange venues are uniformly reflected as having occurred in intervals of 1,000 microseconds (i.e., 1 millisecond). We interpret this pattern as reflecting the fact that most non-exchange venues have continued to record transactions at the level of the millisecond.[18] As we discuss below, the delay in transaction reporting for non-exchange trades is so large it can be measured in milliseconds—and hence microsecond precision is not necessary to get an accurate sense of latency for these transactions.

Across exchanges, we find a mean (median) reporting latency for quote updates of 1,128 (557) microseconds. Given our definition of reporting latency, this delay reflects both the time it takes for a quote update to travel from a reporting venue to the appropriate SIP, as well as the

[18] Since 2014, FINRA has required that firms report a trade's execution time in milliseconds when reporting trades to the FINRA facilities if the firm's system captures time in milliseconds. See FINRA Regulatory Notice 14-21 (May 2014), available at http://www.finra.org/sites/default/files/NoticeDocument/p506337.pdf. The new timestamp requirements permit FINRA to convert to microseconds any transaction times submitted in milliseconds by a FINRA member. See NasdaqTrader.com, UTP Vendor Alert #2015 - 7 : New Timestamp Definitions for July 2015 Release, available at https://www.nasdaqtrader.com/TraderNews.aspx?id=UTP2015-07.

time it takes for the SIP to process the message and place it on its multicast feed for distribution. As we are unaware of any prior work utilizing the new SIP timestamps, we provide an extended analysis of quote and trade reporting latencies by venue in an Internet Appendix.[19] As we document there, median reporting latencies for both quotes and trades are tightly correlated with the geographic distance between the reporting exchange and the relevant SIP. This finding provides confidence that the participant timestamps accurately reflect the microsecond at which a quote update occurs on an exchange.

An inevitable consequence of these SIP reporting latencies is for the price of the SIP NBBO to lag changes in the price of the Direct NBBO. For instance, across all securities in our sample, the NBB from the SIP NBBO and that of the Direct NBBO differed on average 10,133 times per day. These differences—which, following Ding et al. (2014), we refer to as "dislocations"—ranged from a daily minimum of 334 for General Electric to a maximum of 159,986 for Apple.[20] However, as one would expect from the median reporting latency of quote updates, the duration of these dislocations was typically short-lived. Across all dislocations of the NBB, for example, the mean (median) duration was 975 (474) microseconds. A standard deviation of 523,411, however, highlights the existence of outliers. In Figure 1, we present a histogram of the duration of NBB dislocations which illustrates the thick-tailed nature of this distribution.[21]

[Insert Figure 1]

With regard to the size of these dislocations, mean and median dislocations for the NBB were \$0.0138 and \$0.01, respectively, with a 99th percentile of \$0.04. Dislocations of the NBO were

[19] In the Internet Appendix, we also analyze the extent to which a venue's choice of the SIP NBBO rather than the Direct NBBO affects its trade execution statistics and explore how the use of millisecond timestamps by non-exchange venues might affect our empirical estimates in Section 5.
[20] Across stocks, the mean number of dislocations of the NBO was approximately 10,140, ranging from a minimum of 325 for GE to a maximum of 163,071 for Apple.
[21] The duration of dislocations for the NBO are similar to those of the NBB. In the interest of space, we present results for the NBB only.

similarly slight, having a mean, median, and 99th percentile measure of \$0.0114, \$0.01, and \$0.04, respectively. These figures are consistent with the fact that securities in the sample often traded at or near penny spreads. Figure 2 shows the histogram of the magnitude of NBB dislocations, which emphasizes how tightly clustered around a penny these dislocations are. Penny dislocations are well over 90% of all dislocations. Dislocations of two, three, and four pennies occur, but are rare. Dislocations of a nickel or above occur so infrequently they cannot be discerned in the graph.

[Insert Figure 2]

Before estimating the trading losses associated with these dislocations, a preliminary question concerns the extent to which the potential profitability of these dislocations can be predicted by a strategic trader. As noted previously, dislocations are associated with NBBO updates; therefore, one possible strategy might be to focus on trading environments that are likely to exhaust the displayed liquidity (thus forcing a change to the NBBO), such as those with a limited level of inside depth. Another might be to focus on stocks where there are many quote updates or trades. Yet another might be to focus on stocks of relatively high volatility. Panels A and B of Table 2 present the results of regression models that estimate the relation between the natural log of the total number of daily dislocations observed for stock *i* on day *t* and these measures.[22] We focus on a limited set of covariates that we expect to be associated with NBBO dislocations under the type of reasoning outlined above. Panel A presents the results of

[22] In principle the number of daily dislocations could refer to the number of dislocations on the ask or the bid side of the NBBO or some mixture between them. As a practical matter, the number of ask-side NBBO dislocations and bid-side NBBO dislocations have a correlation of over 0.99 in both levels and in logs. The table presents results for the log of ask-side dislocations.

regression models with no fixed effects, and Panel B presents the results of regression models including fixed effects for the stock and for the trading day.[23]

[Insert Table 2]

Column (1) of Panel A presents the results of a simple bivariate regression in which we regress the natural log of daily dislocations on the natural log of daily average inside depth.[24] This simple bivariate model explains over half the variation in the log of daily dislocations, with an R^2 of 55%. The coefficient is negative, indicating that stocks with high inside depth are associated with fewer dislocations. Since both the dependent variable and covariate are measured in logs, the coefficient of approximately -0.5 admits an interpretation as an elasticity, suggesting that every 10% increase in the daily average of inside depth is associated with a 5% reduction in daily dislocations. We additionally control for the log of total daily quote updates and present the results in Column (2). As might be expected since more liquid stocks have more quotes and more inside depth, the elasticity of daily dislocations with respect to daily inside depth increases in magnitude to -0.8. The estimated elasticity of daily dislocations with respect to daily quote updates is 0.93, indicating a near one-to-one relationship between daily quote updates and daily dislocations. We next add the log of total daily exchange trades to the regression model given that exchange trades have the potential to lead to a dislocation mechanically since they can exhaust inside depth. The results of this model are presented in

[23] The main entries in the table are point estimates, with standard errors in parentheses below. The standard errors cluster on stock, which is accurate if the number of stocks is large. Clustering on a modest number of groups is better than making no adjustment but is known to lead to understated standard errors (Angrist and Pischke, 2009; Chapter 8). To account for that understatement, we additionally present in brackets a confidence interval based on the wild cluster percentile-t bootstrap with Rademacher weights, which is known to be more accurate in this context (Cameron et al., 2008; MacKinnon, 2015). We perform our wild bootstrap calculations in Stata version 13 using David Roodman's package, boottest, which uses bisection to find the endpoints of the confidence region (Roodman et al., 2019). We use 99,999 replications and a seed value of 789747786 (we took this seed from random.org, specifying endpoints of zero and one billion, timestamp 2018-08-21 16:56:36 UTC). We wrote our own (much slower) code to verify the results of boottest and can attest to the accuracy of the program.

[24] We measure inside depth as the daily average of the inside depth on the bid and the ask side.

Column (3). The log of total daily exchange trades has little effect on the inside depth elasticity, which remains approximately -0.8. If we interpret quote updates and exchange trades as proxies for daily stock activity, then it is natural to consider the sum of the coefficients on the two as a measure of the elasticity of daily dislocations with respect to that activity.[25] The sum of these two estimated coefficients is approximately one, suggesting a simple model in which daily dislocations rise proportionately with daily stock activity.[26] Despite the parsimony of this regression model, it explains fully 95% of the variation in log daily dislocations.

In column (4) we present results of a regression model that adds the log of daily off-exchange trades. Exchange trades and off-exchange trades both proxy for trading activity, but only exchange trades can exhaust inside depth. In contrast to exchange trades, the number of daily off-exchange trades for a stock has little predictive power for the number of daily dislocations.

The remaining columns in Panel A present the results for regression models that add as covariates the log of the daily average trade size (column (5)), the log of daily volatility (column (6)), the log of average quoted spread (column (7)), and the fraction of daily trades that are intermarket sweep orders (ISOs) (column (8)). The results in column (5) indicate that a 10 percent increase in the daily average trade size is associated with a 5% increase in daily dislocations, consistent with the fact that larger trades are more likely to exhaust displayed liquidity. Including the log of the daily average trade size in the model also increases the magnitude of the inside depth coefficient elasticity, from -0.83 in column (4) to -0.92 in column (5). Column (6) shows that the daily volatility of a stock has a modest but precisely estimated

[25] This approach is formalized in Lubotsky and Wittenberg (2006), who show that the sum of the coefficients on the two proxies is the best available proxy-based estimator in a specific sense. See their equation (4) and associated discussion.
[26] In unreported results, such a confidence interval always brackets one for models that exclude fixed effects and is above one for those that include fixed effects.

elasticity of about 0.1.[27] The remaining variables we consider, the log of average daily quoted spread and the fraction of daily trades that are ISOs, have generally small and insignificant effects on the log of daily dislocations, as shown in columns (7) and (8).

In columns (1) through (8) of Panel B of Table 2, we present the results of the same regression models used for Panel A of Table 2 except that we additionally include fixed effects for stock and for trading day. Robustly across columns (1)-(8) in Panel B of Table 2, the inside depth elasticity remains large in magnitude and precisely estimated. The quote update elasticity rises slightly relative to models that exclude fixed effects, rising to about 0.7 from 0.6, while the total exchange trades elasticity fluctuates slightly from column to column. The net effect is that the sum of the two elasticities exceeds one in columns (2) through (8) of Panel B with a confidence interval to the right of one, whereas the sum of the two elasticities is more nearly one in columns (2) through (8) of Panel A with a confidence interval that includes one. The elasticity for total off-exchange trades continues to be small and insignificant when fixed effects are included, and the average daily trade size elasticity and daily volatility elasticities are notably smaller in magnitude, though these latter two estimates remain highly statistically significant. Columns (7) and (8) of Panel B show that the final two covariates we consider, the log of average daily quoted spread and the fraction of daily trades that are ISOs, continue to be insignificant predictors of the log of total daily dislocation. Most of the predictive models that include fixed effects have a high R^2 of 98%.

We next examine the extent to which this model can predict not only the number of daily dislocations but also stocks where direct feed arbitrage should yield the greatest profitability. In theory, a trader seeking to exploit the existence of dislocations between the SIP and Direct

[27] We define volatility as the square root of the average squared trade-to-trade returns.

NBBO would seek to know not only those stocks where dislocations are likely to occur but also, conditional on a dislocation occurring, which dislocations will be the greatest and which will last for the longest period of time, since those dislocations will be worth the most money and be the easiest to exploit, respectively. Accordingly, we construct for each stock i on day t a measure called *microsecond pennies*. We construct this measure by first computing, for each dislocation, the product of its dollar value (in pennies) and its duration (in microseconds). The daily sum of these microsecond pennies, which we call *aggregate microsecond pennies*, constitutes our proxy for the potential profitability of direct feed arbitrage for that stock-day.

In Panel C of Table 2, we present the results of a regression model where the outcome variable is the natural log of aggregate microsecond pennies and the regressors include the full set of covariates used in Panels A and B. Column (1) in Panel C presents the results of our baseline model, while column (2) presents the results of a model where we additionally control for time and stock fixed effects. Once again, the model fit is notable with an R^2 of 85% in column (1) and 92% in column (2). The signs of the covariates are generally consistent with those of Panels A and B, indicating that the same factors that are associated with the overall number of dislocations are also associated with their daily profit potential. The primary exception is the average daily quoted spread: Conditional on a dislocation occurring, stocks with wider average daily quoted spreads were strongly associated with having greater profit potential for a trader seeking to engage in direct feed arbitrage.

Finally, in implementing a direct feed arbitrage strategy, a trader would also have to assess the profit potential based on whether a trader planned to target active orders or passive orders. For instance, a trader targeting passive orders would seek to pick off stale quotes priced at the SIP NBBO. The direction of the dislocation accordingly affects the profitability of the strategy:

When seeking to pick off stale quotes at the ask, the trader would seek out those dislocations where the SIP NBO is less than the Direct NBO. Conversely, when seeking to pick off stale quotes at the bid, the trader would seek out those where the SIP NBB is greater than the Direct NBB. Within our sample, the incidence of each form of dislocation is roughly equal. For instance, across the approximately 24 billion microseconds that comprise a trading day, the SIP NBB exceeded the Direct NBB for a total of 5,096,810 microseconds on average, compared to an average of 4,785,411microseconds when the SIP NBB was less than the Direct NBB (t=-0.4). Moreover, regressing the number of times the SIP NBB exceeded the Direct NBB per day on the number of times the Direct NBB exceeded the SIP NBB per day across the 6,840 security-days in our sample yields a coefficient of 0.993 and an R^2 of 0.986. For a trader focused on picking off stale quotes at the SIP NBB, our model would provide a sufficient means to identify profitable trading opportunities.

In contrast, a trader focused on targeting active orders would be faced with a more challenging situation. Such a trader would effectively be in the position of a wholesale market maker looking to fill incoming marketable orders at stale SIP prices, while profiting from prices observable on direct feeds.[28] As with the previous trader, the direction of the dislocation would accordingly matter. For instance, a trader offering liquidity at the bid would profit from direct feed arbitrage when the SIP NBB is less than the Direct NBB. The ability to profit, however, would be conditional on the arrival of incoming marketable sell orders. Yet incoming sell orders should, all other things equal, be associated with a declining NBB as the sell orders absorb depth

[28] The strategy is analogous to a dealer's attempt to exploit the "lookback option" examined in Stoll and Schenzler (2006). In their setting, a dealer holding a marketable order has a positive period of time to fill it. If the price moves in favor of the dealer, the dealer can decide to execute the order at the quote in existence when the order arrived; if the price moves against the dealer, the order can be filled at the new quote. A market maker engaged in direct feed arbitrage has a similar option insofar that it can choose whether to fill the order at the Direct NBBO (reflecting the new quote) or the SIP NBBO (reflecting the old quote).

at the NBB, which should cause the Direct NBB to change lower before the SIP NBB given SIP reporting latencies. As such, while dislocations appear to be roughly split between those where the SIP NBB exceeds the Direct NBB and vice versa, the fact that the trader is dependent on inbound marketable orders to profit from direct feed arbitrage suggests the profitability of this strategy should be considerably less than a strategy that targets passive liquidity. Moreover, to the extent liquidity providers update bids and asks jointly in response to market conditions, the SIP NBBO is likely to be favorable to a liquidity provider when marketable orders are arriving on the wrong side of the market. For instance, to profit in a market where the SIP NBB is less than the Direct NBB, a liquidity provider would need to fill in-bound marketable sell orders. However, a situation where the SIP NBB is less than the Direct NBB is more likely to occur when a series of marketable buy orders exhausts the NBO, inducing liquidity providers to upwardly adjust both sides of the market.

The difficulty of profiting from direct feed arbitrage is illustrated in Figure 3, where we examine a one second trading window in Apple on November 13, 2015. The figure plots the number of sell trades and buy trades per millisecond over the course of this one second window. In addition, two line graphs are presented showing for each millisecond (a) the number of dislocations of the NBB in which the SIP NBB is less than the Direct NBB and (b) the number of dislocations where the NBBs are reversed. A liquidity provider would seek to fill the sell orders when the SIP NBB is less than the Direct NBB (i.e., case (a)). However, the figure illustrates that the sell orders occur only in the millisecond when the Direct NBB dips below the SIP NBB, which should be expected to occur given that sell orders can force the NBB downward as they exhaust displayed liquidity. Moreover, the figure illustrates that when the SIP NBB is lower than the Direct NBB, the trader observes the arrival of buy orders rather than sell orders. This is consistent with liquidity providers updating both sides of the market as buy orders arrive, inducing the SIP NBB to lag the Direct NBB. These considerations suggest that overall profit opportunities should be considerably less for a direct feed arbitrage strategy focused on targeting active orders.

[Insert Figure 3]

5. Trading losses from trading at the SIP NBBO

In this section, we use the microsecond timestamps to investigate the extent to which traders during our sample period could have been adversely affected by SIP reporting latencies.

5.1 Estimated losses to liquidity takers

We first examine claims that direct feed arbitrage can be used to harm retail traders whose market orders are filled by retail market makers having access to direct data feeds. As an example, suppose a direct feed showed the NBBO changing from $10.00 x $10.01 to $9.99 x $10.00, while the SIP's NBBO remained at $10.00 x $10.01. The SEC's recent investigation of Citadel provides evidence that Citadel' market-making desk might fill buy orders by selling to them at $10.01 (the stale NBO reflected in on the SIP feed) rather than at $10.00 (the NBO shown in its direct feed). Citadel could then cover by buying at $10.00 (the actual NBO), earning $0.01 of risk-free profit. However, as the investigation reveals, Citadel ended this strategy in January 2010 and it affected only 0.4% of Citadel's retail order flow, calling into question the extent to which retail traders remain subject to this form of direct feed arbitrage.[29] Likewise, our analysis in Section 4 suggests that exploitable trading opportunities of this nature should be rare.

[29] The investigation also notes that when Citadel filled an order at a stale SIP price, it sought to cover in the market for less than 6.9% of the shares filled in this fashion. *See* In the Matter of Citadel Securities LLC, Jan. 13, 2017, available at https://www.sec.gov/litigation/admin/2017/33-10280.pdf.

Because our data does not discriminate between retail and non-retail orders, we estimate trading losses for all liquidity taking orders that are filled at a price equal to the SIP NBBO rather than the Direct NBBO. To do so, we exploit the fact that our dataset includes both the SIP NBBO as well as the Direct NBBO prevailing for every trade in our sample. This basic structure permits us to estimate investor losses in a two-step process. In step one, we identify those trades that match the SIP NBBO by defining an indicator variable *SIP priced* that equals 1 when the trade price matches either the NBB or NBO as reflected on the SIP NBBO, and equals 0 otherwise.[30] Trades that are SIP priced represent purchase and sale transactions that place the liquidity taker in the same position as if the venue priced the order using the SIP NBBO. Second, because trades priced at the SIP NBBO represent liquidity-taking orders that might have been at risk of direct feed arbitrage, we next compare how these SIP-priced trades would have been priced had they been priced at the Direct NBBO. We then measure whether a trade priced at the SIP NBBO rather than the Direct NBBO resulted in a loss or a profit for the trader placing the liquidity taking order.

In Table 3, we illustrate this two-part process using 35 trades occurring in Apple, Inc. over a 15 millisecond period on November 13, 2015. The time set forth in column (2) is the participant timestamp, which is the timestamp reported by the trading venue for when the trade occurred. We use the participant timestamp to place trades in chronological order. The participant timestamp gives us the ability to sort quotes and trades according to the moment they occurred, conferring knowledge of the actual quoting environment surrounding trades. For comparison, column (3) presents the SIP timestamp for the trade. Note that several pairs of trades, such as the

[30] As noted in Bartlett & McCrary (forthcoming), trading venues also frequently use the NBBO to price trades at its midpoint. However, because we require trade direction to evaluate a trade's profitability, we focus only on those trades priced at exactly the NBB or NBO, which allows us to assign trading direction based on the side of the market that the orders hits.

fifth and sixth, are in chronological order according to the participant timestamp (reflecting the actual sequence in which they occurred) but not in chronological order according to the SIP timestamp.

[Insert Table 3]

Columns (4) and (5) present the NBB and NBO in effect at the time of the trade as reflected in the Direct NBBO, while columns (6) and (7) present the NBB and NBO as reflected in the SIP NBBO. As shown in the table, the trades commenced when the Direct and SIP NBBOs match at $113.37 x $113.38. At that time, however, the market data suggest an inter-market sweep order (ISO) to buy approximately 6,000 shares with a limit price of $113.39 was submitted to all exchanges sitting at the NBO (BATS, Direct Edge A, NASDAQ, and the NYSE Arca).[31] Evidence of this order can be seen by the manner in which the first 30 trades (each marked with code "F" in column (8) for an ISO) sweep through these four exchanges (column (9)), buying all shares on the venues that are offered for less than $113.40 (column (10)). This order results in the Direct NBBO changing to $113.39 x $113.40 by 11:37:47.465000, at which time an apparently unrelated trade occurs in a non-exchange venue (Exchange Code="D"). At the time of this latter trade, however, the SIP NBBO reflects a stale NBBO of $113.37 x $113.38. Following this non-exchange trade, the SIP NBBO updates to reflect the true NBBO so that the Direct NBBO and SIP NBBO match one another by the time of the last three trades.

For purposes of analyzing this sequence of trades, we focus on those trades whose price matched the SIP NBBO, identified in column (13) as "SIP Priced." Were these trades actually

[31] An order marketed as an ISO is exempt from the Order Protection Rule of Reg. NMS, which prohibits a venue from filling an in-bound order if superior prices rest at other exchanges. As such, a trading venue receiving an inbound liquidity-taking ISO can fill it without checking other venues for better prices. However, the broker sending the ISO is responsible for sending simultaneous orders that sweep all venues with better prices. As such, ISO orders allow a trader to sweep through multiple levels of a venue's order book, as occurs in this example.

priced off the SIP, the SIP's delay has an economic effect only for the non-exchange trade (Trade #31, highlighted in bold) occurring immediately after the ISO order finished sweeping through the market and inducing a mismatch between the Direct NBBO and the SIP NBBO. Note, however, that we make no formal assumption about the data feeds actually used by a trading venue: Because we seek to investigate the economic costs of trading at stale SIP prices, our focus is on trades that are priced at the SIP NBBO regardless of whether it was because a venue actually used the SIP to price the trade or whether it used direct feeds that were simply too slow to update the venue's perception of the NBBO before the trade occurred.[32] This focus on SIP-priced trades allows us to estimate the maximum economic consequence of trading at stale SIP prices even in an environment where traders and venues seek to minimize direct feed arbitrage.

Based on the price of this SIP-priced trade, it appears to have been the result of a marketable buy order; therefore, the fact that the trade was filled at $113.38 (the stale NBO) rather than at $113.40 (the new NBO) allowed the originator of the order to save two pennies per share acquired, or $2.00 for the total order. Because we are testing for whether liquidity takers (such as the originator of this order) were harmed by trading at SIP prices, we record this trade as realizing negative "lost profits" of $2.00 (-$0.02 per share) because the liquidity taker *gained* rather than lost by trading at the SIP NBBO. The SIP NBBO and the Direct NBBO matched one another for all other trades, so the choice of NBBO had no effect on trade profitability for these other trades.

[32] In other words, this assumption has the effect of maximizing the number of trades that could be affected by direct feed arbitrage. For instance, if the non-exchange venue handling Trade #31 utilized direct feeds and was co-located with the NYSE, the $113.38 price for Trade #31 could reflect a midpoint execution.

We next generalize this type of analysis to our full sample of Dow Jones 30 trades. In Panel A of Table 4, we summarize by exchange the percentage of trades that we classify as SIP priced (weighted by trade size) and their aggregate transaction value. We find approximately 74% of all shares traded were traded at prices that exactly match the SIP NBBO, representing transaction volume of approximately $3.7 trillion of the $5.2 trillion of observed trading volume. Excluding shares traded in non-exchange venues, this percentage increases to approximately 87% and transaction volume declines to approximately $3 trillion.[33]

[Insert Table 4]

Panel B of Table 4 presents by trading venue the mean amount of lost profits per share that liquidity takers experienced by having their trades priced at the SIP NBBO rather than at the Direct NBBO. For each venue, means are size-weighted based on the number of shares traded. As noted previously, SEC investigations of direct feed arbitrage have focused primarily on whether retail market-makers are exploiting direct data feeds to the detriment of liquidity-taking retail orders. We therefore present results separately for trades reported to a FINRA TRF, which include all trades made by retail market-makers and trades made in dark pools. For completeness, however, we also present results for each exchange.[34]

[33] As discussed in the Internet Appendix, the NASDAQ SIP may have printed timestamps on messages approximately 200 microseconds before it finished processing transaction reports in Tape C securities. All results in Table 4 are virtually identical when we conduct analyses after reducing the SIP timestamps in Tape C securities by 200 microseconds.

[34] Although concerns that retail traders are disadvantaged by direct feed data have focused on retail market-makers, an exchange's use of slow data feeds to calculate the NBBO can also matter for retail traders for at least two reasons. First, exchanges generally permit limit orders to be pegged to the bid, ask, or midpoint of the NBBO, as calculated by the exchange. Second, Rule 611 of Reg. NMS prohibits an exchange from trading-through a protected quotation; therefore, exchanges must route in-bound marketable orders to an exchange holding the NBBO if the exchange is unable to fill the order at a price that is at least as good as the NBBO. An exchange using slow data feeds might accordingly be at risk of filling in-bound market orders at stale NBBO prices by letting them hit pegged orders or by failing to route them to markets holding the actual NBBO.

Overall, the results in Panel B indicate that liquidity-taking trades priced at the SIP NBBO had average lost profits of approximately -$0.0002 per share. As indicated in our Apple illustration, lost profits are defined as the Direct NBB minus the SIP NBB for sell orders, and the SIP NBO minus the Direct NBO for buy orders. As such, these negative lost profits suggest that liquidity takers, on average, saved $0.0002 per share by having their trades priced at the SIP NBBO rather than at the Direct NBBO—the opposite of what would be expected if liquidity takers were systematically receiving inferior pricing due to SIP reporting latencies. The average savings for liquidity-taking trades occurring on a non-exchange venue were slightly higher, having an average of $0.0003 per share.

To understand better the potential profits available from trading at these stale NBBO prices, in Panel C we expand the analysis to the full distribution of lost profits per share based on the number of trades in the sample. Given heightened concerns about direct feed arbitrage within dark pools, we present the distribution separately by exchange and non-exchange venues. We note, however, that estimation of lost profits for non-exchange trades is complicated by the fact that non-exchange venues have continued to record transactions at the millisecond level. For purposes of presenting Panel C in Table 4, we rely on the timestamps attached to each non-exchange trade for estimating the lost profits for these trades. In the Internet Appendix, we also explore how these estimates would change were we to estimate lost profits as the time-weighted lost profits per trade across the 1,500 microseconds at which the trade could have actually occurred, as well as if every trade occurred at the microsecond that would either minimize lost profits per share or maximize lost profits per share.[35]

---

[35] We lack knowledge as to whether clocks used by non-exchange venues round or truncate timestamps to arrive at the reported millisecond timestamp. Accordingly, we assume that a non-exchange trade marked with a timestamp such as 9:45.00.005000 could have actually occurred over 1,500 microseconds beginning with 9:45.00.004500 (if clocks round) and ending with 9:45.00.005999 (if clocks truncate). For every non-exchange trade, we estimate the

Column (1) in Panel C provides the percentage of trades producing the specified amount of lost profits per share. Columns (2)-(4) provide the percentage of trades (on a size-weighted basis), the aggregate transaction value, and the aggregate number of shares associated with these trades, respectively. Column (5) presents the aggregate lost profits for these trades (i.e., Total Shares x Lost Profits Per Share). While the results in Panel B suggest liquidity-taking orders, on average, benefited when their trades were priced at the SIP NBBO, the results in column (5) permit insight into the maximum amount a retail market maker might have captured in our sample by exploiting this form of direct feed arbitrage.[36]

A notable finding reflected in Panel C concerns the extremely high frequency of trades having zero lost profits per share, which occurs when the SIP NBBO and Direct NBBO coincide at the time of a trade. As reflected in the distribution, a trade priced at the SIP NBBO rather than at the Direct NBBO had no economic effect for approximately 97% of trades priced at the SIP NBBO. As is apparent in our Apple illustration, it is only when the SIP and Direct NBBOs differ that the choice of NBBO matching can affect transaction prices. Accordingly, the high percentage of shares traded with zero lost profits reflects the simple fact that the SIP and Direct NBBO typically match one another at the time of a trade.

For those trades that produced non-zero lost-profits per share, Panel C shows that nearly 94% of the trades (weighted by shares traded) produced better pricing for liquidity takers when the trade's price matched the SIP NBBO rather than the Direct NBBO. Specifically, among trades

---

lost profits for each of these hypothetical trade times based on the SIP NBBO and Direct NBBO that prevailed at each of these hypothetical trade times. Our estimate of the time-weighted lost profits per trade is calculated as the time-weighted average of lost profits for these 1,500 hypothetical trade times. In contrast, our estimate of the maximum (minimum) lost profits per trade is the maximum (minimum) lost profit we observe in any of the 1,500 hypothetical trade times irrespective of the probability that this estimate accurately reflected the trading environment when the trade occurred.

[36] For ease of presentation, Panel C presents results with no adjustment to the NASDAQ SIP's timestamp; results using the adjusted timestamp are virtually identical to those shown in Panel C.

priced at the SIP NBBO, approximately 3% of shares traded on non-exchange venues and 2% of shares traded on exchanges had negative lost profits. Moreover, almost all of these instances cluster at -$0.01 lost profits per share.

In general, we attribute this distribution of lost profits to the fact that the NBBO will commonly change in response to serial buy (sell) orders so that late-arriving buy (sell) orders benefit from hitting stale quotes. For instance, in the Apple illustration above, the dark venue's delay in updating the NBBO to reflect the ISO buy order that induced a change in the Direct NBBO allowed the later-arriving non-exchange buy order to benefit by purchasing at the lower, stale NBO. This phenomenon also explains how even exchanges that use direct data feeds can be at risk of filling buy (sell) orders at bargain prices when marketable orders arrive as the market is in the process of moving higher (lower).

Reflecting this logic, the results in Panel C highlight the low likelihood that a marketable order priced at the SIP NBBO received poorer pricing than it would have received had it been priced at the Direct NBBO. For non-exchange trades, the results indicate that, among trades priced at the SIP NBBO, just 0.285% of shares traded had a positive measure of lost profits. This estimate drops to 0.112% for exchange trades. The low incidence of trades with a positive measure of lost profits relative to those with negative lost profits is consistent with our prediction that, because this strategy is dependent on the arrival of marketable orders, market makers pursuing this strategy of direct feed arbitrage will be able to exploit only a fraction of profitable NBBO dislocations.

Finally, regardless of whether a liquidity-taking trade benefited or suffered when priced at the SIP NBBO rather than the Direct NBBO, the results in Panel C highlight that the overall economic significance of either result is manifestly small. For instance, summing aggregate lost profits for all trades where lost profits is negative amounted to $3.8 million for non-exchange trades and $10.6 million for exchange trades. As such, the maximum benefit to liquidity takers of having their trades priced at the SIP NBBO rather than the Direct NBBO was just $14.4 million across both types of venues, notwithstanding the fact that the aggregate value of SIP-priced trades was $3.7 trillion. The aggregate cost to liquidity-taking orders for having their trades priced at the SIP NBBO rather than the Direct NBBO was even smaller. Specifically, the aggregate value of trades having a positive measure of lost profits was just $1,725,994. This figure drops to $530,134 for trades occurring on non-exchange venues, which constitute the venues where retail market makers (along with dark pools) would report their trades.

5.2 Estimated losses to liquidity providers

While the foregoing results suggest liquidity takers generally benefited when trades were priced at the SIP NBBO, the fact that there are two sides to every trade would suggest the reverse conclusion applied to liquidity providers. In our Apple illustration, for example, the buy order completed at the stale SIP NBO of $113.38 rather than at the new NBO of $113.40 meant the seller in the non-exchange venue who had posted the resting liquidity lost $0.02 per share by selling at the stale SIP NBO rather than selling at the Direct NBO. The mean measure of lost profits of approximately -$0.0002 per share accordingly highlights that to the extent trading at SIP prices adversely affects traders, these costs are more likely to be borne by liquidity providers than by liquidity takers.

Depending on the identity of the trader taking liquidity in these trades, this latter finding may point to the presence of the second form of direct feed arbitrage occurring in the market. In particular, we have largely assumed that marketable orders reflect uninformed order flow, such as orders submitted by retail investors. Our basic finding that liquidity takers benefit from being priced at the SIP NBBO, however, is in principle also consistent with claims made by those such as IEX and the Chicago Stock Exchange that HFT firms routinely engage in a separate type of direct feed arbitrage. Under this alternative strategy, HFT firms submit marketable orders to "pick off" stale quotes posted by liquidity providers to earn risk-free profits.

The sequence of Apple trades in Table 3 provides an example of how such a strategy might work in practice. After having secured a "buy" trade at $113.38 (the stale SIP NBO) rather than at $113.40 (the new NBO), the trader submitting the buy order need only sell at the new NBB of $113.39 to realize an immediate, risk-free profit of $0.01 per share (excluding trading fees). Because the ensuing four trades each reflected "sell" transactions at this price, our Apple example—and the results in Table 4 more generally—may simply reflect the strategic use of marketable orders by HFT firms to pick off stale limit orders posted in venues that use slow quote feeds for pricing orders that are pegged to the NBBO. Moreover, the trader would have done so without incurring any market risk. And as highlighted in the introductory quotation to this article, it is the risk-free nature of direct feed arbitrage that has loomed large in debates surrounding the appropriateness of allowing exchanges to sell direct data feeds and the related emergence of "speed bumps" at exchanges such as IEX. This focus on the risk-free character of HFT profits arising from quote reporting latencies similarly informs theoretical accounts of the trading strategy. As summarized by Wah and Wellman (2013: pp. 2-3), "By anticipating [the] future NBBO, an HFT algorithm can capitalize on cross-market disparities before they are reflected in the public price quote, in effect jumping ahead of incoming orders to pocket a small but sure profit." The risk-free character of HFT firms "sniping" stale quotes also distinguishes direct feed arbitrage from other forms of high-speed trading that target stale limit orders to make risky bets on price trends.[37]

To explore the extent to which this form of direct feed arbitrage occurs in our sample, we leverage the participant timestamp data and the fact that an HFT firm following such a strategy would need to make a pair of trades. To see how this works, consider trades immediately subsequent to those trades where trading at the SIP NBBO yielded more favorable pricing for the liquidity-taking order than trading at the Direct NBBO—that is, where the trade produced a negative measure of lost profits. For each of these potential first-leg trades, suppose the trade originated from an HFT firm submitting to a venue an immediate-or-cancel buy or sell order after having observed a change in the Direct NBBO. Because the trading strategy we test assumes risk-free profits from direct feed arbitrage, the success of this HFT strategy requires an off-setting second-leg trade, which one should be able to see in the data.

To execute this pairing strategy, we sort trades based on the participant timestamp and identify each potential first-leg trade based on whether it produced a negative measure of lost profits. We then search forward for a matching second-leg trade until a window of 1,000 microseconds from the first-leg trade timestamp has been exhausted. For a trade to match the first-leg trade, it must satisfy criteria with respect to both its trade direction and its price. In particular, for first-leg buy orders, we require a matching second-leg trade to be a marketable sell order at a price that is higher than the first-leg purchase price; conversely, for first-leg sell orders, we require a matching second-leg trade to be a marketable buy order at a price that is less than

[37] For instance, Harris and Schultz (1998) study "SOES bandits" who used NASDAQ's (now decommissioned) Small Order Execution System (SOES) to trade quickly in and out of positions. The SOES permitted immediate execution of trades; therefore, if market makers began updating quotes, a SOES bandit could use the system to enter a trade in hopes of hitting a slower market maker's quote that had yet to update. Harris and Schultz note that SOES bandits established positions "when they observe short-term trends" and close positions when "they feel a trend has run its course." In contrast to claims about HFT firms engaged in direct feed arbitrage, Harris and Schultz accordingly note that "SOES trading is risky" and that the strategy often resulted in sizeable losses.

the first-leg sales price. We impose a 1,000 microsecond trading window following each first-leg trade to ensure there is sufficient time for a trader to receive a trade confirmation on the first-leg trade before placing the second-leg trade at either an exchange or non-exchange venue.[38] For each second-leg match, we estimate the net profits from the two trades as (a) the trade size of the first-leg trade multiplied by (b) the difference between the second-leg price and the first-leg price where the first-leg trade was identified as a buy order and the difference between the first-leg price and the second-leg price where the first-leg trade was identified as a sell order. Where more than one second-leg trade could be matched to a first-leg trade, we used the second-leg trade price that maximized estimated net profits.

Before presenting our results, it is worth emphasizing that this simple empirical strategy almost certainly over-estimates—potentially by a wide margin—the actual incidence of second-leg matches. Among other things, for instance, our strategy disregards order size and transaction fees and focuses purely on identifying subsequent transactions that are priced higher (lower) than first-leg buy (sell) orders. Moreover, our approach seeks to identify matching second-leg trades independently for each first-leg trade, creating the possibility that the same second-leg trade can be matched to two different first-leg trades. Finally, our strategy also permits second-leg trades to occur on any venue, including both the venue of the first-leg trade and any non-exchange venue, even though the absence of displayed liquidity in non-exchange venues would introduce considerable risk for the trading strategy.[39]

---

[38] We suspect a 1,000 microsecond trading window is most likely too generous for first-leg transactions occurring on stock exchanges. For instance, a trader subscribing to exchanges' fastest fiber optic data feeds and co-located at NASDAQ would receive a trade confirmation of a first-leg trade occurring at the NYSE (the furthest exchange from NASDAQ) in approximately 200 microseconds based on the reporting latencies presented in the Internet Appendix, allowing it to execute a second-leg trade even at the NYSE in approximately 400 microseconds from the time of the first-leg trade.

[39] Even for second-leg trades aimed at hitting an exchange's displayed liquidity, this strategy would appear to involve non-trivial execution risk. As noted in Fox et al. (2015: p. 267), an HFT firm attempting to profit from this form of direct feed arbitrage "must be able to transact against the new best quote before anyone else can."

We present the results of this analysis in Table 5. To facilitate comparison with Panel C of Table 4, we divide the table separately for first-leg trades occurring in non-exchange and exchange venues. Because the trading strategy relies on first-leg trades having negative lost profits, the results in columns (2) and (3) reflect the share-weighted percentage of SIP-priced trades and the raw number of shares traded in non-exchange venues, respectively, that occurred with the negative lost-profits per share set forth in column (1). Columns (6) and (7) present the same figures for exchange trades. Despite the bias in favor of finding second-leg matches, the results in Table 5 reveal an extremely low incidence of matches between the first- and second-legs of this type of trading pairs. Of the 363 million shares traded in non-exchange venues that were traded at SIP prices producing negative lost profits, approximately 3.5% were matched with a second-leg trade for total net profits of just $84,417.[40] A slightly lower match rate applied to the nearly 967 million shares traded on exchanges at SIP prices that produced negative lost profits. For these shares, second-leg matches could be found for 2.65% of first-leg shares, resulting in net profits of $181,004. In all, we estimate total risk-free profits from this trading strategy to be less than $300,000 across the nearly $3.7 trillion of trades occurring at the SIP NBBO.

[Insert Table 5]

We next relax the requirement that a trader picking-off stale quotes priced at the SIP NBBO seeks to earn risk-free profits by using two marketable orders. The assumption of risk-free profits has played a large role in public discourse surrounding direct feed arbitrage; however, imposing a requirement that the second leg trade represents a market order occurring within 1,000 microseconds of the first trade may not realistically reflect HFT trading behavior if HFT

[40] As noted, we disregard trade size and calculate net profits on the assumption that the second-leg trade was sufficiently large to cover all first-leg shares.

firms also utilize passive orders. Accordingly, we re-estimate the net profits from picking off SIP-priced trades assuming that a trader's second-leg trade consists of a passive order. We further permit the second-leg trade to occur during the 1-second window following the first-leg trade, and we repeat the same analysis assuming a 1-minute window following the first-leg trade. For first-leg buy orders, we search within the relevant window for passive sell orders priced at the NBO having a price greater than the first-leg price; for first-leg sell orders, we search within the window for passive buy orders priced at the NBB having a price lower than the first-leg price. Where more than one second-leg match is identified during the window, we calculate the price of the second-leg trade as the size-weighted average price of all potential second-leg trades. In the event no eligible trades occur within the window, we assume the HFT firm exits the position by means of trading at the NBBO midpoint in effect at the end of the window.

The results of this alternative analysis appear in Table 6. In Panel A, we present estimates of net profits from this strategy assuming a trader liquidates its position within one-second of the first-leg trade; in Panel B, we present estimates of net profits assuming a one-minute window. As in Tables 4 and 5, we divide both panels according to whether a first-leg trade occurred in a non-exchange or exchange venue. Likewise, we again calculate net profits by disregarding the trade size of the second-leg match, and we assume that any trade that can be matched on price is sufficiently large to cover all shares of the first-leg trade. In Panel A, aggregate net profits for this strategy amount to approximately $1.69 million on first-leg trades occurring in non-exchange venues and $3.37 million on first-leg trades occurring on exchanges. These estimates increase to approximately $2.5 million and $6.0 million in Panel B, indicating that the total profitability from this strategy is increasing in the amount of time a trader has to execute the second-leg strategy. Share-weighted returns similarly increase from 0.01% for non-exchange trades and 0.006% for exchange trades to 0.06% and 0.06%, respectively. In general, widening the trading window increases the probability a trader will have its second-leg trade filled at the NBBO, thereby averting the need to exit the position at the midpoint of the NBBO. However, it also increases the market risk associated with the trader's open position, as reflected in the increase in the standard deviation of returns when moving from a 1-second window to a 1-minute window. A similar increase in risk can be seen in the increase in the 5th percentile return associated with the second-leg window. These latter considerations underscore how a trader adopting such a strategy is effectively assuming the market risk of a market-maker, albeit one whose trading profits are being subsidized by the amount of the negative lost profits obtained on its market orders. In this respect, for a trader targeting stale quotes at the SIP NBBO, one can view the aggregate negative lost profits of $14.4 million presented in Panel C of Table 4 as the core economic benefit of engaging in direct feed arbitrage for traders targeting passive orders in our sample.

[Insert Table 6]

6. Annual market-wide profits from direct feed arbitrage

A natural question is whether we can approximate how much money might be at stake from exploiting trades priced at stale SIP prices across the entire market over the course of a year, as opposed to just our sample period for just our sample stocks. Turning first to an annualized estimate of our sample, our sample period is from August 6, 2015 to June 30, 2016, or 228 trading days. As noted above, the aggregate value of trades having positive lost profits was approximately $1.7 million, which represents the total amount of money liquidity takers lost during this time period due to having their trades priced at the SIP NBBO rather than the Direct NBBO. There are 253 trading days in a calendar year, so we can estimate the annualized value of these trading losses for trades occurring on both exchange and non-exchange venues at approximately $1,915,248 (i.e., 253/228 x $1,725,994). Focusing on trades in non-exchange venues (which includes retail market makers), the annualized amount of positive lost profits would be just $588,263 (i.e., 253/228 x $530,134). This figure represents our estimate of the aggregate amount liquidity takers in non-exchange venues lost by having their trades priced at the SIP NBBO rather than the Direct NBBO when trading in the Dow Jones 30 over the course of a year. To the extent these losses reflected profits to retail market makers, this amount would represent an outer estimate of the annual gross profits to retail market makers of exploiting quote reporting latencies given that non-exchange trades include trades made by retail market makers as well as those made in dark pools.

Turning to the gross profits to liquidity takers from picking off stale quotes at the SIP NBBO, as shown in Panel C of Table 4, we estimated these during our sample period to be approximately $14.4 million across all trades on exchange and non-exchange venues. While traders would still need to monetize these gains through a second-leg trade, we estimate the annualized value of these gross profits using the same approach, arriving at $15,996,725 ($14,416,021 x 253/228). This figure aggregates trades made both on and off exchanges, which may overstate the profitability of this strategy if traders on exchanges are less likely to use orders pegged to the NBBO. Focusing on those trades that occur on non-exchange venues, Panel C of Table 4 reports gross profits to liquidity takers of $3,845,206, implying annualized gross profits of $4,266,830.

A second consideration is that the Dow Jones 30 covers only 30 stocks, and most of these stocks have quoted spreads at or near a penny. These small spreads diminish the profitability of these two forms of direct feed arbitrage to the extent average dislocations are likely to be at or near $0.01 per trade. Accordingly, to examine the robustness of our results, we re-estimate the annual profitability of direct feed arbitrage to liquidity providers and to liquidity takers for the entire U.S. listed equities market.

Because there are nearly 8,000 exchanged-traded stocks observed over this time period, we estimate these figures using a computationally efficient process that leverages the fact that a small number of stocks account for a large volume of shares traded.[41] In particular, the top-traded 257 stocks correspond to half the trading volume during our sample period, and the top-traded 872 stocks correspond to three-quarters of the trading volume during our sample period. By measuring the profitability of direct feed arbitrage strategies within these two groups during our sample period, we can obtain market-wide estimates by scaling the actual measure by a factor of 2 and 4/3, respectively. Finally, we can convert these estimates to annualized figures by multiplying by 253/228, as noted previously.

We present the results in Table 7. As shown in column (1), there were approximately $25.9 trillion in trades during our sample period in the 257 most traded securities. Of these, approximately $19.1 trillion were priced at the SIP NBBO. Focusing on the top 872 traded securities, these figures are $42.1 trillion and $30.2 trillion, respectively. Grossing up these figures by 2 and 4/3, respectively, and scaling to 12 months yields annual market-wide estimates of approximately $57-62 trillion in total securities traded, of which approximately $42-45 trillion were priced at the SIP NBBO.

[Insert Table 7]

[41] We ignore stocks with any suffixes on their trading symbol. Stocks with no suffixes correspond to 98.5% of trading volume over our study period, so this is of little consequence for the calculations we report here. Ignoring suffixes is computationally advantageous because of the database index structure.

Of these SIP-priced trades, we first calculate aggregate profits to liquidity providers engaged in direct feed arbitrage. Because gains to liquidity providers constitute losses to liquidity takers, we estimate these profits as the total trading loses for liquidity taking orders that were priced at the SIP NBBO rather than at the Direct NBBO for all trades within the top 257 and 872 stocks. In light of our definition of lost profits, we accomplish this by summing lost profits for every trade observed where lost profits were greater than zero. This has the effect of aggregating trades where the liquidity taker could have earned just a penny more per share by having the trade priced at the Direct NBBO with those trades where the dislocation between the SIP NBBO and Direct NBBO was far greater than a penny. (For instance, we observe occasional dislocations of more than $1.00 per share traded.)

As shown in Table 7, the maximum amount of profits to liquidity providers from this form of direct feed arbitrage for trades among the top 257 stocks is approximately $8 million for exchange trades and $3.5 million for non-exchange trades. Multiplying these figures by 2 and annualizing yields annual market-wide estimates of approximately $17.8 million and $7.7 million, respectively. Applying this approach to the top 872 stocks, our estimates of these annual market-wide profits to liquidity providers is $21.6 million for exchange trades and $9 million for non-exchange trades. Because concerns about this form of direct feed arbitrage have focused on retail market makers, we view the estimates for non-exchange trades as particularly relevant for assessing the profitability of this strategy in today's markets. Moreover, using data from FINRA's Order Audit Trail System (OATS), Tuttle (2014) estimates that non-ATS trades reported to FINRA comprise 60% of non-exchange trades. To the extent this reflects the fraction of off-exchange trades at risk to this form of direct feed arbitrage, our market-wide estimate of annual profits decreases to a range of approximately $4.6 million to $5.4 million.

Turning to the gross profits available to liquidity takers from picking-off stale quotes priced at the SIP NBBO, we present our estimates of the gross arbitrage profits from this strategy in the final two rows of Table 7. Again, given our definition of lost profits, we calculate the maximum available profits from this strategy by summing the total value of lost profits for every trade observed where lost profits were less than zero. In general, this approach assumes a single trader is capable of picking off *every* quote in the market that is priced at the stale SIP NBBO and doing so only in those cases where it resulted in a profit to the liquidity taking trader.

As shown in the table, the aggregate value of these gross profits to liquidity-takers for the top 257 stocks during our sample period was approximately $70 million for exchange trades and $35 million for non-exchange trades. For the top 872 stocks, these figures were approximately $145 million and $56 million, respectively. Grossing up all figures by 2 and 4/3, respectively, and annualizing yields annual market-wide estimates of total gross profits to liquidity takers of between $155 million and $214 million for exchange trades and $77 million to $83 million for non-exchange trades. Because concerns about HFT sniping of stale quotes have focused on dark pools that peg orders to the NBBO, these profit opportunities may be confined to the subset of non-exchange trades occurring in dark pools. Turning again to estimates from Tuttle (2014), estimating that approximately 40% of non-exchange trades occur in venues classified as alternative trading systems, we approximate the maximum annual gross profits to liquidity takers from this form of Direct feed arbitrage as $46 million to $50 million. Again, these are gross profits to liquidity takers. Therefore, to the extent a liquidity taker seeks to monetize these gains, this estimate would be further reduced by the costs of the off-setting trade, as well as trading commissions and exchange fees.

Finally, we note that Virtu Financial—a single firm utilizing HFT subject to Exchange Act reporting obligations—had in its 2016 fiscal year "communication and data processing" costs of $71 million and "brokerage, exchange and clearance fees" of $221 million on trading revenue of $665 million.[42] In combination with our findings in Table 7, these disclosures further confirm that the profitability of HFT arises from sources other than these two forms of direct feed arbitrage, notwithstanding their prominence in contemporary debates concerning market structure.

7. Conclusion

Using recently released data from the two SIPs, we examine claims that high-frequency trading (HFT) firms use direct feeds to exploit traders who rely on SIP prices. Across $3.7 trillion of trades in the Dow Jones 30 from August 6, 2015 through June 30, 2016, we find that liquidity-taking orders gain on average $0.0002 per share when priced at the SIP-reported NBBO rather than the Direct NBBO, which reflects the NBBO calculated without reporting latencies. This finding reflects the fact that dislocations between the SIP NBBO and Direct NBBO can occur in response to serial buy and sell orders, allowing late-arriving market orders to benefit if they are priced at an NBBO that has yet to reflect the new trading interest. To the extent this is the case, concerns about the latency of SIP prices would seem more relevant for traders providing liquidity in venues that price limit orders by pegging them to SIP prices. Yet while these concerns are consistent with claims that HFT firms pick off mispriced limit orders in these venues, we estimate gross profits from this strategy to be just $14.4 million across all trades in the Dow Jones 30 during our sample period.

[42] Virtu Financial, Inc, Form 10-K for the Fiscal Year Ending December 31, 2016, available at https://www.sec.gov/Archives/edgar/data/1592386/000155837017001698/virt-20161231x10k.htm.

Overall, our findings reveal that transacting at prices that match the SIP NBBO can benefit liquidity takers to the detriment of liquidity providers; however, the incidence of these gains and losses between these two forms of trading interest appears to be primarily a product of chance rather than of HFT design. Because our data commence in August 2015, however, we emphasize that these findings may very well reflect a new market environment in which the profitability of direct feed arbitrage is less than in the past. Among other things, for instance, the increasing processing speed of the SIPs, enhanced regulatory scrutiny of HFT, and the emergence of venues such as IEX that shield traders from HFT trading may have simply made these SIP-oriented arbitrage strategies increasingly infeasible.

Even with this caveat, our findings yield new insights about the socially costly arms-race for trading speed described in Budish, Cramton & Shim (2015). While our findings are consistent with the incentive of liquidity providers to invest in low latency trading data to avoid being adversely selected through direct feed arbitrage, our results suggest these incentives play, at most, a subsidiary role in promoting this arms-race. To the extent traders participate in this arms race, the primary incentives today would accordingly appear to rest outside a desire either to exploit direct feed arbitrage or to avoid the costs of trading at stale SIP prices.

References

Angrist, J., Pischke, J., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton.

Bartlett, R., McCrary, J., forthcoming. Dark trading at the midpoint: pricing rules, SEC enforcement policy, and latency arbitrage. Journal of Law, Finance, and Accounting.

Brogaard, J., Hagstromer, B., Norden, L., Riordan, R., 2015. Trading fast and slow: colocation and liquidity. Review of Financial Studies 28, 3407–43.

Budish, E., Cramton, P., Shim, J., 2015. The high-frequency trading arms race: frequent batch auctions as a market design response. Quarterly Journal of Economics 130, 1547-1621.

Cameron, A. C., Gelbach, J., Miller D., 2008. Bootstrap-based improvements for inference with clustered errors. Review of Economics and Statistics 90, 414-427.

Ding, S., Hanna, J., Hendershott, T., 2014. How slow is the NBBO? A comparison with direct exchange feeds. Financial Review 49, 313–332.

Egginton, J., Van Ness, B., Van Ness, R., 2016. Quote stuffing. Financial Management 45, 583-608.

Foucault, T., Hombert, J., Rosu, I., 2016. News trading and speed. Journal of Finance 61, 335–82.

Fox, M., Glosten, L., Rautererberg, G., 2015 The new stock market: sense and nonsense. Duke Law Journal 65, 191-277.

Harris, J., Schulz, P., 1998. The trading profits of SOES bandits. Journal of Financial Economics 50, 39-62.

Hendershott, T., Moulton, P., 2011. Automation, speed, and stock market quality: The NYSE's hybrid. Journal of Financial Markets 14, 568–604.

Hoffmann, P., 2014. A dynamic limit order market with fast and slow traders. Journal of Financial Economics 113, 159–69.

Holden, C., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions. Journal of Finance 69, 1747-1785.

Jones, C., 2018. Understanding the market for U.S. equity market data. Manuscript, Columbia Business School.

Jovanovic, B., Menkveld, A., 2012. Middlemen in limit order markets. Manuscript, University of Amsterdam.

Levinson, C., 2016. U.S. investigates market-making operations of Citadel, KCG. Reuters, May 10, 2016.

Lewis, M., 2014. Flash Boys: A Wall Street Revolt, W.W. Norton & Co.

Lubotsky, D., Wittenberg, M., 2006. Interpretation of regressions with multiple proxies. Review of Economics and Statistics 88, 549-562.

MacKinnon, J., 2015. Wild cluster bootstrap confidence intervals. L'Actualité Économique 91, 11-33.

Menkveld, A.J., 2013. High frequency trading and the new market makers. Journal of Financial Markets 16, 712-740.

Menkveld, A.J., Zoican, M., 2017. Need for speed? Exchange latency and liquidity. Review of Financial Studies 30, 1188-1228.

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

Roodman, D., MacKinnon, J., Nielsen, M.O., Webb, M., 2019. Fast and wild: bootstrap inference in Stata using boottest, The Stata Journal 19, 4-60.

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

Tuttle, L., 2014. Description of non-ATS OTC trading in national market system stocks. Manuscript, Securities Exchange Commission.

Wah, E., Wellman, M., 2013. Latency arbitrage, market fragmentation, and efficiency: A two-market model. Proceedings of the fourteenth ACM Conference.

White, M.J., 2015. Testimony on "Examining the SEC's Agenda, Operations and FY 2016 Budget Request", Before the United States House of Representatives Committee on Financial Services, available at https://www.sec.gov/news/testimony/2015-ts032415mjw.html.

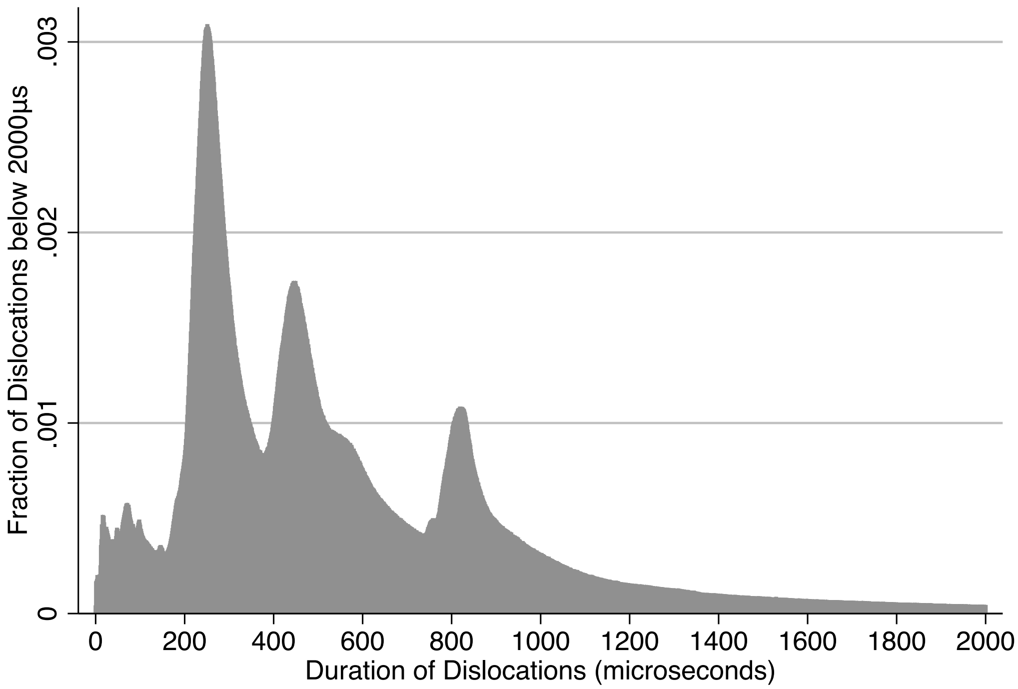

Figure 1: Distribution of dislocation durations. Histogram of the duration in microseconds of observed price dislocations between the SIP NBB and the Direct NBB.

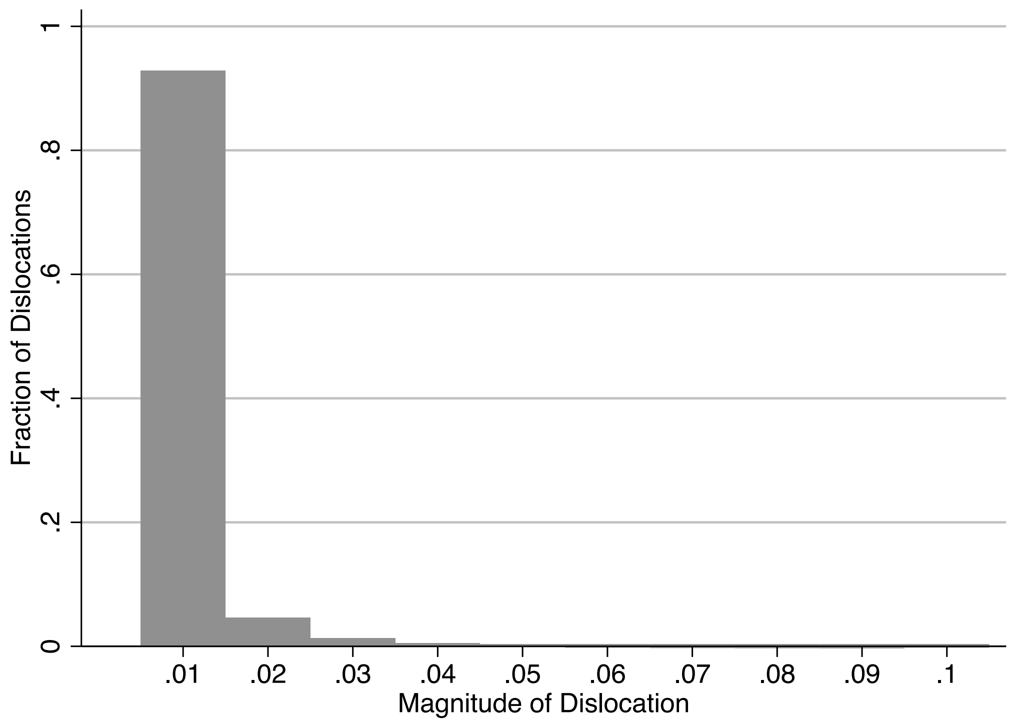

Figure 2: Distribution of dislocation magnitudes. Histogram of the economic magnitude (in dollars) of observed price dislocations between the SIP NBB and the Direct NBB.

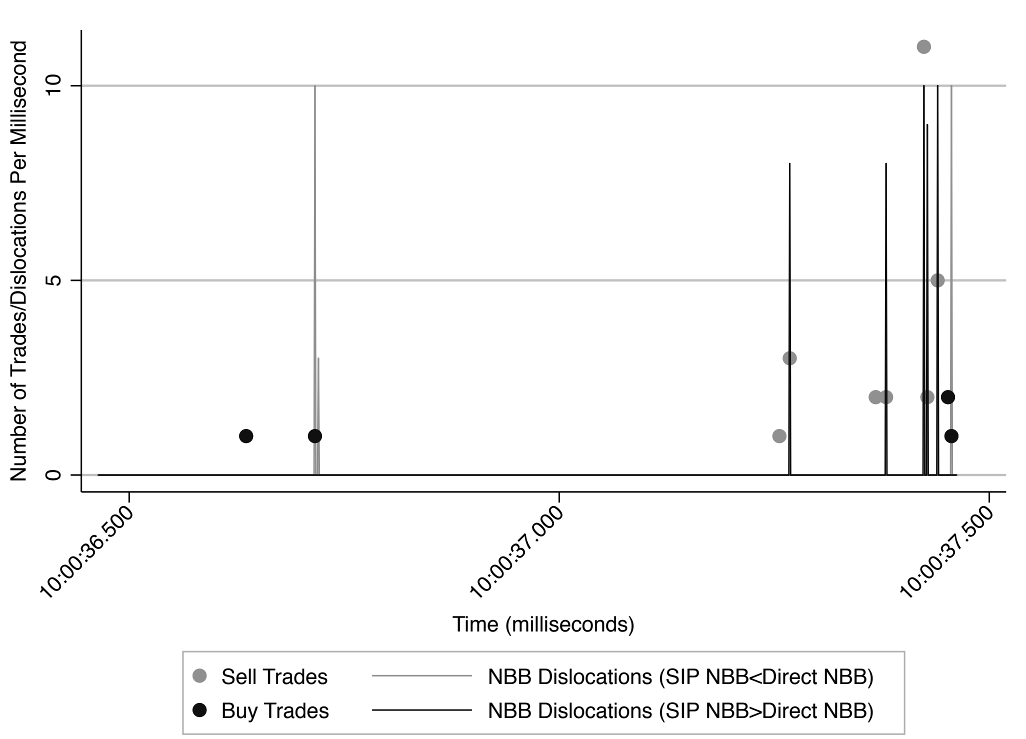

Figure 3: Trading and the direction of NBBO dislocations. One second trading window in Apple on November 13, 2015 showing observed trades in a millisecond (sell orders marked as grey circles; buy orders, black circles) against (i) observed NBB dislocations per millisecond (grey line) where it would be profitable for a liquidity provider to fill an active sell order at the stale SIP NBB (i.e., when the SIP NBB<Direct NBB) and (ii) NBB dislocations per millisecond (black line) where it would be unprofitable for a liquidity provider to fill an active sell order at the stale SIP NBB (i.e., when the SIP NBB>Direct NBB).

Table 1: SIP Processing Times

This table presents the processing times reported by the two SIPs for trade and quote data. Latencies are measured from the moment a trade or quote report is received by a SIP to the moment the SIP completes processing the record. Data for Tape A and Tape B securities can be found at https://www.nyse.com/publicdocs/ctaplan/notifications/trader-update/CTA%20SIP%202Q16%20Consolidated%20Data%20Operating%20Metrics%20Report%20(7-13-16%20Update).pdf. Data for Tape C securities can be found at http://www.utpplan.com/DOC/UTP%202015-Q4%20Stats%20with%20Processor%20Stats.pdf.

Panel A: SIP Processing Time for Trades

	Tape A&B Trade metrics						Tape C Trade metrics					
	Peak Messages per 100 Milliseconds (thousands)	Capacity Messages per 100 Milliseconds (thousands)	Capacity vs Peak Ratio	Average Latency	Median Latency	90th percentile latency	Peak Messages per 100 Milliseconds (thousands)	Capacity Messages per 100 Milliseconds (thousands)	Capacity vs Peak Ratio	Average Latency	Median Latency	90th percentile latency
1q14	21.80	60.00	2.75	0.51	n/a	0.71	19.30	39.40	2.04	1.32	1.25	1.67
2q14	23.50	60.00	2.55	0.51	n/a	0.66	20.50	39.40	1.92	0.82	0.54	0.74
3q14	22.70	65.00	2.86	0.51	n/a	0.66	17.60	48.50	2.76	0.59	0.49	0.68
4q14	24.20	65.00	2.69	0.45	n/a	0.60	19.40	48.50	2.50	0.59	0.49	0.67
1q15	22.10	70.00	3.17	0.45	n/a	0.59	20.10	68.70	3.42	0.53	0.45	0.60
2q15	31.80	70.00	2.20	0.34	n/a	0.43	22.80	132.80	5.82	0.54	0.46	0.62
3q15	27.10	75.00	2.77	0.32	0.24	0.41	16.10	132.80	8.25	0.58	0.47	0.64
4q15	43.70	75.00	1.72	0.31	0.24	0.41	18.60	132.80	7.14	0.62	0.47	0.66
1q16	42.40	86.00	2.03	0.33	0.25	0.43	19.40	132.80	6.85	0.77	0.49	0.76
2q16	37.40	96.00	2.57	0.34	0.24	0.45	28.20	132.80	4.71	0.63	0.48	0.68
mean	29.67	72.20	2.53	0.41	0.24	0.54	20.20	90.85	4.54	0.70	0.56	0.77

Panel B: SIP Processing Time for Quotes

	Tape A&B Trade metrics						Tape C Trade metrics					
	Peak Messages per 100 Milliseconds (thousands)	Capacity Messages per 100 Milliseconds (thousands)	Capacity vs Peak Ratio	Average Latency	Median Latency	90th percentile latency	Peak Messages per 100 Milliseconds (thousands)	Capacity Messages per 100 Milliseconds (thousands)	Capacity vs Peak Ratio	Average Latency	Median Latency	90th percentile latency
1q14	121.10	300.00	2.48	0.45	n/a	0.90	51.50	70.70	1.37	1.20	1.08	1.62
2q14	131.70	300.00	2.28	0.44	n/a	0.76	51.20	70.70	1.38	0.69	0.48	0.70
3q14	121.10	325.00	2.68	0.45	n/a	0.88	49.80	83.80	1.68	0.59	0.43	0.79
4q14	141.80	325.00	2.29	0.41	n/a	0.75	95.40	83.80	0.88	0.55	0.43	0.66
1q15	146.40	350.00	2.39	0.39	n/a	0.68	85.50	166.90	1.95	0.50	0.44	0.62
2q15	142.60	350.00	2.45	0.46	n/a	1.02	48.00	215.00	4.48	0.65	0.44	0.69
3q15	158.40	375.00	2.37	0.51	0.23	1.13	37.10	215.00	5.80	0.80	0.45	0.79
4q15	162.30	375.00	2.31	0.44	0.21	0.93	41.00	215.00	5.24	0.81	0.45	0.81
1q16	163.30	392.00	2.40	0.49	0.22	1.08	60.10	215.00	3.58	0.92	0.47	1.04
2q16	168.40	400.00	2.38	0.49	0.22	1.09	83.00	215.00	2.59	0.80	0.46	0.93
mean	145.71	349.20	2.40	0.45	0.22	0.92	60.26	155.09	2.90	0.75	0.51	0.87

Table 2: Modeling NBBO Dislocations and Their Profit Potential

This table presents regression results for models of NBBO dislocations (Panels A and B) and potential profitability from exploiting NBBO dislocations (Panel C). The dependent variable in Panels A and B is the natural log of the total number of dislocations observed for a stock on a trading day. The independent variables are daily measures for the natural log of: the average of the inside depth on the bid and the ask (*Log Inside Depth*)*,* the total number of BBO updates for a stock (*Log Quotes*), the total number of trades on exchanges (*Log Exchange Trades*), the total number of trades on non-exchange venues (*Log Off Exchange Trades*), the average trade size (*Log Average Trade Size*), daily volatility (*Log Volatility*), and average quoted spread. *Fraction of ISO* is the fraction of daily trades marked as Intermarket Sweep Orders. The dependent variable in Panel C is the natural log of aggregate microsecond pennies, which we define in the text. Independent variables in Panel C are the same as those used in Panels A and B. Standard errors clustered on stock are given in parentheses. Confidence intervals based on inverting the wild cluster percentile-t bootstrap two-sided *p*-values are given in brackets.

Panel A: Log Number of Dislocations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Inside Depth	-0.506	-0.791	-0.827	-0.826	-0.919	-0.891	-0.880	-0.878
	(0.044)	(0.025)	(0.022)	(0.022)	(0.021)	(0.020)	(0.022)	(0.024)
	[-0.598 ; -0.392]	[-0.887 ; -0.736]	[-0.889 ; -0.777]	[-0.893 ; -0.775]	[-0.970 ; -0.873]	[-0.939 ; -0.849]	[-0.935 ; -0.832]	[-0.938 ; -0.827]
Log Quotes		0.931	0.585	0.583	0.609	0.595	0.622	0.622
		(0.034)	(0.048)	(0.053)	(0.047)	(0.047)	(0.045)	(0.045)
		[0.839 ; 1.001]	[0.480 ; 0.696]	[0.467 ; 0.703]	[0.504 ; 0.720]	[0.490 ; 0.704]	[0.521 ; 0.728]	[0.521 ; 0.728]
Log Exchange Trades			0.454	0.467	0.467	0.438	0.422	0.415
			(0.052)	(0.105)	(0.088)	(0.085)	(0.085)	(0.086)
			[0.340 ; 0.567]	[0.238 ; 0.703]	[0.270 ; 0.672]	[0.249 ; 0.638]	[0.232 ; 0.622]	[0.224 ; 0.614]
Log Off Exchange Trades				-0.011	-0.076	-0.020	-0.022	-0.014
				(0.067)	(0.057)	(0.053)	(0.054)	(0.058)
				[-0.162 ; 0.134]	[-0.204 ; 0.048]	[-0.142 ; 0.096]	[-0.148 ; 0.098]	[-0.144 ; 0.112]
Log Average Trade Size					0.522	0.458	0.470	0.469
					(0.088)	(0.079)	(0.074)	(0.074)
					[0.288 ; 0.708]	[0.250 ; 0.625]	[0.265 ; 0.620]	[0.265 ; 0.619]
Log Volatility						0.111	0.101	0.101
						(0.019)	(0.017)	(0.017)
						[0.072 ; 0.152]	[0.068 ; 0.135]	[0.068 ; 0.136]
Log Spread							0.057	0.057
							(0.035)	(0.035)
							[-0.029 ; 0.148]	[-0.029 ; 0.148]
Fraction ISO								0.071
								(0.218)
								[-0.389 ; 0.541]
Constant	10.34	-1.494	-1.523	-1.527	-3.584	-2.271	-2.402	-2.427
	(0.164)	(0.430)	(0.246)	(0.249)	(0.317)	(0.318)	(0.302)	(0.309)
Observations	6,836	6,836	6,836	6,836	6,836	6,836	6,836	6,836
R^2	0.547	0.932	0.954	0.954	0.959	0.962	0.962	0.962
Stock Effects?	No	No	No	No	No	No	No	No
Day Effects?	No	No	No	No	No	No	No	No
Number of clusters	30	30	30	30	30	30	30	30

Panel B: Log Number of Dislocations with Stock and Day Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Inside Depth	-0.596	-0.767	-0.843	-0.843	-0.871	-0.854	-0.843	-0.845
	(0.079)	(0.042)	(0.032)	(0.032)	(0.030)	(0.029)	(0.029)	(0.029)
	[-0.789 ; -0.435]	[-0.875 ; -0.685]	[-0.926 ; -0.783]	[-0.926 ; -0.782]	[-0.942 ; -0.813]	[-0.923 ; -0.797]	[-0.905 ; -0.785]	[-0.907 ; -0.786]
Log Quotes		0.925	0.703	0.702	0.706	0.692	0.699	0.698
		(0.046)	(0.037)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
		[0.825 ; 1.024]	[0.624 ; 0.782]	[0.617 ; 0.789]	[0.621 ; 0.793]	[0.608 ; 0.779]	[0.614 ; 0.786]	[0.613 ; 0.784]
Log Exchange Trades			0.425	0.427	0.415	0.423	0.427	0.442
			(0.025)	(0.053)	(0.054)	(0.054)	(0.053)	(0.053)
			[0.373 ; 0.475]	[0.312 ; 0.536]	[0.299 ; 0.525]	[0.306 ; 0.535]	[0.313 ; 0.536]	[0.328 ; 0.552]
Log Off Exchange Trades				-0.002	-0.020	-0.015	-0.019	-0.036
				(0.043)	(0.040)	(0.039)	(0.038)	(0.039)
				[-0.091 ; 0.088]	[-0.103 ; 0.063]	[-0.096 ; 0.066]	[-0.099 ; 0.061]	[-0.119 ; 0.047]
Log Average Trade Size					0.159	0.153	0.156	0.160
					(0.037)	(0.036)	(0.038)	(0.038)
					[0.089 ; 0.240]	[0.083 ; 0.231]	[0.083 ; 0.241]	[0.087 ; 0.244]
Log Volatility						0.064	0.063	0.062
						(0.013)	(0.013)	(0.013)
						[0.038 ; 0.092]	[0.035 ; 0.091]	[0.035 ; 0.091]
Log Spread							0.048	0.046
							(0.040)	(0.040)
							[-0.029 ; 0.154]	[-0.033 ; 0.153]
Fraction ISO								-0.173
								(0.150)
								[-0.486 ; 0.139]
Constant	11.185	-1.441	-2.801	-2.801	-3.253	-2.566	-2.517	-2.453
	(0.223)	(0.697)	(0.384)	(0.386)	(0.403)	(0.448)	(0.468)	(0.476)
Observations	6,836	6,836	6,836	6,836	6,836	6,836	6,836	6,836
R^2	0.927	0.966	0.975	0.975	0.976	0.976	0.976	0.976
Stock Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	30	30	30	30	30	30	30	30

Panel C. Log Aggregate Microsecond-Pennies

	(1)	(2)
Log Inside Depth	-0.538	-0.595
	(0.034)	(0.040)
	[-0.625 ; -0.465]	[-0.676 ; -0.505]
Log Quotes	0.615	0.655
	(0.072)	(0.045)
	[0.454 ; 0.787]	[0.560 ; 0.751]
Log Exchange Trades	0.693	0.395
	(0.116)	(0.047)
	[0.444 ; 0.957]	[0.301 ; 0.494]
Log Off Exchange Trades	-0.219	-0.092
	(0.079)	(0.037)
	[-0.395 ; -0.044]	[-0.169 ; -0.017]
Log Average Trade Size	-0.040	0.082
	(0.097)	(0.051)
	[-0.284 ; 0.151]	[-0.024 ; 0.184]
Log Volatility	0.142	0.080
	(0.027)	(0.013)
	[0.083 ; 0.200]	[0.053 ; 0.107]
Log Spread	0.585	0.443
	(0.058)	(0.054)
	[0.458 ; 0.739]	[0.313 ; 0.550]
Fraction ISO	0.414	-0.311
	(0.363)	(0.168)
	[-0.382 ; 1.161]	[-0.657 ; 0.028]
Constant	12.141	12.401
	(0.619)	(0.547)
Observations	6,836	6,836
R^2	0.854	0.920
Stock Effects?	No	Yes
Day Effects?	No	Yes
Number of clusters	30	30

Table 3: Apple Trades Ordered by Participant Timestamp, November 13, 2015

This table illustrates how trades in the sample are matched to the prevailing SIP NBBO and Direct NBBO. *Participant Timestamp* is the time in microseconds at which a venue reports executing a trade. *SIP Timestamp* is the time the SIP placed the trade report on its multicast line for dissemination, which incorporates transit and SIP-processing latencies. The *NBB Direct* and *NBO Direct* are calculated using the participant timestamp for quote updates, which reflects the time an exchange matching engine processed a quote. The *NBB SIP* and *NBO SIP* are calculated using the traditional SIP timestamp assigned to quotes, which reflects the time a SIP disseminated a quote update. The Direct NBBO is matched to each trade based on the participant timestamp of the trade and the participant timestamp of the Direct NBBO. The SIP NBBO is matched to each trade based on the participant timestamp of a trade and the SIP timestamp of the SIP NBBO. Columns (8)-(10) report the trade condition, the exchange code for where the trade occurred, and the trade size, respectively, as reported on the trade record. (Exchange codes are: Z=BATS; K=Direct Edge A; Q=Nasdaq; P=NYSE Arca; D=FINRA TRF). Column (12) indicates the trade direction. *SIP Priced* in Column (13) is coded as 1 where trade price equals the SIP NBBO and zero otherwise. *Lost profits* in Column (13) are calculated where SIP Priced=1 as the Direct NBB minus SIP NBB for sell orders, and the SIP NBO minus the Direct NBO for buy orders.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Trade No.	Participant Timestamp	SIP Timestamp	NBB Direct	NBO Direct	NBB SIP	NBO SIP	Trade Cond.	Exch.	Trade Price	Trade Size	Buy Order	SIP Priced	Lost Profits
1	11:37:47.464119	11:37:47.464616	113.37	113.38	113.37	113.38	@F	Z	113.38	2,500	1	1	0
2	11:37:47.464119	11:37:47.464706	113.37	113.38	113.37	113.38	@F	Z	113.38	100	1	1	0
3	11:37:47.464119	11:37:47.464762	113.37	113.38	113.37	113.38	@F	Z	113.39	100	1	0	
4	11:37:47.464119	11:37:47.464792	113.37	113.38	113.37	113.38	@F	Z	113.39	100	1	0	
5	11:37:47.464119	11:37:47.464848	113.37	113.38	113.37	113.38	@F	Z	113.39	200	1	0	
6	11:37:47.464135	11:37:47.464743	113.37	113.38	113.37	113.38	@F	K	113.38	100	1	1	0
7	11:37:47.464135	11:37:47.464820	113.37	113.38	113.37	113.38	@F	K	113.38	200	1	1	0
8	11:37:47.464135	11:37:47.464861	113.37	113.38	113.37	113.38	@F	K	113.39	100	1	0	
9	11:37:47.464135	11:37:47.464889	113.37	113.38	113.37	113.38	@F	K	113.39	100	1	0	
10	11:37:47.464135	11:37:47.464916	113.37	113.38	113.37	113.38	@F	K	113.39	100	1	0	
11	11:37:47.464298	11:37:47.464673	113.37	113.38	113.37	113.38	@F	Q	113.38	100	1	1	0
12	11:37:47.464298	11:37:47.464727	113.37	113.38	113.37	113.38	@F	Q	113.38	100	1	1	0
13	11:37:47.464298	11:37:47.464777	113.37	113.38	113.37	113.38	@F	Q	113.38	100	1	1	0
14	11:37:47.464298	11:37:47.464806	113.37	113.38	113.37	113.38	@F	Q	113.38	100	1	1	0
15	11:37:47.464315	11:37:47.464834	113.37	113.38	113.37	113.38	@F	Q	113.38	200	1	1	0
16	11:37:47.464315	11:37:47.464875	113.37	113.38	113.37	113.38	@F	Q	113.39	100	1	0	
17	11:37:47.464315	11:37:47.464903	113.37	113.38	113.37	113.38	@F	Q	113.39	100	1	0	
18	11:37:47.464315	11:37:47.464929	113.37	113.38	113.37	113.38	@F	Q	113.39	100	1	0	
19	11:37:47.464315	11:37:47.464943	113.37	113.38	113.37	113.38	@F	Q	113.39	100	1	0	
20	11:37:47.464360	11:37:47.465298	113.37	113.38	113.37	113.38	@F	P	113.38	100	1	1	0
21	11:37:47.464360	11:37:47.465320	113.37	113.38	113.37	113.38	@F	P	113.38	100	1	1	0
22	11:37:47.464360	11:37:47.465337	113.37	113.38	113.37	113.38	@F I	P	113.38	73	1	1	0
23	11:37:47.464360	11:37:47.465352	113.37	113.38	113.37	113.38	@F	P	113.38	200	1	1	0
24	11:37:47.464397	11:37:47.465380	113.37	113.39	113.37	113.38	@F	P	113.39	500	1	0	
25	11:37:47.464397	11:37:47.465423	113.37	113.39	113.37	113.38	@F	P	113.39	100	1	0	
26	11:37:47.464397	11:37:47.465441	113.37	113.39	113.37	113.38	@F	P	113.39	100	1	0	
27	11:37:47.464397	11:37:47.465456	113.37	113.39	113.37	113.38	@F	P	113.39	100	1	0	
28	11:37:47.464397	11:37:47.465472	113.37	113.39	113.37	113.38	@F	P	113.39	100	1	0	
29	11:37:47.464397	11:37:47.465487	113.37	113.39	113.37	113.38	@F	P	113.39	100	1	0	
30	11:37:47.464397	11:37:47.465502	113.37	113.39	113.37	113.38	@F I	P	113.39	72	1	0	
31	**11:37:47.465000**	**11:37:47.467422**	**113.39**	**113.40**	**113.37**	**113.38**		**D**	**113.38**	**100**	**1**	**1**	**-0.02**
32	11:37:47.466000	11:37:47.511814	113.39	113.40	113.39	113.40		D	113.39	100	0	1	0
33	11:37:47.466018	11:37:47.466459	113.39	113.40	113.39	113.40		Z	113.39	100	0	1	0
34	11:37:47.475000	11:37:47.478795	113.39	113.40	113.39	113.40		D	113.40	245	1	1	0
35	11:37:47.479000	11:37:47.482618	113.39	113.40	113.39	113.40		D	113.40	805	1	1	0

Table 4: Gross Profits to Liquidity Takers and Liquidity Providers from Direct Feed Arbitrage

This table presents estimates of gains and losses to liquidity takers and to liquidity providers for transacting at prices equal to the SIP NBBO rather than the Direct NBBO across trades in the Dow Jones 30. Panel A presents the fraction of trades (weighted by trade size) in the sample where the trade price matched the SIP NBBO and the aggregate transaction value of these SIP-priced trades. Panel B presents estimates of the mean amount of lost profit per share on all SIP-priced trades (weighted by trade size) that liquidity takers experienced by having their trades priced at the SIP NBBO rather than at the Direct NBBO. Lost profits are defined as the Direct NBB minus SIP NBB for sell orders, and the SIP NBO minus the Direct NBO for buy orders. Therefore, positive measures reflect losses to liquidity takers and gains to liquidity providers; negative measures reflect gains to liquidity takers and losses to liquidity providers. Panel C presents the distribution of lost profits per share across all trades (weighted by trade size) that were priced at the SIP NBBO and provides estimates of the aggregate gross gains to liquidity takers and liquidity providers. In Panel B, robust standard errors are in parentheses. *** $p < .01$, ** $p < .05$, * $p < .1$

Panel A: Fraction of Trades Priced at the SIP NBBO

(1)	(2)	(3)
Exchange	*% of Trades Matching SIP NBBO*	*Transaction Value of SIP-Priced Trades*
NYSE	87.31%	$718,312,000,000
NYSE MKT	80.31%	$4,331,127,627
NYSE Arca	89.87%	$400,242,000,000
Nasdaq OMX BX	90.47%	$89,912,200,000
NASDAQ OMX PSX	93.45%	$58,900,900,000
NASDAQ	91.21%	$730,940,000,000
BATS	88.31%	$327,115,000,000
BATS Y	92.11%	$164,237,000,000
Direct Edge A	94.34%	$115,206,000,000
Direct Edge X	93.43%	$390,303,000,000
Chicago	11.60%	$8,708,697,546
NSX	95.11%	$113,532,078
FINRA TRF	45.46%	$740,324,000,000
All venues:	73.56%	$3,748,645,457,251
All Exchanges:	87.37%	$3,008,321,457,251

Panel B: Mean Lost Profits from Trading at SIP Prices	
(1)	(2)
Exchange	Lost Profit Per Share
NYSE	-0.0001***
	(0.00001)
NYSE MKT	-0.0001***
	(0.00001)
NYSE Arca	-0.0001***
	(0.00001)
Nasdaq OMX BX	-0.0001***
	(0.00001)
NASDAQ OMX PSX	-0.0002***
	(0.00002)
NASDAQ	-0.0003***
	(0.00003)
BATS	-0.0003***
	(0.00003)
BATS Y	-0.0001***
	(0.00001)
Direct Edge A	-0.0002***
	(0.00002)
Direct Edge X	-0.0002***
	(0.00002)
Chicago	0.0000
	(0.0000)
NSX	-0.0001***
	(0.00003)
FINRA TRF	-0.0003***
	(0.00003)
All Venues	-0.0002***
	(0.00001)
All Exchanges	-0.0002***
	(0.00001)

Panel C: Distribution of Lost Profits from Trading at SIP Prices

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Non-Exchange Venues				Exchange Venues			
Lost Profit Per Share Traded	Percent of Trades	Transaction Value	Total Shares	Aggregate Lost Profits	Percent of Trades	Transaction Value	Total Shares	Aggregate Lost Profits
<-0.1	0.00%	$11,401,892	87,772	-$13,852	0.00%	$61,111,277	460,926	-$76,500
-0.1	0.00%	$4,006,829	29,916	-$2,992	0.00%	$18,025,715	132,785	-$13,279
-0.09	0.00%	$5,570,101	41,163	-$3,705	0.00%	$26,809,831	199,255	-$17,933
-0.08	0.00%	$8,921,207	64,200	-$5,136	0.00%	$40,078,041	302,120	-$24,170
-0.07	0.00%	$15,349,541	110,641	-$7,745	0.00%	$61,268,832	452,318	-$31,662
-0.06	0.00%	$19,741,363	151,470	-$9,088	0.00%	$105,229,329	772,974	-$46,378
-0.05	0.00%	$58,361,039	444,040	-$22,202	0.00%	$261,167,391	1,909,145	-$95,457
-0.04	0.01%	$104,095,266	788,367	-$31,535	0.01%	$477,430,349	3,564,324	-$142,573
-0.03	0.02%	$276,454,150	2,182,136	-$65,464	0.02%	$1,204,069,839	9,244,983	-$277,350
-0.02	0.07%	$1,008,461,887	8,986,507	-$179,730	0.07%	$4,089,576,550	34,978,000	-$699,560
-0.01	2.87%	$22,291,000,000	350,375,751	-$3,503,758	1.92%	$67,706,900,000	914,595,338	-$9,145,953
0	96.74%	$713,644,000,000	11,808,600,000	$0	97.86%	$2,929,500,000,000	46,696,900,000	$0
0.01	0.27%	$2,730,452,450	33,258,715	$332,587	0.10%	$4,323,649,088	48,771,859	$487,719
0.02	0.01%	$77,996,965	774,803	$15,496	0.00%	$199,097,262	1,949,659	$38,993
0.03	0.00%	$17,138,040	173,823	$5,215	0.00%	$50,507,833	498,047	$14,941
0.04	0.00%	$5,700,650	56,826	$2,273	0.00%	$24,512,628	257,847	$10,314
0.05	0.00%	$3,324,105	30,342	$1,517	0.00%	$13,249,008	141,737	$7,087
0.06	0.00%	$2,356,419	30,594	$1,836	0.00%	$9,830,251	112,900	$6,774
0.07	0.00%	$2,080,765	25,657	$1,796	0.00%	$8,380,148	106,651	$7,466
0.08	0.00%	$2,328,248	26,418	$2,113	0.00%	$5,762,064	75,339	$6,027
0.09	0.00%	$964,133	13,208	$1,189	0.00%	$6,324,178	80,756	$7,268
0.1	0.00%	$951,678	11,921	$1,192	0.00%	$4,669,417	57,374	$5,737
>0.1	0.00%	$33,567,567	426,336	$164,920	0.00%	$124,877,045	1,561,264	$603,534
Total:	100.00%	$740,324,224,294	12,206,690,606	$3,315,072	100.00%	$3,008,322,526,075	47,717,125,601	$9,374,955
Total Gains to Liquidity Takers (|Sum| of All Negative Lost Profits):				$3,845,206				$10,570,815
Total Gains to Liquidity Providers (Sum of All Positive Lost Profits):				$530,134				$1,195,860

Table 5: Estimates of Net Profits from Picking Off Stale SIP Quotes Using Active Orders to Monetize Gains

This table presents estimates of the total net profits a trader would earn from a two-trade strategy in which (a) the trader first uses a marketable order to buy (sell) at the SIP NBBO when doing so results in better pricing than trading at the Direct NBBO (i.e., a trade would have a negative value of lost profits) and (b) immediately places an active order to sell (buy) at the Direct NBBO. The table assumes that all trades in the sample with a negative value of lost profits reflect the trader's successful execution of the first trade in the strategy. Net profits were calculated based on observed trades within 1,000 microseconds for each of these first-leg trades that could represent profitable second-leg transactions. *Lost Profit Per Share Traded* is the lost profit of the first-leg trade observed in the sample. *Percent of SIP-Priced Trades* is the percentage of the number of shares from the first-leg trades as a percentage of all shares that were priced at the SIP NBBO. *Total Shares* is the total number of shares from the first-leg trades having the specified Lost Profit Per Share Traded and *Total Shares Having a Second-Leg Match* is the number of these shares that could be matched to a qualifying second-leg trade. *Net Profits* are estimates of the profitability of this strategy based on all trades in the sample.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Non-Exchange Venues				Exchange Venues			
Lost Profit Per Share Traded	Percent of SIP-Priced Trades	Total Shares	Total Shares Having a Second-Leg Match	Net Profits	Percent of SIP-Priced Trades	Total Shares	Total Shares Having a Second-Leg Match	Net Profits
<-0.1	0.00%	87,772	10,371	$337	0.00%	460,926	40,907	$1,632
-0.1	0.00%	29,916	1,700	$62	0.00%	132,785	5,951	$155
-0.09	0.00%	41,163	4,160	$94	0.00%	199,255	11,492	$322
-0.08	0.00%	64,200	4,477	$100	0.00%	302,120	18,172	$436
-0.07	0.00%	110,641	7,575	$192	0.00%	452,318	23,637	$540
-0.06	0.00%	151,470	9,652	$188	0.00%	772,974	39,668	$772
-0.05	0.00%	444,040	44,260	$1,092	0.00%	1,909,145	89,558	$1,815
-0.04	0.01%	788,367	70,327	$1,356	0.01%	3,564,324	164,036	$2,804
-0.03	0.02%	2,182,136	161,128	$2,719	0.02%	9,244,983	379,114	$5,445
-0.02	0.07%	8,986,507	706,656	$8,380	0.07%	34,978,000	1,752,351	$20,550
-0.01	2.87%	350,375,751	11,743,656	$68,897	1.92%	914,595,338	23,043,321	$146,532
Total Net Profits:				$83,417				$181,004
Total First-Leg Shares Matched to Second-Leg Trade:				3.51%				2.65%

Table 6: Estimates of Net Profits from Picking Off Stale SIP Quotes Using Passive Orders to Monetize Gains

This table presents estimates of the total net profits a trader would earn from a two-trade strategy in which (a) the trader first uses a marketable order to buy (sell) at the SIP NBBO when doing so results in better pricing than trading at the Direct NBBO (i.e., a trade would have a negative value of lost profits) and (b) the trader next places a passive order to sell (buy) that rests for specified period of time before trading out of the position at the midpoint of the Direct NBBO prevailing at the end of the time period. Panel A presents estimates if the passive trading window is one second. Panel B presents estimates if the passive trading window is one minute. The table assumes that all trades in the sample with a negative value of lost profits reflect the trader's successful execution of the first trade in the strategy. Net profits were calculated based on observed trades within the specified trading window for each of these first-leg trades. Where more than one passive trade was observed that could reflect the second-leg trade, net profits for the first-leg trade were calculated based on the size-weighted average across all possible matched trades. *Lost Profit Per Share Traded* is the lost profit of the first-leg trade observed in the sample. *Percent of SIP-Priced Trades* is the number of shares from the first-leg trades as a percentage of all shares that were priced at the SIP NBBO. *Mean Return* and *Std. Dev.* are the mean return and standard deviation across all first-leg trades having the specified lost profit per share traded. *5 Percentile Return* is the fifth percentile return for all first-leg trades having the specified lost profit per share traded. *Net Profits* are estimates of the profitability of this strategy based on all trades in the sample.

Panel A: One Second Window

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Non-Exchange Venues					Exchange Venues				
Lost Profit Per Share Traded	Percent of SIP-Priced Trades	Mean Return	Std. Dev.	5 Percentile Return	Net Profits	Percent of SIP-Priced Trades	Mean Return	Std. Dev.	5 Percentile Return	Net Profits
<-0.1	0.00%	0.065%	0.172%	-0.020%	$5,996	0.00%	0.032%	0.134%	-0.039%	$16,814
-0.1	0.00%	0.024%	0.044%	-0.009%	$622	0.00%	0.012%	0.034%	-0.034%	$1,374
-0.09	0.00%	0.010%	0.062%	-0.068%	$658	0.00%	0.013%	0.034%	-0.028%	$2,996
-0.08	0.00%	0.012%	0.028%	-0.019%	$740	0.00%	0.006%	0.056%	-0.042%	$2,780
-0.07	0.00%	0.012%	0.039%	-0.016%	$1,786	0.00%	0.010%	0.029%	-0.025%	$4,718
-0.06	0.00%	0.008%	0.035%	-0.044%	$1,301	0.00%	0.008%	0.028%	-0.025%	$6,368
-0.05	0.00%	0.010%	0.023%	-0.021%	$4,779	0.00%	0.004%	0.026%	-0.030%	$6,968
-0.04	0.01%	0.009%	0.019%	-0.015%	$7,476	0.01%	0.005%	0.018%	-0.020%	$18,829
-0.03	0.02%	0.007%	0.017%	-0.012%	$16,781	0.02%	0.005%	0.015%	-0.016%	$44,258
-0.02	0.07%	0.008%	0.014%	-0.007%	$60,708	0.07%	0.005%	0.012%	-0.011%	$149,820
-0.01	2.87%	0.010%	0.008%	0.000%	$1,587,275	1.92%	0.006%	0.008%	-0.005%	$3,111,814
Total Gains to Liquidity Takers (Sum of Net Profits):					$1,688,122					$3,366,740
Average Share-Weighted Returns:					0.010%					0.006%
Share-Weighted Standard Deviation of Returns:					0.009%					0.009%

Panel B: One Minute Window

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Non-Exchange Venues					Exchange Venues				
Lost Profit Per Share Traded	Percent of SIP-Priced Trades	Mean Return	Std. Dev.	5% Return	Net Profits	Percent of SIP-Priced Trades	Mean Return	Std. Dev.	5% Return	Net Profits
<-0.1	0.00%	0.225%	0.67%	-0.29%	$24,602	0.00%	0.049%	0.709%	-0.300%	$33,812
-0.1	0.00%	0.063%	0.15%	-0.11%	$1,611	0.00%	0.043%	0.165%	-0.160%	$5,572
-0.09	0.00%	0.072%	0.24%	-0.20%	$3,290	0.00%	0.035%	0.265%	-0.182%	$8,485
-0.08	0.00%	0.053%	0.25%	-0.10%	$4,079	0.00%	0.027%	0.269%	-0.198%	$12,757
-0.07	0.00%	0.040%	0.17%	-0.14%	$4,476	0.00%	0.029%	0.235%	-0.159%	$18,987
-0.06	0.00%	0.019%	0.45%	-0.16%	$6,590	0.00%	0.033%	0.259%	-0.146%	$31,707
-0.05	0.00%	0.027%	0.15%	-0.13%	$15,096	0.00%	0.028%	0.160%	-0.126%	$66,929
-0.04	0.01%	0.027%	0.15%	-0.09%	$26,116	0.01%	0.023%	0.139%	-0.111%	$98,394
-0.03	0.02%	0.024%	0.11%	-0.08%	$57,347	0.02%	0.017%	0.110%	-0.096%	$194,582
-0.02	0.07%	0.016%	0.10%	-0.08%	$143,139	0.07%	0.014%	0.102%	-0.089%	$548,337
-0.01	2.87%	0.011%	0.09%	-0.06%	$2,227,958	1.92%	0.008%	0.060%	-0.073%	$5,041,915
Total Gains to Liquidity Takers (Sum of Net Profits):					$2,514,304					$6,061,478
Average Share-Weighted Returns:					0.06%					0.06%
Share-Weighted Standard Deviation of Returns:					0.09%					0.07%

Table 7: Estimates of Annual Profit Opportunities from Direct Feed Arbitrage for the Entire Equities Market

This table provides estimates of the annual profitability of the two primary direct feed arbitrage strategies across all trades in every listed security. Estimates are obtained by extrapolating from all trades in the top-traded 257 securities and the top-traded 872 traded securities during our 10 month sample period. These securities represent half and three-quarters of all trading volume in all listed securities, respectively. All figures are obtained from TAQ data. *Total Trading Value* is measured as the dollar value of all trades in the relevant group of securities. *Value of All SIP-Priced Trades* is the value of these trades where the trade is priced at the SIP NBBO. *Maximum Available Profits to Liquidity Providers* is an estimate of the total profitability to market makers seeking to exploit direct feed arbitrage. It is calculated as the dollar value liquidity takers lost on all trades by having trades priced at the SIP NBBO rather than the Direct NBBO, summed across all trades where liquidity takers received inferior pricing at the SIP NBBO. *Maximum Available Profits to Liquidity Takers* is an estimate of the gross profitability to liquidity takers from picking off stale quotes priced at the SIP NBBO. It is calculated as the dollar value liquidity providers lost on trades by having trades priced at the SIP NBBO rather than the Direct NBBO, summed across all trades where liquidity providers received inferior pricing at the SIP NBBO. Annualized estimates for the entire market are obtained by multiplying measured values by 506/228 (i.e., 2 x 253/228) in the case of the top-traded 257 securities or 1012/684 (i.e., 4/3 x 253/228) in the case of the top-traded 872 securities.

	Trades in 257 Most Traded Securities		Trades in 872 Most Traded Securities	
	As Measured During Sample Period	Annualized Estimate of Whole Market	As Measured During Sample Period	Annualized Estimate of Whole Market
Total Trading Value	$25.9 trillion	$57.5 trillion	$42.1 trillion	$62.3 trillion
Value of All SIP-Priced Trades	$19.1 trillion	$42.4 trillion	$30.2 trillion	$44.7 trillion
Maximum Available Profits to Liquidity Providers - Exchange Trades	$8,029,143	$17,819,062.91	$14,594,889	$21,593,608
Maximum Available Profits to Liquidity Providers – Non-Exchange Trades	$3,462,051	$7,683,324.69	$6,114,556	$9,046,683
Maximum Available Profits to Liquidity Takers - Exchange Trades	$70,144,999	$155,672,673	$144,787,657	$214,217,996
Maximum Available Profits to Liquidity Takers - Non-Exchange Trades	$34,788,793	$77,206,707	$56,166,367	$83,099,946