UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

Department of Finance
College of Business
340 Wohlers Hall, MC-706
1206 South Sixth Street
Champaign, IL 61820-6980

December 22, 2014

Mr. Brent J. Fields, Secretary
Securities and Exchange Commission
100 F Street NE
Washington, DC 20549-1090
Re: "Plan to Implement a Tick Size Pilot Program"

Dear Mr. Fields:

As academic researchers on financial market microstructure, we appreciate the opportunity to comment on the proposed pilot program to widen the quoting and trading increments for certain small capitalization stocks (Tick Size Pilot hereafter). We understand that proponents to increase tick size argue that larger tick size increases market-making revenue and supports sell-side equity research and, eventually, increases the number of IPOs. Economic theories suggest that constrained prices should facilitate non-price competition, but we doubt that non-price competition would take the form of competing on proving the best research or more IPOs. I understand that both tick size and IPO have decreased since 1990, but their correlation does not imply causality. Currently, I have never seen any empirical results that demonstrate the causality of tick size on IPO based on clean identification strategy. Two of our recent papers, however, suggest two potential effects of widening tick size. First, it will encourage liquidity provision from high frequency traders (HFT). Second, it will encourage competition for maker-taker fee across different trading venues and market fragmentation.

Summary of Our Two Papers

Our empirical design is based on the cross-sectional variations in relative tick size. The uniform one-cent tick size implies that the relative tick size, or one-cent divided by the nominal price, is higher for low priced securities. Our paper "Tick Size Constraints, High-Frequency Trading, and Liquidity" find that stocks with large relative tick size attract HFT liquidity provision. The economic mechanism is as follows. 1) A large relative tick size hinders price competition of liquidity providers by imposing high cost on establishing price priority, thus it can force liquidity providers, who would have differentiated themselves by quoting at difference prices under a small relative tick size, to quote the same price. 2) A large relative tick size can encourage speed competition of liquidity providers to establishing time priority, as the precedence of execution for limit orders at the same price is determined by their time of submission. To establish the causality from relative tick size to the liquidity provision by HFT, we use splits/reverse splits of ETFs as exogenous shocks to the relative tick size. An ETF that experiences a split event witnesses a decrease in liquidity, at the same time, an increase HFT liquidity provision, compared with paired ETF that tracks the same index but does not experience split/reverse split events. Reverse splits, on the other hand, reduce HFT liquidity provision but increase liquidity.

The other paper "Two-Sided Markets, Make-Take Fees and Competition between Stock Exchanges" shows that tick size leads to complex make-take fee games among stock exchanges. The economic intuition is as follows. The minimum price variation prohibits liquidity providers and demanders from negotiating price increments of less than a tick size. The make-take fees set by the exchanges, however, are not subject to the tick size regulation. The nature of the fee game reflects competition between exchanges for orders based on proposing sub-penny prices for makers and takers. We expect that an increase in tick size would generate more intense competition for make-take fees. Particularly, an increase in tick size would potentially increase the market share of the taker/maker market, or the market that charges liquidity providers and subsidizes liquidity demanders.

Policy Implications

Our research generates the following policy implications:

1. The U.S. market structure has already been very complex and interrelated, and many policy proposals should not be evaluated in isolation. Besides the policy debate on tick size, we also see independent policy debate on HFT and make-take fee. However, we need to be aware these policy issues are related, and the change in one dimension may affect the other.

2. Current policy debates on tick size, HFT and make-take fee focus on whether additional regulation is required. Sometimes, *deregulation* can be a solution. We argue that we should consider *decreasing* tick size for low priced liquid stocks. At the minimum, the first step towards further regulations is to evaluate current policies.

3. According to spirit of point 2, we suggest SEC consider a pilot program that *decreases* tick size for low-priced liquid stocks.

4. Make-take fee setting by stock exchange effectively changes the tick size, and we have heard that NASDAQ is considering implementing a pilot program for make-take fee. We

believe it would be beneficial for SEC and NASDAQ to coordinate these two pilots to create overlaps for the sample of stocks.

5. The economically meaningful tick size is relative tick size, but not the nominal tick size. Therefore, we need to consider the possibility that some firms in the pilot group will reverse split their stocks to undo a larger nominal tick size. At this point, it is hard to predict the extent of such reverse splits. On one hand, previous empirical evidence suggests that firms do not actively manage their relative tick size. For example, the tick size in NYSE changed from $1/8 to 1 cent, but we have not seen the nominal share price fall by a factor of 12.5. However, these changes in tick size applied to all stocks. It is hard to predict the outcome when a firm knows that it has been treated relative to its peers. We should definitely consider the possibility of reverse splits in the design of the pilot.

We attach our two papers on tick size to these comments. Please feel free to contact Mao Ye at maoye@illinois.edu, Chen Yao at Chen.Yao@wbs.ac.uk or Yong Chao at yong.chao@louisville.edu if you have any questions.

Respectfully submitted,

Mao Ye
Assistant Professor of Finance
University of Illinois, Urbana-Champaign

Attachments
Paper 1: Tick Size Constraints, High-Frequency Trading, and Liquidity
Chen Yao: University of Warwick
Mao Ye: University of Illinois, Urbana-Champaign

Paper 2: Two-Sided Markets, Make-Take Fees and Competition between Stock Exchanges
Yong Chao: University of Louisville
Chen Yao: University of Warwick
Mao Ye: University of Illinois, Urbana-Champaign

Tick Size Constraints, High-Frequency Trading, and Liquidity[1]

November 18, 2014

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. We thank Jim Angel, Shmuel Baruch, Robert Battalio, Dan Bernhardt, Jonathan Brogaard, Jeffery Brown, Eric Budish, John Campbell, John Cochrane, Amy Edwards, Robert Frank, Thierry Foucault, Harry Feng, Slava Fos, George Gao, Paul Gao, Arie Gozluklu, Joel Hasbrouck, Frank Hathaway, Terry Hendershott, Björn Hagströmer, Yesol Huh, Pankaj Jain, Tim Johnson, Charles Jones, Andrew Karolyi, Nolan Miller, Katya Malinova, Steward Mayhew, Albert Menkveld, Maureen O'Hara, Neil Pearson, Richard Payne, Andreas Park, Josh Pollet, Ioanid Rosu, Gideon Saar, Ronnie Sadka, Jeff Smith, Duane Seppi, Chester Spatt, Clara Vega, Haoxiang Zhu, 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, HEC Paris, the AFA 2014 meeting (Philadelphia), the NBER market microstructure meeting 2013, the Midway Market Design Workshop (Chicago Booth), the 8th Annual MARC meeting, the 6th annual Hedge Fund Conference, the Financial Intermediation Research Society Conference 2013, the 3rd MSUFCU Conference on Financial Institutions and Investments, the Northern Finance Association Annual Meeting 2013, The Warwick Frontiers of Finance Conference, the China International Conference in Finance 2014, 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, Rukai Lou, Tao Feng, Yingjie Yu, and Chao Zi for their excellent research assistance.

Abstract

We demonstrate a tick size constraints channel of speed competition. Liquidity provision from HFT is more active for low priced stocks, because the large relative tick size (one cent divided by price) of low priced stocks constrain non-HFTs from establishing price priority but help HFTs establish time priority. We use splits/reverse splits of ETFs as exogenous shocks to the relative tick size, with paired ETFs that track the same index as controls, finding that an increase in the relative tick size decreases liquidity but increases HFT liquidity provision. Profits from liquidity provision are higher for lower-priced stocks.

JEL Classification: G10, G14, G18

Key words: High Frequency Trading, Tick Size, Liquidity, Price Priority, Time Priority

1. Introduction

The literature on high-frequency trading (HFT) identifies two overarching effects of speed competition. 1) price competition effect: speed allows high-frequency traders (HFTers) to be the low-cost providers of liquidity; 2) information effect: speed enables HFTers to trade on advance information and adversely select slow traders (Jones, 2013, Biais and Foucault, 2014). We contribute to the HFT literature by establishing one additional channel for speed competition: tick size constraints. From this perspective, speed enables HFTers to achieve execution priority over non-HFTers by establishing time priority when price-priority rule is unable to differentiate two types of traders due to tick size regulation. This newly identified channel for speed competition, in turn, leads to a new perspective in the policy debate on market structure: the current policy debate focuses on whether and how to pursue additional regulation of HFT; our paper, however, demonstrates that HFT, particularly the liquidity-supplying behavior, can be a causal response to an *existing* regulation on tick size.

An important yet often neglected assumption for Walrasian equilibrium is infinitely divisible price. In reality, price competition is constrained by the tick size regulation. SEC rule 612 (the Minimum Pricing Increment) of regulation NMS prohibits stock exchanges from displaying, ranking, or accepting quotations for, orders for, 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.[2] The tick size imposes a constraint for price competition in liquidity provision, because it stops traders from bidding a securities price up or down to its marginal valuation (Foucault, Pagano, and Röell, 2013). Rockoff (2008) summarizes four possible responses when controls prevent price from adjusting to natural level: queuing, black markets,

[2] There are some limited exemptions, such as the Retail Price Improvement (RPI) Program and mid-point peg orders.

evading, and rationing. We observe speed competition as a form of queuing through which traders with the capacity to trade at high speeds compete for a position at the front of the queue.

Most limit order book markets impose price-time-priority on order executions. Price priority means that limit orders offering better terms of trade – limit sells at lower prices and limit buys at higher prices – execute ahead of limit orders at worse prices. Time priority means that for limit orders offered at the same price, the one entered earliest execute ahead of others. The uniform one-cent tick size implies that the relative tick size, defined as the tick size (one-cent) divided by the nominal price, is larger for stocks with lower prices. A large relative tick size increases the cost to establish price priority. For example, in January 2010, Citigroup had a relative tick size of around 30 basis points (nominal price level of around $3.30), and HSBC has a relative tick size of around 1.69 basis points (nominal price level of around $59). This implies that the cost to undercut the price of other traders by 1 nominal tick is 30 basis points for Citi, but it is only 1.69 basis points for HSBC. Limit order submitters, who would have differentiated themselves by quoting at different prices, can be forced to quote the same price due to a large relative tick size. Time-priority rule serves as the secondary rule to allocate supplies for quotes the same price.

A large relative tick size generates two interrelated effects: it hinders price competition and encourages speed competition. We first demonstrate that price competition between HFTer and non-HFTers is more constrained under larger relative tick sizes. We find that for stocks with larger relative tick sizes, the best quotes provided by HFTers and non-HFTers are more likely to be the same, which indicates lower incidences of the two types of traders diverging on the best quote they provide, implying a less vigorous quote competition.

One major finding in this paper regarding the price competition between HFTers and non-HFTers is that a small relative tick size facilitates non-HFTers to establish price priority. A recent editorial report by Chordia, Goyal, Lehmann, and Saar (2013)) raises the concern that *"HFTs use their speed advantage to crowd out liquidity provision when the tick size is small and stepping in front of standing limit orders is inexpensive"*. From this perspective, HFTers should take a more important role in providing best quotes for stocks with a smaller relative tick size, since for those stocks, traders face less constraints when undercutting prices. Our results, however, indicates the opposite: stocks with a smaller relative tick size are associated with higher incidences of non-HFTers establishing best quotes relative to that of HFTers. The findings are consistent with the view that non-HFTers move to the front of the price queue, especially when the relative tick size is small. As the relative tick size increases, non-HFTers face constraints when seeking to undercut HFTers, so the two types of traders are likely to quote the same price.

We next demonstrate that a large relative tick size encourages HFTers to establish time priority. The time priority of HFTers refers to the scenario that though the top of the limit order book contain quotes from both HFTers and non-HFTers, only orders from HFTers execute, because orders from non-HFT do not occupy the top of the queue at the same price. To this end, we first classify each execution into two categories: 1). Execution due to the liquidity provider achieving time priority; 2). Execution due to the liquidity provider achieving price priority. We find that HFTers enjoy higher percentage of volume due to their limit orders obtaining time priority relative to price priority than that of non-HFTers. As the relative tick sizes increase, the difference between HFTers' and non-HFTers' percentage of volume due to their limit orders obtaining time priority widens.

The causal impact of the relative tick size on HFT liquidity provision is demonstrated using a diff-in-diff analysis. ETF splits/reverse splits as regarded as exogenous shocks to relative tick sizes. ETFs that experience splits/reverse splits are classified in the treatment group, while ETFs that track the same index but experience no splits/reverse splits are classified in the control group. We find that HFT market marking activities increases/decreases after splits/reverse splits.

The same identification strategy also shows that a larger relative tick size decreases liquidity. The results for liquidity, in turn, provide an intuition for understanding the economic mechanism that drives speed competition. We find an increase in proportional quoted spread, depth and effective spread after splits. This result is consistent with the findings by Conroy, Harris and Benet (1990), Schultz (2000) and Kadapakkam, Krishnamurthy, and Tse (2005) that splits harm liquidity, but our identification cleanly addresses the endogeneity concern. Our main contribution to the literature on tick size, however, is showing that an increase in the relative tick size leads to a change in HFT behaviors. Stock splits lead to a coarser price grid. Therefore, traders who were able to differentiate each other in price can be forced to quote the same price after splits, which leads to an increase in proportional spread and the length of the queue to supply liquidity at the more constrained best bid and offer. Therefore, splits favors HFTers by discouraging price competition and encouraging speed competition. Reverse splits, on the other hand, creates new eligible price levels (in a proportional sense), which implies that liquidity providers who were forced to quote the same price can now differentiate themselves by price, shortening the queue at the best price. Therefore, reverse splits encourages price competition but discourages speed competition in market making activities.

The tick size channel differs from both price competition effect and information effect of speed competition documented in the literature. The price competition effect assumes that a speed advantage enables faster traders to provide better quotes, while the tick size channel presupposes that speed competition can be resulted from constrained price competition [3]. The clean identification involving ETF splits/reverse splits reveals that the tick size constraint channel is different from the information effect.[4] A relative change in HFT activity after ETF splits/reverse splits cannot be ascribed to information events, because the events should not change the information content of affected ETFs relative to their controls.

Competition on speed appeals to a general finance and economics audience because of its potential impact on the real economy. First, HFT may indirect impact the real economy through liquidity channel, through which liquidity affects asset prices and, by extension, affects a firm’s cost of capital and real decisions. Second, the rents from investing in high speed technologies induce speed arms races and further investment in speed, which may directly impact physical and human capital allocation. Admittedly, instant access or quick responses to information may create rents (Biais, Foucault and Moinas, 2013, Budish, Cramton and Shim, 2013), however, we argue that tick size regulation and time-priority rule, can also be utilized to extract rents. The literature has documented that a large tick size leads to higher profit of market making (Harris (1994) and Foucault, Pagano, and Röell (2013)) and the profit for traders with higher time priority is higher (Sandås (2001)) and Biais, Hillion and Spatt (1995)). We find empirical evidence consistent with these two predictions in our data, and we argue tick size regulation and time-priority rule can be another driver for arms race in speed.

[3]The literature has identified three venues for price improvements due to faster speed: avoiding adverse selection (Hendershott, Jones and Menkveld, 2011), better inventory management (Brogaard, Hagströmer, Nordén and Riordan, 2013), and low operations costs (Carrion, 2013).

[4] The information channel has been modeled extensively (Biais, Foucault and Moinas, 2013; Martinez and Rosu, 201; Foucault, Hombert, and Rosu, 2012). The empirical work focuses on the type of information used by HFTers, such as dislocation of separate financial instruments that track the same index (Budish, Cramton and Shim, 2013) triangular arbitrage opportunities (Chaboud, Chiquoine, Hjalmarsson and Vega, 2014; Foucault, Kozhan and Tham, 2014), order flow (Hirschey, 2013; Brogaard, Hendershott and Riordan, 2013), and macro news announcements (Brogaard, Hendershott and Riordan, 2013).

Current debate focuses on whether additional regulation is required for HFT, and if so, how to pursue it. Our paper is also the first study to point out that HFT can be regarded as a market design response to existing regulations. One possible solution is tick size *deregulation*. At the minimum, the first step towards further regulations is to evaluate current policies. Admittedly, literature shows that optimal tick size is non-zero (Cordella and Foucault (1996), Seppi (1997), Harris (1994) and Foucault, Kadan and Kandel (2005)), but our results indicate that current tick size can be too wide for liquid low priced stocks. We are concerned that U.S. regulation is moving to the opposite direction. Encouraged by the Jumpstart Our Business Startups Act (the JOBS Act), SEC has announced a pilot program to increase the tick size to five cents for small stocks.[5] We suggest SEC considering a pilot program that decreases tick size for low-priced liquid stocks. Proponents of increasing tick size argue that a larger tick size controls the growth of HFT (Weild, Kim and Newport, 2012), but our results indicate that larger tick size can *encourage* HFT. The second argument for increasing tick size is that a larger tick size increases liquidity, but this paper demonstrates that a large relative tick size reduces liquidity. The final argument for increasing tick size is that it increases market-making revenue and supports sell-side equity research and, eventually, increases the number of IPOs (Weild, Kim and Newport, 2012). Economic theories suggest that constrained prices should facilitate non-price competition,[6] but we doubt that non-price competition would take the form of providing better research, especially when speed exists as a more direct form of non-price competition.

This paper is organized as follows. Section 2 discusses the study's hypotheses as well as the empirical strategy for testing them. Section 3 describes the data used in the study. Section 4 presents preliminary results pertaining to the relationship between the relative tick size and HFT

---

[5] "SEC Provides Details of 5-Cent Tick Test," *Wall Street Journal*, June 25, 2014.
[6] Airlines, for example, offer better service when price competition is constrained (Douglas and Miller, 1974).

liquidity provision using double sorting. Section 5 examines the determinants of HFT liquidity provision and profits using regression analysis. Section 6 examines causal relationships between the relative tick size, liquidity, and HFT using a diff-in-diff test. Section 7 provides additional robustness check of our results. Section 8 concludes the paper and discusses the policy implications.

2. Testable hypothesis and empirical design

The fundamental hypothesis underlying the price competition effect is that the speed advantage enjoyed by HFTers allows them to provide better prices for liquidity. The improvement in liquidity may come from HFTers' lower operation costs (Carrion, 2013), lower adverse selection risk (Hendershott, Jones and Menkveld, 2011), or better inventory management (Brogaard, Hagströmer, Nordén and Riordan, 2013). Though these channels for liquidity improvement are plausible, the literature, however, overlooks the possibility that speed advantage may lessen HFTers' incentives to improve quotes. The U.S. market prioritizes price over time. HFTers can achieve time priority when they quote the same price as non-HFTers, which reduces their incentive to undercut the price; non-HFTers are likely to lose time priority when they quote the same price as HFTers, which gives them a greater incentive to improve the price to gain price priority. Therefore, to draw a conclusion about who is the best quote provider, we need to examine whether one type of trader is more likely to provide a better price than the other type, particularly when the constraints on price competition are less binding.

The strategy to test price competition effect involves relative tick size. Suppose that HFTers are better providers of liquidity. In that case a finer tick size reduces the constraints that enable HFTers to undercut non-HFTers, so their market share in liquidity provision should increase when tick size decreases. A recent survey conducted by Chordia, Goyal, Lehmann, and Saar (2013) raises the concern that non-HFTers may be crowded out by HFTers when the tick size is small. We thus formulate the first hypothesis in line with best liquidity providers corresponding to the relative tick size.

Hypothesis 1: The smaller the relative tick size, the larger the probability that HFTers provide better quotes than non-HFTers.

Hypothesis 1 presupposes that speed competition encourages price competition. The tick size channel, however, presupposes that speed competition is a consequence of constrained price competition. The main hypothesis of this paper is that a lower nominal price, or a larger relative tick size, leads to higher levels of HFT market marking:

Hypothesis 2: a larger relative tick size causes more HFT liquidity provision.

Hypothesis 2 is the main causal relationship this paper aims to establish. We refer to this causal relationship as the *tick size constraints hypothesis*. The key challenge in testing the tick size constraints hypothesis is addressing possible endogeneity. We examine the issue by using diff-in-diff regression involving EFT splits/reverse splits as exogenous shocks to their nominal prices. ETFs that split/reverse split are in the pilot group and the non-split ETFs that track the same index are in the control group.

The test is particularly interesting not only because the shocks are exogenous but also because they differentiate the tick size channel hypothesis from the information-based explanation. ETFs tracking the same index should have the same fundamental information. Therefore, the change in HFT activity after splits/reverse splits should not be driven by information when controlling for their pairs. We acknowledge the importance of the information channel, and the diversity of HFTers (Hagströmer and Nordén, 2013) implies that HFT activities can be driven by different initiatives. This paper is designed to reveal an important new channel of HFT activity.

The ETF tests also provide a clean environment within which to identify the impact of splits/reverse splits on liquidity, an interesting question in its own right. The literature finds mixed results on the impact of splits on liquidity (Berk and DeMarzo, 2013). One possible reason for such inconclusiveness is a counterfactual: it is hard to know what happens to a security if it does not split. We contribute to this literature by finding securities with identical fundamentals that do not split. Our conjecture on liquidity is motivated by the theoretical argument of Foucault, Pagano, and Röell (2013), which shows that a large tick size increases spread as well as the depth at a given price level. Therefore, we have our third hypothesis:

Hypothesis 3: Splits increase the quoted spread, depth, and effective spread as well as HFT market-making activity; reverse splits reduce the quoted spread, depth, and effective spread as well as HFT market-making activity.

The hypothesis pertaining to reverse splits is the mirror image of the hypothesis pertaining to splits under tick size constraints.

The final hypothesis relates to the profits of HFTers. The question is important for two reasons. First, higher profit of market making is another proxy for constrained price competition. Second, the profits from market making can be one of the reasons that drive arms race in speed, the success in which requires investments in both equipment and talent. One fundamental question for finance is whether activities in financial markets are sideshows or affect real resource allocation decisions. The literature has revealed three indirect mechanisms through which the market microstructure affects real decisions: liquidity, information risk, and ambiguity (O'Hara, 2007;

Easley and O'Hara, 2010).[9] Speed competition attracts the attention of a broader finance and economics audience because of its direct real impact. A recent article in the *Financial Times* estimates that a 1-millisecond advantage is worth up to $100 million in annual gains.[10] The common belief is that profits are produced by fast access to information, while this paper aims to show that profits can be extracted from a binding tick size, a non-informational source. This hypothesis is motivated by two lines of literature. The first line of literature relates profit with tick size: Harris (1994) and Foucault, Pagano, and Röell (2013) shows that profit of market making profits increase with tick size. The second line of literature relates market making profits with time priority: Sandås (2001) argues that orders with higher time priority earns larger expected profit than orders at the end of the queue. Therefore, we form the following two hypotheses:

Hypothesis 4: Market-making profits increase with tick size. The market making profit of HFTers are higher than non-HFTers.

3. Data and institutional details

This paper uses three main datasets: a NASDAQ HFT dataset, the NASDAQ TotalView-ITCH with a nanosecond time stamp, and Bloomberg. CRSP and Compustat are also used to calculate stock characteristics. The sample period for our analysis is October 2010 unless indicated otherwise.

3.1. Sample of stocks and NASDAQ HFT data

---

[9] For liquidity, see Pastor and Stambaugh (2003), Acharya and Pedersen (2005), Amihud and Mendelson (1986), Chordia, Roll, and Subrahmanyam (2000), Chordia, Sarkar, and Subrahmanyam (2005). The role of information risk in asset pricing is demonstrated in Easley, Hvidkjaer, and O'Hara (2003), O'Hara (2003), and Easley and O'Hara (2004). For the ambiguity channel, see Easley and O'Hara (2010).

[10] "Speed fails to impress long-term investors," *Financial Times*, September 22, 2011.

The NASDAQ HFT dataset provides information on limit-order books and trades for 120 stocks selected by Hendershott and Riordan. The original sample includes 40 large stocks from the 1000 largest Russell 3000 stocks, 40 medium stocks ranked from 1001–2000, and 40 small stocks ranked from 2001–3000. Among these stocks, 60 are listed on the NASDAQ and 60 are listed on the NYSE. Since the sample was selected in early 2010, three stocks disappeared from our sample period. Panel A of Table 1 contains the summary statistics for the 117 stocks.

Insert Table 1 about Here

The limit-order book data offer one-minute snapshots of the book with an indicator that breaks out liquidity providers into HFTers and non-HFTers at each price level, which facilitates the analysis of best quotes and depth provided by HFTers and non-HFTers. The trade file provides information on whether the traders involved in each trade are HFTers or non-HFTers. In particular, trades in the dataset are categorized into four types, using the following abbreviations: “HH”: HFTers who take liquidity from other HFTers; “HN”: HFTers who take liquidity from non-HFTers; “NH”: non-HFTers who take liquidity from HFTers; and “NN”: non-HFTers who take liquidity from other non-HFTers.

We are aware of three limitations to our data. The first limitation of the NASDAQ HFT data is that the NASDAQ identifies a firm as an HFT firm if it engages only in proprietary trading. HFT desks in large and integrated firms (e.g. Goldman Sachs and Morgan Stanley) may be excluded because these large institutions also act as brokers for customers and engage in proprietary low-frequency strategies; thus their orders cannot be uniquely identified as HFT or non-HFT business. The other omission involves orders from small HFTers that route their orders through these integrated firms. Nevertheless, the inclusion of some HFTers in the non-HFTers group tends to bias the estimate of their differences towards zero. NASDAQ HFT data thus underestimate the true differences between HFTers and non-HFTers and attenuates our findings. The fact that we still detect economically and statically significant differences between their activities demonstrates the robustness of our results.

Second, the data based on snapshots do not provide a dynamic view of the limit order book. Therefore, we do not know whether non-HFTers becomes the provider of the best price because they actively improve the quotes, or because the HFTers withdraw from the best quotes. Fortunately, NASDAQ also provide us with 5 days of data based on all the quote updates from HFTers and non-HFTers from February 22, 2010 to February 26, 2010. The robustness check in section 7 using this small sample of data show consistent results.

Third, the data only contains trades in NASDAQ, but not other exchanges. NASDAQ is a traditional maker/taker market, where liquidity providers get rebate by providing liquidity. Effectively, a liquidity provider can undercut the price at the maker/taker market by trading in taker/maker market where he needs to pay to provide liquidity (Yao and Ye, 2014 and Chao, Yao and Ye, 2014). Two exchanges, Boston Stock Exchange and Direct Edge A, provide such a function throughout our sample period. However, the share of volume of taker/maker market is small relative to NASDAQ. In our sample period, the volume of Direct EDGE A is 12.67% of NASDAQ, and the volume of Boston is only 5.46% of NASDAQ. Also, our paper focuses on the cross-sectional variation of HFT activity for stocks with different price level. Because the maker/taker fee also does not vary with the price level for stocks above $1, the role of maker/taker fee is similar to tick size. It is relatively more costly to undercut the price by moving from the maker/taker market to a taker/maker market for low priced stocks.

3.2. Sample of stocks and NASDAQ ITCH data

We use a diff-in-diff test to clearly identify the causal impact of the relative tick size on HFT liquidity provision. The test uses Leveraged ETFs that have undergone splits/reverse splits, and treats them as the pilot group, whereas the control group contains Leveraged ETFs that track the same indexes and do not split/reverse split. Leveraged ETFs are issued prominently by Proshares and Direxion, and often appear in pairs that track the same index but in opposite directions. For example, the ETFs SPXL and SPXS both track the S&P 500, but SPXL amplifies S&P 500 returns by 300% while SPXS does so by -300%. These twin Leveraged ETFs usually have identical nominal price when launched for IPO, but the return amplification results in frequent divergence of their nominal prices after issuance. The issuers often use splits/reverse splits to keep their nominal prices aligned with each other. The shocks to nominal prices, or relative tick sizes, caused by the splits/reverse splits are exogenous after control for their past returns.

We search the Bloomberg and ETF Database to collect information on leveraged ETF pairs that track the same index with an identical multiplier, and the data are then merged with CRSP to identify their splitting/reverse splitting events. We identify 5 splits and 21 reverse splits from January 2010 through November 2011. Reverse splits occur more frequently, because their issuers are often concerned about the higher trading cost of low-priced ETFs.[11] Our empirical analysis provides evidence that supports this concern.

Since the NASDAQ HFT dataset does not provide HFT information for ETFs, we compute HFT activities based on methodologies introduced by Hasbrouck and Saar (2013) using NASDAQ ITCH data, which is a series of messages that describe orders added to, removed from, or executed on the NASDAQ. We also use ITCH data to construct a limit-order book at nanosecond-scale

[11] "Why has ProShares decided to reverse split the shares of these funds?" (http://www.proshares.com/resources/reverse_split_faqs.html)

resolution, upon which the calculation of liquidity is based. Details on how to link these messages can be found in Gai, Yao, and Ye (2013) and Gai, Choi, O'Neal, Ye, and Sinkovits (2013). Summary statistics for leveraged ETFs are presented in Panels B of Table 1.

4. Double sorting

Our preliminary analysis starts with 3-by-3 double sorting based first on the market cap and then on the relative tick size of the stock. We sort the 117 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. Section 4.1 contains the results for best quotes. Section 4.2 provides the results for depth at best quotes. Section 4.3 presents the share of volume with HFTers as liquidity providers.

4.1. Provision of the Best Quotes

A fundamental question in the HFT literature is whether the speed advantage enjoyed by HFTers enables them to provide better prices relative to non-HFTers. The cost of liquidity comes from three sources: information asymmetry, inventory risk, and order-processing costs (Stoll, 2000). The *price competition effect* presupposes that HFTers can reduce these costs.[12] The literature, however, overlooks the possibility that speed advantage may lessen HFTers' incentives to improve quotes. The U.S. market prioritizes price over time. HFTers can achieve time priority when they quote the same price as non-HFTers, which reduces their incentive to undercut the price;

[12] See Hendershott, Jones, and Menkveld (2011) for information asymmetry, Brogaard, Hagströmer, Nordén, and Riordan (2014) for inventory risk, and Carrion (2013) for the argument on order-processing costs.

non-HFTers are likely to lose time priority when they quote the same price as HFTers, which gives them a greater incentive to improve the price to gain price priority.

The one-minute snapshots of the limit-order book in the NASDAQ high-frequency book enable us to examine this question in detail. The data indicate the depth provided by both HFTers and non-HFTers at each bid and ask price. Our analysis starts by categorizing the best price (the bid and the ask are treated independently) in these 391 minute-by-minute snapshots for each stock and each day into three groups according to the following criteria: 1) best price is displayed by HFTers only, 2) best price is displayed by non-HFTers only, and 3) best price is displayed by both HFTers and non-HFTers. The number is then averaged across all the stocks in each portfolio for each day, resulting in 21 daily observations for each of the 3-by-3 portfolios. Column 1 of Table 2 presents the percentage of time that HFTers are unique providers of the best quotes, Column 2 presents the percentage of time that non-HFTers are unique providers of the best quotes, and column 3 presents the percentage of time that HFTers and non-HFTers both display the best quotes. Columns 1, 2, and 3 sum to 100 percent. Column 4, defined as column 2 minus column 1, shows the differences in percentage between non-HFTers as unique providers of the best price and HFTers as the unique providers of the best price. Column 5 contains the statistical inferences for column 4 based on 21 daily observations.

Insert Table 2 about Here

Table 2 shows that HFTers and non-HFTers are more likely to quote identical prices for stocks with large relative tick sizes. The best quotes provided by HFTers and non-HFTers are the same 95.9% of the time for low-priced large stocks, while best quotes provided by HFTers and non-HFTers are the same only 45.5% of time for high-priced large stocks. In other words, price priority is sufficient to differentiate two types of traders only 4.1% of the time for low-priced large stocks, but price priority is sufficient 54.5% of the time for high-priced large stocks. This result implies that a large relative tick size is associated with lower level of price differentiation, while a small relative tick size is associated with more competition on price improvement between the two types of traders. This evidence suggests that price competition is indeed more constrained for stocks with a large relative tick size.

As noted, a small relative tick size encourages price differentiation and alleviates constraints on establishing price priority. If speed facilitates HFTers to establish price priority over non-HFTers, they should then be more likely to be the unique best price provider as the relative tick size decreases. However, we find as the relative tick size decreases, non-HFTers play a more and more prominent role than HFTers in providing the best pricecolumn 5). The finding is inconsistent with Hypothesis 1 that the smaller the relative tick size, the larger the likelihood that HFTers provide better quotes than non-HFTers. This result also belies the common belief that a smaller relative tick size leads to penny-jumping behavior on the part of HFTers when stepping in front of standing limit orders is inexpensive (Chordia, Goyal, Lehmann and Saar, 2013). Our results are instead consistent with the view that non-HFTers are more likely to establish price priority when the relative tick size is small.

When non-HFTers and HFTers quote the same price, it is natural to expect that HFTers are more likely to establish time priority at the top of the queue. The result is formally established in section 7.2. To summarize, a small relative tick size helps non-HFTers achieve price priority, and a large relative tick size leads non-HFTers and HFTers to quote the same price, but HFTers use their speed advantage to establish time priority. This intuition helps in understanding the results reported in section 4.3, which show that the volume share with HFTers as liquidity providers is the highest for low-priced stocks.

4.2. Percentage of Depth at BBO Provided by HFTers

This section analyzes the best depth, or the quantity provided at the best price. We denote the depth provided by HFTers and non-HFTers as $\{HFTdepth_{itm}, NonHFTdepth_{itm}\}$, where i is the stock, t is the date, and m is the time of day. The share-weighted average first sums the HFT liquidity provision for all stocks in the portfolio and then divides the number by the total liquidity provision for all stocks in the portfolio for each day.[13]

The average depths from HFTers and non-HFTers for stock i on day t are:

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

The depth provided by HFTers relative to the total depth of portfolio J on day t is:

$$HFTdepthshare_{Jt} = \frac{\sum_{i\in J} HFTdepth_{it}}{\sum_{i\in J}(HFTdepth_{it} + NonHFTdepth_{it})} \quad (2)$$

Table 3 shows the average percentage of depth provided by HFTers for each of the market cap by relative tick size portfolios. The result indicates that the percentage of depth at BBO provided by HFTers increases monotonically with the relative tick size. The depth from HFTers is as high as 55.66% for large stocks with a large relative tick size, while the figure is only 35.07% for large stocks with a small relative tick size. The difference is 20.59% and the t-statistic based on the 21 observations runs as high as 22.10. The depth percentage provided by HFTers is 39.73% for mid-cap stocks with a large relative tick size, while the figure is 24.61% for mid-cap stocks with a small relative tick size. The difference is 15.13% with a t-statistic of 22.88. As the percentage of depth at BBO offered by non-HFTers is 1 minus the percentage of depth at BBO

[13] The depth result is share weighted. We also try Equal-weighted depth and the results are similar.

offered by HFTers, it is easy to see that stocks with a smaller relative tick size have a larger ratio of the depth at BBO offered by non-HFTers to that offered by HFTers. Taken as a whole, section 4.1 shows that a smaller tick size is associated with a greater chance that non-HFTers offer the best price. Quantity of shares offered at the best price by HFTers holds the same pattern.

Insert Table 3 about Here

4.3. Tick size constraints and volume

Sections 4.1 and 4.2 show that non-HFTers are more and more likely to quote better prices to achieve price priority over HFTers as the relative tick size goes down. A large relative tick size, however, discourages price differentiation and increases the likelihood that these two types of traders quote the same price, thus leaving time priority to determine the execution sequence. In this section, we demonstrates the percentage of volume with HFTers as the liquidity providers is the highest for the portfolio where HFTers and non-HFTers are most likely to quote the same price, or the portfolio for which time priority is the most important.

The NASDAQ high-frequency data indicate, for each trade, the maker and taker of liquidity. Recall that NH_{it}, HH_{it}, HN_{it}, and NN_{it} are the four types of share volume for stock *i* on each day *t*. For each portfolio *J* on day *t*, the volume with HFTers as liquidity providers relative to total volume is defined as:[14]

$$HFTliqmake_{Jt} = \frac{\sum_{i \in J}(NH_{it} + HH_{it})}{\sum_{i \in J}(NH_{it} + HH_{it} + HN_{it} + NN_{it})} \tag{3}$$

Table 4 shows the average percentage of volume with HFTers as the liquidity providers for each of the market cap-by-relative tick size portfolios. The result demonstrates a clear pattern

[14] The equal-weighted average yields similar results. As a falsification test, we also perform double sorting for the percentage of volume with HFTers as liquidity takers. We do not find it increase with the relative tick size.

according to which the volume with HFTers as liquidity providers increases with the relative tick size. For example, about half of the volume is due to HFTers' being liquidity providers for large stocks with a large relative tick size, but only 35.93% of the volume is due to HFTers' being liquidity providers for large stocks with a small relative tick size. The difference is 14.03% with a t-statistic of 15.54 for the 21 observations. Across all 3-by-3 observations, volume with HFTers as liquidity providers is highest for large stocks with a large relative tick size, in which case HFTers and non-HFTers quote the same price 95.9% of time (Table 2). In summary, our results show that liquidity provision by HFTers is more active for stocks with a large relative tick size, or stocks that face more tightly constrained price competition.

Insert Table 4 about Here

5. Regression analysis

The multivariate regressions in this section further confirm that stocks with large relative tick size generate more HFT liquidity provision. In addition, we demonstrate that the profit for liquidity provision is also higher for stocks with large relative tick size. We control for additional variables to overcome omitted variable bias. A sufficient condition for omitted variable bias to occur is that the missing variables are correlated with both the nominal price and HFTer market making. We are not aware of any papers that have touched on the variables correlated with both of these variables. Therefore, we start from the necessary condition that omitted variables be correlated with at least one of these two variables. The search for control variables is guided by the nominal price literature and the HFT literature.

5.1. Control variables

The nominal price literature suggests that the industry norm is important in choosing the nominal price. Benartzi, Michaely, Thaler, and Weld (2009) find that a firm may split/reverse split if its price deviates from the industry average. The advantage of regression analysis is that we can control for this average using an industry fixed effect, where industries are classified using the Fama and French classification of 48 industries. Although Benartzi, Michaely, Thaler, and Weld (2009) argue that other hypotheses cannot explain nominal prices, to run a robustness check we nevertheless take five lines of studies in the nominal price literature into consideration, three of which suggest additional control variables for our analysis.[15] The summary of control variables from these hypotheses is presented in Table 1.

The optimal tick size hypothesis argues that firms choose the optimal tick size through splits/reverse splits (Angel, 1997). This hypothesis implies that firm characteristics can determine both HFT market making and the relative tick size. However, the optimal tick size hypothesis has been rejected by the following experiment. If firms could choose their optimal relative tick sizes, they would aggressively split their stocks when the tick size changes from 1/8 to 1/16 and then to one cent. Such aggressive splits have not occurred in reality (Benartzi, Michaely, Thaler, and Weld, 2009). Nevertheless, we include the idiosyncratic risk, and the number of analysts that may affect the choice of the optimal tick size, from this study.[16]

[15]Two other lines of research do not suggest additional variables to control for in our study. The catering hypothesis proposed by Baker, Greenwood, and Wurgler (2009) discusses time-series variations in stock prices: firms split when investors place higher valuations on low-priced firms and vice versa, but our analysis focuses on cross-sectional variation. Campbell, Hilscher, and Szilagyi (2008) find that an extremely low price predicts distress risk, but the 117 firms in our sample are far from default, and the distress risk should not affect the ETFs in our sample.

[16] Angel also argues that the relative tick size also depends on whether a firm is in a regulated industry, and this effect has been taken care of by including an industry-by-time fixed effect. Angel uses firm book value as a control for size, which is similar to the market cap for which we have controlled. When book value is included as an additional control, the results are similar.

The marketability hypothesis argues that a lower price appeals to individual traders. Tests of this hypothesis find mixed results.[17] Nevertheless, we include the measure of small investor ownership suggested by Dyl and Elliott (2006), which is equal to the logarithm of the average book value of equity per shareholder.

The signaling hypothesis states that firms use stock splits to signal good news. The empirical literature, however, does not reach a definitive conclusion as to whether splits serve as a signal or, if so, what types of news prompt firms to signal.[18] In addition, the 117 stocks in our sample do not split. Although our ETF sample contains splits, these splits should not be regarded as information driven, particularly when compared with ETFs that track the same index but do not split. Nevertheless, we use PIN offered by Easley, Kiefer, O'Hara and Paperman (1996) to control for information asymmetry.

We then introduce additional control variables from the HFT literature. We include turnover and volatility in our regression following Hendershott, Jones, and Menkveld (2011). PIN is an interesting variable from a HFT prospective as well. The literature has reached a consensus that HFTers' speed advantage allows them to reduce the pick-off risk by cancelling their quotes before being adversely selected, but it is interesting to further test whether HFTers take a higher or lower market share for stocks with higher probability of informed trading. The regression also includes past returns as the independent variable to examine the impact of returns on HFT market

[17] Some papers find that individuals prefer low-priced stocks (Dyl and Elliott, 2006), whereas Lakonishok and Lev (1987) find no long-term relationship between nominal prices and retail ownership. Byun and Rozeff (2003) suggest that if there are any short-term effects of low prices, they are very small.
[18] See Brennan and Copeland (1988), Lakonishok and Lev (1987), and Kalay and Kronlund (2013).

making.[19] The method used for calculating the variables from the HFT literature is also summarized in Table 1.

5.2. Regression results for HFT liquidity provision.

The regression specification is

$$y_{i,t} = u_{j,t} + \beta \times tick_{relative_{i,t}} + \Gamma \times X_{i,t} + \epsilon_{i,t} \qquad (4)$$

Panel A of Table 5 presents the regression results with $y_{i,t}$ as the daily percentage of depth at BBO provided by HFTers (*HFTdepth*), and Panel B presents the regression results with $y_{i,t}$ as the percentage of volume with HFTers as liquidity providers (*HFTvolume*) for each stock *i* on date *t*.[20] . $u_{j,t}$ is the industry-by-time fixed effect.[21] The key variable of interest, $tick_{relative_{i,t}}$, is the daily inverse of the stock price. $X_{i,t}$ are the control variables presented in Table 1.

Insert Table 5 about here

Table 5 confirms that *HFTdepth* and *HFTvolume* both increase with the relative tick size, which is consistent with our tick size constraints hypothesis. Large-cap stocks also show higher *HFTdepth* and *HFTvolume*. Column 3 shows that the sign for retail trading ($logbv_{average}$) is mixed: more retail trading leads to a decrease in *HFTdepth* but an increase in *HFTvolume*. Column 4 indicates that firm age is the only variable that predicts both *HFTdepth* and *HFTvolume* under the relative tick size hypothesis: HFTers tend to provide more liquidity for older firms. Columns 1–7 show that no other variables can consistently predict the market share of HFTers other than the

[19] We thank an anonymous reviewer for the Texas Finance Festival for the suggestion to add past returns.
[20] We also analyze who provides the best price using multinomial logit and probit models and obtain similar results. The results are not reported but are available upon request.
[21] We do not include firm fixed effects because the focus of this paper is on understanding the cross-sectional variation in HFT market-making activity, and firm fixed effects defeat this purpose (Roberts and Whited, 2012).

relative tick size, market cap, and PIN. The insignificant coefficients before other variables do not necessarily imply that these variables have no impact on absolute magnitude of the liquidity provision of HFTers. For example, it is well know that volatility affects the displayed depth of the limit order book (Ahn, Bae and Chan (2001), and Hasbrouck and Saar (2001)), a result confirmed by our data as well (not reported). Therefore, our results can only be interpreted that volatility has similar impact on the liquidity provision of HFTers and non-HFTers and has no statistically significant impact on their ratio.

Columns 5 and 7 show that a higher PIN, alone or combined with other control variables, reduces *HFTvolume* and *HFTdepth*. This intriguing result suggests that the speed advantage enjoyed by HFTers when updating their quotes does not lead to higher market share for stocks with higher information asymmetry. The main takeaway from Table 5 is that HFTers concentrate their activity on stocks with a large relative tick size and lower information asymmetry.

5.3. Relative tick size and profits

Hypothesis 4 relates market making profit to tick size. We test this hypothesis using the following specification.

$$y_{i,t,n} = \beta_1 \times HFTdummy_{i,t,n} + \beta_2 \times tick_{relative_{i,t}} + \beta_3 \times tick_{relative_{i,t}} \times HFTdummy_{i,t,n}$$
$$+u_{j,t} + \Gamma \times X_{i,t} + \epsilon_{i,t,n} \quad (5)$$

In the specification, $y_{i,t,n}$ is the unit profit for each stock *i* on date *t* for trader type *n*. We have two daily observations for each stock: a unit profit for HFTers and a unit profit for non-HFTers. These unit variables depends on a number of controls ($X_{i,t}$) and three key variables of interest: relative tick size, HFT dummy and their interaction. $HFTdummy_{i,t,n}$ equals 1 for the unit profit of HFTers and 0 otherwise. $tick_{relative_{i,t}}$, is the relative tick size of stock *i* at day *t* minus the average of the relative tick size of the sample. We demean the relative tick size to facilitate the interpretation of β_1, which captures the difference in profit between HFTers and non-HFTers for stocks with average relative tick size.[22] β_2 measures the relationship between market making profit and tick size for non-HFTers.[23] The interaction term $tick_{relative_{i,t}} \times HFTdummy_{i,t,n.}$ captures the differences in profits between HFTers and non-HFTers for stocks with different relative tick size. $u_{j,t}$ are industry-by-time fixed effects. $X_{i,t}$ are control variables presented in Table 1.

Our profit measure comes from Brogaard, Hendershott, and Riordan (2014), Menkveld (2013), and Baron, Brogaard, and Kirilenko (2014). The HFT market-marking profit for an individual stock during one day for a certain time interval *t* is defined as

$$\pi^{HFT,t} = \sum_{n}^{N} -(HFT_n^t) + INV\_HFT_n^t \times P_{mid}^t \tag{6}$$

The profit comes from two components. The first term, $\sum_n^N -(HFT_n^t)$, captures total cash flows throughout the interval, with *n* indicating each of the *N* transactions within each interval.[24] The second term, often referred as "positioning profit," cumulates value changes associated with net position. In our analysis, $INV\_HFT_n^t$ or the interval *t* is cleared at the end of the interval midpoint quote P_{mid}^t. The positioning profit is negative when liquidity providers are adversely selected (e.g., Glosten and Milgrom, 1985), or if liquidity providers are willing to mean-revert out of nonzero position (Ho and Stoll, 1981).

[22] Without demeaning the data, β_2 is interpreted as the difference in profit between HFTers and non-HFTers for stocks with zero tick size, or infinite price.
[23] The result is similar if we choose HFTers as the base group and set dummy equal to 1 if the profit comes from non-HFTers.
[24] We also add a liquidity rebate when calculating the first part of the profit. NASDAQ has a complex fee structure and we use a fee of 0.295 cents per share, but the results are similar at other fee levels.

Since we are interested in the market marking profit per dollar volume, we calculate the unit profit as:

$$\pi^{HFT} = \sum_{t}^{T} \pi^{HFT,t} / DolVol \qquad (7)$$

where T is the total number of intervals during trading hours 9:30–15:59 and $DolVol$ is the total dollar volume with HFTers as liquidity providers for each stock on each day. For example, if the interval t is taken to be 30 minutes long, cash flows are calculated for each of the 30-minute intervals and inventories accumulated are emptied at the end of the 30-minute interval; the total number of intervals T equals 13. We calculate multiple daily unit profit measures taking t at varying lengths: five-minute, 30-minute, one-hour, and one-day lengths, respectively.[25] The market-making profit per dollar volume of non-HFTers is calculated analogously.

Table 6 shows that the unit profit for market making increases with the relative tick size for all profit measures. The regression coefficient for profits measured on the assumption that inventories are cleared at the end of the day is 29.734 basis points. The economic magnitude of this coefficient can be interpreted as follows. The stock with the lowest price in our sample has a relative tick size of about 0.192 (price around $5), and the median relative tick size in our sample is about 0.034 (price around $30), and their difference in daily market-making profits is 29.734*0.192-29.734*0.034 = 4.7 basis points per dollar volume.

Insert Table 6 About Here

Whether HFTers enjoy a higher unit profit than non-HFTers, however, depends on the assumption regarding the frequency of inventory clearance. The first column in able 6 shows that the unit profit for HFTers is 0.762 basis points higher than that for non-HFTers if inventory can

[25] Because each day has 6.5 trading hours, the first interval for one-hour profit unit is from 9:30 to 10:00.

be cleared at a frequency of five minutes. The significantly positive coefficient of 7.054 for the interaction term suggests that the difference in the unit profit between HFTer and non-HFTers increases with the relative tick size. The economic and statistical significance of the *HFTdummy* and the interaction term decrease with the time horizon. Indeed, one important feature of HFTers is their “very short time-frames for establishing and liquidating positions” SEC (2010). One possible source for the higher profit HFTers enjoy over non-HFTers at a short horizon can be attributed to their speed advantage and time priority. A limit order at the front of the queue at a given price has a greater expected profit than other limit orders in the queue since its execution probability is higher (Foucault, Kadan and Kandel, 2013; Sandås, 2001).

6. Identifications using ETFs

This section examines the causal impact of the relative tick size on HFT market-making activity (hypothesis 2) using diff-in-diff test using ETF splits. The same identification strategy also facilitates the analysis of the causal relationship between the relative tick size and liquidity (hypothesis 3). The question of liquidity is not only important in its own right, but it also offers additional economic insights to help us understand the relationship between the relative tick size and HFT. Section 6.1 illustrates how HFT activity and liquidity are measured. Section 6.2 presents the diff-in-diff test using the splits/reverse splits of leveraged ETFs as exogenous shocks to the relative tick size.

6.1. Measure of HFT activity and liquidity

Because the NASDAQ HFT dataset does not contain HFT information for ETFs, we use “strategic runs,” proposed by Hasbrouck and Saar (2013), as a proxy for HFT market-making activity. A strategic run is a series of submissions, cancellations, and executions that are likely to form an algorithmic strategy. The link between submissions, cancellations, and executions are constructed based on three criteria: (1) Limit orders with their subsequent cancellations or executions are linked by reference numbers provided by data distributors.[26] (2) When inference is needed in deciding whether a cancellation is linked to either a subsequent submission of a nonmarketable limit order or a subsequent execution that occurs when the same order is re-sent to the market priced to be marketable, we follow Hasbrouck and Saar (2013), and infer such a link when a cancellation is followed within 100 ms by a limit-order submission of the same size and same direction or by an execution of a limit order of the same size but in the opposite direction. (3) If a limit order is partially executed and the remainder is cancelled, we apply criterion (2) based on the cancelled quantity. *RunsInProcess* is the sum of the time length of all strategic runs with 10 or more messages divided by the total trading time of that day (Hasbrouck and Saar, 2013).

To test whether *RunsInProcess* is a good proxy for HFT market-making activity, we calculate *RunsInProcess* for the 117 stocks for which we have both ITCH data and NASDAQ HFT data. Table 7 presents the cross-sectional correlation between *RunsInProcess* and three measures of HFT activity. HFTvolume (making) and HFTdepth are measures of HFT liquidity provision, and HFTvolume (taking) are measures of the percentage of volume with HFTers as takers of liquidity.[27] Table 7 also contains the correlation test for two other widely used HFT proxies: the quote-to-trade ratio (Angel, Harris and Spatt, 2010 and 2013) and negative dollar volume divided by total number of messages (Hendershott, Jones, and Menkeveld, 2011; Boehmer, Fong, and Wu, 2013), both of which are based on the intuition that HFTers tend to cancel more orders than non-HFTers. Surprisingly, these two measures have either low or negative correlations with HFT

[26] We have a more recently updated version of the data relative to those reported in Hasbrouck and Saar (2013). Therefore, if a trader chooses to use the "U" (update) message to cancel an order and add another one, we know that the addition and cancellation comes from the same trader.

[27] The measure is defined as the sum of volume from HN and HH types relative to total volume.

market making.[28] However, *RunsInProcess* has a positive correlation of 0.65 with the percentage of depth provided by HFTers and the correlation runs as high as 0.765 with the volume with HFTers as liquidity providers. A high *RunsInProcess* combines three factors: a fast response (within 100 ms), frequent cancellation (10 messages or more), and persistent interest in supplying liquidity (staying in the queue to provide liquidity conditional on fast and frequent cancellation). In this regard, *RunsInProcess* becomes a good proxy for HFT market making activity.

RunsInProcess has a correlation of 0.283 with HFT market-taking activity, indicating that it may also capture some liquidity-taking activity on the part of HFTers. However, its high correlation with HFT liquidity making activity and low correlation with HFT liquidity taking activity imply that it is a better proxy for liquidity-making activities than for liquidity-taking activities. Indeed, pure submissions of market orders are not considered strategic runs, because all "runs" start with limit orders. The 10 message cut-off and the time weight also increase the correlation of *RunsInProcess* with patient liquidity-providing algorithms. Impatient liquidity-demanding algorithms may use limit orders, but these algorithms are more likely to switch to market orders once the initial limit orders fail to be executed. Therefore, strategic runs that arise from liquidity-demanding algorithms should contain fewer messages. Even if they contain more than 10 messages, it is natural to expect that they span a shorter period of time and carry lower

[28] Both the quote-to-trade ratio and negative dollar volume divided by total number of messages are good proxies for distinguishing trader types if the comparison is made within the same security or to measure time-series variation in HFT activity. Cross-sectional comparison of HFT market-making activity can be affected by the relative tick size. A large relative tick size attracts HFTers to move to the front of the queue, but HFTers are less likely to cancel an order once they are in the queue, since their positions in the queue will be lost by cancellation. A smaller relative tick size discourages HFT liquidity provision, but remaining HFTers cancel more frequently because a smaller relative tick size implies more frequent price movement. One additional contribution of our paper is that we put forward the notion of carefully interpreting results that use quote-to-trade ratios and negative dollar volume divided by total number of messages in cross-sectional comparisons.

time weight in *RunsInProcess*. Therefore, we use *RunsInProcess* as a proxy for liquidity making, though we are aware that it may capture some liquidity-demanding HFT activity to a small degree.

Insert Table 7 about Here

Stock market liquidity is defined as the ability to trade a security quickly at a price that is close to its consensus value (Foucault, Pagano, and Röell, 2013). The 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. These liquidity measures come from a message-by-message limit-order book we construct from ITCH data.[29]

The quoted spread (*Qspread*) is measured as the difference between the best bid and ask at any given time. The proportional quoted spread (*pQspread*) is defined as the quoted spread divided by the midpoint of the best bid and best ask prices. In addition to earning the quoted spread, a market maker also obtains a rebate from each executed share from the NASDAQ. Therefore, we compute two other measures of the quoted spread: $Qspread_{adj}$ and $pQspread_{adj}$, which are spreads adjusted by the liquidity supplier's rebate.[30] Specifically,

$$Qspread_{adj_{i,t}} = Qspread_{i,t} + 2 * liquidity\ maker\ rebate \quad (8)$$

$$pQspread_{adj_{i,t}} = (Qspread_{i,t} + 2 * liquidity\ maker\ rebate)/midpoint_{it} \quad (9)$$

[29] We are aware that measures from limit order book is not a sufficient statics liquidity. See Jones and Lipson (2001) for a discussion.
[30] For each stock on each day, the liquidity maker's rebate is 0.295 cents per execution, but the results are qualitatively similar at other rebate levels.

All four of these quoted spreads are weighted by the duration of a quote to obtain the daily time-weighted average for each stock.

Our main measures of depth are the time-weighted average of displayed dollar depth at the best bid and offer. The effective spread measures the cost to trade against the actual supply of liquidity (SEC, 2012). The effective spread (*Espread*) for a buy is 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 proportional effective spread (*pEspread*) is defined as the effective spread divided by the midpoint. Also, a liquidity demander on the NASDAQ also pays the taker fee.[31] Therefore, we compute the fee-adjusted effective spread and the fee-adjusted proportional effective spread:

$$Espread_{adj_{it}} = Espread_{it} + 2 * liquidity\ taker\ fee \tag{10}$$

$$pEspread_{adj_{it}} = (Espread_{it} + 2 * liquidity\ taker\ fee)/midpoint_{it} \tag{11}$$

6.2. Diff-in-Diff test using leveraged ETF splits

This section establishes the causal relationship between the relative tick size, liquidity and HFT liquidity provision using a diff-in-diff test. ETF splits provide us with a cleaner environment within which to isolate the effects of the tick size constraints channel than stock splits do, since stock splits may also be motivated by information (Grinblatt, Masulis, and Titman, 1984; Brennan and Hughes, 1991; Ikenberry, Rankine and Stice,1996). The splits for ETFs, however, are much less likely to be motivated by informational reasons. Furthermore, the ETFs that track the same index but do not split provide an ideal control even if related splits involve information. Among ETFs, splits/reverse splits are more frequent for leveraged ETFs. The reason that splits occur is

[31] We set the taker fee at 0.3 cents per share.

completely transparent. The bear and bull ETFs for the same index are usually issued by the same company at similar IPO prices, but large cumulative movements of an index result in the divergence of their nominal prices. The issuers of leveraged ETFs usually use splits/reverse splits to align the nominal prices of bull and bear ETFs. Splits/reverse splits can be regarded as exogenous after controlling for past returns.

The regression specification for the diff-in-diff test is:

$$y_{i,t,j} = u_{i,t} + \gamma_{ij} + \rho \times D_{trt_{i,t,j}} + \theta \times return_{i,t,j} + \epsilon_{i,t,j} \tag{13}$$

where $y_{i,t,j}$ is HFT market-making activity or the liquidity measure for ETF j in index i at time t. u_{it}, the index-by-time fixed effects, controls for the time trend that may affect each index. The new element, γ_{ij}, is the ETF fixed effect that absorbs the time-invariant differences between two leveraged ETFs that track the same index. After controlling for index-by-time and ETF fixed effects, the only major differences between the bull and bear ETFs that track the same indexes are returns, $return_{i,t,j}$. The key variable in this regression is $D_{trt_{i,t,j}}$, the treatment dummy, which equals 0 for the control group. For the treatment group, the treatment dummy equals 0 before splits/reverse splits and 1 after splits/reverse splits. Therefore, coefficient ρ captures the treatment effect. The leveraged bull ETF is in the treatment group, and the leveraged bear ETF is in the control group if the leveraged bull ETF splits, and vice versa.

Panel A of Table 8 reports the splits results. Quoted spreads without rebate adjustment (Columns 1) and with rebate adjustment (Columns 2) both decrease by 9.686 cents following splits. Column 3 shows the proportional quoted spread increases by 0.998 basis points. Without the relative tick size frictions, we would expect the nominal quoted spread to decrease by the same percentage as the decrease in the nominal price, keeping the proportional spread unchanged.

However, with the increased relative tick size, the percentage change in the nominal quoted spread is larger than the percentage change in nominal price,[32] which leads to an increase in proportional quoted spread. Column 4 shows that proportional quoted spread with rebate adjustment experiences an even larger increase (1.205 basis points). Indeed, the liquidity rebate is held constant during the events, which imply a higher rebate per stock price after splits. In this sense, the economic impact of liquidity rebate resembles that of tick size, the inclusion of which leads to a further increase in proportional quoted spread.

Column 5 shows that the average dollar depth at BBO increases by 15,000 dollars. A large relative tick size after splits implies that traders who were able to quote a range of prices under a finer grid may have to quote the same price under a coarser grid, which lengthens the queue at the best price. The literature generally agrees that securities with lower quoted spreads and greater depth are more liquid. Because splits lead to higher quoted spreads and greater depth, the key variable of interest turns out to be the effective spread, the measure for the actual transaction costs incurred by liquidity demanders. The nominal effective spread decreases by 5.385 cents following a split, but the proportional effective spread, or the transaction cost for a fixed transaction dollar amount, increases by 0.801 basis points and 1.012 basis points after fee adjustment correspondingly. Therefore, we find liquidity decreases as the cost for actual transaction increases.

Our main results in panel A is presented in Column 10, which shows an increase in HFT activity in the treatment group after splits relative to the control group. Since the analysis has controlled for the index-by-time fixed effect, an increase in HFT activity cannot be attributed to the change in the underlying fundamental of the leveraged ETF. Specifically, the increased HFT

[32] Technically, nominal stock price may also subject to tick size constraints, but the magnitude of relative tick size has a greater impact on nominal quoted spread than on nominal price, because the nominal price is much larger than the nominal quoted spread.

activity cannot be ascribed to information events, because the treatment and the control groups have the same underlying information. The ETF fixed effect and their past returns further control other possible differences between the Bear and Bull ETFs.[33] We argue that the increase in HFT market making activity is driven by changes in the relative tick size. Follow the same intuition as discussed previously, splits lead to coarser price grids, which can force traders who quoted different prices before splits to quote the same price after splits, thus causing an increase in the length of the queue at the new but more constrained best price. Once liquidity providers are in the same price queue, time priority rule determines the execution sequence, which intensifies speed competition. Taken as a whole, splits encourage HFT liquidity provision, as a large relative tick size impedes price competition while encourages speed competition.

Insert Table 8 about Here

Panel B shows reverse splits generate opposite pattern relative to splits do. Column 1-4 shows that quoted spread increases but proportional quoted spread decreases after reverse splits. The patterns hold with or without rebate adjustment. We ascribe this result to the reduction in the relative tick size following the reverse split. Given a fixed nominal spread, after reverse splits, proportional spread has new incremental units; the newly added units of increment encourage price differentiation. The depth at the best price decreases by 324,000 dollars, implying a shorter queue to provide liquidity. This result, along with the reduction in the proportional quoted spread, constitutes evidence that traders who were forced to quote the same price can now choose to differentiate themselv6es by price, leading to a reduction in depth offered at the new BBO. Column 6-9 show that effective spread increases but proportional effective spread decreases, suggesting

[33] One such example is clientele effect. For example, short sellers may prefer the bear ETF because of the short sale constraints for the bull ETF.

that the transaction cost for liquidity demanders falls after reverse splits. Column 10 reports a reduction in HFT activity following the reverse splits. A reduction in the relative tick size enhances price competition and reduces the queue for supplying liquidity, which weakens the incentives to be at the top of the queue at the constrained price. Therefore, we see less HFT market-making activity following reverse splits.

In summary, we find that splits decrease liquidity whereas reverse splits increase liquidity. We also find that splits attract HFT market-making activity whereas reverse splits reduce HFT market-making activity. These results should not be interpreted as HFTers having a negative impact on liquidity provision, because the tests are based on exogenous shocks to the tick size but not HFT activity. The correct economic interpretation is that a large relative tick size reduces liquidity but attracts HFT liquidity provision. Our results, however, do reveal the importance to control for relative tick size when considering the causal relationship between HFT market making and liquidity.

7. Robustness Tests

7.1. Do HFTers have time priority?

The tick size constraints channel argues that speed enables HFTers to achieve execution priority over non-HFTers by establishing time priority when price-priority rule is unable to differentiate two types of traders due to tick size regulation. This section provides robustness test for the claim that HFTers are more likely to establish time priority that non-HFTer, and studies how their difference in establishing time priority varies with relative tick size.

7.1.1. Data

Nasdaq top-of-book of displayed limit orders and trade record for the week of Feb 22 - 26, 2010 for the same 117 stocks are used in the test. The Nasdaq top-of-book of displayed limit orders data contain the collective best HFT quotes, the collective best non-HFT quote and their corresponding aggregate sizes. There is a data record whenever there was a change in the HFT or non-HFT quote. For each record, we are able to identify whether the change to the top-of-book is made by an order added to the book, or an order disappeared from the book. An order can disappear from the limit order book by being either executed or cancelled. Since Nasdaq also provide trading record with trade identifier (HH, HN, NH, NN) for the 117 stocks during this sample period, we match the top-of-book data and the trade data by millisecond timestamp, sign of the order (trade), price of the order (trade), size of the order (trade) and type of the liquidity providers (HFTer or non-HFTer). For the ones which we identify as disappeared from the limit order book, and matches can be found in the trading record, we classify them as order executions. The robustness test in Section 7.1 uses the execution message classified following above steps, and for each execution, we are able identify their liquidity providers (HFTer or non-HFTer), trade price, size, sign, and timestamp.

7.1.2. Methodology

For each trade which we identify as executions in the top-of-book of limit order data, we then classify it into two categories: 1) execution due to the liquidity provider having time priority; 2). execution due to the liquidity provider having price priority. If at the time of execution, both HFTers and non-HFTers provide quotes at the execution price, we define this trade executed due to the liquidity provider having time priority. If at the time of execution, only HFTers or non-HFTers provide quotes at the execution price, we define this trade executed due to the liquidity provider having price priority.

To test which type of trader is more likely to have time priority over the other type, and how this relation changes with relative tick size, we perform the following regression:

$$y_{i,t,n} = \beta_1 \times HFTdummy_{i,t,n} + \beta_2 \times tick_{relative_{i,t}} + \beta_3 \times tick_{relative_{i,t}} \times HFTdummy_{i,t,n}$$

$$+u_{j,t} + \Gamma \times X_{i,t} + \epsilon_{i,t,n} \qquad (14)$$

In the specification, $y_{i,t,n}$ is the percentage of dollar volume due to the liquidity provider having time priority, for stock *i* on date *t* of trader type *n*, calculated as the dollar volume due to limit orders having time priority over dollar volume due to limit orders having time priority plus that having price priority. We have two daily observations for each stock: one for HFTers and one for non-HFTers. The dependent variable is impacted by control variables ($X_{i,t}$) and three key variables of interest: relative tick size, HFT dummy and their interaction. $HFTdummy_{i,t,n}$ equals 1 for HFTers' and 0 otherwise. $tick_{relative_{i,t}}$ is the relative tick size of stock *i* at day *t* minus the average of the relative tick size of the sample. We demean the relative tick size to facilitate the interpretation of β_1, which captures the difference between HFTers and non-HFTers for stocks with average relative tick size. β_2 measures the relationship between the dependent variable and relative tick size for non-HFTers. The interaction term $tick_{relative_{i,t}} \times HFTdummy_{i,t,n.}$ captures the differences between HFTers and non-HFTers' percentage of volume due to having time priority for stocks with different relative tick size. $u_{j,t}$ is the industry-by-time fixed effect. $X_{i,t}$ are control variables presented in Table 1.

Table 9 shows that the larger the relative tick size, the higher the percentage of volume due to liquidity providers having time priority. The economic magnitude of this coefficient can be interpreted as follows. Take column 7 as an example. The stock with the lowest price in our sample has a relative tick size of about 0.192 (price around $5), and the median relative tick size in our sample is about 0.034 (price around $30), and their difference in percentage of volume due to liquidity providers having time priority is 3.374*0.192-3.374*0.034 = 54.3%. The table also shows that HFTers enjoy a higher percentage of volume due to achieving time priority than that of non-HFTers, and the difference is 7.7% for stocks with median tick sizes. Noticeably, the larger the relative tick size, the larger the difference between HFTer's percentage of volume due to obtaining time priority and non-HFTers' percentage of volume due to obtaining time priority, as the positive significant coefficient of the interaction term between the relative tick size and the HFT dummy variable indicates.

Insert Table 9 About Here

7.2. Active Update of Quotes

Using minute-by-minute snapshot to the limit order book, Table 3 shows that as relative tick size goes down, non-HFTers become the more and more likely providers of the best quotes. There are two ways that an order can become the best bid (or ask). The first, the active way, is by actively posting a bid (or ask) within spread and thereby narrowing the spread. The second, the passive way, is through the worsening the collective best quotes of orders, and thus the remaining order becomes the very best bid (or ask) (Blume and Goldstein (1997)). Since orders actively improving BBO or orders passively cancelled or executed can both affect the state of the limit order book, one concern towards the finding in Table 3 is that the finding is due to orders which passively change the top-of-book. To address this concern, we focus only on orders which actively improve the BBO, and examine how their activities vary with relative tick size.

We uses Nasdaq top-of-book of displayed limit orders for the week of Feb 22 - 26, 2010 to calculate the dollar size of each order which actively improves BBO for the 117 stocks on each day. For each update in the top-of-book, we check whether the best bid price after top-of-book update is higher than previous best bid price, or whether the best bid ask price after top-of-book update is lower than the previous ask price. If so, we consider the BBO has been actively improved. Since the data allow us to identify the type of traders who actively improve the BBO, we aggregate the dollar sizes for all orders that actively improved BBO for HFTers and non-HFTers respectively.[34] The specification of the regression is as follows:

$$\Delta y_{i,t} = \beta \times tick_{relative_{i,t}} + \Gamma \times X_{i,t} + u_{j,t} + \epsilon_{i,t} \tag{15}$$

$\Delta y_{i,t}$ is the percentage difference between non-HFTers' and HFTers' aggregate dollar size of orders that actively improve the BBO. It is calculated as $\frac{DolSize_{nonhft} - DolSize_{hft}}{DolSize_{nonhft} + DolSize_{hft}}$ for stock *i* on date *t*, where $DolSize_j$ denotes the aggregate dollar size of orders that improve the BBO for trader type j. The variable of interest, β, measures how the percentage difference between HFTers and non-HFTers active dollar size improvement varies with relative tick size. $u_{j,t}$ is the industry-by-time fixed effect. $X_{i,t}$ are the control variables presented in Table 1.

Table 10 displays the regression output. The result shows a decrease in relative tick size leads to an increase in the percent of active dollar size improvement from non-HFTs relative to HFTs. When the tick size decreases from the stocks with the lowest price to stocks with the median tick size, the percent of active dollar size improvement from non-HFTs in Column 7 increases by -2.061*(0.034-0.192) =32.6%. Therefore, our result that HFTs provides a larger proportion of best prices when relative tick size is large is robust when we only consider active improvement on quotes.

[34] When weighted by size, the case where traders only improve best quotes by 1 share carries less weight. The aggregated dollar size is not weighted by time. This is due to the fact that for those orders, the duration of them staying in the limit order book cannot be determined. The top-of-book data provide the information on the cumulative depth, but without the information on order ID.

Insert Table 10 About Here

7.3. Stocks with one-cent spread

The previous sections establish the result that a large relative tick size encourages HFT liquidity provision. One concern towards the finding is that the result is applicable only to stocks with one-cent quoted spread, for which stocks with less than one-cent natural quoted spread are forced to have a quoted spread equal to the tick-size wide. In this section, we provide evidence showing that the effect of relative size on HFT liquidity provision persists among the cross-sectional stocks and it is not exclusive to stocks with one-cent quoted spread.

To achieve this end, we re-perform the liquidity provision analysis in Table 5 with a newly added dummy variable capture the effect induced by the *one-cent* quoted spread. The regression takes the following form:

$$y_{i,t} = u_{j,t} + \beta_1 \times tick_{relative_{i,t}} + \beta_2 \times onecent_{dummy_{i,t}} + \Gamma \times X_{i,t} + \epsilon_{i,t} \qquad (16)$$

Specially, the $onecent_{dummy_{i,t}}$ equal to 1 if the time weighted quoted spread calculated using ITCH is less than 1.05 cents for stock *i* on day *t*.[35] All other variables are as described in Table 5.

It is worth mentioning that the regression may suffer from endogeneity issue, as HFT liquidity provision may affect the quoted spread of a stock. However, the purpose of this regression is not to show how one-cent quoted spread casually impact HFT liquidity provision, but to shed light on when controlling for stocks with one-cent quoted spread, whether tick size constraints still play a role in affecting HFT liquidity provision.

[35] We try other cut-offs close to 1 and the results are similar.

Panel A of Table 11 presents the regression results with $y_{i,t}$ as the daily percentage of depth at BBO provided by HFTers (*HFTdepth*), and Panel B presents the regression results with $y_{i,t}$ as the percentage of volume with HFTers as liquidity providers (*HFTvolume*) for each stock *i* on date *t*.[36] The positively significant coefficients for $onecent_{dummy_{i,t}}$ in both Panel A and B indicate that stocks with one-cent quoted spread are associated with higher HFT liquidity provision. Panel A shows that after controlling for stocks with one-cent, stocks with large relative tick size attract more HFTers to supply liquidity at BBO, though the coefficient on $tick_{relative_{i,t}}$carry less weight and significant power compared with the one in Panel A of Table 5. However, as Panel B displays, the effect of relative tick size on percentage of volume with HFTers as liquidity providers remain strongly significantly.

Insert Table 11 About Here

7.4. Cross-sectional variation of HFT activity with relative tick size

The previous sections have established the results that stocks with large relative tick sizes attract more HFT liquidity provision. Questions are raised on whether the finding is associated with HFTer's preference on trading low-priced stocks, and they actively engage in not only liquidity making, but also liquidity taking activities on these stocks. To address this concern, we follow the methodology in Table 3, and use the Nasdaq HFT trade file for 117 stocks in Oct 2010 to compute the volume with HFTers as liquidity takers relative to total volume for each portfolio *J* on day *t*:

$$HFTliqtake_{Jt} = \frac{\sum_{i\in J}(HN_{it} + HH_{it})}{\sum_{i\in J}(NH_{it} + HH_{it} + HN_{it} + NN_{it})} \tag{17}$$

36 We also analyze who provides the best price using multinomial logit and probit models and obtain similar results. The results are not reported but are available upon request.

Table 12 Panel A shows the average trading volume percentage due to HFT liquidity takers for each of the market-cap by relative-tick-size portfolios. The result demonstrates a clear pattern according to which the volume with HFTers as liquidity taker decreases with the relative tick size.

In addition, to examine the whether HFTers exhibit any preference on trading low-priced stocks, we compute the volume with HFTers engaged either as liquidity providers or liquidity takers relative to total volume for each portfolio *J* on day *t*:

$$HFTliqboth_{Jt} = \frac{\sum_{i \in J}(HN_{it} + NH_{it} + HH_{it})}{\sum_{i \in J}(NH_{it} + HH_{it} + HN_{it} + NN_{it})} \tag{18}$$

Table 12 Panel B shows the average trading volume percentage due to HFT liquidity providers or HFT liquidity takers for each of the market-cap by relative-tick-size portfolios. The panel does not exhibit clear pattern of HFT activity with respect to the relative tick size. The percentage of volume with HFTers engaged as either liquidity makers or takers increases with the relative tick size for large-cap stocks, but deceases with relative tick size for small-cap stocks, and displays a U-shaped pattern for stocks with medium market-cap.

Insert Table 12 about Here

8. Conclusion

We contribute to the HFT literature by providing a tick size–based explanation of speed competition in liquidity provision. Contrary to the common belief that HFTers provide liquidity at lower cost, we find that non-HFTers play a more prominent role in providing liquidity at better prices than HFTers do, particularly when the relative tick size is small. An increase in the relative tick size, however, constrains non-HFTers' abilities to undercut price and helps HFTers achieve price priority when they quote the same price as non-HFTers. As a consequence, HFTers provide a larger fraction of liquidity for low-priced stocks in which a one-cent uniform tick size implies a larger relative tick size. For ETFs that track the same index, HFTers are more active in trading lower-priced ETFs. In addition, splits increase HFT liquidity-providing activity and the quoted spread and lengthen the queue for providing liquidity; reverse splits decrease the quoted spread and reduce depth and also decrease HFT liquidity-providing activity. We also find that HFTers are less active in market making for stocks with higher PIN, suggesting that quick access to information does not give market-making HFTers a more prominent role in liquidity provision with a higher probability of informed trading.

The tick size constraints channel provides a new possibility for explaining the results reported in the extant literature. The literature on HFT finds that speed improvement increases liquidity and the common belief about the source of improvement is that it comes from traders with higher speed, which is at odds with the results of the current paper. There are two possible reconciliations of this discrepancy. First, the extant literature is based on technology shocks at millisecond or full-second scale, whereas speed competition in our sample is recorded at nanosecond scale (Gai, Yao and Ye, 2013), implying a diminishing return for speed (Jones, 2013). A more intriguing conjecture is that the source of liquidity improvement may come from traders who decline to improve their speed in the face of those technology shocks. Traders who pay for technology enhancement enjoy time priority and a lesser need to undercut the price, whereas traders who choose not to pay for speed need to undercut the price more often. A test for this conjecture based on account-level data would be very interesting.

The tick size constraints channel provides new insights into the policy debate over HFT. The current policy debate focuses on whether additional regulation is required, and if so how to pursue it. Our paper points to a new direction: HFT may simply be a consequence of the existing tick size regulation and one possible policy solution would be deregulation instead of additional regulation. At the minimum, the first step in pursuing additional regulation would involve due diligence to evaluate the impact of the existing tick size regulation on HFT.

This paper also provides a benchmark for evaluating the economic consequences of increasing tick size. The JOBS Act encourages the SEC to examine the possibility of increasing tick size, and a pilot program is under way for less liquid stocks. Proponents of a wider tick size have offered three rationales for this position (Weild, Kim and Newport, 2012). First, a larger tick size controls the growth of HFT. Second, a larger tick size should increase liquidity. Our paper shows that a larger tick size *encourages* HFT without improving liquidity. The third argument for increasing tick size is that a larger tick size increases market-making revenue, supports sell-side equity research, and increases the number of IPOs. The economic argument that controlling prices leads to non-price competition is valid, but we doubt that non-price competition would take the form of stock research or more IPOs. The causal impact of the tick size on IPOs has never been proved by academic research, but speed competition under constrained price competition is well established by this paper. We believe that an increase in tick size would create more rents for time priority, and would fuel another round of the arms race in speed.

Our paper can be extended in various ways. First, current theoretical work on speed competition focuses on the role of information. Our paper points out another channel for speed competition: tick size constraints. Models using discrete prices can be constructed to indicate the value of speed and the impact of tick size constraints on market quality. Second, speed competition is not the only market design response to tick size constraints (Buti, Consonni, Rindi, Wen and Werner, 2014). In a companion paper, we examine the causal impact of tick size constraints on the taker/maker fee market or the market that charges liquidity providers but subsidizes liquidity demanders (Yao and Ye, 2014). The literature on market microstructure focuses on liquidity and price discovery under certain market designs, but market design is also endogenous, and it should prove fruitful to examine why certain market designs exist in the first place. Finally, the SEC recently announced a pilot program for increasing tick size for a number of small stocks, believing that tick size may need to be wider for less liquid stocks. We encourage the SEC to consider decreasing tick size for liquid stocks in the pilot program, particularly for large stocks with lower prices.

References

Acharya, V., Pedersen, L., 2005, Asset pricing with liquidity risk, Journal of Financial Economics 77, 375–410.

Amihud, Y., Mendelson, H., 1986, Asset pricing and bid-ask spread, Journal of Financial Economics 17, 223-249.

Angel, J, 1997, Tick size, share prices, and stock splits, Journal of Finance 52, 655–681.

Angel, J., Harris, L., Spatt, C., 2010, Equity trading in the 21st Century, Unpublished working paper, University of South California, Los Angeles, CA.

Angel, J., Harris, L., Spatt, C., 2013, Equity trading in the 21st Century: an update, Unpublished working paper, University of South California, Los Angeles, CA.

Avramovic, A., 2012, Manufacturing volume: the stock split solution, Credit Suisse, available from Edge, accessed Dec. 18, 2013.

Baker, K., Gallagher, P., 1980, Management's view of stock splits, Financial Management 9, 73-77.

Baker, M., Greenwood, R., Wurgler, J., 2009, Catering through nominal share prices, Journal of Finance 64, 2559-2590.

Baron, M., Brogaard, J., Kirilenko, A., Risk and return in high frequency trading, Unpublished working paper, Princeton University, Princeton, NJ.

Benartzi, S., Michaely, R., Thaler, R., Weld, W., 2009, The nominal share price puzzle, Journal of Economic Perspectives 23,121-142.

Berk, J., DeMarzo, P., 2013, Corporate finance, Prentice Hall Press, New Jersey.

Biais, B., Foucault, T., 2014, HFT and market quality, Bankers, Markets&Investors 128, 5-19.

Biais, B., Foucault, T., Moinas, S., 2013, Equilibrium high-frequency trading, Unpublished working paper, HEC Paris, France.

Boehmer, E., Fong, K., Wu, J., 2014, International evidence on algorithmic trading, Unpublished working paper, University of Georgia, GA.

Brennan, M., Copeland, T., 1988, Stock splits, stock prices, and transaction costs, Journal of Financial Economics 22, 83-101.

Brennan, M., Hughes, P., 1991, Stock prices and the supply of information, Journal of Finance 46, 1665-1691.

Brogaard, J., Hagströmer, B., Nordén, L., Riordan, R., 2014, Trading fast and slow: collocation and market quality, Unpublished working paper, Washington University, WA.

Brogaard, J., Hendershott, T., Riordan, R., 2014, High frequency trading and price discovery, Review of Financial Studies, forthcoming.

Budish, E., Cramton, P., Shim, J., 2013, The High-frequency trading arms race: frequent batch auctions as a market design response, Unpublished working paper, University of Chicago, Chicago, IL.

Buti, S., Consonni, F., Rindi, B., Wen, Y., Werner, I., 2014, Tick size: theory and evidence, Unpublished working paper, Monterey, CA.

Byun, J., Rozeff, M., 2003, Long-run performance after stock splits: 1927 to 1996, Journal of Finance 58, 1540–6261.

Campbell, J., Hilscher, J., Szilagyi, J., 2008, In search of distress risk, Journal of Finance 63, 2899-2939.

Carrion, A., 2013, Very fast money: high-frequency trading on the NASDAQ, Journal of Financial Markets 16, 680-711.

Chordia, T., Goyal, A., Lehmann, B., Saar, G., 2013, High-frequency trading, Journal of Financial Markets 16, 637-645.

Chordia, T., Sarkar, A., Subrahmanyam, A., 2005, An empirical analysis of stock and bond market liquidity, Review of Financial Studies 18, 85-130.

Chordia, T., Roll, R., Subrahmanyam, A., 2000, Commonality in liquidity, Journal of Financial Economics 56, 3-28.

Douglas, G., Miller, J., 1974, Quality competition, industry equilibrium, and efficiency in the price-constrained airline market, The American Economic Review 64, 657-669.

Dyl, E., Elliott, W., 2006, The share price puzzle, Journal of Business 79, 2045-2066.

Easley, D., Kiefer, N., O’Hara, M., Paperman, J., 1996, Liquidity, information and infrequently traded stocks, Journal of Finance 51, 1405-1436.

Easley, D., O’Hara, M., 1987, Price, trade size, and information in securities markets, Journal of Financial Economics 19, 69-90.

Easley, D., O’Hara, M., 2004, Information and the cost of capital, Journal of Finance 59, 1553-1583.

Easley, D., O'Hara, M., 2010, Microstructure and ambiguity, Journal of Finance 65, 1817–1846.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002, Is information risk a determinant of asset returns?, Journal of Finance 57, 2185-2221.

Foucault, T., Hombert, J., Rosu, L., 2013, News trading and speed, Unpublished working paper, HEC Paris, Paris, France.

Foucault, T., Kadan, O., Kandel, E., 2013, Liquidity cycles and make/take fees in electronic markets, Journal of Finance 68, 299-341.

Foucault, T.,Pagano, M., Röell, A., 2013, Market liquidity: theory, evidence and policy, Oxford University Press, Oxford.

Foucault, T., Kozhan, R., Tham, W., 2014, Toxic arbitrage, Unpublished working paper, HEC Paris, France.

Gai, J., Ye, M., Yao, C., 2013, The externalities of high frequency trading, Unpublished working paper, University of Illinois at Urbana, Champaign, Urbana, IL.

Gai, J., Choi, D., O'Neal, D., Ye, M., Sinkovits, R., 2013, Fast construction of nanosecond level snapshots of financial markets, In Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery.

Glosten, L., Milgrom, P., 1985, Bid, ask and transaction prices in a specialist market with heterogeneously informed traders, Journal of Financial Economics 14, 71-100.

Grinblatt, M., Masulis, R., Titman, S., 1984, The valuation effects of stock splits and stock dividends, Journal of Financial Economics 13, 461-490.

Hagströmer, B., Nordtr, L., 2013, The diversity of high-frequency traders, Journal of Financial Markets 16, 741-770.

Hasbrouck, J., Saar, G., 2013, Low-latency trading, Journal of Financial Markets 16, 646-679.

Hendershott, T., Riordan, R., 2013, Algorithmic trading and the market for liquidity, Journal of Financial and Quantitative Analysis 48, 1001-1024.

Hendershott, T., Jones, C., Menkveld, A., 2011, Does algorithmic trading improve liquidity?, Journal of Finance 66, 1–33.

Hirschey, N., 2013, Do high-frequency traders anticipate buying and selling pressure?, Unpublished working paper, London Business School, London, UK.

Ho, T., Stoll, H., 1981, Optimal dealer pricing under transactions and return uncertainty, Journal of Financial Economics, 9, 47-73.

Ikenberry, D., Rankin, G., Stice, E., 1996, What do stock splits really signal?, Journal of Financial and Quantitative Analysis 31., 357-375.

Jones, C., 2013, What do we know about high-frequency trading?, Unpublished working paper, Columbia University, New York, NY.

Kadapakkam, P., Krishnamurthy, S., Tse, Y., 2005, Stock splits, broker promotion and decimalization, Journal of Financial and Quantitative Analysis 40, 873-895.

Kalay, A., Kronlund, M., 2014, The market reaction to stock split announcements: earnings information after all, Unpublished working paper, Columbia University, New York, NY.

Lakonishok, J., Lev, B., 1987, Stock splits and stock dividends: why, who and when, Journal of Finance 42, 913–932.

Martinez, V., Rosu, L., 2013. High-frequency traders, news and volatility, Unpublished working paper, City University of New York, New York, NY.

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

O'Hara, M., 2003, Presidential address: liquidity and price discovery, Journal of Finance 58, 1335-1354.

O'Hara, M., 2007, Optimal microstructures, European Financial Management 13, 825-832.

O'Hara, M., Saar, G., Zhong, Z., 2014, Relative tick size and the trading environment, Unpublished working paper, Cornell University, Ithaca, NY.

O'Hara, M., Yao, C., Ye, M., 2014, What's not there: the odd-lot bias in market data, Journal of Finance, forthcoming.

Parkinson, M., 1980, The extreme value method for estimating the variance of the rate of return, Journal of Business 53, 61-65.

Pastor, L., Stambaugh, R., 2003, Liquidity risk and expected stock returns, Journal of Political Economy 111, 642-686.

Roberts, M., Whited, T., 2012, Endogeneity in empirical corporate finance, Handbook of the Economics of Finance 2, 493-572

Rockoff, H., 2008, Price Controls, The concise encyclopedia of economics, Indianapolis: The Liberty Fund, 409-412.

Sandås, P., 2001, Adverse selection and competitive market making: empirical evidence from a limit order market, Review of Financial Studies 14, 705-734.

Schultz, P., 2000, Stock splits, tick size, and sponsorship, Journal of Finance 55, 429-450.

Stoll, H., 2000, Friction, Journal of Finance, 55.4, 1479-1514.

U.S. Securities and Exchange Commission (SEC), 2010, Concept release on equity market structure.

Weild, D., Kim, E., Newport, L., 2012, The trouble with small tick sizes, Grant Thornton, available at www.GrantThornton.com.

Yao, C., Ye, M., 2014, Tick size constraints, market structure, and liquidity, Unpublished working paper, University of Illinois at Urbana, Champaign, Urbana, IL.

Table 1. Summary Statistics

This table presents the summary statistics for all stocks and ETFs. Panel A contains stocks in the NASDAQ HFT sample. Panel B provides the summary statistics for the non-leveraged ETF sample used in the twin ETFs test. Panel C lists the summary statistics for the leveraged ETF sample used in the diff-in-diff test. All the variables are measured for each stock per day unless otherwise indicated. *HFTdepth* is the percentage of depth at the best bid/ask provided by HFTers. *HFTvolume* is the percentage of trading volume with HFTers as liquidity providers. $tick_{relative}$ is the reciprocal of price. *logmcap* is the log value of market capitalization. *turnover* is the annualized share turnover. *volatility* is the standard deviation of open-to-close returns based on the daily price range, that is, high minus low, proposed by Parkinson (1980). $logbv_{average}$ is the logarithm of the average book value of equity per shareholder at the end of the previous year (December 2009). *idiorisk* is the variance on the residual from a 60-month beta regression using the CRSP Value Weighted Index. *age* (in 1k days) is the length of time for which price information is available for a firm on the CRSP monthly file. *numAnalyst* is the number of analysts providing one-year earnings forecasts calculated from I/B/E/S. *pastreturn* is the past one-month return. *PIN* is the probability of informed trading. *return* is the contemporaneous daily return. *Qtspd* represents the time-weighted quoted spread. *pQtspd* represents the time-weighted proportional quoted spread. $Qtspd_{adj}$ and $pQtspd_{adj}$ represent the time-weighted quoted spread and the time-weighted proportional quoted spread after adjusting for fees. *SEffspd* represents the size-weighted effective spread and *pSEffspd* represents the size-weighted proportional effective spread. $SEffspd_{adj}$ and $pSEffspd_{adj}$ represent the size-weighted effective spread and the size-weighted proportional effective spread after adjusting for fees. *Depth* represents the time-weighted dollar depth at the best bid and ask. *RunsInProcess* is the proxy for HFT market-making activity. The sample period for Panel A and Panel B is October 2010 and the sample period for Panel C is 30 days before leveraged ETFs splits/reverse splits during 2010 and 2011.

Panel A. NASDAQ HFT Sample

	Mean	Median	Std.	Min.	Max.
HFTdepth(in pcg)	0.321	0.298	0.166	0.006	0.744
HFTvolume(in pcg)	0.285	0.273	0.134	0	0.728
$tick_{relative}$	0.048	0.034	0.038	0.002	0.192
logmcap	22.01	21.473	1.893	19.371	26.399
turnover	2.367	1.801	2.179	0.074	33.301
volatility	0.015	0.013	0.009	0.002	0.099
$logbv_{average}$	13.148	13.196	2.112	8.822	17.881
idiorisk	0.013	0.008	0.017	0.001	0.139
age (in 1k days)	9.751	7.671	7.839	0.945	30.955
numAnalyst	13.731	12	10.048	1	48
pastreturn	0.105	0.096	0.073	-0.077	0.336
PIN	0.118	0.111	0.052	0.021	0.275

Panel B. Leveraged ETF Sample				
	Mean		Median	
	Treatment	Control	Treatment	Control
Split Sample				
return	-0.002	0	0.006	-0.007
Qtspd (in cent)	20.259	2.062	18.232	1.938
pQtspd (in bps)	11.544	9.797	9.657	10.082
$Qtspd_{adj}$ (in cent)	20.849	2.652	18.822	2.528
$pQtspd_{adj}$ (in bps)	11.881	12.868	9.994	13.476
SEffspd (in cent)	11.846	1.526	10.732	1.423
pSEffspd (in bps)	6.673	7.477	6.039	7.156
$SEffspd_{adj}$ (in cent)	12.446	2.126	11.332	2.023
$pSEffspd_{adj}$ (in bps)	7.015	10.6	6.38	9.835
Depth (in mn dollars)	0.179	0.188	0.141	0.114
RunsInProcess	15.607	37.570	15.042	13.401
Reverse Split Sample				
return	0	0	-0.002	0.002
Qtspd (in cent)	1.362	4.415	1.066	2.333
pQtspd (in bps)	13.839	9.464	12.6	9.645
$Qtspd_{adj}$ (in cent)	1.952	5.005	1.656	2.923
$pQtspd_{adj}$ (in bps)	20.224	11.837	17.958	11.644
SEffspd (in cent)	1.179	3.031	1	1.635
pSEffspd (in bps)	12.191	7.185	10.815	6.389
$SEffspd_{adj}$ (in cent)	1.779	3.631	1.6	2.235
$pSEffspd_{adj}$ (in bps)	18.683	9.598	16.124	8.105
Depth (in mn dollars)	1.304	0.512	0.412	0.095
RunsInProcess	106.230	63.938	66.973	36.252

Table 2. Who Provides the Best Quotes?

This table displays the percentage of time HFTers and non-HFTers provide the best bid and ask quotes to the NASDAQ limit-order book. The sample includes 117 stocks in the NASDAQ HFT data from October 2010. Stocks are sorted first into 3-by-3 portfolios by average market cap and then by average price from September 2010. For each portfolio and each trading day, we calculate the percentage of time that HFTers are the sole providers of the best quotes, the percentage of time that non-HFTers are the sole providers of the best quotes, and the percentage of time that both provide the best quotes. Column (1) presents the average percentage of time that HFTers are the sole providers of the best quotes and column (2) presents the average percentage of time that non-HFTers are the sole providers of the best quotes. Column (3) presents the average percentage of time that both HFTers and non-HFTers provide the best quotes. Column (4) shows the difference between column (1) figures and (2) figures. *t*-statistics for column (4) based on 21 daily observations are presented in column (6). *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively.

		(1)	(2)	(3)	(4)	(5)
	Relative Tick Size	HFT Only	Non-HFT Only	HFT & Non-HFT	Non-HFT minus HFT	*t*-stat
Large Cap	Large (Low Price)	1.6%	2.5%	95.9%	0.9%***	7.27
	Medium (Medium Price)	11.9%	18.6%	69.6%	6.7%***	14.01
	Small (High Price)	16.8%	37.7%	45.5%	20.9%***	37.81
Middle Cap	Large (Low Price)	18.0%	15.2%	66.8%	-2.9%***	-3.52
	Medium (Medium Price)	20.0%	56.6%	23.4%	36.6%***	36.03
	Small (High Price)	20.7%	63.7%	15.7%	43.0%***	67.15
Small Cap	Large (Low Price)	11.3%	54.7%	34.1%	43.4%***	27.55
	Medium (Medium Price)	20.2%	55.8%	24.0%	35.7%***	30.11
	Small (High Price)	18.6%	70.7%	10.7%	52.1%***	66.79
Total		15.4%	41.7%	42.9%	26.3%***	18.31

Table 3. Market Share of BBO Depth Provided by HFTers

This table presents the percentages of depth at the NASDAQ best bid and offer (BBO) provided by HFTers. The sample includes 117 stocks in the NASDAQ HFT data from October 2010. The stocks are sorted first by average market cap and then by average price from September 2010. To calculate the share-weighted average for each portfolio on each day, we aggregate the number of shares provided by HFTers at the BBO and then divide it by the total number of shares at the BBO for that portfolio. *t*-statistics are calculated based on 21 daily observations. *, **, and *** represent statistical significance of large-minus-small differences at the 10%, 5%, and 1% levels, respectively.

	Large Relative Tick Size (Low Price)	Medium Relative Tick Size (Medium Price)	Small Relative Tick Size (High Price)	Large-Small Relative Tick Size (Low-High Price)	*t*-stat
Large Cap	55.66%	45.44%	35.07%	20.59%***	22.10
Middle Cap	39.73%	29.24%	24.61%	15.13%***	22.88
Small Cap	25.78%	23.02%	20.78%	5.00%***	3.18
L-S Cap	29.88%***	22.43%***	14.29%***		
t-statistics	18.84	17.92	16.80		

Table 4. Percentage of Volume with HFTers as the Liquidity Providers

This table presents the trading volume percentage due to HFTers as liquidity providers. The sample includes 117 stocks in the NASDAQ HFT data from October 2010. The stocks are sorted first by average market cap and then by average price from September 2010 into 3-by-3 portfolios. To calculate the volume-weighted average for each portfolio on each day, we aggregate the volumes due to HFT liquidity providers and then divide that figure by the total volume for that portfolio. *t*-statistics are calculated based on 21 daily observations. *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels of large-minus-small differences, respectively.

	Large Relative Tick Size (Low Price)	Medium Relative Tick Size (Medium Price)	Small Relative Tick Size (High Price)	Large-Small Relative Tick Size (Low-High Price)	*t*-stat
Large Cap	49.96%	38.23%	35.93%	14.03%***	15.54
Middle Cap	39.30%	24.03%	24.33%	14.97%***	18.74
Small Cap	24.11%	18.88%	18.49%	5.62%***	5.38
L-S Cap	25.84%***	19.35%***	17.43%***		
t-statistics	21.33	21.76	18.22		

Table 5. HFT Liquidity Provision

This table presents the results of the regressions of HFT liquidity provision on relative tick size (nominal tick size divided by price). The regressions use the NASDAQ HFT data sample as of October 2010 and merges it with all the other variables calculated from databases including CRSP, COMPUSTAT, etc. Panel A presents the results for the daily percentage of depth provided by HFTers. Panel B contains the results for the daily percentage of trading volume with HFTers as liquidity providers. The regression specification is:

$$y_{i,t} = u_{j,t} + \beta \times tick_{relative_{i,t}} + \Gamma \times X_{i,t} + \epsilon_{i,t}$$

where $tick_{relative_{i,t}}$ is the daily inverse of the stock price. $u_{j,t}$ represents industry-by-time fixed effects. The definitions for the control variables $X_{i,t}$ are presented in Table 1. *t*-statistics are shown in parenthesis; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A. HFT Liquidity Provision as in HFT Trading Depth

Dep. Var	*HFTdepth* (in percentage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$tick_{relative}$	0.678***	0.699***	0.673***	0.691***	0.641***	0.677***	0.658***
	(5.48)	(5.60)	(5.40)	(5.16)	(5.26)	(5.44)	(4.98)
logmcap	0.038***	0.037***	0.038***	0.032***	0.032***	0.038***	0.023***
	(16.57)	(15.13)	(16.56)	(7.86)	(12.15)	(16.37)	(4.90)
turnover		0.001					0.001
		(0.62)					(0.83)
volatility		-0.837*					-0.531
		(-1.75)					(-1.13)
$logbv_{average}$			-0.003				0.002
			(-1.48)				(0.99)
idiorisk				-0.290			-0.308
				(-1.33)			(-1.46)
age				0.005***			0.005***
				(7.40)			(7.83)
numAnalyst				-0.001			-0.001
				(-0.96)			(-1.09)
PIN					-0.358***		-0.435***
					(-4.05)		(-4.71)
pastreturn						-0.069	-0.117**
						(-1.33)	(-2.26)
R^2	0.461	0.462	0.462	0.485	0.466	0.462	0.494
N	2268	2268	2268	2268	2268	2268	2268
Industry*time FE	Y	Y	Y	Y	Y	Y	Y

Panel B. HFT Liquidity Provision as in HFT Trading Volume

Dep. Var	*HFTvolume* (in percentage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$tick_{relative}$	0.937***	0.940***	0.943***	0.992***	0.893***	0.937***	0.956***
	(9.88)	(9.86)	(10.10)	(9.35)	(9.72)	(9.87)	(9.58)
logmcap	0.051***	0.048***	0.051***	0.042***	0.044***	0.051***	0.031***
	(33.54)	(29.38)	(33.52)	(13.68)	(23.63)	(33.47)	(9.43)
turnover		0.008***					0.007***
		(5.50)					(5.44)
volatility		-0.803**					-0.536
		(-2.34)					(-1.64)
$logbv_{average}$			0.003***				0.005***
			(2.68)				(4.01)
idiorisk				-0.495***			-0.547***
				(-3.00)			(-3.76)
age				0.002***			0.003***
				(3.77)			(6.37)
numAnalyst				0.001**			0.001**
				(2.25)			(2.04)
PIN					-0.423***		-0.407***
					(-6.64)		(-6.51)
pastreturn						-0.007	-0.034
						(-0.22)	(-1.09)
R^2	0.632	0.641	0.634	0.639	0.643	0.632	0.664
N	2268	2268	2268	2268	2268	2268	2268
Industry*time FE	Y	Y	Y	Y	Y	Y	Y

Table 6. Tick size and profits

This table presents the results of regressions of HFT unit profit on relative tick size (nominal tick size divided by price). The regression uses the NASDAQ HFT data sample as of October 2010.The regression specification is:

$$y_{i,t,n} = \beta_1 \times tick_{relative_{i,t}} + \beta_2 \times HFTdummy_{i,t,n} + \beta_3 \times tick_{relative_{i,t}} \times HFTdummy_{i,t,n}$$
$$+u_{j,t} + \Gamma \times X_{i,t} + \epsilon_{i,t,n}$$

Columns (1)–(4) present regression results for daily unit profit measured on the assumption that inventories are emptied at 5-minute, 30-minute, 60-minute intervals and at daily closing. $tick_{relative_{i,t}}$ is the daily inverse of the stock price. $HFTdummy_{i,t,n}$ is equal to 1 if the profit measure is from HFTers and 0 otherwise. $u_{j,t}$ represents the industry-by-time fixed effects. The definitions for the control variables $X_{i,t}$ are presented in Table 1. *t*-statistics are shown in parenthesis; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Dep. Var	*Unit Profit (bps)*			
	(1)	(2)	(3)	(4)
$tick_{relative}$	15.497***	17.513***	18.977***	29.734**
	(5.45)	(3.97)	(3.43)	(2.31)
HFTdummy	0.762***	0.421**	0.243	-0.094
	(6.43)	(2.29)	(1.06)	(-0.18)
$tick_{relative}$ * *HFTdummy*	7.054**	2.328	-0.261	-16.344
	(2.21)	(0.47)	(-0.04)	(-1.13)
logmcap	-0.160*	-0.129	-0.144	-0.233
	(-1.67)	(-0.87)	(-0.77)	(-0.54)
turnover	-0.085**	0.038	0.062	-0.178
	(-2.37)	(0.68)	(0.89)	(-1.11)
volatility	-19.509**	-91.450***	-105.805***	-179.865***
	(-1.97)	(-5.95)	(-5.48)	(-4.01)
$logbv_{average}$	0.036	0.017	0.082	0.314*
	(0.98)	(0.30)	(1.15)	(1.92)
idiorisk	2.883	11.624*	16.507*	26.079
	(0.65)	(1.70)	(1.92)	(1.31)
age	-0.008	-0.028	-0.019	-0.033
	(-0.61)	(-1.37)	(-0.76)	(-0.57)
numAnalyst	0.040**	0.033	0.033	0.063
	(2.54)	(1.33)	(1.07)	(0.89)
PIN	0.797	2.895	7.518**	9.835
	(0.44)	(1.02)	(2.11)	(1.19)
pastreturn	-1.658*	1.373	4.769**	9.321**
	(-1.66)	(0.89)	(2.45)	(2.06)
R^2	0.220	0.214	0.203	0.194
N	4484	4484	4484	4484
Industry*time FE	Y	Y	Y	Y

Table 7. Correlation Test

This table presents the cross-sectional correlations between HFT proxies and the HFT activity measures which are calculated from the NASDAQ HFT dataset. The HFT activity measures include the percentage of depth provided by HFTers, the percentage of volume with HFTers as liquidity providers, and the percentage of volume with HFTers as liquidity takers. The HFT proxies include *RunsInProcess* by Hasbrouck and Saar (2013), the Quote-to-Trade Ratio and the Dollar Volume (in $100)–to–Message Ratio multiplied by -1. *P*-values are shown under correlation coefficients; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

HFT Measures	HFTdepth (liq. making)	HFTvolume (liq. making)	HFTvolume (liq. taking)
RunsInProcess	0.650***	0.765***	0.283***
	<.0001	<.0001	0.002
Quote / Trade Ratio	-0.385***	-0.390***	-0.082
	<.0001	<.0001	0.378
-Trading Vol. (in $100) / Message Ratio	0.031	-0.252***	-0.274***
	0.7403	0.0061	0.003

Table 8. Diff-in-Diff Test Using Leveraged ETF Splits

This table presents the results of a diff-in-diff test using leveraged ETF split (and reverse-split) events in 2010 and 2011. Panel A reports the results for 5 splits and Panel B reports the results for 21 reverse splits. The event windows are 30 days before and 30 days after splits/reverse splits. The regression specification is:

$$y_{i,t,j} = u_{i,t} + \gamma_j + \rho \times D_{trt\,i,t,j} + \theta \times return_{i,t,j} + \epsilon_{i,t,j}$$

$u_{i,t}$ is the index by time fixed effects and γ_j is the ETF fixed effects. The treatment dummy D_{trt}, equals 0 for the control group. For the treatment group, D_{trt} equals 0 before splits/reverse splits and 1 after the splits/reverse splits. *return* is the daily return. The regression includes both an index-by-time fixed effect and an ETF fixed effect. Column (1) - (9) present regression results for liquidity. *Qtspd* represents the time-weighted quoted spread. *pQtspd* represents the time-weighted proportional quoted spread. $Qtspd_{adj}$ and $pQtspd_{adj}$ represent the time-weighted quoted spread and the time-weighted proportional quoted spread after adjusting for fees. *SEffspd* represents the size-weighted effective spread and *pSEffspd* represents the size-weighted proportional effective spread. $SEffspd_{adj}$ and $pSEffspd_{adj}$ represent the size-weighted effective spread and the size-weighted proportional effective spread after adjusting for fees. *Depth* represents the time-weighted dollar depth at the best bid and ask in millions of dollars. Column (10) presents regression results for HFT activity, which is proxied by *RunInProcess* proposed by Hasbrouck and Saar (2013). *t*-statistics are shown in parenthesis; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A. Split Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Qtspd (in cent)	$Qtspd_{adj}$ (in cent)	*pQtspd* (in bps)	$pQtspd_{adj}$ (in bps)	*Depth* (in mn)	*SEffspd* (in cent)	$SEffspd_{adj}$ (in cent)	*pSEffspd* (in bps)	$pSEffspd_{adj}$ (in bps)	*RunInProc.*
D_{trt}	-9.686***	-9.686***	0.998*	1.205**	0.015	-5.385***	-5.385***	0.801*	1.012**	3.372***
	(-15.83)	(-15.83)	(1.89)	(2.23)	(1.41)	(-11.96)	(-11.96)	(1.96)	(2.40)	(3.28)
return	-9.307**	-9.307**	-6.954**	-8.112**	0.000	-5.235*	-5.235*	-5.482**	-6.660**	-3.725
	(-2.48)	(-2.48)	(-2.15)	(-2.45)	(-0.00)	(-1.89)	(-1.89)	(-2.19)	(-2.57)	(-0.59)
R^2	0.911	0.911	0.741	0.717	0.915	0.868	0.868	0.637	0.731	0.978
N	597	597	597	597	597	597	597	597	597	597
Index*time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ETF FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B. Reverse Split Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Qtspd (in cent)	*Qtspd*$_{adj}$ (in cent)	*pQtspd* (in bps)	*pQtspd*$_{adj}$ (in bps)	*Depth* (in mn)	*SEffspd* (in cent)	*SEffspd*$_{adj}$ (in cent)	*pSEffspd* (in bps)	*pSEffspd*$_{adj}$ (in bps)	*RunInProc.*
D_{trt}	1.198***	1.198***	-2.611***	-5.030***	-0.324***	0.692***	0.692***	-2.958***	-5.418***	-32.108***
	(8.45)	(8.45)	(-13.33)	(-18.27)	(-5.98)	(5.92)	(5.92)	(-12.45)	(-17.18)	(-9.72)
return	-1.478	-1.478	-3.551**	-4.449**	0.842**	-0.696	-0.696	-1.748	-2.661	18.536
	(-1.39)	(-1.39)	(-2.42)	(-2.16)	(2.08)	(-0.80)	(-0.80)	(-0.98)	(-1.13)	(0.75)
R^2	0.832	0.832	0.883	0.856	0.789	0.768	0.768	0.785	0.798	0.852
N	2517	2517	2517	2517	2517	2517	2517	2517	2517	2517
Index*time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ETF FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 9. Trades due to Time Priority with Respect to Tick Size

This table presents the regressions of percentage of dollar volume due to the limit orders having time priority on relative tick size (nominal tick size divided by price). The regression uses the NASDAQ HFT data sample of February 22-26, 2010.The regression specification is:

$$y_{i,t,n} = \beta_1 \times tick_{relative_{i,t}} + \beta_2 \times HFTdummy_{i,t,n} + \beta_3 \times tick_{relative_{i,t}} \times HFTdummy_{i,t,n}$$
$$+u_{j,t} + \Gamma \times X_{i,t} + \epsilon_{i,t,n}$$

$y_{i,t,n}$ is the percentage of dollar volume due to limit orders having time priority, for stock *i* on date *t* of trader type *n*, calculated as the dollar volume due to limit orders having time priority over total dollar volume (dollar volume due to limit orders having time priority plus that having price priority). $tick_{relative_{i,t}}$ is the daily inverse of the stock price. $HFTdummy_{i,t,n}$ equals to 1 if the profit measure is from HFTers and 0 otherwise. $u_{j,t}$ represents the industry-by-time fixed effects. The definitions for the control variables $X_{i,t}$ are presented in Table 1. *t*-statistics are shown in parenthesis; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Dep. Var	*HFTers Dollar Volume Due to Time Priority* (in percentage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$tick_{relative}$	3.062***	3.262***	3.051***	3.252***	3.069***	3.071***	3.374***
	(16.49)	(17.34)	(16.45)	(17.40)	(16.53)	(16.51)	(17.94)
HFTdummy	0.076***	0.076***	0.076***	0.077***	0.076***	0.076***	0.077***
	(8.05)	(8.17)	(8.06)	(8.49)	(8.03)	(8.05)	(8.53)
$tick_{relative}$ * *HFTdummy*	0.442**	0.441**	0.442**	0.436**	0.444**	0.442**	0.439**
	(1.99)	(2.02)	(2.00)	(2.06)	(2.01)	(2.00)	(2.09)
logmcap	0.091***	0.086***	0.091***	0.052***	0.089***	0.092***	0.047***
	(25.79)	(23.40)	(25.75)	(8.47)	(22.42)	(25.78)	(7.19)
turnover		0.007***					0.007***
		(3.17)					(3.05)
volatility		-4.119***					-3.141***
		(-5.21)					(-3.97)
$logbv_{average}$			-0.005**				0.002
			(-2.12)				(0.72)
idiorisk				-0.886***			-0.720**
				(-2.99)			(-2.34)
age				0.007***			0.007***
				(7.80)			(7.15)
numAnalyst				0.006***			0.006***
				(5.61)			(5.27)
pin					-0.156		-0.147
					(-1.43)		(-1.35)
pastreturn						-0.062	-0.019
						(-0.94)	(-0.30)
R^2	0.657	0.667	0.659	0.688	0.658	0.658	0.695
N	1074	1074	1074	1074	1074	1074	1074
Industry*time FE	Y	Y	Y	Y	Y	Y	Y

Table 10. Active Quote Improvement

This Table presents the regression of the difference between HFTers' and non-HFTers' dollar sizes of active quote improvement relative to total dollar size of active quote improvement on relative tick size (nominal tick size divided by price). The regression uses the NASDAQ HFT data sample as of February 22-26, 2010 and merges it with all the other variables calculated from databases including CRSP, COMPUSTAT, etc. The regression specification is:

$$\Delta y_{i,t} = \alpha + \beta \times tick_{relative_{i,t}} + \Gamma \times X_{i,t} + u_{j,t} + \epsilon_{i,t}$$

where $\Delta y_{i,t}$ is non-HFTers' minus HFTers' dollar sizes of active quote improvement relative to total dollar size of active quote improvement for stock *i* on day *t*. $tick_{relative_{i,t}}$ is the daily inverse of the stock price. $u_{j,t}$ represents the industry-by-time fixed effects. The definitions for the control variables $X_{i,t}$ are presented in Table 1. *t*-statistics are shown in parenthesis; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Dep. Var	*Dollar Depth of non-HFTers Minus HFTers' Active Quote Improvement* (in Percentage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$tick_{relative}$	-1.867***	-1.910***	-1.904***	-1.918***	-1.898***	-1.831***	-2.061***
	(-3.57)	(-3.52)	(-3.66)	(-3.26)	(-3.54)	(-3.54)	(-3.38)
logmcap	-0.101***	-0.096***	-0.102***	-0.101***	-0.085***	-0.100***	-0.065***
	(-9.05)	(-8.00)	(-8.99)	(-4.54)	(-6.55)	(-9.02)	(-2.84)
turnover		-0.016*					-0.013
		(-1.77)					(-1.51)
volatility		2.312					2.131
		(0.83)					(0.84)
$logbv_{average}$			-0.018*				-0.033***
			(-1.89)				(-2.95)
idiorisk				0.391			0.578
				(0.48)			(0.66)
age				-0.003			-0.007**
				(-0.82)			(-2.02)
numAnalyst				0.002			0.001
				(0.38)			(0.24)
pin					0.976***		1.326***
					(2.64)		(3.63)
pastreturn						-0.268	-0.286
						(-1.27)	(-1.38)
R^2	0.460	0.466	0.467	0.463	0.471	0.463	0.498
N	540	540	540	540	540	540	540
Industry*time FE	Y	Y	Y	Y	Y	Y	Y

Table 11. HFT Liquidity Provision and Stocks with One-cent Quoted Spread

This table presents the regressions of HFT liquidity provision on relative tick size, controlling for stocks with one-cent quoted spread. The regressions use the NASDAQ HFT data sample as of October 2010 and merges it with all the other variables calculated from databases including ITCH, CRSP, COMPUSTAT, etc. Panel A presents the results for the daily percentage of depth provided by HFTers. Panel B contains the results for the daily percentage of trading volume with HFTers as liquidity providers. The regression specification is:

$$y_{i,t} = u_{j,t} + \beta_1 \times tick_{relative_{i,t}} + \beta_2 \times onecent_{dummy_{i,t}} + \Gamma \times X_{i,t} + \epsilon_{i,t}$$

where $tick_{relative_{i,t}}$ is the daily inverse of the stock price. $onecent_{dummy_{i,t}}$ is a dummy variable equal to one if the time-weighted quoted spread calculated using ITCH is less than 1.05 cents for stock *i* on day *t*, and equal to zero otherwise. $u_{j,t}$ represents industry-by-time fixed effects. The definitions for the control variables $X_{i,t}$ are presented in Table 1. *t*-statistics are shown in parenthesis; *, **, and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

Panel A. HFT Liquidity Provision as in HFT Trading Depth

Dep. Var	*HFTdepth* (in percentage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$tick_{relative}$	0.224*	0.241**	0.216*	0.231*	0.208*	0.227**	0.240**
	(1.96)	(2.08)	(1.89)	(1.89)	(1.82)	(1.98)	(1.97)
onecent_dummy	0.109***	0.108***	0.109***	0.097***	0.105***	0.108***	0.089***
	(9.23)	(9.14)	(9.29)	(8.13)	(8.94)	(9.15)	(7.39)
logmcap	0.027***	0.026***	0.027***	0.023***	0.023***	0.027***	0.017***
	(10.66)	(9.81)	(10.55)	(5.37)	(7.84)	(10.62)	(3.58)
turnover		0.001					0.001
		(0.66)					(0.81)
volatility		-0.552					-0.408
		(-1.14)					(-0.85)
$logbv_{average}$			-0.003*				0.001
			(-1.92)				(0.56)
idiorisk				0.072			0.026
				(0.34)			(0.12)
age				0.004***			0.004***
				(6.78)			(6.86)
numAnalyst				-0.001			-0.001
				(-1.03)			(-1.12)
pin					-0.293***		-0.364***
					(-3.39)		(-4.19)
pastreturn						-0.035	-0.083*
						(-0.73)	(-1.71)
r2	0.487	0.488	0.488	0.504	0.491	0.488	0.511
N	2268	2268	2268	2268	2268	2268	2268
Industry*time FE	Y	Y	Y	Y	Y	Y	Y

Panel B. HFT Liquidity Provision as in HFT Trading Volume

Dep. Var	*HFTvolume* (in percentage)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$tick_{relative}$	0.557***	0.551***	0.565***	0.585***	0.536***	0.556***	0.591***
	(7.41)	(7.36)	(7.53)	(7.22)	(7.22)	(7.39)	(7.46)
onecent_dummy	0.091***	0.091***	0.090***	0.086***	0.087***	0.091***	0.078***
	(11.76)	(11.95)	(11.71)	(10.81)	(11.32)	(11.78)	(9.95)
logmcap	0.041***	0.039***	0.042***	0.034***	0.036***	0.041***	0.026***
	(24.76)	(22.38)	(24.93)	(11.70)	(19.03)	(24.76)	(8.56)
turnover		0.007***					0.007***
		(6.63)					(6.16)
volatility		-0.561*					-0.428
		(-1.78)					(-1.38)
$logbv_{average}$			0.003***				0.004***
			(2.92)				(3.86)
idiorisk				-0.174			-0.256*
				(-1.21)			(-1.83)
age				0.001***			0.002***
				(3.06)			(5.09)
numAnalyst				0.001**			0.001**
				(2.56)			(2.33)
pin					-0.369***		-0.345***
					(-6.57)		(-6.12)
returns						0.021	-0.004
						(0.66)	(-0.12)
r2	0.660	0.669	0.662	0.663	0.669	0.660	0.683
N	2268	2268	2268	2268	2268	2268	2268
Industry*time FE	Y	Y	Y	Y	Y	Y	Y

Table 12. Percentage of Volume with HFTers as the Liquidity Takers and Percentage of Volume with HFTers as the Liquidity Takers or Liquidity Makers

This table presents the trading volume percentage due to HFTers as liquidity takers and trading volume percentage due to HFTers as liquidity providers or liquidity takers. The sample includes 117 stocks in the NASDAQ HFT data from October 2010. The stocks are sorted first by average market cap and then by average price from September 2010 into 3-by-3 portfolios. Panel A displays the volume-weighted percentage of trading volume with HFTers liquidity takers. To calculate the volume-weighted average for each portfolio on each day, we aggregate the volumes due to HFT liquidity takers and then divide that figure by the total volume for that portfolio. Panel B displays the volume-weighted percentage of trading volume with HFTers engaged as either liquidity providers or takers. To calculate the volume-weighted average for each portfolio on each day, we aggregate the volumes due to HFT liquidity providers or HFT liquidity takers and then divide that figure by the total volume for that portfolio.*t*-statistics are calculated based on 21 daily observations. *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels of large-minus-small differences, respectively.

Panel A: Percentage of Trading Volume with High-frequency Traders as Liquidity Takers

	Large Relative Tick Size (Low Price)	Medium Relative Tick Size (Medium Price)	Small Relative Tick Size (High Price)	Large-Small Relative Tick Size (Low-High Price)	*t*-stat
Large Cap	44.11%	47.57%	48.28%	-4.17%***	-5.31
Middle Cap	39.19%	42.40%	46.97%	-7.78%***	-6.25
Small Cap	27.65%	33.36%	39.64%	-11.99%***	-10.69
L-S Cap	16.46%***	14.21%***	8.64%***		
t-statistics	13.76	13.04	8.47		

Panel B: Percentage of Trading Volume with High-frequency Traders as Liquidity Makers or Takers

	Large Relative Tick Size (Low Price)	Medium Relative Tick Size (Medium Price)	Small Relative Tick Size (High Price)	Large-Small Relative Tick Size (Low-High Price)	*t*-stat
Large Cap	73.72%	69.03%	68.14%	5.58%***	9.30
Middle Cap	64.46%	57.39%	61.45%	3.01%**	2.55
Small Cap	45.61%	46.51%	51.41%	-5.80%***	-3.92
L-S Cap	28.11%***	22.53%***	16.73%***		
t-statistics	20.03	18.75	15.13		

Two-Sided Markets, Make-Take Fees and Competition between Stock Exchanges

Yong Chao Chen Yao Mao Ye*

November 15, 2014

Abstract

This paper examines competition between stock exchanges for order flow by setting make fees for limit orders and take fees for market orders. We find that exchanges can use make-take fees to create sub-tick prices and facilitate trades that are blocked by the tick size regulation. The discrete tick size generates two-sided markets in which the charge on each side matters even for the same total charge. Our two-sided market model explains several anomalies relative to a standard one-sided market. First, the breakdown of make-take fees is not neutral for social welfare, and the equilibrium fee structure always involves one side being subsidized and the other side being charged. Second, the price competition of two identical exchanges does not lead to Bertrand outcome, but to mixed strategy equilibrium with positive profits. This justifies the diversity of fee structures and their frequent adjustments, as well as the entry of exchanges with new fee structures. Third, the model predicts that liquidity makers prefer being charged (subsidized) when the tick size is large (small), and the market becomes more fragmented under a larger tick size. We find empirical evidence consistent with these two predictions using reverse splits of ETFs as exogenous shocks to the relative tick size, with paired ETFs that track the same index but do not reverse split as controls.

---

* Yong Chao is from the University of Louisville. E-mail: yong.chao@louisville.edu. Telephone: (502) 852-2573. Chen Yao is from the University of Warwick. E-mail: Chen.Yao@wbs.ac.uk. Mao Ye is from the University of Illinois at Urbana-Champaign. E-mail: maoye@illinois.edu. Telephone: (217)244-0474. We thank Jim Angel, Robert Battalio, Jean Colliard, Shane Corwin, Thierry Foucault, Chester Spatt, Bart Yueshen Zhou and seminar participants at Baruch College and University of Notre Dame for their suggestions. This research is supported by National Science Foundation grant 1352936. We also thank Bei Yang, Yingjie Yu, Fan Yang and Chao Zi for their excellent research assistance.

"It was not obvious to Brad why some exchanges paid you to be a taker and charged you to be a maker, while others charged you to be a taker and paid you to be a maker. No one he asked could explain it, either. To Brad this all just seemed bizarre and unnecessarily complicated—and it raised all sorts of questions. "Why would you pay anyone to be a taker? I mean, who is willing to pay to make a market? Why would anyone do that?"

Michael Lewis: Flash Boys

Nowadays, stock prices are determined in organized exchanges through the interaction of buyers and sellers. However, the mechanisms through which exchanges set prices for their services are not well understood, particularly when competing exchanges provide identical services. This paper aims to fill this gap in our understanding by theoretically and empirically examining price competition between stock exchanges in the United States when they compete for order flow.

Currently, traditional dealers and specialists play a minimum role in stock exhanges, and trading happens through the direct interaction of buyers and sellers. A trader in these markets can either post a (buy or sell) limit order by specifying the price and quantity, or accept the terms of a previously posted limit order by submitting a market order. Once a trade occurs, limit orders pay make fees and market orders pay take fees to the particular exchange, and the exchange profits from the total fee, or the sum of the make and take fees. At first glance, competition between exchanges on fees should lead to two simple market outcomes. First, exchanges should compete on the total fee, but how the total fee breaks down into make and take fees should not matter, because standard economic models show that, for a given tax level, it does not matter who, the buyer or the seller, is charged for the tax. Second, the competitive environment for these exchanges seems to follow the form of a Bertrand model: exchanges compete on price (contrary to the Cournot game) and the services they provide are barely differentiated (contrary to the Hotelling's model). Therefore, competition should result in pure strategy equilibrium with a total fee of zero and zero profit for the exchange, thus deterring entry into the game if establishing a new exchange would incur fixed costs. The canonical predictions of tax neutrality and Bertrand equilibrium, however, are inconsistent with stylized facts regarding exchange competition.

Two facts in particular are inconsistent with fee neutrality. First, table 1 shows that exchanges can charge the same total fee but with varying breakdowns of take and make fees. This fact applies not only to competing exchanges, such as the NASDAQ and the NYSE, but also to platforms operated by the same parent company, such as EDGA and EDGX. Notably, EDGX operates under the maker/taker model, in which the liquidity maker pays a negative make fee (obtains a rebate), whereupon the liquidity taker pays a positive take fee. EDGA operates under the taker/maker market, in which the liquidity taker pays a negative take fee (obtains a rebate), whereupon the liquidity maker pays a positive make fee. The non-consolidation of the two platforms with the same total fee strongly indicates that the breakdown of the total fee serves its own purposes. Second, the make fee and the take fee always carry opposite signs. Table 1 reveals that all exchanges charge one side while subsidizing the other side. The adjustment of the fee structure documented by Cardella, Hao and Kalcheva (2013) provides further evidence for these two facts. Exchanges can switch from charging liquidity makers to charging liquidity takers but hold the total fee fixed, suggesting that the breakdowns of fees matter. However, none of the major exchanges has ever switched to charging both sides.

Figure 2 demonstrates evidence that is inconsistent with Bertrand equilibrium. The total fee does not inevitably move towards zero; occasionally it bounces back. We continue to observe entry of platforms with new fee structures. For example, on October 22, 2010, BATS created a new trading platform, BATS Y, with a taker/maker fee structure.

This paper explains the deviation from neutrality and Bertrand equilibrium based on the burgeoning literature exploring two-sided platforms. Two-sided platforms are markets in which the volume of transactions depends not only on the overall level of the fees charged by a given platform, but also on the structure of the total-fee breakdowns between two sides of the market (Rochet and Tirole (2006)). In a one-sided market, price allocation can be neutralized by end-users. A market becomes two-sided when end-users face constraints on neutralizing the allocation of prices. In stock exchanges, such constraints come from tick size regulations. Securities and Exchange Commission (SEC) rule 612 (the Minimum Pricing Increment) of regulation NMS 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."[1] Due to this rule, liquidity providers and demanders cannot negotiate price increments of less than a penny. The make-take fees set by the exchanges, however, are not subject to the tick size regulation. The nature of the fee game reflects competition between exchanges for orders based on proposing sub-penny prices for makers and takers.

Consider the following game involving two stock exchanges, a buyer, and a seller. All of them are risk neutral. The seller's valuation is uniformly distributed on [0, 0.5] and the buyer's valuation is uniformly distributed on [0.5, 1]. Due to the unequal valuations, efficiency requires that a trade occur. The game plays out in three stages. At Date 0, the exchanges set their make-take fees. At Date 1, nature decides whether the buyer or the seller arrives first. The trader arriving first (the liquidity maker) chooses whether to submit a limit order to exchange 1 or exchange 2, or

[1] There are some limited exemptions such as the Retail Price Improvement (RPI) Program and mid-point peg orders.

submits no limit order. Consider a case in which the tick size is equal to 1.[3] Then, the price of the limit order needs to be either 0 or 1, due to the tick size regulation. At Date 2, the liquidity taker arrives, and decides whether to submit a market order to the exchange selected by the liquidity maker. Exchanges profit from charging the make fee and take fee conditional on execution.

The non-neutrality of the fee can be demonstrated using one exchange, but the same intuition follows with two exchanges. When exchanges charge neither side, trades fail to happen because the buyer is willing to post a limit order only at price grid 0 and the seller is willing to post a limit order only at grid price 1, both of which would be rejected by the liquidity taker.[4] A positive charge for both sides again leads to no-trade equilibrium. Fees with opposite signs, however, are able to create effective buy and sell prices within the same tick. To see this, consider the following taker/maker structure that subsidizes the liquidity taker 0.5 and charges the liquidity maker 0.5 and assume the buyer arrives first. The buyer knows that the seller is willing to accept a limit order at price 0 after the 0.5 subsidy on Date 2. The buyer thus posts a limit buy order at price 0 on Date 1, although he is charged 0.5 for providing liquidity. Therefore, the nominal buy and sell prices are both 0, but the effective buy and sell prices are 0.5. Under such a fee structure, trades always happen and efficiency is restored. This example demonstrates that make-take fees are not neutral: with a total fee equal to 0, the exchange can move from no-trade equilibrium to socially optimal equilibrium by taxing one party to the transaction and subsidizing the other.

We demonstrate the non-existence of pure-strategy equilibrium between competition exchanges under a positive tick size, which fundamentally differentiates the two-sided market

---

[3] Such an assumption not only facilitates our analysis, but it also captures the reality of the market. All fees in all U.S. exchanges have an absolute value below one tick, and one main purpose of the fees is to create price increments of less than a penny. Therefore, it is important to focus on buyers and sellers whose valuations fall within the same tick.

[4]The exception is the knife-edge case in which the valuation of the buyer is exactly equal to 1 or the valuation of the seller is exactly equal to 0.

studied in this paper from the one-sided market studied in Colliard and Foucault (2012). We first show that pure strategy equilibrium with positive total fees do not exist, because competing exchanges have incentives to undercut each other toward a zero total fee. The surprising insight from the two-sided market, however, is that the zero-total-fee outcome cannot be sustained in equilibrium. When both exchanges set the total fee at zero, one can always find a profitable deviation by *increasing* the total fee. One strategy in response would be to charge liquidity maker ε more while charging liquidity taker $\mu\varepsilon$ less (with $0<\mu<1$).[5] Such a deviation reduces the liquidity maker's profit conditional on execution, but increases the liquidity taker's incentive to accept the limit order, which attracts liquidity makers with a higher trading surplus. The other strategy is to charge maker $\mu\varepsilon$ less but increase the charge to taker ε more, which appeals to liquidity makers with low gains from trade. Therefore, exchanges have incentives to increase the total fee when their competitors move toward zero total fees, consistent with what Figure 1 exhibits. We then demonstrate the existence of mixed-strategy equilibrium between competing exchanges, which justifies the frequent fee changes documented by Cardella, Hao and Kalcheva (2013). Importantly, mixed-strategy equilibrium generates positive profit, and the profit increases linearly with tick size, which explains the new entry into the fee game.

The model generates two predictions that are supported by empirical results. The first prediction relates to the competition between the maker/taker market and the taker/maker market. Negative fees for liquidity providers are generally thought to encourage liquidity provision (Malinova and Park (2013)). Therefore, the intuition suggests that liquidity makers should prefer a market that pays a rebate. Surprisingly, we show that liquidity providers prefer being subsidized instead of being charged when the tick size is small, and they prefer being charged instead of being

[5] We assume that this small change does not move the make fee from negative to positive.

subsidized when the tick size is large. This surprising result is, however, supported by the following diff-in-diff test. We use ETF reverse splits as exogenous shocks to the relative tick size (1 divided by price), with ETFs that reverse-split as the treatment group and with ETFs that track the same index but experience no reverse splits as the control group. The same diff-in-diff specification is used to support another prediction of our model: a large relative tick size encourages entry of new trading platforms and market fragmentation. We find that trading in the treatment group becomes more consolidated and that market share for those in the treatment group in the taker/maker market decreases relative to that in the maker/taker market.

Our paper contributes to the literature on competition for order flow between stock exchanges. The seminal work by Stigler (1961, 1964) argues that trading tends to centralize in major market centers. Yet a recent paper by O'Hara and Ye (2011) demonstrates significant fragmentation of trading volume. A number of theoretical studies examine whether competing trading platforms can co-exist, but trading platforms in most models are passive with no optimization behavior or strategy. (Glosten (1994), Parlour and Seppi (2003), Hendershott and Mendelson (2000), Foucault and Menkveld (2008)). In addition, competing exchanges in this literature usually provide differentiated services or apply alternative trading rules (Pagnotta and Philippon (2013), Santos and Scheinkman (2001), Rust and Hall (2003)). We contribute to the literature by proposing a price competition model for otherwise-identical trading platforms. The closest model to ours is that of Colliard and Foucault (2012), which predicts fee neutrality and Bertrand equilibrium under a zero tick size. We predict fee non-neutrality and mixed-strategy equilibrium under a positive tick size, which implies that tick size can drive fee competition and market fragmentation. In this regard, we also contribute to the literature on tick size (Chordia and Subrahmanyam (1995) and Yao and Ye (2014), O'Hara, Saar and Zhong (2014), and Buti, Rindi, Wen and Werner (2014)), particularly market making and tick size (Chordia and Subrahmanyam (1995) and Anshumman and Kalay (1998)), and provide insight into recent policy debates. Encouraged by the Jumpstart Our Business Startups Act (the JOBS Act), the SEC has announced a pilot program to increase the tick size to five cents for small stocks.[6] The proposed argument is that increasing tick size can increase market-making revenue and support sell-side equity research and, eventually, increase the number of IPOs (Weild, Kim and Newport (2012)). Our results, however, indicate that an increase in tick size may intensify fee competition between exchanges.

This paper contributes theoretically and empirically to the literature on two-sided platforms. The lion's share of papers in the two-sided market literature are based on cross-side externality (Evans and Schmalensee (2013), and Rochet and Tirole (2006)). We are the first to show that a market without cross-side externalities can be two-sided because of the regulation on trading prices. Empirical analyses of two-sided markets are scarce, and our paper provides one of the first empirical analyses of competition between two-sided markets with clean identification.

Finally, this paper contributes to the literature on make-take fees, and provides insight into the current policy debate. Regulators are interested in make-take fees because they are transaction costs for traders. The take fee is capped by the SEC at thirty cents per round lots in U.S. equity markets, and more aggressive initiatives are underway on Capitol Hill to completely ban these fees.[7] Ours is the first paper to show that the fee competition can be a natural consequence of *existing* tick size regulations. One argument to ban the fee is based on fairness, because the fee leads to wealth transfer from one side of the market to the other side. We show that the make-take fee structure can provide Pareto improvement of social welfare. Even the side being charged can benefit. One other argument to ban the fees cites their complexity, while the complexity can be

[6] "SEC Provides Details of 5-Cent Tick Test," *Wall Street Journal*, June 25, 2014.
[7] "Make-take fees in spotlight on Capitol Hill." http://marketsmedia.com/make-take-fees-spotlight/.

explained by the mixed-strategy equilibrium found in this paper. The last argument to ban the fee is based on agency issues. Recently, Battalio, Corwin and Jennings (2014) find that broker/dealers have a strong incentive to route customers' limit orders to the market offering the highest rebate, because brokers/dealers are permitted to pocket such rebates. This conflict of interest leads to two policy proposals: 1) passing the rebate back to customers; 2) eliminating the fee structure (Angel, Harris, and Spatt (2010, 2013)). In our opinion, passing the rebate back to customers is a direct solution to the agency issue, while eliminating the fee might hinder the would-be efficiency of trading.

The rest of the paper is organized as follows. Section I describes the setup of the model. Section II considers a monopoly exchange's profit maximization fee structure. Section III examines the fee competition between two exchanges. Section IV presents the empirical test of the theoretical predictions of the model. Section V concludes the paper and discusses the policy implications. The appendix contains proofs of the lemmas and propositions that are used or derived in the paper.

I. Model

Our model has three types of risk-neutral players: a buyer b, a seller s, and a stock exchange (or two competing stock exchanges, 1 and 2). Denote the tick size (or the minimum pricing increment) as d.[8] The grid starts at p_o, and the regulation dictates that the order can be priced *only* at $p_o + n \cdot d$ $(n = 0, 1, 2, 3, \dots)$. Without loss of generality, we focus on one grid interval, in which the trade can occur only at either p_o or $p_o + d$. Buyer b's valuation of the stock is $v_b \sim U[p_o + d/2,\ p_o + d]$, and seller s's valuation of the stock is $v_s \sim U[p_o,\ p_o + d/2]$. That is,

[8] SEC Rule 612 requires that $d = 1$ cent for any stock with a price level set above 1 dollar. However, the relative tick size can be heterogeneous for stocks with varying price levels (Yao and Ye (2014)). Here we allow for a tick of any magnitude.

the buyer's valuation is above the quote midpoint whereas the seller's valuation is below the quote midpoint.[9] v_b and v_s are the buyer's and seller's private information, respectively.

We consider a three-stage game. At Date 0, each stock exchange sets its fee structure at $F = (f_m, f_t)$, where f_m denotes the make fee for the liquidity provider and f_t denotes the take fee for the liquidity taker. Fees are charged only upon trade execution. At Date 1, nature either draws a buyer to arrive at the market first with probability p, or draws a seller to arrive at the market first with probability $1 - p$.[10] Whoever arrives at the market first can propose a trading price P on one of the exchanges after observing the make-take fees, or he can choose not to submit a limit order at all. *Due to the tick size regulation, trading can occur only at a price that falls on either of the two pre-specified endpoints of the grid.* That is,

$$P = \{p_0, p_0 + d\}. \quad (1)$$

At Date 2, the counterparty arrives. The counterparty observes the make and take fees as well as the price proposed by the maker, and then decides whether to trade. If he decides to trade, he must join the platform that the liquidity provider chooses at Date 1 and trade at the proposed trading price P.[11]

Exchanges profit through the total fees they charge. To ensure that stock exchange(s) continue to survive, it is reasonable to assume that make-take fees have to be at least balanced (i.e., the net sum of the make and take fees is non-negative).[12] That is,

$$f_m + f_t \geq 0. \quad (2)$$

---

[9] The prediction of the model does not change as long as the valuations of some buyers and sellers are within the same tick.
[10] It turns out that the analysis of the model does not depend on p.
[11] In reality, a market order can trade with a limit order on another platform due to regulation NMS. However, there is a routing fee for cross-platform execution.
[12] Figure 1 shows that some trading platforms can charge negative total fees for short periods of time.

The nature of this model is to characterize the competition between exchanges. Therefore, the model is parsimonious for limit and market orders. Traders do not choose the order type, the order book is empty when the maker arrives, the market has only one tick, and none of the traders has private information. Theoretical studies on order-placing strategy generally provide a richer structure within which to address these issues (Rosu (2009), Parlour (1998), and Parlour and Seppi (2003)). These four simplifications, however, allow us to examine more complex competitions between stock exchanges. Section IV demonstrates that the game between exchanges reaches complex mixed-strategy equilibrium even granted these four simplifications. The purpose of this paper is to build a simple model to capture the nature of exchange competitions through make-take fees. We show that the model is able to explain the stylized facts regarding non-neutrality and a non-Bertrand outcome, and that empirical data support the model's predictions.

II. One Exchange

The result for non-neutrality can be established under a monopoly exchange, and the intuition regarding competing exchanges follows. The analysis of the game with one exchange helps to build intuitions and establish intermediary results that simplify the analysis of the game with competing exchanges. Section II.A examines the buyer's and seller's optimal behaviors given fees (f_m, f_t). Section II.B endogenizes the fee structure and solves the profit maximization fee for the monopoly exchange.

A. Buyer/Seller Behavior with one Exchange

After the natural draw at Date 1, the subgame is a sequential-move game. We can solve the subgame by backward induction for any fixed fee structure (f_m, f_t).

Without loss of generality, suppose the buyer arrives at Date 1 and the seller arrives at Date 2.[13] Given the make and take fees (f_m, f_t) at Date 2, the seller's willingness-to-sell (WTS) will be $v_s + f_t$. That is, the seller will trade if and only if

$$P \geq v_s + f_t. \quad (3)$$

At Date 1, the buyer's willingness-to-buy (WTB) will be $v_b - f_m$. This implies that the buyer will propose a price such that

$$P \leq v_b - f_m. \quad (4)$$

Combining these results with the fact that $p_0 \leq v_s$ and $v_b \leq p_0 + d$, (3) and (4) can be rewritten as

$$p_0 \leq v_s \leq P - f_t \quad (3')$$

$$P + f_m \leq v_b \leq p_0 + d. \quad (4')$$

Recall our balanced fee assumption (2): $f_m + f_t \geq 0$ is equivalent to

$$P - f_t \leq P + f_m. \quad (2')$$

Thus, in order for a trade to occur, (2), (3), and (4) (or equivalently, (2'), (3') and (4')) together require:

$$p_0 \leq v_s \leq P - f_t \leq P + f_m \leq v_b \leq p_0 + d. \quad (*)$$

Insert Figure 2 Here

Through (*), we can see the impact of the tick size requirement on trading.

In the presence of the tick size requirement, the buyer's proposed price P is restricted to only two discrete points: either p_0 or $p_0 + d$. Without make-take fees, the buyer would be willing to buy only at p_0, except for the knife-edge case of $v_b = p_0 + d$; similarly, the seller would be willing to sell only at $p_0 + d$, except for the knife-edge case of $v_s = p_0$. Hence, (*) can never hold

[13] The solution to the case when the seller arrives first is similar.

for any P in the price grid that is enforced by the tick size regulation, except when $v_s = p_0$ or $v_b = p_0 + d$. Therefore, the would-be efficient trade cannot occur when the buyer's and seller's valuations fall within the same tick.

Lemma 1: *When the tick size requirement is enforced, an otherwise efficient trade is blocked by the tick size requirement if no make-take fees are imposed.*

We can understand Lemma 1 by considering the bid-ask spread. With zero fee charged to both liquidity makers and liquidity takers, the buyer will post a limit buy order at price p_0 if he arrives first, and the seller will post a limit sell order at price $p_0 + d$ if he arrives first. In limit-order book markets, the bid price is set by the limit buy order and the ask price is determined by the limit sell order. Therefore, the market without fees has a bid-ask spread exactly equal to the tick size d. Empirically, a large number of low-priced liquid stocks have bid-ask spreads equal to one tick. A recent study by Credit Suisse demonstrates that 50% of S&P stocks priced below $100 per share have one-penny quoted spreads (Avramovic (2012)). The clustering of quoted spreads on one penny suggests that many of those stocks should have an equilibrium bid-ask spread of less than one penny in the absence of tick size constraints. We show that exchanges can use make-take fees to create a real spread of less than 1 penny, thus realizing gains from such a trade, and to extract a portion of the gains by proposing the solution that circumvents the tick size constraints.

Hereafter, we take the tick size requirement (1) as given, and consider the case in which stock exchange(s) charge make-take fees.

If $f_m > 0$, then (*) implies that the buyer must propose $P = p_0$, because otherwise $P + f_m > p_0 + d$. It then follows from (*), again, that $f_t < 0$, because otherwise we would have $P - f_t < p_0$. Similarly, if $f_m < 0$, then (*) implies that the buyer must propose $P = p_0 + d$ and $f_t >$

0. Thus, a necessary condition for the occurrence of a trade is charging one side while subsidizing the other side.

Meanwhile, (*) leads to

$$\begin{cases} p_0 \le P - f_t \\ P + f_m \le p_0 + d \end{cases},$$

which is equivalent to

$$\begin{cases} f_t \le P - p_0 \\ f_m \le p_0 + d - P \end{cases}.$$

Thus, for (*) to hold, we must have:

$$f_m + f_t \le d.$$

The results are summarized in Lemma 2 below.

Lemma 2 (Fee Structure, Trading Price and Participation with One Exchange): *With tick size requirement (1), given make-take fees* (f_m, f_t)*, the following consequences hold.*

(1) In order for a trade to happen, the exchange must charge one side while subsidizing the other side. Moreover, the total fee cannot exceed the tick size.

That is,

$$f_m \cdot f_t < 0. \tag{5}$$

and

$$f_m + f_t \le d. \tag{6}$$

(2) Trading prices under alternate fee structures:

a. if the liquidity provider is a buyer, then the proposed trading price is

$$P = \begin{cases} p_0 & when\ f_m > 0\ (so\ that\ f_t < 0) \\ p_0 + d & when\ f_m < 0\ (so\ that\ f_t > 0) \end{cases}. \tag{7}$$

b. if the liquidity provider is a seller, then the proposed trading price is

$$P = \begin{cases} p_0 + d & when\ f_m > 0\ (so\ that\ f_t < 0) \\ p_0 & when\ f_m < 0\ (so\ that\ f_t > 0) \end{cases}. \qquad (8)$$

(3) Participation:

a. *if the liquidity provider is a buyer, then a buyer with* $v_b \geq max\{P + f_m, p_0 + d/2\}$ *will trade with a seller with* $v_s \leq min\{P - f_t, p_0 + d/2\}$.

b. *if the liquidity provider is a seller, then a buyer with* $v_b \geq max\{P + f_t, p_0 + d/2\}$ *will trade with a seller with* $v_s \leq min\{P - f_m, p_0 + d/2\}$.

Lemma 2 provides an important intuition that helps to explain why exchanges always charge liquidity makers and liquidity takers fees with opposite signs, although competing exchanges subsidize liquidity makers and takers in the reverse order. Trades fail to occur when the valuations of buyers and sellers fall within the same tick. Compared with charging zero fees to both liquidity makers and liquidity takers, charging positive fees to both sides can make the spread between the effective buy and sell prices wider than the grid size, which again leads to zero-trade equilibrium. By charging one side and subsidizing the other side, the exchange shifts the effective buy and sell prices to a region that is below the valuations of some buyers but above the valuations of some sellers, enabling trades to occur. To be sure, a profit-optimizing exchange would not set effective buy and sell prices that facilitate all trades, but a necessary condition for profit maximization is that fees be set so as to allow some trades to occur. We include a formal characterization of the profit maximization fee in section II.B.

Lemma 2 also reveals the nature of the fee game. By setting the make-take fees, the exchange essentially determines *effective buy and sell prices*. In particular, under $F = (f_m, f_t)$, if the liquidity provider is a buyer, then the effective buy price is $P + f_m$, and the effective sell price is $P - f_t$; similarly, if the liquidity provider is a seller, then the effective buy and sell prices are $P + f_t$ and $P - f_m$, respectively.

Denote the *effective* buy and sell prices (p_b, p_s) as

$$(p_b, p_s) \equiv \begin{cases} (P + f_m, P - f_t) & \textit{if the liquidity provider is a buyer} \\ (P + f_t, P - f_m) & \textit{if the liquidity provider is a seller} \end{cases} \tag{9}$$

Lemma 3 reveals that, for any fee structure in the taker/maker market, there always exists a fee structure in the maker/taker market that achieves exactly the same market outcome. This maker/taker fee decreases the charge to the liquidity maker by one tick and increases the charge to a liquidity taker by one tick relative to what occurs under the taker/maker model.

Lemma 3 (Twin Fees): *Given any* $F = (f_m, f_t)$ *with* $f_m > 0 > f_t$ *and* $0 \leq f_m + f_d \leq d$, *there always exists a unique* $\tilde{F} = (\tilde{f}_m, \tilde{f}_t)$ *resulting in the same effective buy and sell prices under* F, *and*

$$\begin{cases} \tilde{f}_m = f_m - d < 0 \\ \tilde{f}_t = f_t + d > 0 \end{cases} \tag{10}$$

As discussed above, we can always transform each exchange's choice variables from fees (f_m, f_t) to the effective buy and sell prices (p_b, p_s). Such relabelings can avoid a tedious discussion of trading price P contingent on the sign of f_m and f_t. So, in the following analysis, we consider each exchange's decision variables as the effective buy and sell prices (p_b, p_s).

Given effective buy and sell prices (p_b, p_s), the marginal buyer's and seller's valuations are given by

$$\tilde{v}_b \equiv \max\{p_b, p_0 + \tfrac{d}{2}\} \text{ and } \tilde{v}_s \equiv \min\{p_s, p_0 + \tfrac{d}{2}\}. \tag{11}$$

Then the number of buyers and sellers are given by

$$q_b = \Pr(v_b \geq \tilde{v}_b) \text{ and } q_s = \Pr(v_s \leq \tilde{v}_s). \tag{12}$$

B. Socially Optimal Fees and Surplus Divisions

This section considers the socially optimal fees in this model, which serve as benchmarks for future comparison.

Proposition 1 (Socially Optimal Fees and Surplus Divisions): *With tick size requirement (1), no matter who arrives first, the socially optimal make-take fees are given by*

$$\begin{cases} f_m^{so} = \frac{1}{2} \cdot d \\ f_t^{so} = -\frac{1}{2} \cdot d \end{cases}, \qquad Or \qquad \begin{cases} f_m^{so} = -\frac{1}{2} \cdot d \\ f_t^{so} = \frac{1}{2} \cdot d \end{cases}. \tag{13}$$

The surplus divisions are

$$BS^{so} = v_b - p_0 - \frac{1}{2} \cdot d, SS^{so} = p_0 + \frac{1}{2} \cdot d - v_s, \pi^{so} = 0. \tag{14}$$

To see this, consider the case in which the buyer arrives first. From parts (2a) and (3a) of Lemma 2, we know that under the fee structure $f_m^{so} = \frac{1}{2} \cdot d$ and $f_t^{so} = -\frac{1}{2} \cdot d$, the buyer will propose trading at price p_o. The effective buy price after the fee, however, is $p_o + \frac{d}{2}$. So all buyers with $v_b \geq p_o + \frac{d}{2}$ would like to buy. The effective sell price is also $p_o + \frac{d}{2}$, and all sellers with $v_s \leq p_o + \frac{d}{2}$ would like to sell. Efficiency is achieved. When the seller arrives first, he will propose a sell price of $p_o + d$, but the effective sell price is $p_o + \frac{d}{2}$. The effective price for the buyer is also $p_o + \frac{d}{2}$ after the rebate of $-\frac{1}{2} \cdot d$. Therefore, the effective buy and sell prices are both $p_o + \frac{d}{2}$ with $f_m^{so} = \frac{1}{2} \cdot d$ and $f_t^{so} = -\frac{1}{2} \cdot d$, no matter whether the buyer or the seller arrives first. Similarly, we can see that a fee structure of $f_m^{so} = -\frac{1}{2} \cdot d$ and $f_t^{so} = \frac{1}{2} \cdot d$ can also realize efficiency. When the buyer arrives first, he will propose a buy price of $p_o + d$. The effective buy and sell prices after the fee, however, are also $p_o + \frac{d}{2}$. When the seller arrives first, he will propose a sell price of p_o, which again leads to effective buy and sell prices of $p_o + \frac{d}{2}$. Therefore, both the buyer and the seller are willing to trade, again leading to market efficiency.

Proposition 1 has two interesting properties. First, combined with Lemma 1, Proposition 1 shows that the tick size regulation leads to two-sided markets. Rochet and Tirole (2006) define two-sided markets as platforms in which both the total fees charged and the fee structure breakdown matter. The fee structures in Lemma 1 and Proposition 1 both indicate zero total fees, but the zero total fees are broken down differently between the maker and the taker in these two settings, leading to completely different outcomes: Lemma 1 implies the no-trade equilibrium while Proposition 1 implies the socially optimal equilibrium. Second, the fees imposed by the exchange Pareto improve the market: even the side being charged benefits from the fees exchange. Therefore, regulators should exercise caution when evaluating the policy proposal to eradicate the make-take fees, because such an aggressive policy might reduce total gains from trading.

Another intuitive way of understanding Proposition 1 is to consider the bid-ask spread. Under socially optimal fees, the effective buy and sell prices are always $p_0 + \frac{1}{2} \cdot d$, no matter whether the buyer arrives first or the seller arrives first, and no matter whether the fees are set in the maker/taker market or the taker/maker market. The effective bid-ask spread is always zero under socially optimal fees. A necessary condition for zero spread is a zero total fee, but a zero total fee does not always lead to zero spread, since the breakdown of the total fee between liquidity makers and takers also matters. For example, if the exchange charges zero fees to both liquidity makers and liquidity takers, the spread becomes d, as illustrated by Lemma 1. The true bid-ask spread in this model is determined by three factors: the tick size, the total fee, and the breakdown of the total fee.

Of course, the "socially optimal" fee should not be understood literally. If the valuation of the buyer and the seller is not separated by the midpoint, we may not have the same socially optimal fee. Nevertheless, the socially optimal fee sets a benchmark for comparison for this model. Also, trading volume is maximized under the socially optimal fee, which also has practical implications. Stock exchanges can obtain additional benefits from volume maximization. For example, data revenue from the consolidated tape is allocated based on the market shares taken by competing exchanges (Caglio and Mayhew (2012)). The pursuit of data revenue implies that an exchange's object function is a weighted average of maximized direct revenues from the make-take fee and maximized volume. Our paper focuses on competition on direct revenues from the make-take fees, but further exploration of volume maximization would also be interesting.

C. Profit Maximizing Fees for Monopoly Exchange

Given the optimal buyer and seller behaviors summarized in Lemma 2, we can now determine a monopoly exchange's optimal make-take fees (f_m, f_t). The structure of the game is illustrated in Figure 3.

Insert Figure 3 Here

Suppose that the buyer arrives at Date 2. According to part (4a) of Lemma 1, the monopoly exchange's profit is given by

$$\pi = (f_m + f_t) \cdot \Pr(v_b \geq \max\{P + f_m, p_0 + d/2\})$$
$$\cdot \Pr(v_s \leq \min\{P - f_t, p_0 + d/2\})$$
$$= c^2 \cdot (f_m + f_t) \cdot \left(p_0 + d - \max\left\{P + f_m, p_0 + \frac{d}{2}\right\}\right) \cdot \left(\min\left\{P - f_t, p_0 + \frac{d}{2}\right\} - p_0\right), \quad (15)$$

where c is a constant equal to $\frac{2}{d}$ or the value of the probability density function of the uniform distribution. Since π increases with f_m (or f_t) whenever $\max\left\{P+f_m, p_0+\frac{d}{2}\right\}=p_0+d/2$ (or $\min\left\{P-f_t, p_0+\frac{d}{2}\right\}=p_0+d/2$), it is easy to see that the monopoly exchange will always set (f_m, f_t) such that $\max\left\{P+f_m, p_0+\frac{d}{2}\right\}=P+f_m$ and $\min\left\{P-f_t, p_0+\frac{d}{2}\right\}=P-f_t$. So (15) becomes

$$\pi = c^2\cdot(f_m+f_t)\cdot(p_0+d-P-f_m)\cdot(P-f_t-p_0)$$

$$=\begin{cases} c^2\cdot(f_m+f_t)\cdot(d-f_m)\cdot(-f_t) & if\ f_m>0 \\ c^2\cdot(f_m+f_t)\cdot(-f_m)\cdot(d-f_t) & if\ f_m<0 \end{cases} \tag{16}$$

Note that the $f_m>0$ and $f_m<0$ cases are isomorphic. Suppose the seller arrives first; by parallel argument, we can find exactly the same profit function. Therefore, the optimal fee structure and the surplus divisions do not depend on whether the buyer or the seller arrives first. So the equilibrium outcomes of a monopoly stock exchange can be solved from (16), and are summarized in the proposition below.

Proposition 2 (Optimal Monopoly Fees and Equilibrium Surplus Divisions): *With tick size requirement (1), no matter who arrives first, the monopoly stock exchange's optimal make-take fees are given by*

$$\begin{cases} f_m^*=\frac{2}{3}\cdot d \\ f_t^*=-\frac{1}{3}\cdot d \end{cases}, \quad or \quad \begin{cases} f_m^*=-\frac{1}{3}\cdot d \\ f_t^*=\frac{2}{3}\cdot d \end{cases}. \tag{17}$$

The buyer surplus, seller surplus, and profit for the stock exchange are

$$BS^*=v_b-p_0-\frac{2}{3}\cdot d, SS^*=p_0+\frac{1}{3}\cdot d-v_s, \pi^*=\frac{4}{27}\cdot d. \tag{18}$$

The core of the game under the monopoly exchange has the following outcome. The two monopoly fee structures impose an effective buy price of $p_o+\frac{2}{3}d$ and an effective sell price of

$p_o + \frac{1}{3}d$, no matter whether the buyer arrives first or the seller arrives first. The effective bid-ask spread is $\frac{1}{3}d$ under the monopoly fee. The buyer pays $p_o + \frac{2}{3}d$ when his valuation is at $\left[p_o + \frac{2}{3}d,\ p_o + d\right]$, and he does not participate in trades once his valuation falls below $p_o + \frac{2}{3}d$. The seller obtains $p_o + \frac{1}{3}d$ when his valuation is at $\left[p_o, p_o + \frac{1}{3}d\ \right]$ and he does not trade when his valuation rises above $p_o + \frac{1}{3}d$. The exchange obtains $\frac{1}{3}d$ once a trade happens, and 0 otherwise. This fee structure excludes buyers with low valuations (i.e., $[p_o + \frac{1}{2}d,\ p_o + \frac{2}{3}d]$) or sellers with high valuations (i.e., $[p_o + \frac{1}{3}d, p_o + \frac{1}{2}d]$), each of which comprises one-third of the buyer or seller population. By excluding traders with low incentive to trade, the exchange enjoys monopoly profits by attracting only high-valuation buyers or low-valuation sellers. The effective buy price under the monopoly fee is $\frac{1}{6}d$ higher than the effective buy price under the socially optimal fee, and the effective sell price is $\frac{1}{6}d$ lower than the effective sell price under the socially optimal fee. These differences are reflected in the $\frac{1}{3}d$ real bid-ask spread set by the exchange, as compared with a zero bid-ask spread under the socially optimal fee. However, the real spread is lower here than in the case with zero make and take fees, which is exactly one tick. Nevertheless, the monopoly fees achieve Pareto improvement over the outcome with zero fee charged to each side because the maker, the taker, and the exchange all benefit.

III. Competing Exchanges

In this section, we consider two stock exchanges, 1 and 2, with the tick size requirement in place. The structure of the game is illustrated in Figure 4.

Insert Figure 4 Here

To the best of our knowledge, no existing studies examine the optimal fee structures of competing exchanges with the tick size requirement. Foucault, Kadan, and Kandel (2013) consider the impact of the tick size when the total fee is exogenously given and the exchange is monopolistic. Colliard and Foucault (2012) consider competition between two competing exchanges without the tick size requirement and show that two competing exchanges will end up with a Bertrand outcome: the competition leads the two to undercut each other to zero total fees and zero profit. Nevertheless, we find that the presence of the tick size requirement dramatically changes the nature of the equilibrium. Indeed, we find that there is no pure-strategy Nash equilibrium, but we do find symmetric mixed-strategy equilibrium in which *both competing exchanges set non-zero make-take fees and earn strictly positive profit*.

A. Liquidity Maker's Fee Structure Preference

This section examines a liquidity maker's choice between two exchanges given their fee structures, and the next section endogenizes the fees. In this game, liquidity takers seem to play a passive role: takers can trade only in markets selected by liquidity makers, because an unchosen exchange has an empty book. Therefore, it seems that the priority of the exchanges is to attract liquidity makers, and a natural way to do so is to subsidize the makers. This intuition, however, is incorrect because a liquidity maker profit only when a liquidity taker accepts his limit order. Therefore, a liquidity maker needs to take into account the liquidity taker's decision when selecting the exchange. Our model captures two mechanisms through which the maker may prefer a charge instead of a subsidy. First, the liquidity maker has to post a more aggressive limit order in a market that subsidizes him. Consider the case in which the buyer arrives first. In a market that subsidizes liquidity makers, the buyer can post a limit order only at price $p_o + d$, because no seller would trade with a limit buy order at p_o when he has to pay a positive take fee. In a market that charges liquidity makers, the buyer will post a limit buy order at a price p_o, because some seller can accept the limit order of p_o after the subsidy. Therefore, the charge leads to a more favorable nominal buy price for the maker. Second, charging a higher fee to the liquidity maker allows the exchange to provide a more aggressive subsidy to the liquidity taker, which increases the probability of execution for the liquidity maker. Therefore, we show in Proposition 3 that subsidizing the maker can actually reduce his expected profit. Therefore, the maker may prefer a market that charges him over a market that subsidizes him.

Table 2 demonstrates the buyer's and seller's segmentation given the fee set by two competing exchanges when the buyer arrives first. The buyer may always go to one of the exchanges independently of his valuation, or he may choose one exchange when he has a low valuation and choose the other when he has a high valuation. One surprising result in the segmentation is that the liquidity maker may choose a market that charges him instead of one that subsidizes him.

Insert Table 2 Here

Proposition 3 (Liquidity Provider's Fee Structure Preferences): *Suppose exchange 1 adopts fee structure (f_m, f_t), and exchange 2 adopts fee structure (f_t, f_m), where*

$$d > f_m > -f_t > 0.$$

With tick size requirement (1), no matter who the liquidity provider is, he must prefer exchange 1 (or 2) when $|f_m| + |f_t| < d$ (or $|f_m| + |f_t| > d$). The liquidity provider is indifferent when $|f_m| + |f_t| = d$.

Proof: *See the appendix.*

Proposition 3 argues that, when the tick size is large relative to the level of the make-take fees, the liquidity maker prefers the market that charges him and subsidizes the liquidity taker.

Fixing the level of make-take fees, as the tick size decreases the liquidity maker gradually shifts his preference to the market that subsidizes him and charges liquidity takers. This counterintuitive result, however, dovetails nicely with stylized facts as well as with the empirical results to be established in Section IV.

Proposition 3 explains the existence of the taker/maker market. The taker/maker market enjoys a comparative advantage when trading stocks with large relative tick sizes. The emergence of the charging of makers by markets is a puzzle, particularly when several other regulations or practices put taker/maker markets at a disadvantage. One such policy is the trade-through rule. In the United States, orders are routed to the market with the best nominal price. This regulation favors markets that subsidize makers. To see this, start with the model of Colliard and Foucault (2012). Their model predicts that the taker/maker market and the maker/taker market can co-exist when they have the same total fees. The taker/maker market has a wider nominal quoted spread and the maker/taker market has a narrower nominal quote spread, although the spread after the fee is the same. The trade-through rule, however, is imposed on the nominal price, which implies that the taker/maker market cannot win in competition with the maker/taker market because the latter has a better nominal price, ceteris paribus, where orders are routed. To the best of our knowledge, there is no theoretical explanation of the comparative advantage of a market that charges liquidity makers when it competes with a market that subsidizes liquidity makers, and our paper fills this gap.

Empirically, Proposition 3 predicts that the taker/maker market should be more active for securities with relative large tick sizes, a prediction supported by empirical evidence in Section IV. It also provides an explanation of the fact that many exchanges operate two trading platforms with opposite fee structures: they use the taker/maker market to attract volume for low-priced securities and use the maker/taker market to attract volume for high-priced securities.

B. Non-existence of a Pure-Strategy Equilibrium

The main result reported in this sub-section is the non-existence of pure-strategy equilibrium in the game. This result contrasts starkly with the case involving zero tick size in Colliard and Foucault (2012), which predicts pure-strategy equilibrium with zero fees and profit.

To simplify the proof, we start directly from the effective buy and sell price on both exchanges when the buyer arrives first. Given the effective prices from both exchanges, $\left(p_b^i, p_s^i\right)(i = 1{,}2)$, if the liquidity provider is a buyer, his surplus to choose exchange i is

$$BS^i = \left(v_b - p_b^i\right) \cdot q_s^i = c \cdot (v_b - p_b^i) \cdot (p_s^i - p_0). \tag{19}$$

Without loss of generality, consider $p_s^1 \geq p_s^2$. When $p_s^1 > p_s^2$, let $\varphi^1 \equiv p_b^1 + (p_b^1 - p_b^2) \cdot \frac{p_s^2 - p_0}{p_s^1 - p_s^2}$. The buyer segmentation is summarized in Table 2.[14]

Insert Table 2 Here

Proposition 4 (No Pure-strategy Equilibrium): *There is no pure-strategy equilibrium when two exchanges compete.*

Proof: *See the appendix.*

The intuition underlying the proof is as follows. The first part of the proof demonstrates the non-existence of pure strategy equilibrium with exchanges earning positive profit. The proof follows the Bertrand intuition in Colliard and Foucault (2012). Suppose that exchanges 1 and 2

[14] The derivation of the segmentation can be found in the appendix.

have differing expected profit. The exchange with the lower expected profit can increase its expected profit by decreasing the effective buy price by ϵ and increasing its effective sell price by ϵ relative to those of its competitor. Suppose that they have the same positive profit. One of the exchanges can obtain the entire market by the same deviation. Therefore, there is no pure strategy equilibrium with both exchanges earning positive profit.

A new insight added by the two-sidedness of the market is the non-existence of pure strategy equilibrium with zero total fee. The proof of Proposition 4 demonstrates that when both exchanges charge a zero total fee, one can find a profit deviation by *increasing* the total fee. The formal proof of this result can be found in the appendix. Here we offer an intuitive example to show the mechanism of the deviation. Let us assume that both exchanges start with the socially optimal fee, which implies that the effective buy and sell prices are both $p_o + \frac{d}{2}$. Then consider a deviating strategy which adjusts the make fee to $0.4d$ and the take fee to $-0.3d$. With a total fee of $0.1d$, this fee structure leads to positive profit as long as some buyers and sellers prefer this new fee structure. Suppose that the buyer arrives first. The new fee structure implies an effective buy price of $p_o + 0.4d$ and an effective sell price of $p_o + 0.3d$. Figure 5 demonstrates that buyers with low valuation prefer the new fee structure to the socially optimal fee structure. This result rests on a trade-off between the profit conditional on execution and the execution probability. Under the socially optimal fee structure, the buyer's profit conditional on execution is $v_b - (p_o + \frac{d}{2})$ and the execution probability is 1. Therefore, the expected profit for the buyer is $\left[v_b - \left(p_o + \frac{d}{2}\right)\right] \cdot 1$. Under the new fee structure, the buyer's profit conditional on execution is $v_b - (p_o + 0.4d)$, which is higher than the buyer's profit conditional on execution under the socially optimal fee. However, the execution probability decreases to 0.6, because only sellers with valuations lower than $p_o + 0.3d$ accept the offer. Therefore, the buyer's expected profit is $[v_b - (p_o + 0.4d)] \cdot 0.6$.

It is easy to show that $[v_b - (p_o + 0.4d)] \cdot 0.6 > \left[v_b - \left(p_o + \frac{d}{2}\right)\right]$ when $v_b \in \left[p_o + \frac{d}{2}, p_o + 0.65d\right)$. Therefore, buyers with low valuations prefer an exchange with a lower effective buy price even though the execution probability is lower in that exchange as well.

Insert Figure 5 Here

The above example shows that the socially optimal fee cannot be sustained in a pure strategy equilibrium. The proof of Proposition 4 shows that, generally, no zero total fee structures can lead to pure strategy equilibrium because of two types of deviations that *increase* the total fee. Suppose again that the buyer arrives first. One deviating strategy of the exchange reduces the effective buy price by ϵ but reduces the effective sell price by $\mu\epsilon$ with $\mu > 1$. This strategy increases the buyer's profit conditional on execution, but decreases the execution probability because the seller is less likely to accept the offer. The reduction in effective buy price caters to buyers with low valuations. The other deviating strategy increases the effective buy price by ϵ but increases the effective sell price by $\mu\epsilon$ with $0 < \mu < 1$. This strategy reduces the buyer's profit conditional on execution, but increases execution probability because the seller is more likely to accept the offer. Surprisingly, buyers with high valuations may prefer this solution. Even if they face a worse price conditional on execution, the increase in execution probability can compensate for their lower profit conditional on execution. The proof of Proposition 4 shows that at least one of these two deviations is profitable, which implies that a zero total fee cannot be sustained in a pure strategy equilibrium.

An interesting insight from this proof is that the total fee tends to move towards zero when it is too high, but if it moves too close to zero a profitable deviation will induce an increase in the total fee. This mechanism provides justifications for the total fee fluctuations displayed in Figure 1.

C. Mixed-Strategy Equilibrium and the Exchanges' Profit

The no pure-strategy equilibrium result in Proposition 4 is surprising, because the standard Bertrand argument for a one-sided market usually leads to a unique pure-strategy equilibrium with marginal cost pricing and zero profit for both firms. However, we find that, due to the two-sidedness of the market and the heterogeneity of buyer/seller valuations, one exchange can always find a profitable deviation strategy given the pure strategy chosen by the other exchange. From now on, we investigate the random nature of competing fee structures. We focus on characterizing a symmetric mixed-strategy equilibrium, in which both exchanges follow the same cumulative distribution function $F(p_b, p_s)$ when deciding their effective buy and sell prices (p_b, p_s).

Proposition 5 (Mixed-strategy equilibrium): *A symmetric mixed-strategy equilibrium exists, such that*

(i) (p_b, p_s) *has a convex support on* $\left[p_0 + \frac{d}{2}, p_0 + d\right] \times [p_0, p_0 + \frac{d}{2}]$*;*

(ii) *both exchanges earn strictly positive profit.*

(iii) *the exchanges' profit increase linearly with d.*

Proof: *see the appendix.*

Proposition 5 proves the existence and properties of the mixed-strategy equilibrium. The mixed strategy has a convex support, which implies that there is a connected range of fees in which no specific fee is either better than or inferior to any of its neighbors. This result demonstrates the non-existence of an ideal fee structure that all the exchanges should adopt, even in the sense of probability. This pattern dovetails nicely with the diverse fee structures across exchanges. At first glance, liquidity providers and takers should all prefer the market that offers them the highest rebate, and it is puzzling why some exchanges can survive with neither the highest rebate for makers nor the highest rebate for takers. Proposition 5 justifies fee structures that offer neither the highest rebate for the maker nor the highest rebate for the taker.

Point (ii) of Proposition 5 states that the profit obtained following the mixed strategy are strictly positive. This result differentiates the one-sided market in Colliard and Foucault (2012) from the two-sided market in our model. When the tick size is zero, the competition between two exchanges can drive the profit to zero (Colliard and Foucault (2012)), which implies that any positive cost involved in establishing a new trading platform would deter entry. In reality, however, we continue to witness entries of new trading platforms, such as the abovementioned entrance of BATS Y on October 22, 2010. This paper shows that the tick size regulation can be one factor that attracts the new entry. When the tick size is positive, the competition between exchanges does not lead to zero profit for the exchanges, implying that a positive tick size can cause market fragmentation. Regulators are concerned that the entry of new trading platforms generates greater market fragmentation (O'Hara and Ye (2011)), but the literature have only limited understanding of why the market becomes more and more fragmented.

The literature has documented multiple projected mechanisms that lead to the consolidation of trading platforms. The seminal analysis of Stigler (1961, 1964) on the economics of information and on securities markets argues that trading will tend to centralize in one location because of economies of scale in information production. Chowdry and Nanda (1991) suggest that order flow will gravitate to the market with the lowest execution costs. Madhavan (2000) raises a "network externality puzzle" for market fragmentation. Arnold, Hersch, Mulherin and Netter (1999) and Brown, Mulherin and Weidenmier (2008) question the viability of competition between stock exchanges based on stock exchange mergers before 1995, when the modern electronic limit-order book markets played a minimum role. However, that markets have become fragmented in recent years despite the abovementioned mechanisms suggests that consolidation should occurred. We show that tick size constraints can drive the co-existence of and competition between limit-order book markets.

Finally, Proposition 5 predicts that exchange profit increase with the tick size. An extension of this prediction is that a larger tick size encourages fee game entry and induces greater market fragmentation. These two predictions have cross-sectional implications for the relative tick size. Because normal tick sizes are all one cent for securities priced above one dollar, the relative tick size—the nominal tick size divided by the price of the security—is higher for lower-priced securities. Therefore, Proposition 5 predicts that exchange profit is higher for lower-priced stocks. In addition, lower-priced stocks should have greater fragmentation. These predictions are tested in Section IV.

IV. Empirical Results

Our model yields two testable predictions: 1) a taker/maker market takes a larger market share relative to a maker/taker market for stocks with large tick sizes; 2) a larger tick size leads to higher profit for the exchange. An extension of the second prediction implies that a large tick size encourages entry and thus market fragmentation. Although securities with prices above 1 dollar have a uniform 1-penny tick size, the relative tick size varies with the price of a security: low-priced securities have larger relative tick sizes and vice versa. Section IV is organized as follows: Section IV.A describes the data, Section IV.B provides preliminary results consistent with the two predictions using double sorting, and Section IV.C provides formal tests of the two hypotheses using diff-in-diff analysis involving twin ETFs and twin trading platforms.

The pilot group used in the twin ETFs test consists of the leveraged ETFs that have undergone reverse splits and the control group used in the test consists of the leveraged ETFs that track the same indexes but experience no reverse splits. Leveraged ETFs are often issued in pairs to track the same index, but each pair tracks the same index in opposite directions. For example, the ETF SPXL and SPXS both track the S&P 500, but SPXL amplifies S&P 500 returns by 300% while SPXS does so by -300%. These twin leveraged ETFs usually have identical nominal prices during the IPO, but the amplification effect causes their nominal prices to diverge after issuance. The issuers often conduct reverse splits to keep their nominal prices aligned with each other.[15] Reverse splits can thus be regarded as exogenous shocks to relative tick sizes after controlling for ETF returns.

Stock exchanges compete along multiple dimensions. To provide a clean identification of the impact of the fee structure on exchange competitions, we use data from twin trading platforms operated by Direct Edge, a stock exchange that executes 12% of U.S. equity trading volume. These twin trading platforms, EDGA and EDGX, have identical infrastructure and trading rules, differing only along the fee-structure dimension. In our sample period EDGX, like most exchanges, has a maker/taker fee structure whereby (as noted above) liquidity demanders pay a fee of 0.30 cents per share while liquidity providers receive 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 receive a rebate of 0.015 cents per share.

A. Data and Sample

[15] Reverse splits occur more frequently than splits, because their issuers are often concerned about the higher trading cost of low-priced ETFs. "Why has ProShares decided to reverse split the shares of these funds?" (http://www.proshares.com/resources/reverse_split_faqs.html)

Two securities samples are used in the empirical analyses. Our preliminary analysis uses 120 stocks selected by Hendershott and Riordan. The original sample selected in early 2010 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. Of these 120 stocks, 117 still exist in our sample period of October 2010.The summary statistics on these stocks are presented in Panel A of Table 3. We also construct an ETF sample for our diff-in-diff analysis. To construct the sample, we search Bloomberg to collect information on leveraged ETF pairs that track the same index with opposite return multipliers, and the data are then merged with CRSP data to identify their reverse-splitting events. We identify 35 reverse splits from January 2010 through November 2011. The summary statistics for the ETF sample are presented in Panel B of Table 3.

Insert Table 3 Here

The two independent variables of interest, market fragmentation and market share of the taker/maker market relative to that of the maker/taker market, are constructed using TAQ data. The consolidated trade files of daily TAQ data provide information on executions across separate exchanges for trades greater than or equal to 100 shares (O'Hara, Yao, and Ye (2013)). The Herfindahl index is used as a proxy for market fragmentation, which is defined as

$$Herfindahl\ index_{i,t} = \sum_{j=1}^{14} \left(\frac{Exchangevolume_{j,i,t}}{Totalvolume_{i,t}} \right)^2$$

for security i on day t , where $Totalvolume$ is the total volume of all 14 stock exchanges in our sample (NASDAQ, AMEX, BATS X, BATS Y, BOSX, CINN, EDGA, EDGX, ISE, MWSE, NYSE, ARCX, CBOE, and PHLX). The market share of Direct Edge A is defined as the market share of EDGA relative to that of EDGA and EDGX ($EDGAratio_{it}$).

B. Preliminary Results Based on Double-sorting

We first sort the 117 stocks into three portfolios based on average market cap in September 2010 and, for each of the market cap portfolios, we sort stocks into three portfolios based on their average relative tick sizes in September 2010. The result is preliminary because it has not directly controlled for endogeneity that might originate from simultaneity bias or omitted variables (Roberts and Whited (2012)). Nevertheless, a recent paper by Benartzi, Michaely, Thaler, and Weld (2009) indicates that the average nominal price for a share of stock is nearly exogenous in a way that cannot be explained by any of several popular hypotheses pertaining to nominal prices, such as the marketability hypothesis, the pay-to-play hypothesis, and the signaling hypothesis. The main cross-sectional result established in that paper is that large firms have higher prices.[16] Benartzi, Michaely, Thaler, and Weld (2009) thus provide some empirical grounding for our double-sorting methodology. We defer the presentation of our clean means of identification to section IV.C.

Table 4 presents the cross-sectional variations in the market share of the taker/maker market and market fragmentation in October 2010. For each stock on each day, we calculate their $EDGAratio_{i,t}$ and $Herfindahl\ index_{i,t}$, and take the averages across all stocks in that portfolio. In this way, we have 21 daily observations for each 3-by-3 portfolio. Table 4 presents the averages of these daily observations and statistical inferences based on 21 daily observations.

Panel A reveals that the taker/maker market has a surprisingly large market share for stocks with large relative tick sizes and large or medium market cap. EDGA accounts for 59.12% of the volume for stocks with large relative tick sizes and large market cap, leaving EDGX to account for only 40.88% of the volume. The difference between the EGDA market share for stocks with

[16] Benartzi, Michaely, Thaler, and Weld (2009) also find that a firm splits when its price deviates from those of other firms in the same industry. However, there are no splits in our 117 NASDAQ HFT sample.

large relative tick sizes and that for stocks with small relative tick sizes is 21.08%, with a t-statistic of 30.56 based on 21 observations. The same pattern holds for stocks in the medium and small market cap categories. The results suggest that the taker/maker market takes a relatively higher market share for stocks with large relative tick sizes, whereas the maker/taker market takes a relatively higher market share for stocks with small relative tick sizes. The finding is consistent with the prediction of Proposition 3 that liquidity providers are willing to pay a fee to make the market for stocks with large tick sizes.

Panel B shows that trading for stocks with small relative tick sizes is more consolidated, especially for stocks with large and medium market cap. The Herfindahl index for stocks with large relative tick size and large market cap is 0.224, whereas the Herfindahl index for stocks with small relative tick size and large market cap with is 0.283. The difference is -0.059 with a t-statistic of 29.33. For stocks with medium market cap, the difference in the Herfindahl index between stocks with large relative tick sizes and stocks with small relative tick sizes is -0.044 with a t-statistic of 10.46. [17]

Insert Table 4 Here

C. Reverse Splits as Exogenous Shocks to the Relative Tick Size

This section establishes the causal relationship between the relative tick size and market share of the taker/maker market and the causal relationship between the relative tick size and market fragmentation. Diff-in-diff tests involving leveraged ETF reverse splits are used as a clean means of identification. Leveraged ETFs that have experienced reverse splits are used as the pilot group, and leveraged ETFs that track the same index but have not undergone reverse splits are

[17] The results for small stocks are less clear, probably because small stocks are subject to less trading overall, and it is hard to detect the pattern for 14 exchanges.

used as the control group. The bear and bull ETFs for the same index are usually issued by the same company at similar IPO prices, but large cumulative movements of an index cause their nominal prices to diverge. The issuers of leveraged ETFs usually use reverse splits to align the nominal prices of bull and bear ETFs. Therefore, reverse splits can be regarded as exogenous after controlling for past returns.

The regression specification for the diff-in-diff test is:

$$y_{i,t,j} = u_{i,t} + \gamma_j + \rho \times D_{trt_{i,t,j}} + \theta \times return_{i,t,j} + \epsilon_{i,t,j} \qquad (20)$$

where $y_{i,t,j}$ is the *EDGAratio* or the *Herfindahl index* for ETF j in index i at time t. u_{it} captures the index-by-time fixed effects, which controls for the time trend that may affect each index. γ_j is the ETF fixed effect that absorbs the time-invariant differences between two leveraged ETFs that track the same index. After controlling for the index-by-time fixed effects and the ETF fixed effects, the major difference remaining between the bull and bear ETFs is in their return differences, which are taken into account in the regression specification as well. The key variable in this regression, the treatment dummy $D_{trt_{i,t,j}}$, equals 0 for the control group. For the treatment group, the treatment dummy equals 0 before reverse splits and 1 after reverse splits. Therefore, coefficient ρ captures the treatment effect.

Table 5 displays the regression result. Column 1 shows that the market share of EDGA (the taker/maker market) increases relative to the combined volume of EDGA and EDGX (the maker/taker market), consistent with the prediction of Proposition 3. Column 2 demonstrates an increase in the Herfindahl index after reverse splits, implying that trading becomes more consolidated after the reverse splits.

Insert Table 5 Here

V. Conclusion

We examine the competition for order flow between stock exchanges by proposing make-take fees. When the valuations of the liquidity maker and the liquidity taker fall within the range of the same tick, there is no trade due to the tick size requirement. Exchanges can, however, set sub-penny make-take fees to bypass the tick size regulation. The implementation of fees allows a trade to take place that otherwise would not occur under a binding tick size. For such trades to occur, the make fee and the take fee must carry opposite signs. Therefore, the breakdown of the make and take fee matters, and this apparent violation of the neutrality of taxes and subsidies comes from the discrete tick size. When the price is not continuous, the buyer and the seller cannot negotiate at the sub-penny price level, but the sub-penny fee can create a price increment of less than one tick and thereby facilitate trades.

A positive tick size also leads to competition between exchanges that deviate from Bertrand equilibrium. We demonstrate the non-existence of pure strategy equilibrium in the fee game with a positive tick size. Such a mixed strategy explains the diversity of and frequent changes in fee structures. The mixed-strategy equilibrium, however, generates positive profit, and the profit increases with the tick size. This result explains the entry of platforms with new fee structures. Our model predicts that an increase in the relative tick size leads to a larger market share of the taker/maker market relative to the maker/taker market and greater market fragmentation, both of which are supported by empirical evidence.

Our paper contributes to the literature on competition between exchanges by showing that tick size can drive market fragmentation. Also, we contribute to the literature on two-sided markets by showing that tick size constraints can drive the market's two-sidedness. Finally, this paper contributes to the literature on make-take fees and tick size, which has direct implications for the policy debate on the two topics.

Because the make-take fee is also the transaction cost for the liquidity maker and the liquidity taker, it attracts the attention of regulators. A regulation which caps the take fee at 30 cents per one hundred shares has been implemented, and more aggressive initiatives, such as banning the fee completely, are under discussion among regulators. The first argument for regulating the fee cites its complex structure, but we show in this paper that tick size regulation, which prevents end-users of stock exchanges from negotiating prices at less than one-tick increments, can contribute to the complexity of the fee structure. The second argument for regulating the fee is based on fairness. Because major exchanges always charge one side and subsidize the other side, regulators are concerned about wealth transfer between liquidity makers and takers. We show, however, that such a fee structure is a natural response to the tick size regulation: by charging one side and subsidizing the other, the exchanges create acceptable trading prices for those failed but would-be efficient trades under a positive tick size. In this regard, fees imposed by the exchanges Pareto improve the market relative to the market without fees. The final argument for regulating the fee cites the agency issue. Angel, Harris, and Spatt (2010, 2013) and Battalio, Corwin, and Jennings (2013) argue that brokerage firms have an incentive to route non-marketable limit orders from retail traders to the maker/taker market because broker/dealers usually are able to retain the rebates upon executing transactions. This conflict of interest is certainly a concern, but we need to exercise caution if we contemplate solving the issue through such a regulation. A modest solution to this problem would be to require brokerage firms to hand rebates back to their customers, but a more aggressive proposal to ban the entire fee structure has also been put forward. As we show in this paper, fees can facilitate trading when tick size constraints are binding. Therefore, banning the fee can also reduce social value.

Insofar as the make-take fee structure is a natural response to tick size constraints, a direct solution would be tick size deregulation. At a minimum, we believe the first step to take before considering additional regulations would be to evaluate the impact of current regulations. Interestingly, we have witnessed a reverse trend. Encouraged by the Jumpstart Our Business Startups Act (the JOBS Act), the SEC is proposing a pilot program to increase the tick size.[18] As noted above, the motivation for increasing the tick size is that it may increase market-making revenue and support sell-side equity research and, eventually, increase the number of IPOs (Weild, Kim and Newport (2012)). Our paper suggests, however, that a direct effect of increasing the tick size is the prevalence of the taker/maker market, which charges liquidity providers and subsidizes liquidity takers.

Our model can be extended in several directions. One extension would be to take the trade-through rule into account. This rule favors markets that subsidize liquidity providers and charges liquidity takers, because order-routing decisions are based on nominal prices. It would be interesting to see whether such a rule would generate new game dynamics. Other possible extensions involve adding more exchanges to the model, allowing multiple buyers and sellers, and allowing traders to choose their order types. Also, the competing exchanges in this model maximize profit directly from make-take fees. In reality, exchanges may benefit indirectly from greater volume. For example, a greater trading volume implies more revenue from the consolidated tape. It would be interesting to examine the new dynamics of the game when volume is also included in the objective function. Finally, the SEC is currently implementing a 5-cent tick size

[18] "SEC Provides Details of 5-Cent Tick Test," *Wall Street Journal*, June 25, 2014.

pilot program, which provides a clean identification for testing the model's theoretical predictions. Empirical analysis using data from the pilot program would be fruitful.

REFERENCES

Angel, James, Lawrence Harris, and Chester Spatt, 2010, Equity trading in the 21st century, Working paper, Marshall School of Business.

Angel, James, Lawrence Harris, and Chester Spatt, 2013, Equity trading in the 21st century: an update, Working paper, Marshall School of Business.

Anshuman, V. Ravi, and Avner Kalay, 1998, Market Making with Discrete Prices, *Review of Financial Studies* 11, 81-109.

Avramovic, Ana, 2012, Manufacturing volume: the stock split solution, credit suisse, available from edge, accessed Dec. 18, 2013.

Battalio, Robert H., Shane A. Corwin, and Robert H. Jennings, 2014, Can brokers have it all? on the relation between make take fees & limit order execution quality, Working paper, University of Notre Dame and Indiana University Bloomington.

Buti, Sabrina, Francesco Consonni, Barbara Rindi, Yuanji Wen, and Ingrid M. Werner, 2014, Tick size: theory and evidence, Working paper, University of Toronto.

Caglio, Cecilia, and Stewart Mayhew, 2012, Equity trading and the allocation of market data revenue, Working paper, Federal Reserve Board.

Cardella, Laura, Jia Hao, and Ivalina Kalcheva, 2013, Make and take fees in the U.S. equity market, Working paper, Texas Tech University.

Chordia, Tarun, and Avanidhar Subrahmanyam, 1995, Market Making, the Tick Size, and Payment-for-Order Flow: Theory and Evidence, *Journal of Business* 68, 543-575.

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

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

Evans, David, and Richard Schmalensee,, 2013, The antitrust analysis of multi-sided platform businesses, Working paper, National Bureau of Economic Research.

Foucault, Thierry, and Albert Menkveld, 2008, Competition for order flow and smart order routing systems, *Journal of Finance* 63, 119–158.

Glosten, Lawrence, 1994, Is the electronic open limit order book inevitable? *Journal of Finance* 49, 1127–1161.

Hendershott, Terrence, and Haim Mendelson, 2000, Crossing networks and dealer markets: competition and performance, *Journal of Finance* 55, 2071–2115.

Malinova, Katya, and Andreas Park, 2013, Subsidizing liquidity: the Impact of make/take fees on market quality, *Journal of Finance*, forthcoming.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality? *Journal of Finance Economics* 100, 459-474.

O'Hara, Maureen, G. Saar, and Z. Zhong, 2014, Relative tick size and the trading environment, Unpublished working paper, Cornell University.

Pagnotta, Emiliano S., and Thomas Philippon, 2013, Competing on Speed, Working Paper, New York University.

Parlour, Christine, 1998, Price dynamics in limit order markets, *Review of Financial Studies* 11, 789-816.

Parlour, Christine, and Duane Seppi, 2003, Liquidity-based competition for order flow, *Review of Financial Studies* 16, 301-343.

Parlour, Christine, and Duane Seppi, 2003, Liquidity-based competition for order flow, *Review of Financial Studies* 16, 301-343.

Rochet, Jean-Charles, and Jean Tirole, 2006, Two-Sided Markets: A Progress Report, *RAND Journal of Economics* 37, 645-667.

Rosu, Ioanid, 2009, A dynamic model of the limit order book, *Review of Financial Studies* 22, 4601–4641.

Rust, George, and John Hall, 2003, Simulated minimum distance estimation of a model of optimal commodity price speculation with endogenously sampled prices, Working paper, Yale University and University of Maryland.

Santos, Tano, and José A. Scheinkman, 2001, Competition among Exchanges, *Quarterly Journal of Economics* 116, 1027-1061.

Stigler, George J., 1961, The economics of information, *Journal of Political Economy* 69, 213-225.

Stigler, George J., 1964, A theory of oligopoly, *Journal of Political Economy* 72, 44-61.

Weild, David, Edward Kim, and Lisa Newport, 2012, The trouble with small tick sizes, Working paper.

Yao, Chen, and Mao Ye, 2014, Tick Size Constraints, Market Structure, and Liquidity, Working paper, University of Warwick and University of Illinois at Urbana-Champaign.

Appendix: Mathematical Proofs

Proof of Lemma 3

Suppose the buyer arrives at Date 2 and the seller arrives at Date 3.

By Lemma 2, since $f_m > 0, P = p_0$. It follows that

$$\begin{cases} p_b = p_0 + f_m \\ p_s = p_0 - f_t \end{cases}$$

For $\tilde{f}_m < 0, P = p_0 + d$. Then

$$\begin{cases} \tilde{p}_b = p_0 + d + \tilde{f}_m \\ \tilde{p}_s = p_0 + d - \tilde{f}_t \end{cases}$$

In order for $p_b = \tilde{p}_b$ and $p_s = \tilde{p}_s$, the unique solution is (12). $\blacksquare$

Proof of Proposition 3: Here we prove only the case in which the liquidity provider is the buyer. The case in which the liquidity provider is the seller can be proved by parallel reasoning.

Under exchange 1's fee structure (f_m, f_t), according to parts (2a) and (3a) of Lemma 2, the liquidity provider will propose a trading price $P = p_0$, and trade with the seller with $v_s \leq \min\{p_0 - f_t, p_0 + d/2\}$. So the buyer's surplus when joining exchange 1 is

$$BS^1 = (v_b - p_0 - f_m) \cdot \Pr(v_s \leq \min\{p_0 - f_t, p_0 + d/2\})$$

$$= \begin{cases} c \cdot (v_b - p_0 - f_m) \cdot (-f_t) & if\ -f_t < d/2 \\ v_b - p_0 - f_m & if\ -f_t \geq d/2 \end{cases}.$$

Under exchange 2's fee structure (f_t, f_m), according to parts (2a) and (3a) of Lemma 1, the liquidity provider will propose a trading price $P = p_0 + d$, and trade with the seller with $v_s \leq \min\{p_0 + d - f_m, p_0 + d/2\}$. So the buyer's surplus when joining exchange 2 is

$$BS^2 = (v_b - p_0 - d - f_t) \cdot \Pr(v_s \leq \min\{p_0 + d - f_m, p_0 + d/2\})$$

$$= \begin{cases} c \cdot (v_b - p_0 - d - f_t) \cdot (d - f_m) & if\ f_m > d/2 \\ v_b - p_0 - d - f_t & if\ f_m \leq d/2 \end{cases}.$$

We need to consider the following three possible cases.

Case (i): $f_m > -f_t \geq d/2$

$$BS^1 - BS^2 = v_b - p_0 - f_m - c \cdot (v_b - p_0 - d - f_t) \cdot (d - f_m)$$

$$= c \cdot \left[\left(f_m - \frac{d}{2}\right) \cdot v_b - \frac{d}{2} \cdot (p_0 + f_m) + (p_0 + d + f_t) \cdot (d - f_m)\right].$$

Note that $BS^1 - BS^2$ increases with v_b, because $f_m > d/2$. Hence,

$$BS^1 - BS^2 \leq [BS^1 - BS^2]|when\ v_b = p_0 + d$$

$$= \ c \cdot (d - t_1) \cdot (\frac{d}{2} + f_t) \leq 0.$$

The liquidity provider prefers exchange 2.

Case (ii): $f_m > \frac{d}{2} > -f_t$

$$BS^1 - BS^2 = c \cdot (v_b - p_0 - f_m) \cdot (-f_t) - c \cdot (v_b - p_0 - d - f_t) \cdot (d - f_m)$$

$$= c \cdot (t_1 - f_t - d) \cdot (v_b - p_0 - d).$$

So

$$BS^1 \gtreqless BS^2\ if\ and\ only\ if\ f_m - f_t \lesseqgtr d.$$

Case (iii): $d/2 \geq f_m > -f_t$

$$BS^1 - BS^2 = c \cdot (v_b - p_0 - f_m) \cdot (-f_t) - v_b - p_0 - d - f_t$$

$$= c \cdot \left[\left(-f_t - \frac{d}{2}\right) \cdot v_b - (-f_t) \cdot (p_0 + f_m) + (p_0 + d + f_t) \cdot \frac{d}{2}\right].$$

Note that $BS^1 - BS^2$ decreases with v_b, because $d/2 > -f_t$. Hence,

$$BS^1 - BS^2 \geq [BS^1 - BS^2]|when\ v_b = p_0 + d$$

$$= \ c \cdot (-f_t) \cdot (\frac{d}{2} - f_m) \geq 0.$$

The liquidity provider prefers exchange 1.

Combining the above three cases, the lemma follows. ■

Buyer Segmentation in the Competition Case

Consider the case in which the liquidity provider is a buyer.[19] From (18),

$$BS^1 - BS^2 = c \cdot (v_b - p_b^1) \cdot (\tilde{v}_s^1 - p_0) - c \cdot (v_b - p_b^2) \cdot (\tilde{v}_s^2 - p_0)$$

$$= c \cdot [(\tilde{v}_s^1 - \tilde{v}_s^2) \cdot (v_b - p_b^1) - (p_b^1 - p_b^2) \cdot (\tilde{v}_s^2 - p_0)] \quad \text{(A.1)}$$

When $\tilde{v}_s^1 = \tilde{v}_s^2$, $BS^1 - BS^2 \gtreqless 0 \; if \; and \; only \; if \; p_b^1 \lesseqgtr p_b^2$ follows from (A.1).

Without loss of generality, suppose $\tilde{v}_s^1 > \tilde{v}_s^2$. From (A.1), the comparison between BS^1 and BS^2 can be reduced to a comparison between v_b and φ^1. Note that if $p_b^1 \le p_b^2$, then $\varphi^1 \le p_b^1 \le p_b^2$; if $p_b^1 > p_b^2$, then $p_b^2 < p_b^1 < \varphi^1$.

When $p_b^1 \le p_b^2$, for any $v_b \ge p_b^2$, we have $v_b \ge \varphi^1$. So $BS^1 - BS^2 \ge 0$ for any $v_b \ge p_b^1$.

When $p_b^1 > p_b^2$, there are three possible cases to be considered.

(i) If $\varphi^1 \le p_0 + \frac{d}{2}$, then for any $v_b \ge p_0 + \frac{d}{2}$, we have $v_b \ge \varphi^1$. So $BS^1 - BS^2 \ge 0$ for any $v_b \ge p_0 + \frac{d}{2}$.

(ii) If $p_0 + d \le \varphi^1$, then for any $v_b \le p_0 + d$, we have $v_b \le \varphi^1$. So $BS^1 - BS^2 \le 0$ for any $v_b \le p_0 + d$.

(iii) If $p_0 + \frac{d}{2} < \varphi^1 < p_0 + d$, then for any $\tilde{v}_b^2 \le v_b \le \varphi^1$, $BS^1 - BS^2 \le 0$; for any $\varphi^1 < v_b \le p_0 + d$, $BS^1 - BS^2 > 0$.

All the above possible cases are summarized in Table 2.

Proof of Proposition 4: (By Contradiction) Without loss of generality, suppose the pure-strategy equilibrium exists, and in equilibrium $\pi^1 \ge \pi^2$. There are two possible cases: (i) $\pi^1 > 0$; (ii) $\pi^1 = \pi^2 = 0$.

(i) There are two subcases: (i-a) $\pi^1 > \pi^2 \ge 0$; (i-b) $\pi^1 = \pi^2 > 0$.

(i-a) Exchange 2 can set its fees such that

[19] Seller segmentation when the liquidity provider is a seller is similar.

$$p_b^2 = p_b^1 - \varepsilon \; and \; p_s^2 = p_s^1, \tag{A.2}$$

$where \; \varepsilon > 0$. Then, according to Table 2, no one goes to exchange 1 anymore, and exchange 2's profit becomes

$$\hat{\pi}^2 = (p_b^1 - p_s^1 - \varepsilon) \cdot \; \Pr(v_b \geq \tilde{v}_b^2) \cdot \Pr(v_s \leq \tilde{v}_s^2)$$

$$\geq (p_b^1 - p_s^1 - \varepsilon) \cdot \; \Pr(v_b \geq \tilde{v}_b^1) \cdot \Pr(v_s \leq \tilde{v}_s^1)$$

$$= \pi^1 - \varepsilon \cdot \Pr(v_b \geq \tilde{v}_b^1) \cdot \Pr(v_s \leq \tilde{v}_s^1),$$

where the inequality follows from (A.2). Clearly, as long as $\pi^1 > \pi^2$, exchange 2 can always strictly increase its profit by deviation (A.2) with a sufficiently small ε.

(i-b) B can deviate to (A.2), so that no one goes to exchange 1 anymore, and exchange 2's profit becomes

$$\hat{\pi}^2 = (p_b^1 - p_s^1 - \varepsilon) \cdot \; \Pr(v_b \geq \tilde{v}_b^2) \cdot \Pr(v_s \leq \tilde{v}_s^2)$$

$$\geq (p_b^1 - p_s^1 - \varepsilon) \cdot \; \Pr(v_b \geq \tilde{v}_b^1) \cdot \Pr(v_s \leq \tilde{v}_s^1)$$

$$> (p_b^1 - p_s^1 - \varepsilon) \cdot Some \; of \; \Pr(v_b \geq \tilde{v}_b^1) \cdot \Pr(v_s \leq \tilde{v}_s^1)$$

$$= \pi^1 - \varepsilon \cdot Some \; of \; \; \Pr(v_b \geq \tilde{v}_b^1) \cdot \Pr(v_s \leq \tilde{v}_s^1)$$

where the first inequality follows from (A.2), and the second inequality is derived from the fact that exchange 1 does not get all the buyers with $\Pr(v_b \geq \tilde{v}_b^1)$ in this case, because some buyers go to exchange 2. Thus, exchange 2 can always strictly increase its profit by the deviation (A.2) with a sufficiently small ε.

(ii) There are two subcases: (ii-a) No trade; (ii-b) Trade with $x^i = y^i \in (p_0, p_1), i = 1{,}2$.

(ii-a) No trade implies that $p_b^i \geq p_0 + d$ or $p_s^i \leq p_0$ for both $i = 1{,}2$. Then exchange 2 can set fees such that $p_0 < p_s^2 < p_b^2 < p_0 + d$. Then the buyer with $v_b \geq \tilde{v}_b^2$will trade with the seller with $v_s \leq \tilde{v}_s^2$, and $\hat{\pi}^2 > 0$.

(ii-b) Denote $p_b^1 = p_s^1 = a$. There are three subsubcases: (ii-b-I): $p_0 < a < p_0 + \frac{d}{2}$; (ii-b-II): $a = p_0 + \frac{d}{2}$; (ii-b-III): $p_0 + \frac{d}{2} < a < p_0 + d$.

(ii-b-I): Exchange 2 can set its fees such that

$$p_b^2 = a + \varepsilon \; and \; p_s^2 = a + \mu \cdot \varepsilon, \tag{A.3}$$

$where\ \varepsilon > 0\ and\ 0 < \mu < 1.$ For sufficiently small ε, we will have $\tilde{v}_s^2 = p_s^2 > a = \tilde{v}_s^1, p_b^2 > p_b^1, and$

$$\varphi^2 = p_b^2 + (p_b^2 - p_b^1) \cdot \frac{\tilde{v}_s^1 - p_0}{\tilde{v}_s^2 - \tilde{v}_s^1}.$$

$$= (a - p_0)\left(1 + \frac{1}{\mu}\right) + p_0 + \varepsilon.$$

Clearly, φ^2 decreases with , $\lim_{\mu \to 1} \varphi^2 = 2a - p_0 + \varepsilon < p_0 + d \; and \; \lim_{\mu \to 0} \varphi^1 = a = p_0 + \frac{d}{2}$, where the inequality follows from $a < p_0 + \frac{d}{2}$ and ε is sufficiently small. Hence, for sufficiently small ε, we can always have $p_0 + \frac{d}{2} < \varphi^1 < p_0 + d$. And from the buyer segmentation analysis, we know that the buyer with $v_b \geq \varphi^2$ will go to exchange 2. So the relatively high-valuation buyers will go to exchange 2 and exchange 2 will make strictly positive profit, because $p_b^2 - p_s^2 = (1 - \mu) \cdot \varepsilon$.

(ii-b-II): Exchange 2 can set its fees such that

$$p_b^2 = a - \mu \cdot \varepsilon \; and \; p_s^2 = a - \varepsilon, \tag{A.4}$$

$where\ \varepsilon > 0\ and\ 0 < \mu < 1.$ For sufficiently small ε, we will have $\tilde{v}_s^2 = p_s^2 < a = \tilde{v}_s^1, p_b^2 < p_b^1, and$

$$\varphi^1 = p_b^1 + (p_b^1 - p_b^2) \cdot \frac{\tilde{v}_s^2 - p_0}{\tilde{v}_s^1 - \tilde{v}_s^2}.$$

$$= (a - p_0)(1 + \mu) + p_0 - \mu \cdot \varepsilon.$$

Clearly, φ^1 increases with $\mu, and \lim_{\mu \to 1} \varphi^1 = p_0 + d - \varepsilon < p_0 + d$, where the inequality follows from $a = p_0 + \frac{d}{2}$ and $\varepsilon > 0$. Hence, for sufficiently small ε, we can always have $\varphi^1 < p_0 + d$. And from the buyer-segmentation analysis, we know that the buyer with $v_b \leq \varphi^1$ will go to exchange 2. So the relatively low valuation buyers will go to exchange 2 and exchange 2 will make strictly positive profit, because $p_b^2 - p_s^2 = (1 - \mu) \cdot \varepsilon$.

(ii-b-III): Exchange 2 can set its fees such that

$$p_b^2 = a - \mu \cdot \varepsilon \; and \; p_s^2 = a - \varepsilon, \tag{A.4}$$

$where \; \varepsilon > 0, 0 < \mu < 1 \; and \; a - \varepsilon > p_0 + \frac{d}{2}$. For sufficiently small ε, we will have $\tilde{v}_s^2 = p_0 + \frac{d}{2} = \tilde{v}_s^1, and \; p_b^2 < p_b^1$. Then the buyer with $v_b \geq \tilde{x}^2$ will trade with the seller with $v_s \leq \tilde{v}_s^2$, and $\hat{\pi}^2 > 0$, because $p_b^2 - p_s^2 = (1 - \mu) \cdot \varepsilon$. ■

Proof of Proposition 5: We establish the lemma in the following 4 steps.

Step 1: $p_0 \leq p_s \leq p_b \leq p_0 + d$.

Suppose $p_b > p_0 + d$ or $p_s < p_0$ occurs with some positive probability in equilibrium. Note that these cases result in zero profit for exchanges. One exchange can always deviate by shifting such a probability to a strategy with $p_0 \leq \tilde{v}_s < \tilde{v}_b \leq p_0 + d$, so that it will earn strictly positive profit with that probability.

Step 2: No mass point in the mixed-strategy equilibrium strategy.

There are two possible mass points to be considered: (a) some $(p_b, p_s) \; with \; p_b > p_s$; (b) some $(p_b, p_s) \; with \; p_b = p_s$. In case (a), a profitable deviation is given by (A.2). In case (b), a profitable deviation is given by (A.3), (A.4), and (A.5), respectively, for $p_0 < p_b = p_s < p_0 + \frac{d}{2}, p_b = p_s = p_0 + \frac{d}{2} \; and \; p_0 + \frac{d}{2} < p_b = p_s < p_0 + d$.

Step 3: (p_b, p_s) has a convex support on $\left[p_0 + \frac{d}{2}, p_0 + d\right] \times \left[p_0, p_0 + \frac{d}{2}\right]$.

First, given $p_b^i \geq p_0 + \frac{d}{2}$, any $p_b^j < p_0 + \frac{d}{2}$ is strictly dominated by $p_b^j = p_0 + \frac{d}{2}$ for exchange j, because a lower effective price than $p_0 + \frac{d}{2}$ cannot increase the number of buyers, it can only lower its per-unit profit. Similarly, we can rule out a $p_s > p_0 + \frac{d}{2}$ strategy.

Second, the support of the mixed strategy must be convex. Suppose there is an unconnected support $[\alpha, \beta]$ and $[\gamma, \delta]$. By symmetry, the other exchange would not randomize over the "hole" interval $[\beta, \gamma]$. However, in that case one exchange will not be indifferent between choosing β and γ, which is a necessary condition for it to randomize over these two intervals. Thus, the support must be convex.

Step 4: There exists symmetric mixed-strategy equilibrium, and both firms earn strictly positive profit.

Given our first 3 steps, the existence of symmetric mixed-strategy equilibrium can be established by applying Theorem 6 in Dasgupta and Maskin (1986). The support ranges for x and y in Step 3 imply that both firms earn strictly positive profit in equilibrium. ■

Figure 1. Total Fee Structure of the Eight Exchanges in the U.S.
This figure displays the total fee structure for eight stock exchanges in the U.S.

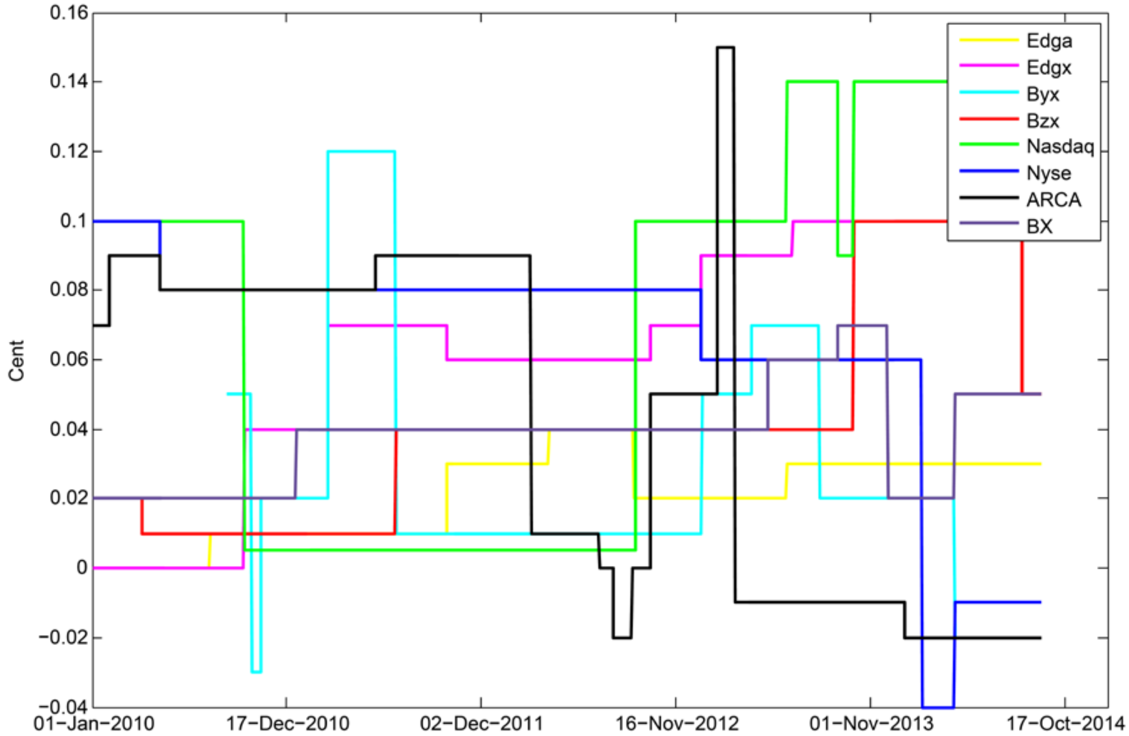

Figure 2. Effective Buy/Sell Prices under Fees (f_m, f_t)

This figure shows that under fees (f_m, f_t), when the liquidity provider is a buyer, the effective buy price is $P + f_m$, and a buyer with valuations higher than that will trade; the effective sell price is $P - f_t$, and a seller with valuations lower than that will trade.

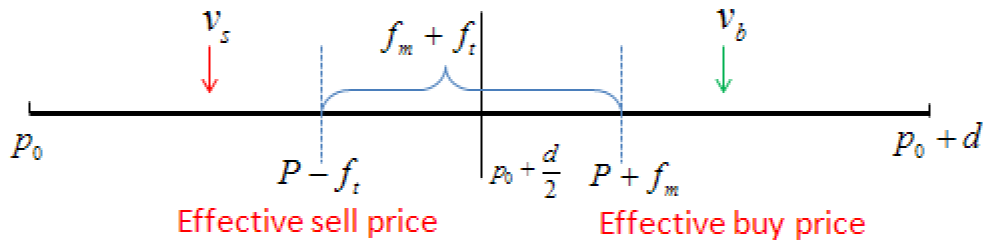

Figure 3. Game Tree with Monopoly Exchange

This figure depicts the timeline of the subgame after the monopoly exchange sets its fees. At Date 2, nature draws either a buyer to arrive at the market first with probability p or a seller to arrive at the market first with probability $1 - p$. Whoever arrives at the market first will propose a trading price after observing the maker/taker fees. At Date 3, the counterparty arrives. The counterparty observes the maker and taker fees as well as the price proposed by the maker, and then decides whether to trade.

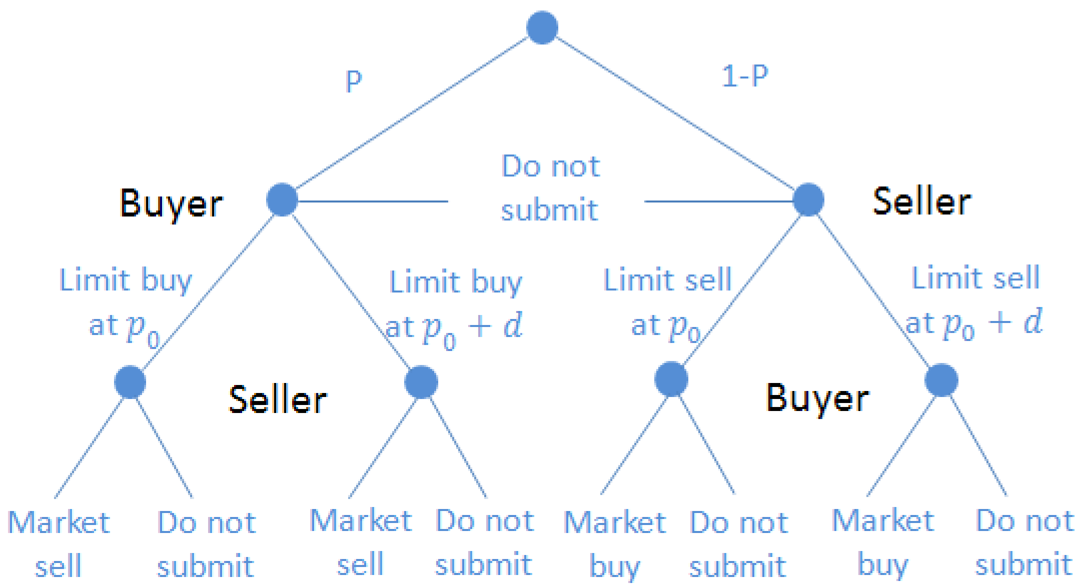

Figure 4. Game with Two Competing Exchanges
This figure depicts the timeline of the subgame after two exchanges set their fees. At Date 2, nature draws either a buyer to arrive at the market first with probability p or a seller to arrive at the market first with probability $1-p$. Whoever arrives at the market first will choose an exchange to join, and propose a trading price after observing the maker/taker fees. At Date 3, the counterparty arrives. The counterparty observes the maker and taker fees as well as the price proposed by the maker, and then decides whether to trade. If he decides to trade, he must join the platform that the liquidity provider chooses at Date 2 and trade at the proposed trading price.

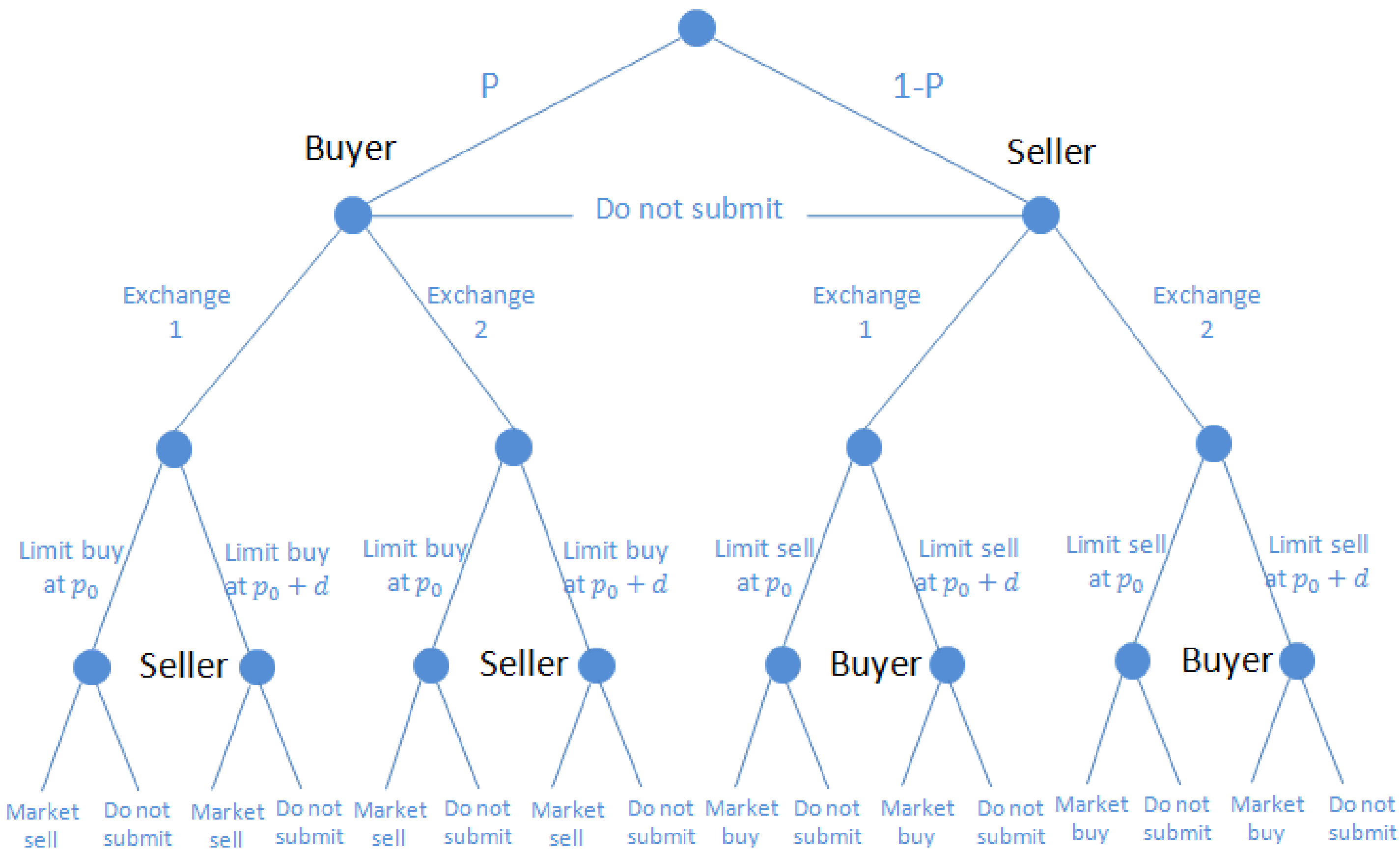

Figure 5: Buyer's Preference

This figure shows the profitable deviation of Exchange 2 when Exchange 1 charges the socially optimal fee. Exchange 2 can charge a make fee of 0.4d and a take fee of -0.3d. The figure shows that a buyer with $v_b \in \left[p_o + \frac{d}{2}, p_o + 0.65\right)$ would enjoy a higher expected profit by joining Exchange 2, leading to positive profit for Exchange 2.

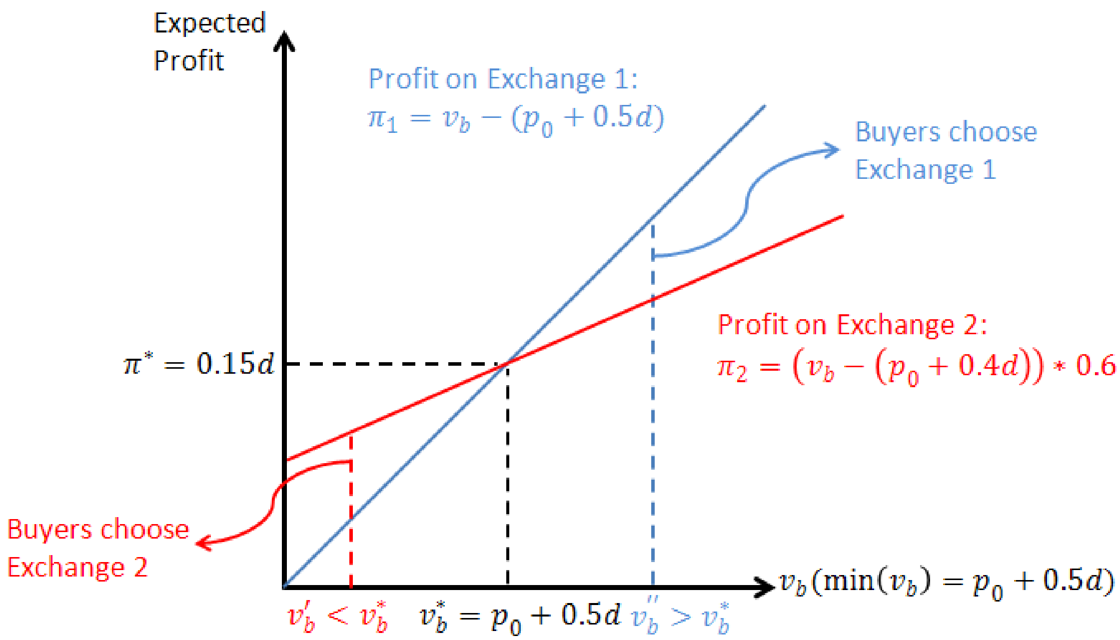

Table 1. Make-Take Fee Structure U.S. Exchanges

This table shows the make-take fee structure of seven exchanges in the U.S. on January 1, 2010. The trading fees for BATS Y (BYX) are not presented because BYX was created after January 1, 2010.

EX	Make Fee	Take Fee	Total Fee
EDGA	0.0002	-0.0002	0
EDGX	-0.0029	0.0029	0
NASDAQ	-0.002	0.003	0.001
NYSE	-0.0015	0.0025	0.001
BZX	-0.0001	0.0003	0.0002
ARCA	-0.0023	0.003	0.0007
BX	0.0003	-0.0001	0.0002

Table 2. Buyer Segmentation under Competition

No buyer goes to *B*: $\begin{cases} q_b^1 = \Pr(v_b \geq \tilde{v}_b^1) \\ q_b^2 = 0 \end{cases}$	$\begin{cases} p_s^1 = p_s^2 \\ p_b^1 < p_b^2 \end{cases}$ or $\begin{cases} \tilde{v}_s^1 > \tilde{v}_s^2 \\ p_b^1 \leq p_b^2 \end{cases}$ or $\begin{cases} \tilde{v}_s^1 > \tilde{v}_s^2 \\ p_b^1 > p_b^2 \\ \varphi^1 \leq p_0 + \frac{d}{2} \end{cases}$
No buyer goes to *A*: $\begin{cases} q_b^1 = 0 \\ q_b^2 = \Pr(v_b \geq \tilde{v}_b^2) \end{cases}$	$\begin{cases} p_s^1 = p_s^2 \\ p_b^1 > p_b^2 \end{cases}$ or $\begin{cases} \tilde{v}_s^1 > \tilde{v}_s^2 \\ p_b^1 > p_b^2 \\ p_0 + d \leq \varphi^1 \end{cases}$
A and *B* equally share the buyer: $\begin{cases} q_b^1 = \Pr\left(v_b \geq \tilde{v}_b^1\right)/2 \\ q_b^2 = \Pr\left(v_b \geq \tilde{v}_b^2\right)/2 \end{cases}$	$\begin{cases} p_s^1 = p_s^2 \\ p_b^1 = p_b^2 \end{cases}$
High-valuation buyer goes to *1*, Low-valuation buyer goes to *2*: $\begin{cases} q_b^1 = \Pr(\varphi^1 < v_b \leq p_0 + d) \\ q_b^2 = \Pr(\tilde{v}_b^2 \leq v_b \leq \varphi^1) \end{cases}$	$\begin{cases} \tilde{v}_s^1 > \tilde{v}_s^2 \\ p_b^1 > p_b^2 \\ p_0 + \dfrac{d}{2} < \varphi^1 < p_0 + d \end{cases}$

Table 3. Sample Summary Statistics

This table presents summary statistics on the sample data used in the paper. Panel A presents summary statistics on the 117-stock sample as of October 2010 using CRSP market data. Panel B presents descriptive statistics for 35 pairs of leveraged ETFs using CRSP market data. Leverage ETFs that have undergone reverse splits are classified in the treatment group. Leveraged ETFs that track the same index as the ones in the treatment group but experience no reserve split are classified in the control group. The reserve splitting events occur between January 2010 and November 2011.

Panel A: CRSP Summary Statistics of 117 Stocks as of October 2010

	Market Cap ($Million)	Avg Closing Price ($)	Avg Daily Volume (1000s)
Mean	19352	41.5	5095
Medium	2032	27.22	573
Std	42246	64.41	11176
Min	282	5.72	24
Max	275000	575.94	67028

Panel B: Summary Statistics of Split Sample & Matched Firms

	Mean		Median	
	Treatment	Control	Treatment	Control
Reverse Split Sample				
return	0	0	-0.002	0.002

Table 4. Cross-sectional Variation in the Market Share of the Taker/maker Market and Cross-sectional Variation of Market Fragmentation

This table presents the ratio of Direct Edge A (EDGA) volume to the total volume of Direct Edge and the Herfindahl Index, respectively. The sample includes 117 stocks in NASDAQ HFT data in October 2010. The 117 stocks are first sorted into three portfolios by their average market cap in September 2010 and then, for each of the market cap portfolios, stocks are sorted into three portfolios by their average relative tick sizes in September 2010. Panel A presents the average of the daily equal-weighted market share of EDGA. We compute, for each stock on each day, the ratio of the EDGA volume to the total volume of Direct Edge (EDGA volume plus the EDGX volume). The daily equal-weighted market share for each portfolio is the average of the market share for stocks in the portfolio. Panel B presents the average of the daily equal-weighted Herfindahl Index. We compute the Herfindahl Index for each stock on each day. The daily equal-weighted Herfindahl Index for each portfolio is the average of the Herfindahl Index for stocks in the portfolio. t-statistics are calculated based on the 21 daily observations. *, ** and *** represent statistical significance at the 10%, 5%, and 1% levels of large-minus-small differences, respectively.

Panel A: Equal-weighted EDGA Volume / (EDGA Volume + EDGX Volume)

	Large Relative Tick Size (Low Price)	Medium Relative Tick Size (Medium Price)	Small Relative Tick Size (High Price)	Large-Small Relative Tick Size (Low-High Price)	*t*-stat
Large Cap	59.12%	47.39%	38.04%	21.08%***	30.56
Middle Cap	57.43%	49.45%	41.54%	15.89%***	11.65
Small Cap	35.47%	27.43%	25.64%	9.83%***	6.98
L-S Cap	23.65%***	19.95%***	12.40%***		
t-statistics	19.4	20.18	11.81		

Panel B: Equal-weighted Herfindahl Index

	Large Relative Tick Size (Low Price)	Medium Relative Tick Size (Medium Price)	Small Relative Tick Size (High Price)	Large-Small Relative Tick Size (Low-High Price)	*t*-stat
Large Cap	0.224	0.247	0.283	-0.059***	29.33
Middle Cap	0.287	0.304	0.331	-0.044***	10.46
Small Cap	0.340	0.360	0.342	-0.002	0.33
L-S Cap	-0.116***	-0.112***	-0.059***		
t-statistics	-13.56	-14.48	-7.43		

Table 5. Impact of Change in Relative Tick Size on Taker/maker Market Share and Market Fragmentation

This table presents the impact of changes in the relative tick size on the market share of the taker/maker market and on market fragmentation, respectively. Reserve splits of leveraged ETFs are regarded as exogenous shocks to the relative tick size. Leveraged ETFs that have undergone reverse splits are classified in the treatment group, and leveraged ETFs that track the same index as the ones in the treatment group but experience no reserve split are classified in the control group. The reverse-splitting events are between January 2010 and November 2011. The event window includes 30 trading days immediately before and after the reverse-splitting date. The estimation is based on the following diff-in-diff regression:

$$y_{i,t,j} = u_{i,t} + \gamma_j + \rho \times D_{trt_{i,t,j}} + \theta \times return_{i,t,j} + \epsilon_{i,t,j}$$

where $y_{i,t,j}$ is the ratio of the EDGA volume to the total volume of Direct Edge (EDGA volume plus the EDGX volume) for ETF *j* in index *i* on day *t* for Column (1), and is the Herfindahl Index for ETF *j* in index *i* on day *t* for Column (2). $u_{i,t}$ represents the index-by-time fixed effects and γ_j represents the ETF fixed effects. $return_{i,t,j}$ is the return for ETF j in index i on day t. The treatment dummy $D_{trt_{i,t,j}}$ equals 0 for the control group, and equals 0 before reverse splits and 1 after reverse splits for the treatment group. Standard errors are in parentheses, and *, **, and *** represent statistical significance at the 10%, 5%, and 1% levels, respectively.

	(1) *EDGAratio*	(2) *Herfindahl Index*
D_{trt}	-0.023*	0.029***
	(-1.823)	(4.241)
Returns	0.086	0.012
	(0.742)	(0.192)
R^2	0.833	0.675
Index*time FE	Y	Y
ETF FE	Y	Y