Price Constraints, Speed Competition, and Liquidity[1]

November 18, 2013

Chen Yao and Mao Ye

[1]Chen Yao is from the University of Warwick and Mao Ye is from the University of Illinois at Urbana-Champaign. Please send all correspondence to Mao Ye: University of Illinois at Urbana-Champaign, 340 Wohlers Hall, 1206 South 6th Street, Champaign, IL, 61820. E-mail: maoye@illinois.edu. Telephone: 217-244-0474. This paper is the first part of a paper circulated under the title "The Externalities of High Frequency Trading." We thank Jim Angel, Shmuel Baruch, Robert Battalio, Dan Bernhardt, Jonathan Brogaard, Jeffery Brown, John Campbell, John Cochrane, Robert Frank, Slava Fos, George Gao, Paul Gao, Arie Gozluklu, Joel Hasbrouck, Frank Hathaway, Pankaj Jain, Tim Johnson, Charles Jones, Andrew Karolyi, Nolan Miller, Katya Malinova, Maureen O'Hara, Neil Pearson, Richard Payne, Andreas Park, Josh Pollet, Gideon Saar, Ronnie Sadka, Jeff Smith, Duane Seppi, an anonymous reviewer for the FMA Napa Conference, and seminar participants at the University of Illinois, the SEC, CFTC/American University, the University of Memphis, the University of Toronto, the 8th Annual MARC meeting, the Financial Intermediation Research Society Conference 2013, the 3rd MSUFCU Conference on Financial Institutions and Investments, the Northern Finance Association Annual Meeting 2013, the China International Conference in Finance 2013, and the 9th Central Bank Workshop on the Microstructure of Financial Markets for their helpful suggestions. We also thank NASDAQ OMX for providing the research data. This research is supported by National Science Foundation grant 1352936. This work also uses the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. We thank Robert Sinkovits and Choi Dongju of the San Diego Supercomputer Center and David O'Neal of the Pittsburgh Supercomputer Center for their assistance with supercomputing, which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program. We also thank Jiading Gai, Chenzhe Tian, Tao Feng, Yingjie Yu, and Chao Zi for their excellent research assistance.

Abstract

When price competition over nominal bid-ask spreads in a stock market is constrained by tick size, liquidity providers either compete on trading speed due to the time priority rule or they are willing to pay a fee to make the market. We argue that price constraints created by regulation (SEC rule 612, or the Minimum Pricing Increment Rule) is one of the factors driving high-frequency trading and the proliferation of taker/maker-fee markets. We find greater high-frequency liquidity provision for lower-priced stocks with higher market caps, whereby the one-cent tick size exerts a higher constraint on price competition. The same stocks also achieve higher market share in markets with inverted fees, whereby liquidity providers must pay a fee instead of receiving a rebate. A reduction in the nominal share price due to a stock split increases the trading speed of a stock and leads to a migration of trading volume to the market with inverted fees. Price constraints also prevent speed competition from improving the price of liquidity, although time priority can determine which traders are able to provide liquidity at the constrained price. We find that exogenous technology shocks that improve speed at the millisecond, microsecond, or nanosecond levels does not improve quoted spreads, effective spreads, or depth. Our results suggest that deregulation of tick size can reduce speed competition and increase price competition.

Key Words: price constraints, high-frequency trading, maker/taker fees, liquidity, tick size non-price competition

1. Introduction

Under standard Walrasian equilibrium, price is infinitely divisible but time is not; all agents are assumed to arrive at the market at the same time. However, the reality regarding financial markets is exactly the opposite: time becomes divisible at the nanosecond level but price is restricted by tick size. This paper shows that two sources of friction, discrete prices and (almost) continuous time under canonical Walrasian equilibrium, help to explain two important features of the U.S. stock market: high-frequency trading and taker/maker fees.

Recently, the *Wall Street Journal* stated that trading entered the nanosecond age when Fixnetix, a London-based trading technology company, announced that "it has the world's fastest trading application, a microchip that prepares a trade in 740 billionths of a second, or nanoseconds." Since "investment banks and proprietary trading firms spend millions to shave ever smaller slivers of time off their activities . . . [as] the race for the lowest 'latency' [continues], some market participants are even talking about picoseconds—trillionths of a second." [2] Because high-frequency traders aggressively invest in technologies that reduce latency, speed should create private benefit.[3] A more important question, however, is whether investment in speed creates social benefit by improving liquidity. Empirical research on the speed of trading prior to the sub-millisecond era finds that speed decreases both quoted spreads and effective spreads.[4] Put differently, the enhanced speed with which liquidity is provided or canceled reduces the cost of liquidity.

[2] "Wall Street's Need for Trading Speed: The Nanosecond Age," *Wall Street Journal*, June 14, 2011.
[3] A recent article in the Financial Times estimates that a 1-millisecond advantage is worth up to $100 million in annual gains. "Speed fails to impress long-term investors," *Financial Times*, September 22, 2011.
[4] See Boehmer, Fong, and Wu (2013), Hendershott, Jones, and Menkveld (2011), Hendershott and Riordan (2009 and 2011), Hasbrouck and Saar (2013), and Brogaard, Hagströmer, Norden, and Riordan (2013), among others.

However, speed competition, by its nature, is a form of non-price competition. U.S. stock markets observe price, display, and time priority, in that order.[5] If liquidity providers can undercut each other on price in this environment, speed competition is secondary. Therefore, fierce speed competition may not facilitate price competition. We argue, on the contrary, that speed competition in liquidity provision may simply be a consequence of failed price competition. In the United States, tick size is regulated through SEC rule 612 (Minimum Pricing Increment) of regulation NMS. The rule prohibits stock exchanges from displaying, ranking, or accepting quotations, orders, or indications of interest in any NMS stock priced in an increment smaller than $0.01 if the quotation, order, or indication of interest is priced equal to or greater than $1.00 per share. Because the tick size is the same for all stocks priced above $1.00, the natural bid-ask spread for the largest stocks with the lowest prices is more likely to be constrained by tick size, and we find that these stocks have the highest level of high-frequency liquidity provision.[6] After double-sorting stocks by market cap and price, we find that high-frequency traders provide 49.60% of the liquidity for large low-priced stocks, while providing 27.54% of the liquidity for large high-priced stocks.

Although stock exchanges cannot quote sub-penny spreads, this rule can be by-passed through the mechanism of the taker/maker fee. For example, EDGA and EDGX are two trading platforms offered by Direct Edge. They are almost identical except in one dimension. EDGX,

[5] Orders that offer a higher price have the highest execution priority. Among orders that offer the same price, displayed orders take priority over non-displayed orders. Among orders with the same displayed status, orders arriving first have the highest priority.

[6] In some European countries, tick size varies based on the price level of a given stock. Therefore, high-frequency liquidity provision may not have a monotonic relationship with price level. Hagströmer and Norden (2013) analyze the Swedish market.

like most other exchanges, uses a maker/taker fee structure, which pays liquidity makers and charges liquidity takers. EDGA, however, has an inverted taker/maker fee structure whereby the maker of liquidity, or the passive (limit) order, is charged a rebate and the taker of liquidity, or the aggressive (market) order, is paid a fee. The willingness of liquidity providers to pay a fee to make the market implies that the nominal bid-ask spread is constrained. For example, suppose the quoted bid-ask spread is 1 cent; a maker fee of 0.025 cents requires a liquidity provider who is willing to provide liquidity at 0.975 cents. We find that EDGA's market share relative to that of EDGX is also related to the price level of a stock. EDGA takes a relatively high market share in low-priced stocks, whereas EDGX takes a higher market share in high-priced stocks. This result is consistent with the theoretical model of Foucault, Kadan, and Kandel (2013). In their model, stock exchanges find it optimal to charge liquidity makers and subsidize liquidity takers when tick size is large while subsidizing liquidity makers and charging liquidity takers when tick size is small. In reality, relative tick size varies with stock price, but stock exchanges do not adjust fees based on stock characteristics. Therefore, stocks with more severe price constraints generate an active taker/maker fee market whereas stocks with less severe price constraints generate an active maker/taker fee market.

In addition to the results based on cross-sectional price variation, we provide further support of the price constraints hypothesis using stock splits as the identification strategy.[7] Stock splits reduce nominal share prices and increase relative tick size. We document a dramatic increase in trading speed as well as a migration of volume from EDGX to EDGA for stocks that

[7] The identification assumption based on cross-sectional price variation is that nominal share price is exogenous. If trading speed can change share price, or if there are common factors that drive nominal share price and trading speed, our result would be biased. Regarding the concern with reverse causality, we argue that there is no evidence showing that speed competition at the sub-millisecond level can affect cross-sectional variation in stock prices. If there is a link, the channel seems to be through liquidity, but we show that speed does not change liquidity. Regarding the concern with omitted variables, Benartzi, Michaely, Thaler, and Weld (2009) argue that nominal share price is exogenous with respect to firm fundamentals other than the market cap.

split relative to a matched sample of stocks that do not split. Again, these results confirm our intuition that price constraints and time priority drive speed competition and the taker/maker market.

To the best of our knowledge, our paper is the first empirical study that provides a coherent explanation linking the literature on high-frequency trading to the maker/taker fee market.[8] Two common factors, price constraints and time priority, contribute to these two market phenomena. A more severe price constraint implies higher profits for supplying liquidity and an oversupply of liquidity at constrained prices. Traders who achieve higher speeds are able to supply liquidity because of time priority in the queue. A severe price constraint will induce some traders to pay a fee in order to make the market, especially when there is a long queue in which makers obtain rebates. This explains the proliferation of markets with inverted fees. As expected, high-frequency trading is strongly correlated with taker/maker market share relative to maker/taker market share because they are driven by the same two common factors. By sorting the relevant stocks into five portfolios, we find that the highest EDGA-to-EDGX market share quintile experiences twice as much high-frequency market-making activity as the lowest EDGA-to-EGDX market share quintile.

Our paper contributes to the literature by providing a new interpretation of speed competition, which offers new insight into the impact of such competition on liquidity. When speed competition is a consequence of price constraints, speed may determine who provides liquidity but it will not change the price of liquidity, since liquidity providers can no longer undercut each other on price. This result is verified by several major technology stocks that have

[8] On high-frequency trading, see Biais, Foucault, and Moinas (2011), Jovanovic and Menkveld (2010), Pagnotta and Philippon (2012), Chaboud, Chiquoine, Hjalmarsson, and Vega (2009), Hendershott and Riordan (2009 and 2011), Hasbrouck and Saar (2013), and Hendershott, Jones, and Menkveld (2013), among others. For maker/taker markets, see Foucault, Kadan, and Kandel (2013), Colliard and Foucault (2012), Brolley and Malinova (2012), Park and Malinova (2013), and Halmrast, Malinova, and Park (2013), among others.

increased trading speed. First, we identify an enhancement of the matching engine of the NASDAQ in April 2010 and a follow-on technology enhancement on the high-frequency trader side in May 2010. These two shocks have reduced the latency involved in submitting and processing orders from micro- to nanoseconds but they have not improved liquidity measures. We also document technology enhancements that reduce the latency involved in disseminating trading data from 3 milliseconds to 1 millisecond. Interestingly, NASDAQ trading data are disseminated through six identical but independent channels based on the alphabetic order of ticker symbols. Channel 1, which includes all NASDAQ stock ticker symbols beginning with A and B, was upgraded on October 10, 2011. Channels 2 through Channel 6 (covering successive sets of symbols alphabetically) were upgraded on October 17, 2011. These staggered technology enhancements provide a clean test for examining the causal relationship between speed and liquidity. Again, we find that enhancing speed does not improve liquidity.

Our results pertaining to liquidity contrast with those reported in recent work on high-frequency trading, which finds that high-frequency trading improves liquidity.[9] However, existing empirical evidence is based on data reported before the sub-millisecond era. Therefore, we observe our result as supplementary instead of contradictory to results reported in past studies. Speed competition measured in nanoseconds differs not only in magnitude but also in nature from earlier speed competition. The current discussion, either in the academic literature or in the public domain, focuses on the speed advantage that high-frequency traders enjoy over slower traders. Speed competition at the nanosecond level, however, is possible only among high-frequency traders. One question that remains unanswered pertains to the nature of the mechanism by which the interaction of high-frequency trading algorithms improves market

[9] Jones (2013) provides an excellent survey of the literature.

quality (Chordia, Goyal, Lehmann and Saar, 2013). Our results indicate the diminishing return of speed on liquidity. While a reduction in latency can improve liquidity when speed is low, an arms race in speed below the sub-millisecond level confers no observable benefit on liquidity.

Our paper also provides new insight into regulations. Instead of discussing whether and how we should regulate high-frequency trading and maker/taker fees, our paper is the first to propose that these two phenomena may be a consequence of existing regulations. Economic reasoning and our empirical evidence can show, step by step, how various regulations create the world we observe. At infinitely small tick sizes, the breakdown between maker fees and taker fees does not matter because these fees will be neutralized by differences in the nominal bid-ask spread: the maker/taker market will have a lower nominal spread and the taker/maker market will have a higher nominal spread, but the cum fee bid-ask spread is the same for both markets (Angel, Harris and Spatt (2011) and Colliard and Foucault (2012)). However, SEC rule 611 states that orders should be routed to the market with the best displayed (nominal) spread. In that case, all orders should be routed to the maker/taker market and the taker/marker market should be empty. However, rule 612 prohibits sub-penny pricing, so the maker/taker and taker/maker markets can display the same nominal bid-ask spread.[10] In addition, rule 611 imposes price priority only across markets, but time priority is imposed only on the individual market.[11] Under price constraints, the queue in the maker/taker market can be very long and order execution becomes the privilege of liquidity providers who trade at higher speeds. Therefore, some liquidity providers are willing to pay a fee to jump ahead of the queue by trading on the taker/maker fee market.

[10] For example, if the equilibrium spread without tick size is 0.3 cents on the maker/taker market and 0.8 cents on the taker/maker market, both markets will quote a 1-cent spread due to price constraints.
[11] If global time priority had been imposed, liquidity providers would have abandoned the taker/maker fee market to collect rebates on the maker/taker market because the execution probability is the same.

Therefore, reducing tick size will lead to more intense price competition instead of an arms race in speed. Certainly, the benefit of reducing tick size needs to be weighed against the adjustment cost, but we are currently observing a reverse trend. The SEC is now considering increasing tick size.[12] The rationale for this would be that "market makers and dealers need more economic incentives to bring smaller companies public, provide bids and offers and publish stock research."[13] Economic theories suggest that price constraints should facilitate non-price competition. However, we doubt that non-price competition would take the form of providing better services and research. An increase in tick size may facilitate another round of speed competition.

This paper is organized as follows. Section 2 describes the institutional details and data used in the study. Section 3 defines price constraints and examines the relationship between price constraints, high-frequency trading, and taker/maker fees. Section 4 provides a robustness check on the results reported in section 3 based on stock splits. Section 5 demonstrates the impact of speed based on exogenous technology shocks. Section 6 concludes the paper and discusses the policy implications.

2. Data and Institutional Details

2.1 Data

This paper uses four main datasets: NASDAQ TotalView-ITCH with a nanosecond time stamp, daily TAQ data with a millisecond time stamp, a NASDAQ dataset that identifies whether a liquidity maker/taker is a high-frequency trader, and CRSP.

[12] "Exchanges Said to Prepare Pilot Programs for Changing Tick Sizes," *Bloomberg News*, June 13, 2013.
[13] "SEC Weighs Bigger Stock-Price Increments 12 Years After Pennies," *Bloomberg News*, February 24, 2013.

NASDAQ TotalView-ITCH is a series of messages that describe orders added to, removed from, and executed on the NASDAQ. It is used to identify two technology shocks at the matching engine that decrease latency from microseconds to nanoseconds. We also use ITCH data to construct a limit-order book with nanosecond resolution, which is the foundation for calculating liquidity. Table 1 presents a sample of ITCH data. A discussion of efficient construction of the limit-order book using supercomputers can be found in the appendix and in Gai, Choi, O'Neal, Ye, and Sinkovits (2013).

Insert Table 1 about here

Daily TAQ data provide trades and quotes for all issues traded on the NYSE, NASDAQ, and regional exchanges. They come with more detailed information than the usual monthly TAQ data, which are widely used in academic research. Holden and Jacobsen (2013) compare daily TAQ data and monthly TAQ data in detail. The unique feature that facilitates our study is the millisecond time stamp, which helps to identify another technology shock that increases the speed involved in disseminating trading data from 3 milliseconds to 1 millisecond. This technological improvement was adopted in a staggered matter based on the alphabetic order of ticker symbols, providing a clean means of identification in reference to which it is possible to examine the causal effect of speed on market quality.

Also, the consolidated trades file of daily TAQ data provides information on execution across separate exchanges for trades greater than or equal to 100 shares (O'Hara, Yao and Ye, 2013). We use such data to calculate EDGA's market share relative to that of EDGX. These two trading platforms have similar infrastructure but contrasting fee structures. In our sample period, EDGX, like most exchanges, has a maker/taker fee structure whereby liquidity demanders pay a fee of 0.30 cents per share while liquidity providers get a rebate of 0.26 cents per share; EDGA has a taker/maker (or inverted) fee structure whereby liquidity suppliers pay a fee of 0.025 cents per share while liquidity demanders get a rebate of 0.015 cents per share. We argue in section 3 that stocks with more severe price constraints would earn a larger market share in EDGA because more liquidity providers are willing to pay a fee to make the market.

The NASDAQ high-frequency dataset provides information on limit-order books and trades for 120 stocks selected by Hendershott and Riordan. The sample includes 40 large stocks from the 1000 largest Russell 3000 stocks, 40 medium stocks from stocks ranked from 1001–2000, and 40 small stocks from Russell 2001–3000. Among these stocks, 60 are listed on the NASDAQ and 60 are listed on the NYSE. Because the sample was selected in early 2010, three stocks disappear in our sample period so we have 117 stocks. The limit-order book data offer one-minute snapshots of the book with an indicator that breaks out liquidity providers into high-frequency traders and non-high-frequency traders. The trade file provides information on whether the traders involved in each trade are high-frequency traders or non-high-frequency traders. In particular, trades in the dataset are categorized into four types, using the following abbreviations: “HH”: high-frequency traders who take liquidity from other high-frequency traders; “HN”: high-frequency traders who take liquidity from non-high-frequency traders; “NH”: non-high-frequency traders who take liquidity from high-frequency traders; and “NN”: non-high-frequency traders take liquidity from other non-high-frequency traders. The limit-order book and the trade file provide us with two measures of the activity of high-frequency traders.

Our analysis focuses mostly on the sample of 117 stocks because it is the sample that is covered by all four datasets. There are two deviations, though. We are also interested in the impact of stock splits on high-frequency liquidity provision, but these 117 shocks do not provide a large enough sample of splits. Also, the 117 stocks only have 60 NASDAQ-listed stocks, 10 stocks for each of the six channels. Therefore, we need a larger sample to study the impact of channel upgrade. The sample deviations are explained in greater detail below, where we discuss stock splits and channel upgrades.

2.2. Identification of Technology Shocks

Jones (2013) argues that papers that examine high-frequency trading should isolate market structure changes that facilitate such trading. We start our analysis by identifying major technology shocks from 2010–2011. Three corresponding structural breaks are identified in the data and are also verified by NASDAQ. ITCH data measure time at the machining engine, which reveals two major structural breaks in the time needed to submit or process orders. Daily TAQ data measure the time involved in disseminating stock-trading data, which indicates a structural break in the time involved in transmitting the data after matching. Interestingly, both of these structural changes happened on weekends, enabling the stock exchange and traders to test the new technology. We also examine relatively smaller shocks such as increases in bandwidth, but find that these shocks improve neither speed nor liquidity, indicating that technology improvement in the nanosecond era may sometimes merely accommodate existing transaction speeds or provide the potential for increasing speed in the future. The inclusion of various minor technology shocks does not change the main argument of the paper: speed does not improve liquidity.

Figure 1 demonstrates the increase in speed at the matching-engine level using ITCH data. The sample period is January 2010–November 2011. Panel A demonstrates the minimum timestamp difference between two consecutive messages over the course of a day. These two messages do not need to come from the same trader. For example, they can indicate the time difference between one trader's execution and another trader's cancellation. The figure shows a decrease from about 950 nanoseconds to 800 nanoseconds between April 9, 2010 and April 12, 2010 and a dramatic decrease from 800 nanoseconds to 200 nanoseconds between May 21, 2010 and May 24, 2010. Because the ITCH data track the life of each individual order, we know when a cancellation and execution are from the same trader. Panel B of Figure 1 demonstrates, for each day, the quickest execution and cancellation times. Panel B shows the times of the fastest cancellation and execution decreases in the April structural break, along with a dramatic decrease in the volatility of the fastest cancellation and execution across succeeding days. The structural break in May, however, has a dramatic impact on latency. The difference between the fastest cancellation and execution times decreases from about 1.2 microseconds to between 500 and 600 nanoseconds and stays below 1 microsecond for all but seven days after the break. Undoubtedly, NASDAQ entered the realm of nanosecond trading after May 24, 2010. Our conversion with NASDAQ reveals that the first structural break is a consequence of the installment of the Nehalem machine engine, while the second break is more likely to originate on the high-frequency trader side.[14]

Insert Figure 1 about here

The outflow messages on NASDAQ-listed stocks are distributed and processed across six channels in "unlisted trading privileges" (UTP). During our sample period, Channel 1 handles ticker symbols from A through B; Channel 2 handles ticker symbols from C through D; Channel 3 handles ticker symbols from E through I; Channel 4 handles ticker symbols from J through N; Channel 5 handles ticker symbols from O through R; and Channel 6 handles ticker symbols from S through Z. The channel assignments provide us with clean identifications, making it possible to

[14] We thank Frank Hatheway and Jeff Smith for providing us with this insight.

identify the relationship between speed enhancement and market quality. Figure 2 demonstrates the staggered technology enhancements across these six channels. For illustration purposes, we present a 100-millisecond snapshot of the market, which shows that, before Friday, October 7, 2011, the last digits of messages in daily TAQ data for all six channels end with 0, 3, or 7. This implies that information was broadcast in either 3 milliseconds between 0 and 3 and 7 and 0, or 4 milliseconds between 3 and 7. Therefore, Panel A of figure 2 shows a minimum latency of 3 milliseconds for all six channels. Panel B shows that on Monday, October 10, 2011, Channel 1 was enhanced and was able to broadcast information once every millisecond. However, the other five channels were still broadcasting with 3- to 4-millisecond gaps. The same pattern continues until October 14, 2011, but on October 17, 2011, all six channels are able to broadcast information every millisecond.

Insert Figure 2 about Here

The data for the 100-millisecond snapshot is representative of the entire span of trading hours from 9:30:00 to 16:00:00. Figure 3 depicts the median time gap between two broadcasts for six channels. Before October 7, 2011, the median time gap between two broadcasts for all six channels is 3 milliseconds. This is a direct consequence of the fact that all the broadcasts are on milliseconds ending with 0, 3, and 7. On October 10, 2011, the median gap is 1 millisecond for Channel 1 but 3 milliseconds for channels 2–6. On October 17, 2011, all six channels have a median gap of 1 millisecond.

Insert Figure 3 about Here

Our two identification strategies are subject to several limitations. The ITCH data feed is a direct, rapid data feed provided by NASDAQ. The feed, however, is single-thread and does not give us cross-sectional variations in technology enhancement. To address this issue, we follow the methodology of Boehmer, Saar, and Yu (2005) by using fixed effects and controlling for variables known for affecting market quality. The channel upgrades provide cleaner identification because of cross-sectional variation, but the data comes from a relatively slow consolidated feed. According to Durbin (2010), however, even the most aggressive high-frequency trader still listens to consolidated feeds. After all, no market data feed is perfect; a direct feed can sometimes lose packages. Multiple sources of data help verify that an unusual market data tick is genuine by comparing it with a second source. Also, the consolidated feed provides the reference price for dark pools and internalizers, and executed trades for dark pools and internalizers can be obtained only from the consolidated feed. In addition, regulatory actions such as the trade-through rule and limit-up and limit-down are enforced according to consolidated feeds. Finally, most traders subscribe to the consolidated feed, and it provides high-frequency traders with a view of most other traders, which can help high-frequency traders design their own strategies based on other traders' views of the market.

3. Price Constraints, Maker/Taker Fees, and High-Frequency Activity

In this section, we demonstrate the following two results: higher price constraints lead to 1) more speed competition in liquidity provision and 2) higher market share in trading venues where such a price constraint can be weakened by paying a liquidity-maker fee. These two results also provide the economic foundation for the analysis in Section 5. If speed competition is a consequence of failed price competition, an increase in speed cannot change the cost of liquidity.[15]

[15] Nevertheless, it determines who can provide liquidity.

Our results are based on October 2010 data because that is the sample period in which all measures in the section can be computed. NASDAQ high-frequency data provide the market share of high-frequency liquidity provision for 117 stocks for 2008–2009, February 22–26, 2010 and October 2010. EDGA and EDGX volumes are included in the TAQ data from July 2010. Therefore, we have measures of both high-frequency liquidity provision and the market share of the taker/maker market relative to that of the maker/taker market for October 2010.

3.1. Price Constraint Measure

An intuitive measure of price constraints is the price itself. Because stocks with prices above one dollar have a tick size of one cent, the price constraint is stricter for lower-priced stocks, especially for stocks with a higher market cap. This intuition was suggested by the *Wall Street Journal:*[16]

> 'The lower the share price, the more attractive the stock is to high-frequency traders,' said Justin Schack, managing director at Rosenblatt Securities Inc., a brokerage and stock-market research firm. 'And if you can find a stock with a low share price that is also a large-cap stock with a big float, it becomes even more attractive.'

Technically, firms can adjust the nominal price by split or reverse split to achieve optimal relative tick size even though the tick size is one cent. Angel (1997) develops a theory of relative tick size, which predicts that companies tend to split their stocks so that institutionally mandated minimum tick sizes are optimal relative to stock prices. However, the relative tick size was rejected. If relative tick size is optimal, the change in tick size from 1/16 to 1/100 dollar should result in an aggressive split, which is not observed after decimalization. Benartzi, Michaely,

[16] "Citigroup Instantly Becomes a $40 Stock,", *Wall Street Journal*, May 10, 2011

Thaler, and Weld (2009), on the other hand, argue that the nominal share price is a puzzle because it cannot be explained by the marketability hypothesis, the pay-to-play hypothesis, or signaling.[17] The only pattern documented in Benartzi Michaely, Thaler, and Weld (2009) is that large cap stocks have higher prices. Baker, Greenwood, and Wurgler (2009) posit a catering theory of nominal stock prices, in which firms split when investors place higher valuations on low-price firms and vice versa. However, the catering theory focuses more sharply on time-series variations in stock prices while our analysis focuses on cross-sections. Campbell, Hilscher, and Szilagyi (2008) find that prices may predict distress risk when they are very low, but the same paper also acknowledges that the prediction goes away when the price rises above $15. In summary, the prior literature indicates that cross-sectional variations in nominal stock prices are orthogonal to firm fundamentals other than the market cap, which justifies the use of price as a price-constraint measure.

3.2. Price Constraints and High-Frequency Liquidity Provision

This section explores the relationship between price constraints and high-frequency liquidity-making. High-frequency trading is measured in two ways. First, NASDAQ high-frequency trading data provides a snapshot of the limit-order book with flags indicating the market depth provided by high-frequency and non-high-frequency traders. We use the relative market share of depth at the best bid and offer provided by high-frequency traders as the first measure of high-frequency liquidity provision. Second, NASDAQ high-frequency data indicate,

[17] The marketability hypothesis states that low-priced stocks are more attractive to individual investors. (Baker and Gallagher, 1980; Baker and Powell, 1993; Fernando, Krishnamurthy, and Spindt, 1999 and 2004; Lakonishok and Lev, 1987; and Byun and Rozeff, 2003). The Pay-to-play hypothesis posits that firms can split their stocks to achieve optimal relative tick size. A higher relative size motivates more dealers to make markets and investors to provide liquidity by placing limit orders, despite its placing a high floor on the quoted bid-ask spread (Angel, 1997). The signaling hypothesis (Brennan and Copeland, 1988; Lakonishok and Lev, 1987; and Kalay and Kronlund, 2013) states that insiders use stock splits to signal information.

for each trade, the maker and taker of liquidity. There are four types of trades, which we indicate as specified above using NH, HH, HN, and NN. Therefore, the second measure of high-frequency liquidity is the percent of volume with high-frequency liquidity provision.

Insert Table 2 about here

The original 120 stocks selected by Hendershott and Riordan include 40 large stocks from the 1000 largest Russell 3000 stocks, 40 medium stocks from stocks ranked from 1001–2000, and 40 small stocks from Russell 2001–3000. A natural way to conduct the analysis is to sort the stocks 3-by-3 based on the market cap and the price level of the stock. We then sort the 117 remaining stocks first into small, medium, and large groups based on the average market cap of September 2010, and each group is further subdivided into low, medium, and high sub-groups based on the average closing price of September 2010.

We provide two ways aggregating the high-frequency liquidity provision for stocks in each portfolio. The first is volume-weighted average, whereby we first sum the high-frequency liquidity provision for all stocks in the portfolio and then divide it by the total liquidity provision for all stocks in the portfolio. To be more specific, the depth data provide one-minute snapshots of the depth provided by high-frequency traders and non-high-frequency traders, {$HFTdepth_{itm}$, $NonHFTdepth_{itm}$}, where i is the stock, t is the date, and m is the time of day. The average depths provided by high-frequency traders and non-high-frequency traders for each stock on each day are:

$$HFTdepth_{it} = \frac{1}{M}\sum_{i=1}^{M} HFTdepth_{itm} \quad \text{and} \quad NonHFTdepth_{it} = \frac{1}{M}\sum_{i=1}^{M} NonHFTdepth_{itm} \qquad (1)$$

The depth provided by high frequency traders relative to the total depth of portfolio J is then defined as:

$$VWHFTdepthshare_J = \frac{\sum_{i\in J}\sum_{t=1}^{T} HFTdepth_{it}}{\sum_{i\in J}\sum_{t=1}^{T}(HFTdepth_{it} + NonHFTdepth_{it})} \quad (2)$$

Similarly, suppose NH_{it}, HH_{it}, HN_{it}, and NN_{it} are the four types of share volume for each stock on each day; then the volume share with high-frequency traders as liquidity providers relative to total volume is defined as:

$$VWHFTliquidityvolumeshare_J = \frac{\sum_{i\in J}\sum_{t=1}^{T}(NH_{it} + HH_{it})}{\sum_{i\in J}\sum_{t=1}^{T}(NH_{it} + HH_{it} + HN_{it} + NN_{it})} \quad (3)$$

The volume-weighted measure provides the overall impact of high-frequency liquidity provision, with days and stocks with large high-frequency volume having greater impacts. We also provide an equally weighted average of high-frequency activity, whereby each stock and each day have the same weight. For each stock i and day t,

$$HFTdepthshare_{it} = \frac{HFTdepth_{it}}{HFTdepth_{it} + NonHFTdepth_{it}}$$

and

$$HFTliquidityvolumeshare_{it} = \frac{NH_{it} + HH_{it}}{NH_{it} + HH_{it} + HN_{it} + NN_{it}} \quad (5)$$

Next, suppose there are N stocks in portfolio J and T days; the equally weighted high-frequency liquidity provision is defined as

$$EWHFTdepthshare_J = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} HFTdepth_{it}$$

and

$$EWHFTliquidityvolumeshare_J = \frac{1}{N}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} HFTliquidityvolumeshare_{it} \quad (6)$$

Table 2 demonstrates the relationship between high-frequency market-making and the price level of stocks. Panel A shows that the volume-weighted depth provided by high-frequency trading increases monotonically with price. For large-cap stocks, high-frequency traders provide 33.16% of the depth of high-priced stocks, but the liquidity provision is as high as 55.62% for low-priced stocks. The volume percentage with high-frequency liquidity providers (Panel B) is lower than the depth provided by high-frequency traders, but the results are also monotonic. For example, 39.15% of the volume is due to high-frequency liquidity providers for medium stocks with low prices, but the figure is only 22.34% for medium stocks with high prices. For small stocks, 23.40% of the volume is due to high-frequency liquidity providers for high-priced stocks, but the number is only 18.74% for low-priced small stocks. Panels C and D present the equally weighted results. Again, we find that high-frequency liquidity provision decreases in stock prices for large and medium stocks. For example, high-frequency traders provide 49.60% of the depth of low-priced stocks, 43.76% of the depth of medium-priced stocks, and 27.54% of the depth of high-priced stocks. High-frequency liquidity providers supply 35.50% of the volume for medium-cap low-priced stocks; the number is 23.40% for medium-cap medium-priced stocks, and 22.20% for medium-cap high-priced stocks. The results for small-cap stocks are low and are not monotonic for equally weighted averages. Small stocks generally have wider spreads, and the equilibrium spread is less likely to be constrained. Therefore, we observe lower market shares for high-frequency market-making in general and limited variation across distinct price groups.

3.3. Price Constraint and Taker/Maker Market

The one-cent tick size for stocks above one dollar is enforced across exchanges.[18] However, the price constraint can be bypassed. While most exchanges such as the NASDAQ offer rebates to liquidity providers, three exchanges, Boston, BATS-Y, and EDGA, have inverted fee structures that charge liquidity providers. Starting from July 2010, the volumes of EDGA and EDGX can be identified from the TAQ data. These two trading platforms have similar infrastructures and the major difference between them is the breakdown of the maker/taker fees. EDGA charges liquidity makers 0.025 cents per share whereas it provides a 0.015 cents rebate to liquidity takers. EDGX provides liquidity makers with 0.26 cents per share but charges liquidity takers 0.3 cents per share.

Table 3 shows the EDGA market share. We sort stocks 3 by 3 first by average market cap and then by average price on September 2010. As in section 3.2, we also define the volume-weighted and equally weighted EDGA market share relative to the total combined EDGA and EDGX volume for each portfolio. To calculate the volume-weighted (VWEDGAratio) average, we first aggregated the EDGA and EDGX volumes for stocks in the portfolio for all days. VWEDGAratio is defined as the ratio of the aggregated EDGA volume divided by the aggregated EDGA volume plus the aggregated EDGX volume. To define the equally weighted average (EWEDGAratio), we first compute, for each stock i on each day t, the market share of EDGA relative to that of EDGA and EDGX ($EDGAratio_{it}$). EWEDGAratio is the average of $EDGAratio_{it}$ across stock and dates.

Table 3 shows that the EDGA market share decreases in price for both the volume-weighted and equally weighted averages. Panel A is based on the volume-weighted average.

18 Internalized trades are not subject to the sub-penny rule. However, traders who choose the dark pool and internalizers may have reasons other than a price constraint for their actions. Zhu (2013) and Ye (2012) provide models of the selection of trading venues across an exchange and a dark pool.

Large-cap low-priced stocks take 64.28% of the EDGA volume, implying that EDGX accounts for only 35.72% of the volume. However, EDGX beats EDGA in large high-priced stocks. EDGA accounts for only 28.98% of the volume and the remaining 71.02% is in EDGX. For medium- and small-cap stocks, the percentage of the EDGA volume also decreases in price. Panel B presents the results based on equally weighted averages. For example, for large stocks, the EDGA volume is 57.94%, but is only 35.51% for large high-priced stocks. For medium-cap stocks, EDGA takes 56.87% of the trading volume for low-priced stocks, implying that EDGX takes the other 43.13% of the trading volume. For medium-cap high-priced stocks, EDGA takes 43.02% of the volume while EDGX takes 56.98% of the volume. Therefore, the taker/maker fee market takes a relatively greater market share in low-priced stocks, whereas the maker/taker fee market takes a relatively higher market share in high-priced stocks. This demonstrates that liquidity providers are more willing to pay a fee to make a market for low-priced stocks. Our result is consistent with the seminal paper on maker/taker fee by Foucault, Kadan, and Kandel (2013). In their model, the exchange adjusts maker/taker fee to balance the arrival rate of liquidity makers and takers. Therefore, the optimal maker/taker fee is a function of tick size. To increase the trading volume and the total fee charged, a stock exchange should charge liquidity makers and subsidize liquidity takers when the tick size is large, and charge liquidity takers and subsidize liquidity maker when tick size is small. In reality, exchanges cannot adjust maker/taker fees based on stock characteristics. However, economic forces imply that stocks with relatively large tick sizes will be traded at higher volumes in the taker/maker fee market and stocks with relatively small tick sizes will be traded at higher volumes in the maker/taker fee market.

Insert Table 3 about here

3.4. Discussion

Our paper is the first to link the literature on high-frequency trading with the literature on the maker/taker fee, and the link involves price constraints and time priority. When price competition in the nominal spread is constrained, liquidity providers have two options. They can compete on speed to obtain priority of execution, or pay a fee to bypass the price constraint. Time priority is also essential here. First, time priority implies that traders that achieve higher speeds have priority in the queue. A limit order at the front of the queue at the constrained price earns a greater expected profit than other limit orders in the queue, due to higher execution probability. Second, in the United States, time priority is imposed only in each market center but not across separate markets. Therefore, both the taker/maker market and the maker/taker market have their own queues. This is also an important feature that drives the current competitive landscape. If time priority were enforced across all market centers, the taker/maker market would not exist under the trade-through rule because liquidity providers would all go to the maker/taker market to collect the fee. Time priority in each market center implies that liquidity providers can jump ahead of the queue by paying a fee on the taker/maker market.

Because high-frequency liquidity provision and the taker/maker market are driven by the same common factors, they are highly correlated. We sort the 117 stocks into five portfolios. Portfolio 1 stocks have the lowest EDGA-to-EDGX market share ratio and portfolio 5 stocks have the highest. Figure 4 demonstrates that the portfolio with the highest market share in the taker/maker market (quintile 5) has the highest high-frequency liquidity provision. In terms of trade, Panel A shows that 37% of trades have high-frequency liquidity providers, and Panel B shows that 44% of the depth is provided by high-frequency traders. High-frequency liquidity provision decreases monotonically with the market share of the taker/maker fee market. Only

19% of trades have high-frequency liquidity providers for quintile 1, the stocks with the lowest market share in the taker/maker market. High-frequency traders provide only 21% of the depth for stocks in quintile 1. Together these numbers represent only half of the percentage taken by stocks in the highest EDGA-to-EDGX market share ratio quintile.

Insert Figure 4 about here

4. Stock Splits, Speed Competition and Taker/maker Market

Section 3 establishes the relationship between price constraint, speed competition, and the taker/maker market based on the cross-sectional variation in stock prices. We provide a robustness test of the results based on stock splits. In addition, based on the fact that the tick size for stocks with a price lower than $1 is 0.01 cent, a discontinuity test can be conducted. We do not present the results of the discontinuity test because they are subject to alternative explanations.[19]

Unfortunately, only one of the 117 stocks split in the high-frequency trading data. We examine all firms that declared a two-for-one or greater stock split between January 2010 and November 2011 in the CRSP universe. Each of our pre- and post-event windows is comprised of the 30 trading days immediately before the stock-splitting date, and 30 trading days immediately after the stock-splitting date, including the splitting date. We exclude stocks that split more than once during the sample period. Among these stocks, 85 firms have trading data in the ITCH dataset and 66 have data on EDGA and EDGX volumes.[20] We do not have high-frequency

[19] When the price of a stock falls below $1, it is subject to delisting. When stocks have lower levels of high-frequency activity once the price falls below $1, it can either because tick size is smaller or high-frequency traders try to avoid the risk of delisting.
[20] We have fewer data on EDGA and EDGX volumes because the data were not available from January 2010–June, 2010.

liquidity-provision data for these stocks. Fortunately, Foucault, Kadan, and Kandel (2013) provide a proxy for high-frequency liquidity-making. In their model, an increase in tick size increases liquidity makers' profits. As a consequence, the monitoring intensity of liquidity-makers increases. This reduces time in the liquidity-making cycle, which reduces the time gap between an execution of a limit order and the submission of another limit order. Following this model, we estimate the speed of liquidity provision as follows: suppose that an order is executed with an E message; we then look for the closest order submission with the same size and side. Next, we calculate the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th percentiles of resubmission time for each stock on each day. Table 4 shows the correlation of the log of the resubmission speed with our measure of high-frequency liquidity-making for the 117 stocks. The first column shows the correlation between resubmission speed and the percentage of volume with high-frequency traders as liquidity providers, and the second column shows the correlation between resubmission speed and the percentage of depth provided by high-frequency traders. The 1st–75th resubmission speed are negatively correlated with high-frequency liquidity-making, which means that a lower resubmission time, or a higher resubmission speed, is associated with a larger fraction of high-frequency market-making. The correlation becomes insignificant from the 90th percentile to the 99th percentile. Therefore, we chose the 1st–50th resubmission time percentiles as our proxy for high-frequency liquidity-making.

Insert Table 4 About Here

Our sample period is 30 trading days before and 30 days after the split. To address the concern that speed might increase mechanically because of the time trend, we also match splitting stocks one-to-one with stocks that do not split based on price, market cap, and listing exchange. Therefore, for each stock that splits, we match it with a stock listed on the same exchange with minimal matching error D_{ij}, where the matching error is defined as:

$$D_{ij} = \left| \frac{MCAP_i}{MCAP_j} - 1 \right| + \left| \frac{PRC_i}{PRC_j} - 1 \right| \tag{7}$$

Next, we explore the relation between price constraints, speed competition, and the taker/maker market using the diff-in-diff approach. For speed, we run the following regression:

$$Speed_{it} = \alpha + \beta_1 treatment_i + \beta_2 after_t + \beta_3 treatment_i * after_t + e_{it} \tag{8}$$

where $treatment_i$ is equal to 1 for stocks that split and 0 for the matched sample, $after_t$ is equal to 1 after the splitting day for stock i and 0 before the split. The variable of interest is β_3, which measures the impact of the split on trading speed.

We run a similar regression on EDGA's market share relative to EDGX (Edgeratio) as the dependent variable:

$$Edgeratio_{it} = \alpha + \beta_1 treatment_i + \beta_2 after_t + \beta_3 treatment_i * after_t + e_{it} \tag{9}$$

where $Edgeratio_{it}$ is defined as the EDGA volume divided by the sum of the EDGA and EDGX volumes.

The results are presented in Table 5. The first five columns present the results on speed. The dependent variables are the logarithm of the 1^{st}, 5^{th} ,10^{th} ,25^{th} and 50^{th} percentile of limit order resubmission speed for stock i at day t. The constant term is the limit order resubmission speed of the control group before the treatment. For example, the 1^{st} percentile of the resubmission speed is exp (-7.83) =.000397 seconds and the 50^{th} percentile of the resubmission speed is exp (-0.935) = 0.393 seconds. The time scale of the resubmission is too short to have been submitted by humans. Nor would the change in time be driven by economic fundamentals.

We argue that the change is more likely to be driven by speed competition among high-frequency traders. The coefficient β_1 is not significant for the 1st–25th percentiles and only marginally significant for the 50th percentile, meaning that the order resubmission speed in the treatment group is very similar to the resubmission speed of the control group before the split. This result suggests that price, market cap, and listing exchange provide an appropriate control group for trading speed. The variable $after_t$ is not significant, implying that there is no general trend regarding trading speed. The coefficient of interest is β_3, which measures the average treatment effect of the stock split on order resubmission speed. Because the dependent variable representing speed is expressed in logarithmic format, the coefficient can be interpreted approximately as the percentage change in the resubmission speed of the treatment group relative to that of the matched stocks that do not split. For example, the 1st percentile of the resubmission latency of the split stocks decreases by 25% after the split, while the 50th percentile of the resubmission latency decreases by 33.2%. The results show that, after the stock split, there is a dramatic increase in order-resubmission speed. The sixth column demonstrates the results for EDGEratio. We find that Edge A's volume increases by 7.69% relative to that of Edge X.

In summary, the results based on a splitting complement the results presented in section 3. Section 3 shows that low-priced stocks experience more high-frequency liquidity-making and are traded at higher volumes in the taker/maker market relative to the maker/taker market. This section shows that stock splits, or reduced nominal stock prices, increase speed competition. Meanwhile, there is a shift in volume from the maker/taker market to the taker/maker market. These results further strengthen our main argument regarding price constraints. With a regulated minimum tick size of one cent, low-priced stocks suffer from a more severe price constraint. The constrained price spurs speed competition because of time priority, while also resulting in an active taker/maker market.

Insert Table 5 About Here

5. Speed and Market Quality

Jones (2013) argues that the impact of high-frequency trading depends on the nature of competition. The trade-off involves the competition effect and the adverse selection effect. If competition reduces the cost of providing liquidity, liquidity improves. If high-frequency traders increase adverse selection due to the speed at which they process information, liquidity decreases. Meanwhile, stock prices are more efficient because they reflect more information more quickly.

Our paper identifies another channel of speed competition: the price constraints channel. The price constraints channel implies that speed competition has no impact on liquidity. Using exogenous technology shocks as identification strategies, we find that speed does not improve either the quoted spread or the effective spread. Sections 3 and 4 provide the intuition needed to understand this result. When speed competition is a consequence of failed price competition, a reduction in latency cannot decrease the cost of liquidity. Speed may determine which trader provides liquidity, but individual liquidity providers cannot undercut each other on price. Speed may provide a mechanism for allocating the rents created through price constraints, but it does not change the price of the liquidity.

Stock market liquidity is defined as the ability to trade a security quickly at a price close to its consensus value (Foucault, Pagano, and Röell, 2013). Therefore, liquidity has two dimensions: price concession and time. The literature on liquidity proposes three measures of liquidity: spread, depth, and resiliency. Spread is the transaction cost faced by traders, and is often measured by the quoted bid-ask spread or the trade-based effective spread. Depth reflects the market's ability to absorb large orders with minimal price impact, and is often measured by the quoted depth. Spread and depth comprise the price dimension of liquidity. Resiliency is the speed with which a temporary erosion of depth caused by a large uninformative order-flow shock is corrected or neutralized through the flow of new orders.

By their very nature, the technology shocks we have identified improve the speed dimension regarding liquidity. Nanosecond technology enables liquidity suppliers to add orders by nanoseconds instead of microseconds. Still, we know of no paper in the literature on high-frequency trading that has examined this dimension of liquidity. The key question is to determine whether high-frequency trading in the aggregate harms or improves market quality as perceived by long-term investors (Hasbrouck and Saar, 2013). Therefore, the literature on high-frequency trading focuses on the price dimension, finding that speed decreases the spread and increases depth in the millisecond or second era. The literature does not show, however, whether such a benefit extends to the nanosecond era, in which liquidity suppliers and demanders are both computerized.

The results are demonstrated using two tests based on the type of technology shock involved and the associated econometric method. Section 5.1 computes the measures of liquidity. Section 5.2 discusses the impact of technology shocks that reduce the latency involved in submitting or matching orders from microseconds to nanoseconds. Because the NASDAQ matching engine is single-threaded, the technology shocks result in a single time series before and after treatment without cross-sectional differences. We use the method of Boehmer, Saar, and Yu (2005) to control for other variables that might affect liquidity. Section 5.3 demonstrates the effects of a technology shock that reduces the latency of trading data dissemination from 3 milliseconds to 1 millisecond. This technology was first implemented in the channel that handles ticker symbols beginning with A and B. This creates a double difference estimator for our analysis. The methodology follows Hendershott, Jones, and Menkveld (2011). These two types of technology shocks reveal the same effect: an increase in speed at or below the 1-millisecond level does not affect our usual measure of liquidity.

5.1 Measure of Dependent Variables

Our measures of liquidity comes from ITCH data. We construct a message-by-message limit-order book such that the book is updated whenever there is a new message. That is, any order addition, execution, or cancellation leads to a new order book. For example, Microsoft has about 1.08 million messages on an average trading day, and we generate and store all the resulting 1.08 million order books. This provides the most accurate view of the limit-order book at any point in time. The construction is implemented by the Gordon Supercomputer in the San Diego Supercomputing Center.

We calculate four measures of liquidity. Two are spread measures: the time-weighted quoted spread and the size-weighted effective spread. The other two are depth measures: depth at the best bid and ask, and depth within 10 cents of the best bid and ask.[21] Since we construct a full limit-order book, the quoted spread is measured as the difference between the best bid and ask at any given time. Each quoted spread is weighted based on the life of the quoted spread to obtain the daily time-weighted quoted spread for each stock per day. The effective spread for a buy is

[21] The 10-cent cutoff is used by Hasbrouck and Saar (2013).

defined as twice the difference between the trade price and the midpoint of the best bid and ask price. The effective spread for a sell is defined as twice the difference between the midpoints of the best bid and ask and the trade price. The size-weighted effective spread is defined as the size-weighted effective spread of all trades for each stock and each day. The two depth measures, depth at the best bid and ask and depth within 10 cents of the best bid and ask, are weighted using the time for each stock per day.

5.2 Shocks at Matching Engine Level: Single Time-Series Differences

Technology shocks identified through ITCH data have a single time series. We follow the methodology of Boehmer, Saar, and Yu (2005) by running regressions on the event dummy and control variables. In their econometric specification, the liquidity measure for stock i in period t (where t$\in \{before, after\}$), L_{it}, is expressed as the sum of the stock fixed effect (γ_i), an event effect (α), a set of control variables, and an error term (ε_{it})

$$L_{it} = \gamma_i + \alpha After_t + \beta_1 AvgVol_{it} + \beta_2 range_{it} + \beta_3 AvgPrc_{it} + \varepsilon_{it,} \tag{10}$$

where $AvgVol_{it}$ is the average volume of the stock for period t, $range_{it}$ is the volatility measure, which is defined as day high minus day low normalized by the closing price and then is averaged over period t, and $AvgPrc_{it}$ is the average daily price for period t. The equation can be estimated by removing the stock fixed effect:

$$\Delta L_i = \alpha + \beta_1 \Delta AvgVol_i + \beta_2 \Delta range_i + \beta_3 \Delta AvgPrc_i + \eta_{i,} \tag{11}$$

where Δ denotes the difference between the after and before periods.

Panel A of Table 6 demonstrates the results for the technology shock that occurred between April 9, 2010 and April 12, 2010, and Panel B shows the results for the technology shock that occurred between May 21, 2010 and May 24, 2010. For each shock, we use 10 trading days before and after the shock as the event window. Table 6 shows that these two technology shocks do not have a statistically or economically significant impact on the spread. For the first technology shock, the quoted spread decreased by 0.142 cents and the effective spread decreased by 0.011 cents, neither of which is significant. For the second technology shock, the effective spread actually increased by 0.02 cents, although it is also not significant. Depth increases slightly after the second technology shock whereas it decreases after the first technology shock. For example, the depth within 10 cents decreases by 4035 shares after the first technology shock whereas the depth within 10 cents increases by 4494 shares.

Insert Table 6 about here

The fact that speed does not decrease the spread has two natural explanations. First, the exchange follows price priority, whereby the competition to provide liquidity plays out first at the price level. Time priority has a secondary role, coming into play only after the price is determined. The fact that there is intense competition on speed implies that there is little room for price competition at the best bid and ask. As a result, the spread might decrease slightly while speed increases noticeably. Second, one argument for the proposition that speed can increase liquidity is that traders who achieve high speeds can maintain tighter bid-ask spreads because they can quickly update stale quotes before other traders can adversely select them. This argument, however, confirms only that relative speed matters: the trader achieving the highest speed may be able to post the tightest quotes. If traders' speeds increase twofold, the equilibrium level of the spread may not change at all. If the fastest trader is surpassed by the second-fastest trader, the latter may have the ability to quote the tightest spread, but the level of the spread may be the same as it was originally. To summarize, intense speed competition implies that there may be little room for further improving the best bid and offer. Traders with the highest speed may be able to maintain the best bid and ask spread, but the level of bid and ask is unlikely to change.

5.3 Shocks that Increase the Dissemination of Data: The diff-in-diff test

The six channel assignments for NASDAQ stocks provide both cross-sectional and time-series variations of the technology improvements. The six channels all had 3-millisecond latency regarding dissemination of information before October 7, 2011. The technology enhancement that occurred during the weekend of October 8 and October 9, 2011 decreased the latency to 1 millisecond for one channel while leaving the other five channels unaffected. The remaining five channels were upgraded during the following weekend. On October 17, all channels registered 1-millisecond latency.

Those two enhancements enable us to run a diff-in-diff regression following Hendershott, Jones, and Menkveld (2011). The diff-in-diff estimation equation specification uses time fixed effect and firm fixed effect to handle multiple time periods and multiple treatment groups (Roberts and Whited, 2012):

$$L_{it} = \gamma_i + \lambda_t + \delta D_{it} + \beta X_{it} + \varepsilon_{it} \tag{12}$$

where L_{it} is the liquidity measure for stock i on day t, γ_i is the stock fixed effect, λ_t is the time fixed effect, and X_{it} is a vector of control variables. Following Hendershott, Jones, and Menkveld (2011), the control variables include 1 over the daily closing price ($invprc_{it}$), share turnover ($turn\_ratio_{it}$), day high minus day low normalized by the closing price ($volatility_{it}$) and the logarithm of the market cap ($logmktcap_{it}$). D_{it} is the key variable of interest, which equals 1 after a stock move from the 3-millisecond range to the 1-millisecond range. To be more specific, from October 3, 2011 through October 7, 2011, $D_{it} = 0$ for all channels; from October 10, 2011 through October 14, 2011, $D_{it} = 1$ for channel 1 and $D_{it} = 0$ for channels 2–6; and from October 17, 2011 through October 21, 2011, $D_{it} = 1$ for all channels.

The original 120 stocks selected by Hendershott and Riordan include only 60 NASDAQ-listed stocks, which results in 10 stocks per channel. We do not find that speed improves liquidity for these 60 stocks, but that result might be a consequence of the small sample size. Therefore, we followed Hasbrouck and Saar (2013) and selected the top 500 NASDAQ common stocks based on the market cap of September 30, 2011. The results for the 500 stocks are summarized in Table 7, where L_{it} t represents the daily liquidity variable measured as the time-weighted quoted spread, the size-weighted effective spread, the time-weighted depth at the best bid and ask, and the time-weighted depth within 10 cents of the best bid and ask. The results show that the enhancement of the consolidated tape does not change the quoted spread, the effective spread, or depth at the best bid and ask. Depth within 10 cents of the bid-ask spread, however, decreased slightly. Again, our results suggest that speed does not improve liquidity.

Insert Table 7 about here

6. Conclusion

This paper provides a unified empirical explanation of two important features of current stock markets: speed competition and competition between the maker/taker and taker/maker markets. We argue that two deviations from Walrasian equilibrium, discrete price and time priority, are the economic forces that drive these competitions.

Because rule 612 of regulation NMS prohibits price competition at increments less than one cent, speed allocates executions when the equilibrium bid-ask spread is smaller than one tick. Therefore, speed competition in liquidity provision is more intense for large stocks at low prices, which in turn have larger relative tick sizes. These stocks also take a larger market share in the taker/maker market, where the price constraint can be bypassed by paying a fee. We also find that a decrease in the nominal share price due to a stock split is associated with an increase in trading speed and a migration of volume from the maker/taker market to the taker/maker market. These results, again, suggest that price constraints and time priority are the two common factors that drive speed competition and the taker/maker market.

Not surprisingly, if speed competition is a consequence of constrained price competition, an increase in speed will not impact liquidity. We find that, following two specific technology shocks that exogenously increase the speed of matching orders from the microsecond level to the nanosecond level, the spread does not decrease. We also explore technology shocks that increase the speed involved in disseminating trading data. These shocks occurred in a staggered manner based on the alphabetic order of stock names, which provides ideal randomness for the diff-in-diff test. Again, we find similar results: an increase in speed does not improve liquidity.

Policymakers are very concerned about speed competition. Currently, there are several proposed policies—such as imposing a minimum quote life, charging cancellation fees, and levying transaction taxes—that are designed to address these concerns. On the other hand, the finance literature before the sub-millisecond era suggests that an increase in speed benefits liquidity, implying no regulation is needed. Our results, supported by economic reasoning, provide a relatively neutral view of high-frequency trading: high-frequency trading neither improves nor harms liquidity. Regulation 612, which constrains price competition, together with the time-priority rule, encourages speed competition at constrained prices. Our results suggest that speed competition above a certain threshold is a zero-sum game, the existence of which is facilitated by the rent that is created through price constraints. Price constraints also facilitate the co-existence of and competition for markets with varying fee structures.

Therefore, in the debate over whether and how we should regulate high-frequency traders, we should first be aware that high-frequency trading may be a consequence of existing regulations. Therefore, if policymakers are concerned about speed competition, one solution would be to decrease tick size, which would force competition to focus more sharply on price. Interestingly, from an economic perspective, this is deregulation instead of regulation, because it moves the market closer to Walrasian equilibrium. The other solution would be to decrease the importance of time priority below the millisecond level, where orders that arrive in the same millisecond share the same time priority. This policy would differ from imposing a minimum quote life, although it would slow the market as well. A minimum quote-life policy would change the speed of liquidity providers relative to that of liquidity demanders, but the consequence of such a policy is ambiguous. Decreasing the importance of time priority, however, would affect all traders in a similar manner. Budish, Cramton, and Shim (2013) provide a theoretical argument on the benefits of decreasing time priority.

Certainly, the two abovementioned policies need to be justified by both implementation benefits and costs. We are concerned, however, about a recent trend in the other direction. According to Bloomberg, the SEC is considering increasing the tick size for small stocks, reasoning that "[m]arket makers and dealers need more economic incentives to bring smaller companies public, provide bids and offers and publish stock research." The argument that price constraints lead to non-price competition is valid, but we doubt whether non-price competition will involve publishing stock research. A more direct way to create non-price competition is speed competition. Our results suggest that an increase in tick size can facilitate another round of speed competition.

Ours is the first empirical paper to define price constraints and demonstrate their impact on high-frequency trading and the taker/maker fee markets, and it can be extended in various ways. First, current theoretical work on speed competition focuses on the role of information. Our paper points out another speed competition channel: price constraints. Models using discrete prices can be constructed to understand the value of speed and the impact of price constraints on market quality. Second, we explain the market share of taker/maker fees based on price constraints, and theoretical models can be built to understand why there exist separate equilibria for traders on separate trading platforms and how exchanges set fee structures in both markets to maximize total profits. Empirically, the relationship between price constraints, speed competition, and maker/taker versus taker/maker fees can be further explored. For example, the SEC recently announced a pilot program for increasing the tick size for a number of small stocks, and it would be interesting to see the impact of this shock on speed competition and the taker/maker fee market.

REFERENCES

Angel, J., 1997, Tick Size, Share Prices, and Stock Splits, *Journal of Finance* 52, 655-681.

Baker, H., and P.L. Gallagher, 1980, Management's View of Stock Splits, *Financial Management* 9, 73-77.

Baker, H., and G. Powell, 1993, Further Evidence on Managerial Motives for Stock Splits, *Quarterly Journal of Business and Economics* 32, 21-31.

Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2013, Equilibrium High-Frequency Trading, Working paper, Toulouse School of Economics, HEC Paris and Toulouse School of Economics.

Boehmer, Ekkehart, Gideon Saar and Lei Yu, 2005, Lifting the veil: An analysis of pre-trade transparency at the NYSE, *Journal of Finance* 60, 783-815.

Brolley, Michael, and Katya Malinova, 2012, Informed trading and maker-taker fees ina low-latency limit order market, Working paper, University of Toronto.

Brogaard, Jonathan and Hagströmer, Björn and Norden, Lars L. and Riordan, Ryan, 2013, Trading Fast and Slow: Colocation and Market Quality, working paper

Brogaard, Jonathan, Terrence Hendershott and Ryan Riordan, 2013, High frequency trading and price discovery, forthcoming, *Review of Financial Studies*

Budish, Eric; Cramton, Peter; Shim, John (2013) The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response, Working Paper

Byun, J. and M.S. Rozeff, 2003, Long-run Performance after Stock Splits: 1927 to 1996, *Journal of Finance* 58, 1540-6261.

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2009, Rise of the machines: Algorithmic trading in the foreign exchange market, Working paper, Board of Governors of the Federal Reserve System.

Colliard, Jean-Edouard and Foucault Thierry, 2012, Trading fees and efficiency in limit order markets" *Review of Financial Studies*, 25, 3389-3421.

Fernando, C., S. Krishnamurthy and P. Spindt, 1999, Is Share Price Related to Marketability? Evidence from Open-end Mutual Fund Share Splits, *Financial Management* 28(3), 15-31.

Fernando, C., S. Krishnamurthy and P. Spindt, 2004, Are Share Price Levels Informative? Evidence from the Ownership, Pricing, Turnover, and Performance of IPO Firms, *Journal of Financial Markets* 7, 377-403.

Foucault, Thierry, Ohad Kadan and Eugene Kandel, 2013 Liquidity cycles, and make/take fees in electronic market, *Journal of Finance* 68, 299-341.

Foucault, Thierry, Marco Pagano and Alisa Röell, 2013, Market Liquidity, Oxford University Press.

Hasbrouck, Joel, and Gideon Saar, 2009, Technology and liquidity provision: the blurring of traditional definitions, *Journal of Financial Markets* 12, 143-172.

Hasbrouck, Joel, and Gideon Saar, 2013, Low-latency trading, Working paper, *Journal of Financial Markets*, Available online 22 May 2013, ISSN 1386-4181

Hendershott, Terrence, and Ryan Riordan, 2011, Algorithmic trading and information, Working paper, University of Washington and University of California, Berkeley.

Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2011, Does algorithmic trading improve liquidity? *Journal of Finance* 66, 1-33.

Hirschey, Nicholas H., 2012, Do High-Frequency Traders Anticipate Buying and Selling Pressure? Working paper, London Business School.

Jones, Charles, 2013, What Do We Know about High-Frequency Trading? Working Paper, Columbia University

Jovanovic, Boyan, and Albert J. Menkveld, 2011, Middlemen in limit order markets, Working paper, VU University Amsterdam.

Kavajecz, Kenneth A., 1999, A specialist's quoted depth and the limit order book, *Journal of Finance* 54, 747-771.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tuzun Tuzun, 2011, The flash crash: The impact of high-frequency trading on an electronic market, Working paper, MIT Sloan School of Management, University of Maryland, University of North Carolina at Chapel Hill, and Federal Reserve Board.

Lakonishok, J., and B. Lev, 1987, Stock Splits and Stock Dividends: Why, Who and When, *Journal of Finance* 42, 913-932.

Lo, Andrew W., A. Craig MacKinlay, 1988, Stock market prices do not follow random walks: Evidence from a simple specification test, Review of Financial Studies 1, 41-66.

Madhavan, Ananth, 2011, Exchange-Traded Funds, Market Structure and the Flash Crash Working paper

O'Hara, Maureen, Chen Yao, and Mao Ye, 2011. What's not there: The odd-lot bias in market data, Working paper, Cornell University, University of Illinois at Urbana-Champaign, and University of Illinois at Urbana-Champaign.

Malinova, Katya and Andreas Park, Subsidizing liquidity: the Impact of make/take fees on market quality, forthcoming in *Journal of Finance*

Patterson, Scott, 2012, *Crown Business*, New York.

Pagnotta, Emiliano and Thomas Philippon, 2012, Competing on speed, Working paper, New York University.

Roberts, Michael R. and Whited, Toni M., 2012, Endogeneity in Empirical Corporate Finance Simon School Working Paper No. FR 11-29.

Table 1: The Seven Types of Messages Used to Construct the Limit-Order Book

This table provides the format of the seven types of messages used to construct the limit-order book. The sample is from May 24, 2010.

Message Type	Timestamp (second.nanosecond)	Order Reference Number	Buy/Sell	Shares	Stock	Price	Original Order Reference Number	Match Number	Market Participant ID
A	53435.759668667	335531633	S	300	EWA	19.5			
F	40607.031257842	168914198	B	100	NOK	9.38			UBSS
U	53520.367102587	336529765		300		19.45	335531633		
E	53676.740300677	336529765		76				7344037	
C	57603.003717685	625843333		100		32.25		20015557	
X	53676.638521222	336529765		100					
D	53676.740851701	336529765							

Table 2: High-Frequency Trading and Price

This table presents market shares of high-frequency liquidity provision. The 117 stocks in NASDAQ high-frequency data are sorted first by market cap and then by price into 3 by 3 portfolios. Panel A presents the volume-weighted depth at best bid and offer (BBO) provided by high-frequency traders and Panel B presents volume weighted share of volume with high-frequency liquidity providers. To calculate the volume-weighted average, we first aggregated the liquidity provision of high-frequency traders and all traders for stocks in the portfolio for all days. Volume weighted average is defined as the ratio of the aggregated high frequency liquidity provision divided by liquidity provided by all traders. Panel C presents the equally weighted depth provided by high-frequency traders and Panel D presents equally weighted share of volume with high-frequency liquidity providers. To define the equally weighted average, we first compute, for each stock on each day, the liquidity provided by high frequency traders relative to all liquidity provision. The equally weighted average is defined as the average of these stock date observations.

Panel A: Percentage of BBO Depth Provided by High-Frequency Traders (Volume Weighted)			
	Low Price	Medium Price	High Price
Large Cap	55.62%	46.17%	33.16%
Mid Cap	39.60%	30.26%	23.67%
Small Cap	24.92%	23.19%	21.83%

Panel B: Percentage of Trading Provided by High-Frequency Traders (Volume Weighted)			
	Low Price	Medium Price	High Price
Large Cap	49.29%	38.48%	35.53%
Mid Cap	39.15%	23.56%	22.34%
Small Cap	23.40%	19.93%	18.74%

Panel C: Percentage of BBO Depth Provided by High-Frequency Traders (Equally Weighted)			
	Low Price	Medium Price	High Price
Large Cap	49.60%	43.76%	27.54%
Mid Cap	43.05%	27.87%	24.40%
Small Cap	21.24%	25.42%	21.81%

Panel D: Percentage of Trading with High Frequency Liquidity Providers (Equally Weighted)			
	Low Price	Medium Price	High Price
Large Cap	45.40%	37.50%	30.90%
Mid Cap	35.50%	23.40%	22.20%
Small Cap	18.60%	18.90%	18.50%

Table 3: Taker/maker Market and Price

This table presents the ratio of Direct Edge A (EDGA) volume to the total volume of Direct Edge (the volume of EDGA plus Direct Edge X (EDGX)). EDGA charges liquidity providers and pays rebates to liquidity takers, whereas EDGX charges liquidity takers and pays rebates to liquidity makers. The sample includes 117 stocks in the NASDAQ high-frequency data. The stocks are sorted first by market cap and then by price. Panel A is on volume weighted average. We first aggregated the EDGA and EDGX volumes for stocks in the portfolio for all days. Volume weighted average is defined as the ratio of the aggregated EDGA volume divided by the aggregated EDGA volume plus the aggregated EDGX volume. Panel B is on equally weighted average. We first compute, for each stock i on each day t, the market share of EDGA relative to that of EDGA and EDGX. Then we average these observations equally across stocks and dates.

Panel A: Volume weighted Direct Edge A Volume / (Direct Edge A + Direct Edge X Volume)			
	Low Price	Medium Price	High Price
Large Cap	64.28%	54.57%	28.98%
Mid Cap	55.94%	40.95%	38.33%
Small Cap	43.56%	29.79%	24.53%
Panel B: Equally weighted Direct Edge A Volume / (Direct Edge A + Direct Edge X Volume)			
	Low Price	Medium Price	High Price
Large Cap	57.94%	48.48%	35.51%
Mid Cap	56.87%	47.98%	43.02%
Small Cap	36.06%	29.74%	24.04%

Table 4: Correlation between High Frequency Market Making and Resubmission Time

This table presents the correlation between limit order resubmission time and high-frequency liquidity provision. *logpi* is the logarithm of the quickest *i* percentile of a limit order submission subsequent to an execution.

	HFT_Volume	HFT_Depth
logp1	-0.654***	-0.672***
P-Value	(0.000)	(0.000)
logp5	-0.605***	-0.630***
P-Value	(0.000)	(0.000)
logp10	-0.614***	-0.619***
P-Value	(0.000)	(0.000)
logp25	-0.649***	-0.645***
P-Value	(0.000)	(0.000)
logp50	-0.688***	-0.679***
P-Value	(0.000)	(0.000)
logp75	-0.517***	-0.678***
P-Value	(0.000)	(0.000)
logp90	0.028	-0.188**
P-Value	(0.765)	(0.044)
logp95	0.139	-0.045
P-Value	(0.138)	(0.634)
logp99	0.180*	0.086
P-Value	(0.053)	(0.358)

Table 5: Impact of Stock Split on Speed and Taker/maker Market

This table presents the results of the diff-in-diff regression. The specification for column 1-5 is

$$Speed_{it} = \alpha + \beta_1 treatment_i + \beta_2 after_t + \beta_3 treatment_i * after_t + e_{it}$$

where the proxies of the speed is the logarithm of the 1st, 5th 10th, 25th and 50th percentile of resubmission speed of a limit order subsequent to an order execution. $Treatment_i$ is equal to 1 for stocks that split and 0 for the matched sample, $after_t$ equal to 1 after the splitting day for stock 1 and 0 before the stock split.

The specification for column 6 is

$$Edgeratio_{it} = \alpha + \beta_1 treatment_i + \beta_2 after_t + \beta_3 treatment_i * after_t + e_{it}$$

where $Edgeratio_{it}$ is the volume of EDGA relative to the volume of EDGA plus EDGX. Standard errors are in parentheses, and ***, **, and * represent significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) logp1	(2) logp5	(3) logp10	(4) logp25	(5) logp50	(6) Edgeratio
after	0.0495	0.0445	0.0778	0.0876	-0.0221	-0.0172
	(0.0710)	(0.0715)	(0.0749)	(0.0818)	(0.0778)	(0.0116)
pilot	0.0737	0.0637	-0.0164	-0.117	-0.127*	-0.0494***
	(0.0698)	(0.0703)	(0.0736)	(0.0805)	(0.0765)	(0.0114)
pilot_after	-0.250**	-0.273***	-0.257**	-0.269**	-0.332***	0.0769***
	(0.0990)	(0.0997)	(0.104)	(0.114)	(0.108)	(0.0163)
Constant	-7.830***	-7.111***	-6.548***	-4.493***	-0.935***	0.310***
	(0.0501)	(0.0505)	(0.0529)	(0.0578)	(0.0549)	(0.00813)
Observations	9,767	9,767	9,767	9,767	9,767	6,884
R-squared	0.001	0.001	0.002	0.003	0.005	0.005

Table 6: The Impact of Speed Improvement in Submitting or Processing Orders on Liquidity

This table presents the event study of the technology shocks for the four liquidity measures. The regression is based the differences between post and pre-technology shock. The regression specification is

$$\Delta L_i = \alpha + \beta_1 \Delta AvgVol_i + \beta_2 \Delta range_i + \beta_3 \Delta AvgPrc_i + \eta_{i,}$$

where constant term α shows the change of liquidity measure (ΔL_i) after the technology shock. Δ*QSpread* is the difference in the time-weighted quoted spread, Δ*ESpread* is the difference in trade size-weighted effective spreads, ΔWtDepth is the difference in depth at the best bid and ask, Δ*WtDep10* is the difference in cumulative depth for orders 10 cents below the best bid and 10 cents above the best ask. $\Delta AvgVol$ is the difference of average volume, $\Delta Range_{it}$ is the difference between day high and day low normalized by the closing price, and $\Delta AvgPrc_{it}$ is the difference in average daily price. Panel A is based on 10 days before and 10 days after the technology shock between April 9, 2010 and April 12, 2010 and Panel B is based on 10 days before and 10 days after the technology shock between May 21, 2010 and May 24, 2010. Robust standard errors are in parentheses, and ***, **, and * represent significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Technology Shock between April 9, 2010 and April 12, 2010				
	(2)	(1)	(3)	(4)
VARIABLES	ΔQSpread	ΔESpread	ΔWtDep	ΔWtDep10
ΔAvgVol	-0.00955*	-0.00391***	1,371	-27,805
	(0.00486)	(0.00139)	(1,012)	(16,822)
ΔAvgHiLow	0.645*	0.167	-11,344	-74,124
	(0.383)	(0.124)	(9,274)	(98,793)
ΔAvgPrc	0.00201**	0.000597***	-30.58	602.1
	(0.000826)	(9.85e-05)	(30.78)	(466.9)
Constant	-0.00142	-0.000110	-45.25	-4,035***
	(0.00105)	(0.000302)	(123.3)	(1,238)
Observations	117	117	117	117
R-squared	0.338	0.327	0.048	0.133
Panel B: Technology Shock between May 21, 2010 and May 24, 2010				
	(2)	(1)	(3)	(4)
VARIABLES	ΔQSpread	ΔESpread	ΔWtDep	ΔWtDep10
ΔAvgVol	0.0141	-0.00189	-3,765*	-60,726**
	(0.0150)	(0.00267)	(1,934)	(25,682)
ΔAvgHiLow	0.0503	-0.0169	-768.5	5,039
	(0.146)	(0.0898)	(10,192)	(106,862)
ΔAvgPrc	0.000330	0.00102***	43.29	735.3
	(0.00206)	(0.000356)	(33.34)	(522.4)
Constant	-0.000775	0.000256	273.8***	4,494***
	(0.00295)	(0.000808)	(98.42)	(1,362)
Observations	117	117	117	117
R-squared	0.020	0.242	0.137	0.213

Table 7: The Impact of Speed Improvement of the Consolidated Tape on Liquidity

This table presents data that demonstrate the impact of speed enhancement of the consolidated tape on liquidity using diff-in-diff regression. The regression specification is

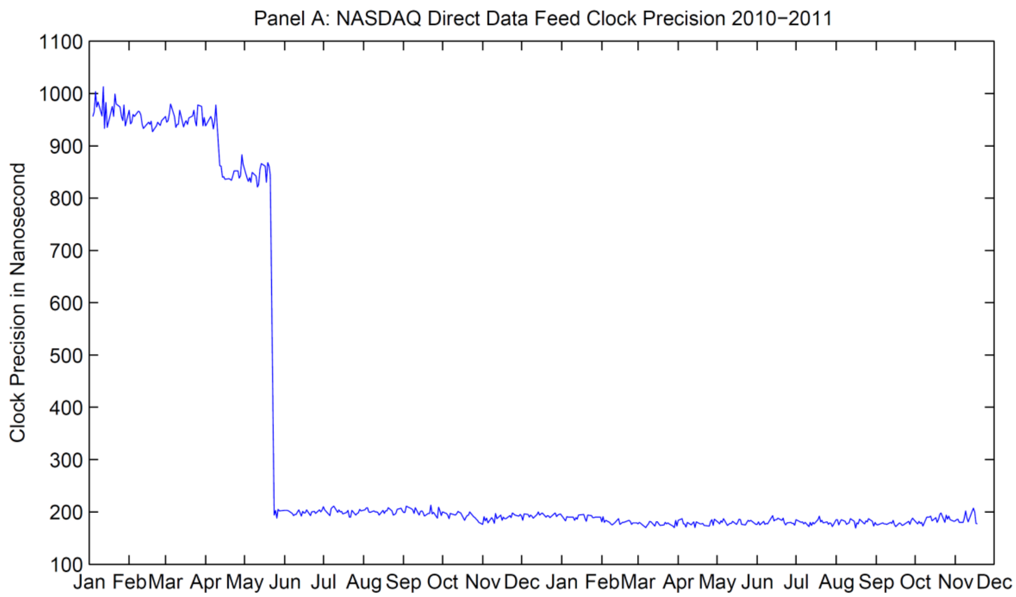
$$L_{it} = \gamma_i + \lambda_t + \delta D_{it} + \beta X_{it} + \varepsilon_{it}$$

where L_{it} is the liquidity measure for stock i on day t, γ_i is stock fixed effect, λ_t is time fixed effect, and X_{it} is a vector of control variables including share turnover, volatility, the inverse of share price, and the log of the market cap. The sample period is from October 3, 2011 to October 21, 2011. D_{it} is dummy variable for speed enhancement that is equal to 1 for channel 1 after October 7, 2011 and for channels 2–6 after October 14. Robust standard errors are in parentheses, and ***, **, and * represent significance at the 1%, 5%, and 10% levels, respectively.

VARIABLES	(1) QSpread	(2) ESpread	(3) WtDep	(4) WtDep10
D_{it}	0.00146	0.000490	-29.00	-2,193**
	(0.00139)	(0.000449)	(154.3)	(1,116)
invprc	-0.818***	-0.197***	87,002***	671,899***
	(0.0889)	(0.0242)	(19,396)	(82,312)
volatility	0.0178	0.0183***	-6,366	-84,690**
	(0.0142)	(0.00460)	(5,648)	(36,651)
logmktcap	-0.0127***	-0.00178*	1,841***	27,803***
	(0.00336)	(0.000946)	(385.6)	(4,648)
turn_ratio	-0.106***	-0.0213***	18,685	139,525
	(0.0291)	(0.00679)	(12,936)	(87,281)
Number of stockid	500	500	500	500
R-squared	0.264	0.164	0.170	0.432

Figure 1: Technology Shocks on Submitting and Cancelling Orders

These figures illustrate the impact of our two technology shocks on latency. The first technology shock occurred between April 9, 2010 and April 12, 2010. The second shock occurred between May 21, 2010 and May 24, 2010. We have two measures of latency. Panel A shows the minimum time differences between two consecutive messages for the NASDAQ market. Panel B shows the fastest cancellation and execution for the NASDAQ market.



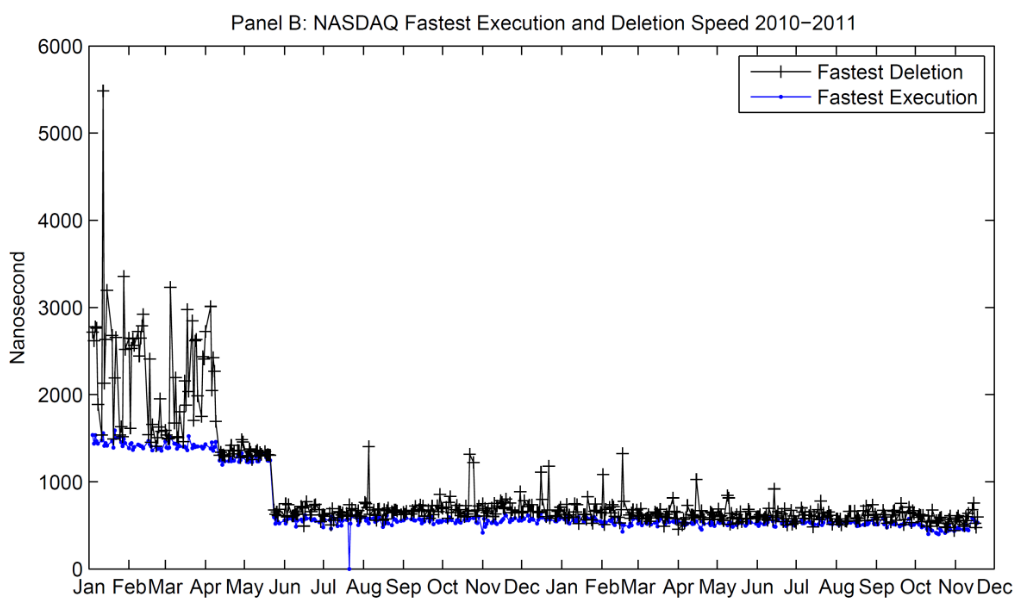


Figure 2: Technology Shocks on Transmitting Trading Data

These graphs illustrate the effects of the technology shocks on the consolidated tape. Before October 7, 2011 messages are broadcast only at milliseconds ending with 0, 3, or 7, implying a minimum gap of 3 milliseconds. On October 10, 2011 Channel 1 was able to broadcast information every millisecond but other channels continued to exhibit 3-millisecond gaps until October 14, 2011. On October 17, 2011, Channels 2–6 are also upgraded and there is no gap to broadcast messages.

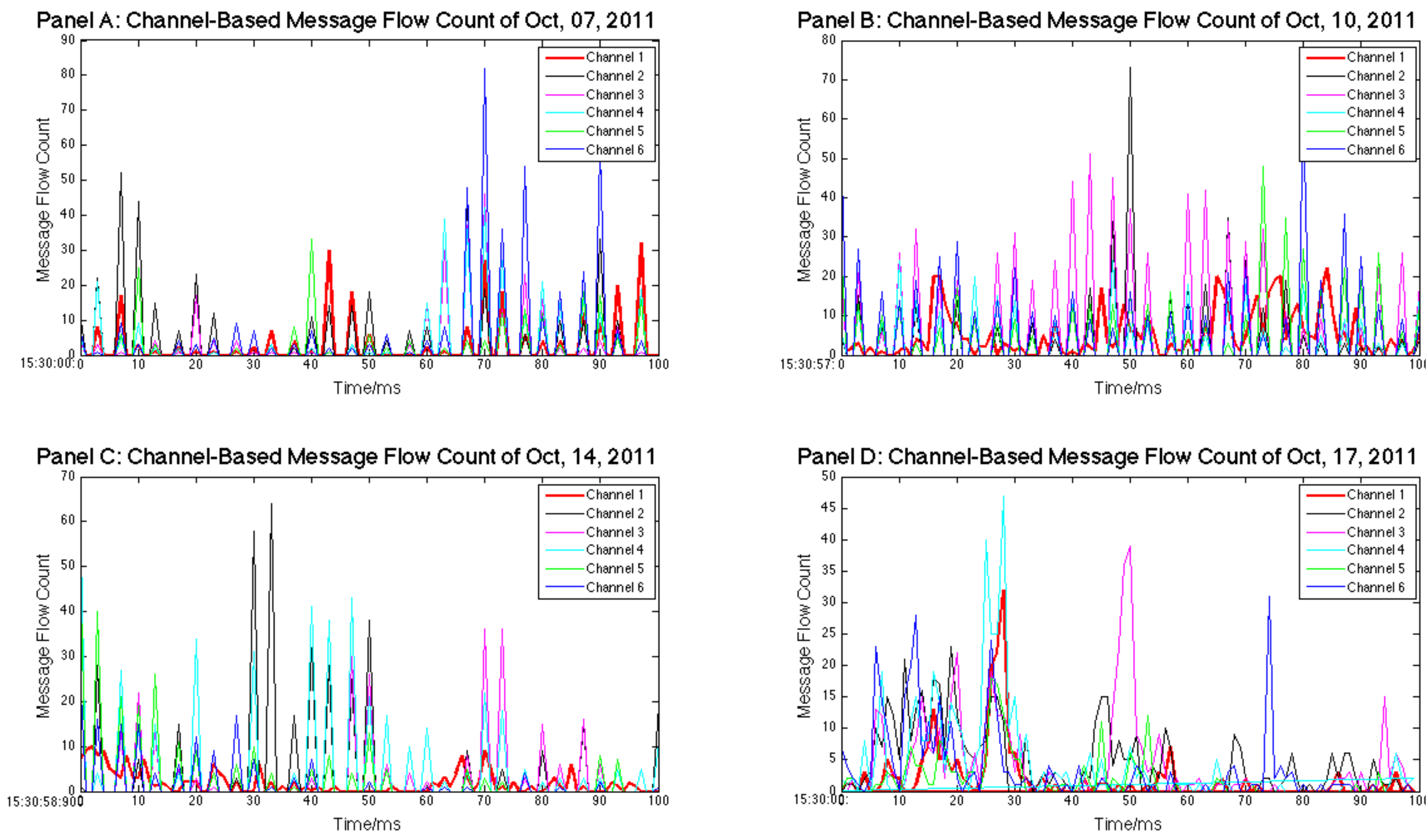


Figure 3: Median Time Gaps between Two Broadcasts

This graph depicts the median time gaps between two broadcasts based on all quotes in the six channels. The horizontal axis represents the trading days between September 26, 2011 and November 1, 2011, and the vertical axis represents the time in milliseconds. Before October 7, 2011, the median time gap between two broadcasts for all six channels is 3 milliseconds. On October 10, 2011, the median gap is 1 millisecond for Channel 1 but 3 milliseconds for Channels 2–6. On October 17, 2011, all six channels have a median gap of 1 millisecond.

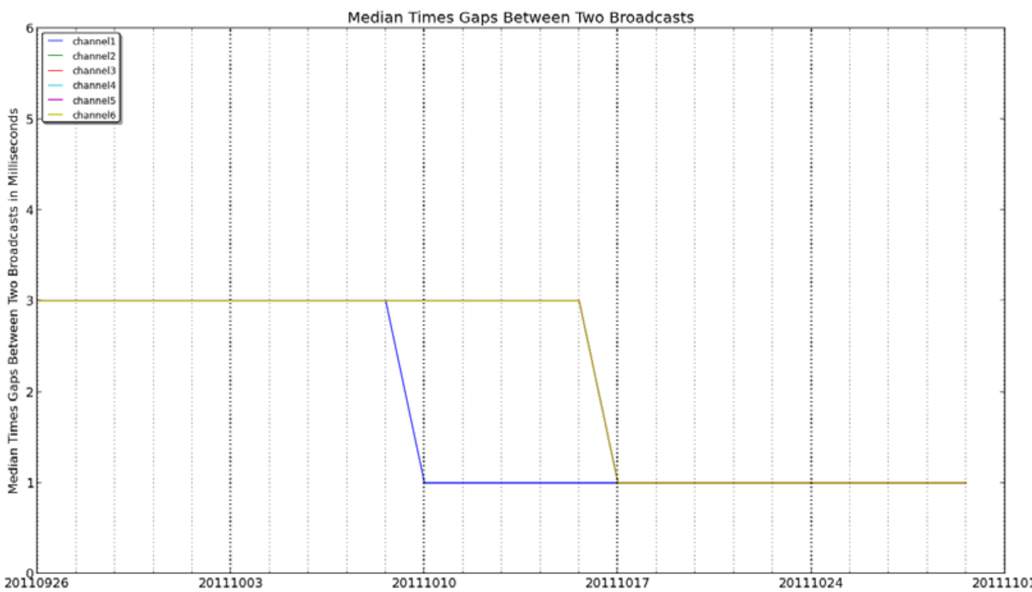
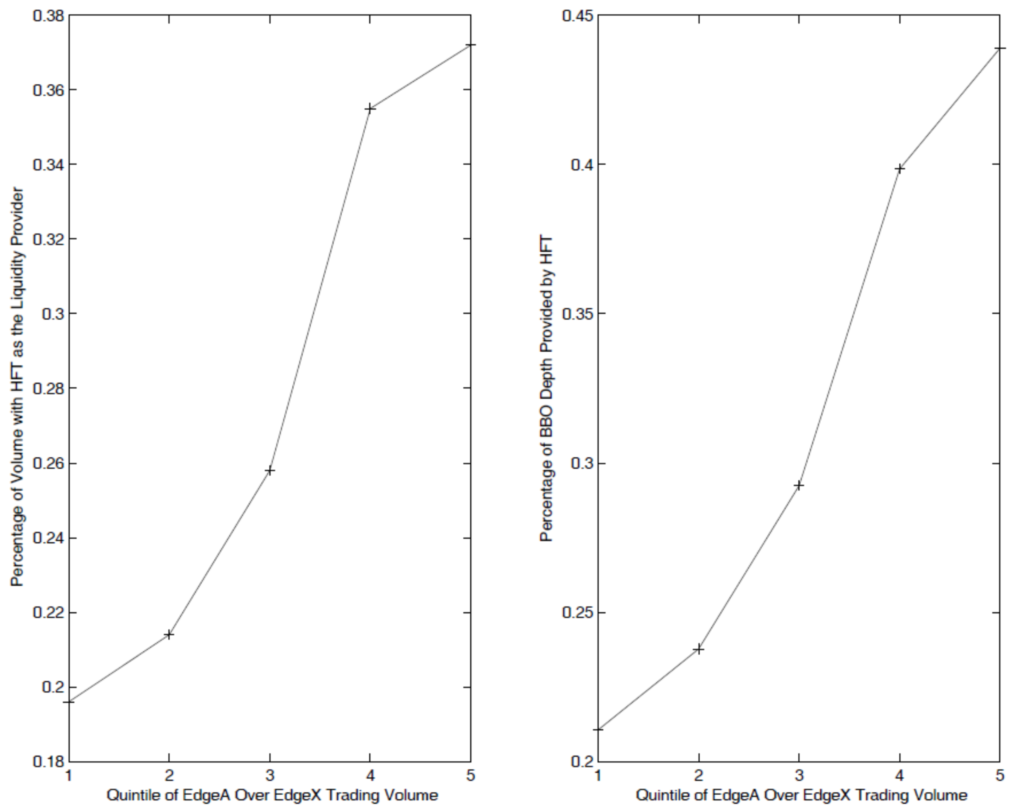


Figure 4: High Frequency Liquidity Provision and Share of Taker/Maker Fee Market

This figure relates high-frequency liquidity provision to our second price-constraint measure: the ratio of the volume of EDGA to that of EDGX. Stocks in quintile 1 exhibit the lowest volume in EDGA (the taker/maker fee market) relative to that of EDGX (the maker/taker fee market) and exhibit the lowest price constraint. Stocks in quintile 5 have the highest price constraint with the largest market share in the taker/maker market. Panel A presents the percent of the volume of high-frequency liquidity providers and panel B provides the percent of depth provided by high-frequency traders.



Appendix: Structure of ITCH Data

ITCH data come in a daily binary file and the first step in using them is to separate order instructions into various types. Each message comes with a timestamp measured in nanoseconds (10^{-9} seconds). Table 1 presents a sample of seven types of main messages from the daily file of May 24, 2010. Messages of types A and F include new orders accepted by the NASDAQ system that are added to the displayable book. A U message means that the previous order is deleted and replaced with a new order. An X message provides quantity information when an order is partially cancelled. An E message is generated when an order in the book is executed in whole or in part. If the order is executed at a price that differs from that of the original order, a C message is generated and the new price is demonstrated in the price field. To save space, some order instructions, such as order deletion, do not indicate the stock symbol but only the reference number of the order to be deleted. It is essential to fill in the redundant details to group the order instructions based on ticker symbols, which is the foundation for the construction of the limit-order book. The construction of the limit-order book is very data-intensive. For example, we find that a message can be deleted and replaced 69,204 times using a U Message.

This appendix provides more information about how to construct the limit-order book from messages in ITCH data. The type of information involved is listed in Table 1.

A and F messages include the new orders that are accepted by the NASDAQ system and added to the displayable book. NASDAQ assigns each message a unique reference number. Messages A and F include a timestamp, a buy/sell reference number, the price, the number of shares, and the stock symbol. The only difference between A messages and F messages is that an F message indicates the market participant identification that is associated with the entered order.

The first message in Table 1 is an A message with a reference number of 335531633 to sell 300 shares of EWA at $19.50 per share. Time is measured as the number of seconds past midnight. Therefore, this order is input at second 53435.759668667, or 14:50:35:759668667. The F message shows a 100-share buy order for NOK at a price of $9.38 per share with UBSS as the market participant. The U message means that the previous order is deleted and replaced with a new order. The update can be on the share price or the quantity of shares. In our example, order 335531633 experiences a change in price from $19.50 to $19.45, generating a new order with reference number 336529765. To conserve space, message U does not indicate the ticker symbol or the buy/sell reference number. Only after the trader finds the reference number for the first time the updated message was deleted can she link the updated message back to the A message or the F message to locate its ticker symbol and buy/sell reference number. In our example, we can link order 336529765 to the original order 335531633 and know that it is a sell order for EWA. We find that a message can be deleted and replaced 69,204 times using a U message. In short, new orders can originate from three message files: messages of types A, F, or U.

An X message provides quantity information when an order is partially cancelled. Orders with multiple partial cancellations share the same reference number. An X message contains only a timestamp, an order number, and the quantity of shares cancelled. We need to link the X message to the original A or F messages in order to find the stock in our sample and update its limit-order book. In our example, the X instruction deletes 100 shares from order 336529765. The U message with reference number 336529765 implies that the size of the order is reduced to 200 shares at a price of $19.45 per share. However, we need to link the U message to the A message to know that the new order is to sell EWA.

An E message is generated when an order in the book is executed in whole or in part. Multiple executions originating from the same order share the same reference number. An E message also has only the order reference number and the quantity of shares executed. Therefore, we need to trace the order to the original A or F message to find the stock and the buy/sell information. In our example, the order reference number first points to the U message (336529765), which then tracks to an A message. Now we know that a sell order for EWA is executed; however, the price information is from the U message, where the price has been updated from $19.50 to $19.45 per share. After matching, the system will generate a matching number of 7344037. If the order is executed at a price that differs from that of the original order, a C message is generated and the new price is demonstrated in the price field.

A D message provides information pertaining to when an order is deleted. All remaining shares are removed from the order book once a D message is sent. In our example, all the remaining shares of order 336529765 are deleted. The order originally had 300 shares, and an X message deletes 100 shares from the book, while an E message leads to an execution for a sale of 76 shares. Therefore, the D message deletes 124 shares from the book. The price level is $19.45 per share, which is known from the U message, and the stock and the buy/sell indicator can be found at the A message.