Tick Size Constraints, Two-Sided Markets, and Competition between Stock Exchanges

Yong Chao     Chen Yao     Mao Ye*
March 14, 2015

Abstract

This paper argues that the one-cent tick size imposed by SEC rule 612 is a driving force in the fragmentation of U.S. stock markets. U.S. stock exchanges compete for order flow by setting make fees for liquidity makers and take fees for liquidity takers. When traders can quote a continuous price, the breakdown of the make-take fees is neutralized and order flow consolidates to the exchange with the lowest total fee. However, under a discrete tick size, the breakdown of the make-take fees cannot be neutralized, which results in two-sided markets. The tick size constraints create an incentive for a monopoly operator to establish multiple exchanges for second-degree price discrimination. It also causes competition between two identical exchanges that precludes a zero-fee, zero-profit Bertrand outcome but reach mixed-strategy equilibria, with both earning positive profits. This prediction justifies a diversity of fee structures and frequent fee adjustments as well as the entry of exchanges with new fee structures. We show that fees can Pareto improve social welfare under tick size constraints. Our model predicts that markets become more fragmented under a larger tick size. We find empirical evidence consistent with this prediction using splits/reverse splits of ETFs as exogenous shocks to the relative tick size, with paired ETFs that track the same index but do not split/reverse split as controls.

Key Words: Market fragmentation, Make-take fee, Tick Size, Two Sided Markets

* 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, Dan Bernhardt, Rohan Christie-David, Jean Colliard, Shane Corwin, Thierry Foucault, Amit Goyal, Andrei Hagiu, Artem Neklyudov, Richard Schmalensee, Chester Spatt, Glen Weyl, Julian Wright, Bart Yueshen Zhou and seminar participants at Baruch College, the University of Notre Dame, University of Illinois at Urbana Champaign, HEC Lausanne and EPFL for their suggestions. This research is supported by National Science Foundation grant 1352936. We also thank Qi Gu, Xin Wang, Bei Yang, Yingjie Yu, Fan Yang and Chao Zi for their excellent research assistance.
We find that the one-cent tick size imposed by SEC rule 612 leads to the fragmentation of U.S. stock markets by changing the nature of price competition from one-sided to two-sided. Figure 1 shows thirteen major exchanges run by five holding companies and their respective pricing structures. These competing exchanges offer nearly homogeneous trading services. First, they trade identical securities.\(^1\) Second, they are all organized as electronic limit-order books, such that trades happen through direct interaction between buyers and sellers. A trader can act as a liquidity maker by posting a (buy or sell) limit order with a specified price and quantity, or as a liquidity taker by accepting the terms of a previously posted limit order through a market order. Third, these exchanges share a similar pricing model by charging a make fee to the liquidity maker and a take fee to the liquidity taker for each executed share. The sum of the make and take fees, or the total fee, is a major source of revenue for stock exchanges.\(^2\)

When price is continuous, the tax neutrality principle predicts that exchanges should compete only on the total fee, not on its breakdown, since liquidity makers and takers can neutralize the fee structure. The one-dimensional competition for the total fee together with homogeneous services leads then to two predictions: (1) No price discrimination: operators have no incentive to open multiple exchanges, because all traders would choose the exchange with the lowest total fee. (2) Bertrand outcome: competition between operators ends in pure-strategy equilibrium with zero total fees and zero profits. These two predictions then imply consolidation if setting up an exchange involves fixed costs.

---

\(^1\) Regulation National Market System (Reg NMS) allows stocks to be traded outside the listing exchange.

\(^2\) For example, in its application for an IPO, the BATS stock exchange reports that about 70% percent of its revenues come from the total fee. BATS S1 registration statement (page F4). O’Donoghue estimate that 34.7% of NASDAQ’s net income is from the fees.
Nevertheless, price competition and the organization of the exchange industry defy the abovementioned intuitive predictions. In terms of pricing strategy, Figure 1 clearly indicates non-neutrality. For example, two exchanges run by the same operator, EDGA and EDGX, have identical total fees but unequal breakdowns. The non-consolidation of EDGA and EDGX strongly suggests that the breakdown of the total fee creates value for the operator. Moreover, Figure 2 exhibits empirically that the total fees do not converge to a stable value, let alone converging to zero as in the Bertrand outcome. In terms of the organization of the industry, Figure 1 reveals that fragmentation can occur along two dimensions: the exchange industry has multiple operators, and most operators offer more than one exchange, while Figure 2 displays new entries in the fee game. For example, on October 22, 2010, BATS created a new trading platform, BATS Y, in addition to its existing BATS X.

We begin by presenting a theoretical model demonstrating that such anomalies can be addressed by the discrete tick size, and the nature of the fee game reflects the ability of the exchange to propose a sub-penny transaction price when end users are constrained. SEC rule 612 (the Minimum Pricing Increment) of regulation NMS prohibits liquidity makers from displaying orders in increments smaller than 1 cent for any stock priced above $1.00 per share. A recent study by Credit Suisse demonstrates that this one-cent tick size is surprisingly binding: 50% of S&P stocks priced below $100 per share have one-cent quoted spreads (Avramovic, 2012). The tick size prevents end users of stock exchanges from negotiating sub-penny price increments. The fees set by exchanges are, however, not subject to the tick size constraints. Consequently, exchanges can use make-take fees to effectively propose sub-penny transaction prices. Because liquidity makers and takers cannot neutralize the sub-penny increment, the price competition between exchanges has now changed from a one-sided (total fee) competition to a two-sided (make fee and
take fee) competition. We then show that this change in the nature of the price competition has affected the structure of the exchange industry.

The following example conveys the intuition behind our theoretical model. Consider a game between exchange operators, a continuum of buyers with valuations uniformly distributed on [0.5, 1] and a continuum of sellers with valuations uniformly distributed on [0, 0.5]. At Date 0, the operators set their make-take fees, with the option of establishing two exchanges with differentiated make-take fees. Without loss of generality, suppose nature draws a buyer at Date 1. As a liquidity maker, the buyer can choose to submit a buy limit order to any exchange, or submit no limit order at all. At Date 2, the seller arrives as a liquidity taker, and decides whether to accept the limit order if the book is not empty. When the liquidity maker can quote a continuous price, the standard tax neutrality principle holds in our model. It follows that it is impossible for any exchange to implement any price discrimination, and competition would result in a Bertrand outcome.

We then introduce the tick size regulation and show that the fee structure is no longer neutral. Suppose that the tick size is 1 and the liquidity maker can only quote integers. Even if the exchange does not charge anything on either side, no trade would occur: a buyer is willing to propose only non-positive integers, which will then be rejected by a seller, because no buyer would buy at 1 or above and no seller would be willing to sell at 0 or below.\(^3\) Next, consider a make fee of 0.5 and a take fee of -0.5, in which case the total fee remains at zero. Under this scenario, the buyer would submit a limit buy order at price 0, but thanks to the subsidy from the exchange, the offer is now always accepted by the seller. The nominal transaction price is zero, but the effective buy price is 0.5 because the buyer has to pay 0.5 to the exchange, and the effective sell price is 0.5

\(^3\) The only 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.
too, because the seller receives a rebate 0.5 from the exchange. It is important to notice that the effective buy and sell prices are proposed by the exchange when buyers’ and sellers’ valuation are within the same tick. Therefore, under a discrete tick size, with total fee fixed at zero, we can move from the no-trade equilibrium to the always-trade-efficient outcome.

The non-neutrality of the fee structure then creates product differentiation for otherwise identical exchanges, which facilitates second-degree price discrimination by the operators. We demonstrate that an operator can propose two exchanges with unequal effective sell prices. Holding the effective buy price constant, the liquidity maker (buyer) prefers to post a limit order to the exchange with higher effective sell price, because the seller arriving at Date 2 is more likely to accept the limit order. We show, however, that the operator also charges a higher effective buy price for the higher execution probability. In the meanwhile, the operator can offer another exchange with lower effective buy price for the liquidity makers, but this exchange has a lower effective sell price and lower execution probability. The liquidity makers then self-select based on the trade-off of profit conditional on execution and execution probability: high-valuation buyers select the exchange with high effective price and high execution probability and low-valuation traders select the exchange with a low effective buy price and low execution probability. Taken together, we have discovered a new form of price discrimination: when end users cannot neutralize the fees, the operator can use one side of the market to price-discriminate against the other side. Surprisingly, such price discrimination improves both buyers and sellers’ welfare because adding another exchange creates more effective transaction prices for end users.

We also show that tick size constraints destroy not only Bertrand equilibrium but also any pure-strategy equilibrium between competing exchanges. First, there exists no pure strategy equilibrium with positive total fees, because competing exchanges have incentives to undercut
each other toward zero total fees. The additional insight from the discrete tick size, however, is that Bertrand equilibrium with zero total fees cannot be sustained either, because the nature of competition has been changed by the tick size constraints. Without the tick size constraints, exchanges compete along one dimension—the total fee; with the tick size constraints, exchanges compete on two dimensions—the profit conditional on execution and the execution probability. It is possible that double deviations destabilize any possible pure-strategy equilibrium. In particular, given one exchange charging a zero total fee, there are two types of profitable deviations for the other exchange to increase total fee. One type of strategy charges a liquidity maker $\varepsilon$ more while charging a liquidity taker $\mu \cdot \varepsilon$ less (with $0 < \mu < 1$).\(^4\) Such a deviation reduces a liquidity maker’s profit conditional on execution, but meanwhile increases the execution probability, which attracts liquidity makers with higher trading surpluses. The other type of strategy attracts liquidity makers with lower valuations, by charging a liquidity maker $\mu \cdot \varepsilon$ less and a liquidity taker $\varepsilon$ more. Importantly, we show that under symmetrical mixed-strategy equilibria both exchanges earn strictly positive profits that increase with tick size, which explains the new entry into the fee game. Another interpretation of this result is that the tick size constraints allow operators to create differentiated services that do not exist when fees can be neutralized by end users, and such differentiation creates non-Bertrand outcomes and encourages new entry to the fee game.

Our model predicts that the tick size constraints encourage fragmentation in stock trading. Such a prediction is tested by the following identification strategy. We use ETF splits/reverse splits as exogenous shocks to the relative tick size (one divided by the price), with ETFs that split/reverse split as the treatment group and with ETFs that track the same index but experience no

\(^4\) We assume that this small change does not move the make fee from negative to positive.
splits/reverse splits as the control group. We find that splits fragment trading volume and reverse splits consolidate trading volume.

Our paper contributes to the literature on exchange competition and market fragmentation. The literature generally suggests that market fragmentation should not arise in equilibrium because of network externality or economies of scale. Yet a recent paper by O’Hara and Ye (2011) demonstrates significant fragmentation of trading volume. We reveal two channels for market fragmentation: operators have incentives to price-discriminate against traders and competition among operators does not lead to zero profit. To the best of our knowledge, the first channel, the price discrimination channel, has never been theoretically examined. For the second channel, we propose the first price competition model for otherwise-identical trading platforms, whereas the existing work on market fragmentation assumes either exogenous exchanges or product differentiation.

We also contribute new insights to the make-take fee literature by suggesting the economic intuition with which to evaluate a recent policy initiative that proposes completely banning these fees. One argument in favor of banning the fees cites their complexity and frequent fluctuations, while the complexity can be justified by the mixed-strategy equilibria documented in this paper. The other argument for eliminating the fee is based on fairness, because the fee leads to wealth transfer from one side of the market to the other side. However, we find that even the side being charged can benefit from the fees. We show that, surprisingly, liquidity providers prefer being charged instead of being subsidized when the tick size is large, and vice versa. This


6 For models based on exogenous exchanges, see Glosten (1994), Parlour and Seppi (2003), Hendershott and Mendelson (2000), Foucault and Menkveld (2008)). For the product differentiation model, see Pagnotta and Philippon (2013), Santos and Scheinkman (2001), Rust and Hall (2003), among others.

counterintuitive result is generated by the “cost” of the subsidy: the rebate to the liquidity provider can force her to quote a more aggressive price. For example, the buyer has to post a limit order at price 1 when the seller is charged, because the seller will reject a limit order at 0 if a positive fee needs to be paid when taking liquidity. However, the buyer can post a limit order at price 0 when she is charged, because the seller can sell at price 0 thanks to the subsidy. With a fixed fee level, the cost of the subsidy is higher when the tick size is large. Such an economic phenomenon also explains the existence of taker/maker markets, exchanges that charges liquidity makers.

Our paper shows that the fee competition serves as a way of bypassing the existing tick size regulations. This result contributes to the literature on tick size and questions the rationale of a recent initiative to increase the tick size of small stocks to five cents. Proponents of increasing the tick size argue that increasing the 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 the tick size would generate higher profits from the fee competition, which might encourage the entry of more stock exchanges.

Lastly, this paper contributes to the burgeoning literature on two-sided markets. Two-sided markets are markets in which “the volume of transactions between end-users depends on the structure and not only on the overall level of the fees charged by the platform” (Rochet and Tirole (R&T hereafter), 2006, p. 646). A fundamental challenge to the two-sided markets literature is the difficulty of demonstrating whether the so-called two-sidedness can generate qualitatively different predictions from those in a setup that is identical except for one-sidedness. To the best of our knowledge, our model is the first that nests a one-sided market model in a two-sided model. We find that price competition in two-sided markets is fundamentally different from such competition in a one-sided market. The fundamental driver of the difference is that the operators
of a two-sided market can engage in a more complex pricing strategy compared with operators of a one-sided market. In our model, two-sidedness creates product differentiation between the two intrinsically homogeneous exchanges, which creates second-degree price discrimination and destroys any pure-strategy equilibrium. The product differentiation also leads to market fragmentation, which leads to a market structure that differs from the one-sided market. These stark contrasts in price competition and the resulting market structure confirm the value of investigating two-sided markets independently. Also, the literature on two-sided markets is overwhelmingly based on network externality with multiple sides (Rysman, 2009). Our paper identifies another competing force to consider with regard to two-sided markets: product differentiation due to non-neutrality. In fact, the network externality model would predict stock market consolidation (Madhavan, 2000), and market fragmentation in reality suggests other economic forces that work in the opposite direction.

The rest of the paper is organized as follows. Section I sets up the model. Section II considers the benchmark case with a continuous tick size. Section III examines the non-neutrality of the fee under a discrete tick size. Section IV examines the product differentiation of otherwise homogeneous exchanges due to the non-neutrality of the fees. Section V considers second-degree price discrimination. Section VI considers competing exchanges with independent operators. Section VII presents the empirical tests of our theoretical model predictions. Section VIII concludes the paper and discusses the policy implications. The appendix contains mathematical proofs of the lemmas and propositions.

I. Model

Our model includes three types of risk-neutral players. A continuum of buyers with valuations of stocks $v_b$ uniformly distributed on $[d/2, d]$, and a continuum of sellers with valuations $v_s$ uniformly distributed on $[0, d/2]$. $v_b$ and $v_s$ are the buyers’ and sellers’ private information, respectively. We consider both the case of one monopoly operator of stock exchanges and the case of two competing operators. Operators have the option of establishing more than one stock exchange. The fixed cost of establishing an exchange is assumed to be infinitely small. The game has three stages. At Date 0, each operator chooses the number of exchanges as well as the fee structure in each exchange $F = (f_m, f_t)$, where $f_m$ denotes the make fee for a liquidity provider and $f_t$ denotes the take fee for a liquidity taker. Fees are charged only upon trade execution. At Date 1, nature draws a buyer to the market, and she can propose a trading price $P$ on one of the exchanges after observing the make-take fees, or she can choose not to submit a limit order at all.\(^9\) At Date 2, a seller arrives. The seller observes the make and take fees as well as the price proposed by the maker and then decides whether to trade. If she decides to trade, she must join the exchange that the liquidity provider chooses at Date 1 and trade at the proposed trading price $P$.\(^{10}\)

Due to the tick size regulation, the proposed price can occur only at an integer grid. That is,

$$P \in \{k \cdot d\}_{k=\infty}^{\infty},$$

(1)

where $k$ is an integer.

Exchanges profit through the total fees they charge. A necessary condition for an exchange to survive is a budget-balanced total fee. That is,

$$f_m + f_t \geq 0.$$  

(2)

---

\(^9\) The analyses and results involved in a case in which a seller arrives first are parallel to those of the case presented here. In an earlier version of the paper, we also considered when nature draws a buyer with probability $p$, and a seller with probability $1 - p$, and derived similar results, but the exposition is tedious.

\(^{10}\) In reality, a market order can trade with a limit order on another exchange due to regulation NMS. However, there is a routing fee for cross-exchange execution.
The purpose of this paper is to examine the impacts of the tick size constraint (1) on market outcomes. In this model, we consider two extreme cases of tick sizes: a continuous tick size of 0 and a discrete tick size of $d$. Other tick sizes can be considered intermediary cases between these two. The large tick size $d$ reflects the following fact about the market: fifty percent of S&P stocks have bid-ask spreads of one tick. Also, our model does not aim to model the valuations of long term investors. Instead, we capture the valuation of two types of short term traders: high-frequency traders or execution desks seeking the best execution of large orders. High-frequency traders care about the tick size because the tick size and fees are significant sources of their profits (Yao and Ye, 2015). These traders account for 70% of U.S. trading volume and are an important factor in stock exchange operations. Brokers or agencies using algorithms to find the best execution for their customers and trading desks inside long-term investment companies executing large orders for portfolio managers also care about the tick size and fees because of their importance in minimizing transaction costs. An important feature of U.S. stock markets is the delegation of trade execution to financial intermediaries (Biais, Glosten and Spatt, 2005). Therefore, we can treat traders in our model as execution desks representing buyers and sellers. Because their goal is to minimize transaction costs, the one-cent binding tick size becomes important for them. The same argument also justifies the division of the valuations of buyers and sellers at the midpoint. The most common benchmark for evaluating the performance of the execution desk is the midpoint of the bid and ask prices (Hasbrouck, 2014). The heterogeneity of valuations can be interpreted as the level of discretion investors give to their agencies. Therefore, even if long-run traders do not directly care about the tick size, the orders representing their trading interest can still be affected by the tick size and fees because of delegated execution.
Our model is parsimonious for limit and market orders. Traders do not choose the order type, the order book is empty when the maker arrives, and our three-stage model involves only one trading round. Therefore, our model does not allow for limit order queuing. Theoretical studies on order-placing strategy generally provide a richer structure of order selection by assuming exogenous stock exchanges (Rosu (2009), Parlour (1998), and Parlour and Seppi (2003)). For example, Foucault and Menkveld (2008) find that two exogenous exchanges can co-exist because of queuing. However, our model demonstrates that two endogenous exchanges can also co-exist in the absence of queuing. Section VI demonstrates that endogenizing the decision of exchanges is not trivial: the game between exchanges reaches complex mixed-strategy equilibria even granting these simplifications. Nevertheless, our model explains several stylized facts that have not been addressed in the literature, and we generate out-of-sample predictions supported by the data.

II. Benchmark: Continuous Tick Size

As a benchmark, we consider the case without tick size constraint (1), so that the buyer can propose any trading price $P$. Denote the total fee as $T \equiv f_m + f_t$.

We solve the model by backward induction. At Date 2, a seller will trade if $v_s \leq P - f_t$. Thus, the probability that a seller accepts an order is $\max\{0, \frac{2}{d} \cdot (P - f_t)\}$. At Date 1, a buyer proposes a price

$$\max_P (v_b - P - f_m) \cdot \Pr(v_s \leq P - f_t) = (v_b - P - f_m) \cdot \max\{0, \frac{2}{d} \cdot (P - f_t)\}$$

which yields
\[ P(v_b) = \frac{v_b + T}{2} - f_m. \]  

Only buyers with \( v_b \geq T \) will trade, and the execution probability is \( \Pr(v_s \leq P(v_b) - f_t) = \frac{v_b - T}{d} \).

So both the participation of both sides depends only on the level of the total fee. Further, a buyer with valuation \( v_b \) will get an expected surplus

\[
BS^c(v_b) = (v_b - P(v_b) - f_m) \cdot \Pr(v_s \leq P(v_b) - f_t) = \frac{(v_b - T)^2}{2d},
\]

which also depends on the total fee \( T \), not its breakdown.

It follows that the exchange’s profit is

\[
\pi^c = T \cdot \Pr(v_b \geq \max\{T, \frac{d}{2}\}) \cdot \int_{\max\{T, \frac{d}{2}\}}^{d} \frac{v_b - T}{d} \cdot \frac{2}{d} dv_b
\]

\[
= T \cdot \frac{4}{d^2} \cdot \left( d - \max\{T, \frac{d}{2}\} \right) \cdot \int_{\max\{T, \frac{d}{2}\}}^{d} \frac{v_b - T}{d} dv_b
\]

\[
= \frac{4}{d^3} \cdot T \cdot \left( d - \max\{T, \frac{d}{2}\} \right) \cdot \int_{\max\{T, \frac{d}{2}\}}^{d} (x - T) dx.
\]

Clearly, \( \pi^c \) depends only on \( T \). It is easy to solve for the optimal \( T \).

The equilibrium can be summarized by the following proposition.

**Proposition 1 (Neutrality of Fees under Continuous Tick Size):** Under a continuous tick size, the makers’ strategy, the takers’ strategy and the exchanges’ strategy only depends on total fee \( T = f_m + f_t \) but not its breakdown.

The optimal total fee is

\[
T^c = \frac{3}{8} \cdot d.
\]

---

11 This is the buyers’ optimal proposed price when \( T < d \). If \( T \geq d \), then no buyer would post any order, and thus there would be no trade. As we will show next, exchanges will always set \( T < d \) in order to earn revenue.
The buyer surplus, seller surplus, and the exchange’s profit are, respectively,

\[ BS^c = \frac{31}{384} \cdot d, \quad SS^c = \frac{31}{768} \cdot d, \quad \pi^c = \frac{9}{64} \cdot d. \]

The intuition underlying Proposition 1 follows the canonical tax-neutrality principle. A direct consequence of the fee neutrality is that all participating buyers would always choose the exchange with the lowest total fee when multiple exchanges offer the same service. Such a result precludes price discrimination by a monopoly operator. Since only the total fee matters to all buyers and sellers, another immediate implication of the fee neutrality is the Bertrand equilibrium that occurs when exchanges owned by different operators compete. These results are summarized in the following corollary.

**Corollary 1:** Under a continuous tick size,

(i) a single operator has no incentive to open more than one exchange;

(ii) competing exchanges belonging to independent operators all choose a zero total-fee and earn zero profits.

Although the neutrality of fees, the absence of price discrimination, and Bertrand competition are consistent with the intuitions implied by canonical principles, they are inconsistent with the stylized facts. Next, we consider the case in which the tick size equals \(d\), and demonstrate how such a small friction can generate results that are dramatically different from conventional wisdom, while nevertheless being consistent with the reality.

## II Non-Neutrality

Beginning in this section, we consider tick size requirement (1), so that liquidity makers can propose a price only at grid \{\ldots, -2d, -d, 0, d, 2d, \ldots \}. A direct consequence of the tick size requirement is the non-neutrality of the fee structure. This non-neutrality resulting from tick size
was first documented by Foucault, Kadan, and Kandell (2013). Our contribution to the literature is to demonstrate the impacts of non-neutrality on exchange competition and the organization of the exchange industry.

We look mainly at exogenous fees throughout this section. This helps us to build intuitions and establish intermediary results that can simplify our analysis when we endogenize fees later. The analysis of the non-neutrality also reveals the nature of fee competition: exchanges are setting sub-penny increments of effective buy and sell prices, and are profiting from the difference. Section II.A. establishes non-neutrality on a monopoly exchange. Section II.B. advances the intuition of non-neutrality by considering competition between two exchanges with an inverted fee structure: a maker/taker market (subsidizing liquidity makers and charging liquidity takers) and a taker/maker market (charging liquidity makers and subsidizing liquidity takers).

A. Buyer/Seller Behavior with One Exchange

In this subsection, we consider a monopoly exchange. The game is a sequential-move game, and we solve it by backward induction.

At Date 2, after observing the make and take fees \((f_m, f_t)\), the seller’s willingness-to-sell (WTS) will be \(v_s + f_t\). Given the buyer’s proposed price \(P\), the seller will trade if and only if

\[
P \geq v_s + f_t. \tag{4}\]

At Date 1, given the make and take fees \((f_m, f_t)\), the buyer’s willingness-to-buy (WTB) will be \(v_b - f_m\). This implies that the buyer will propose a price such that

\[
P \leq v_b - f_m. \tag{5}\]

\(^{12}\)O’Donoghue (2014) and Brolley and Malinova (2013) show that fees can be non-neutral when brokers charge a flat fee to their customers but pay make-take fee to the exchanges.
Combining these results with the fact that \(0 \leq v_s\) and \(v_b \leq d\), (4) and (5) can be rewritten as

\[
0 \leq v_s \leq P - f_t \\
P + f_m \leq v_b \leq d.
\]

(4')

(5')

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

\[
P - f_t \leq P + f_m.
\]

(2')

Thus, in order for a trade to occur, (2), (4), and (5) (or equivalently, (2'), (4') and (5')) together require:

\[
0 \leq v_s \leq P - f_t \leq P + f_m \leq v_b \leq d.
\]

(6)

(6) is a necessary condition for a trade to occur. Based on (6), we can see the impacts of the tick size requirement on trading.

In the presence of tick size requirement (1), the proposed price \(P\) cannot fall within the range of 0 to \(d\). Thus, (6) can never hold when \(f_m = f_t = 0\), except for the knife-edge case of \(v_b = d\) or \(v_s = 0\). Therefore, the would-be efficient trade cannot occur thanks to the tick size constraint.

**Lemma 1:** With tick size requirement (1), 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 fees charged to either liquidity makers or liquidity takers, the buyer’s WTB is 0 while the seller’s WTS is \(d\). 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 (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 the tick size constraints.

Next, we show that there are many fee structures with the same zero total fee that can restore efficient trading. This is in stark contrast to the no-trade equilibrium under zero fees on both sides, and illustrates the non-neutrality of the fee structure.

**Lemma 2 (Volume Maximizing Fees):** With tick size requirement (1), there exist infinitely many volume maximum fees with the following structure:

\[
\begin{align*}
\ell_{m}^{SO} &= (k + \frac{1}{2}) \cdot d \\
\ell_{t}^{SO} &= -(k + \frac{1}{2}) \cdot d
\end{align*}
\]

where \( k \) is an integer.

All these fee structures lead to the same effective buy and sell prices

\[
\begin{align*}
p_{b}^{SO} &= p_{s}^{SO} = \frac{d}{2}
\end{align*}
\]

*Trades always happen under these fees.*

**Proof:** See the appendix.

Lemma 2 conveys four important messages. First, combined with Lemma 1, Lemma 2 indicates that, under the tick size constraints, the fee structure is *not* neutral. The fee structures in both lemmas exhibit zero total fees, but charging zero on both sides leads to the no-trade equilibrium (Lemma 1), while charging the opposite make and take fees gives rise to the socially optimal equilibrium (Lemma 2). Therefore, a discrete tick size makes the stock exchange markets two-sided, according to R&T’s (2006) definition: “the volume of transactions between end-users depends on the structure and not only on the overall level of the fees charged by the platform.”

Second, the fact that infinitely many fees can achieve the same equilibrium outcome suggests that we might need to fine-tune the two definitions of two-sided markets in R&T (2006).
Their alternative definition is that a market is two-sided if and only if the fee structure is designed
to maximize total volume while fixing the level of the total fee is unique (R&T, 2006, p. 658). In
their footnote 31, they refine the definition a bit by mentioning that “a finite number of solutions
is also indicative of two-sidedness.” Nevertheless, the infinitely many solutions of fees in Lemma
2 seems to violate the second definition. Such inconsistency comes from an intriguing fact that is
overlooked in the literature. The literature often focuses on whether end users can neutralize the
fee structure or not, without considering the extent to which they can do so. By contrast, in our
model, end users can neutralize the component of the fees that is in multiples of ticks, but fail to
adjust the residual component that is less than one tick. In particular, under fee structure $f_{m}^{SO} =
\left( k + \frac{1}{2} \right) \cdot d, f_{t}^{SO} = - \left( k + \frac{1}{2} \right) \cdot d$, a buyer will always propose a price $P = -k \cdot d$
to neutralize the $k \cdot d$ component, so that the effective buy and sell prices are $P + f_{m}^{SO} = P - f_{t}^{SO} = \frac{d}{2}$. This
explains why our model has infinitely many fee structures to maximize the volume, while the
equilibrium outcome and effective buy and sell prices are uniquely determined. Because of the
possibility of such partial neutralization, we propose modifying the second definition of the two-
sided markets in R&T (2006) by replacing the fees set by the platform with the effective buy and
sell prices faced by end users.

Third, Lemma 2 establishes that, under the tick size constraints, make-take fees have the
potential to increase social welfare, which questions the rationale of banning the fees. As a
benchmark for future comparison, we refer the fee structure defined in Lemma 2 as socially
optimal fee structure.

Finally, the lemma suggests that, without loss of generality, we can restrict our attention to
the fee structures whose absolute values are no more than one tick, that is, $|f_{i}| \leq d \ (i = m, t)$. This simplification is indeed a stylized fact in practice: no exchange ever charges fees of greater
than one tick. Thus, hereafter, we refine tick size requirement (1) and focus on a two-grid constraint under which buyers can propose a price only at

\[ P = \{0, d\}. \quad (7) \]

Lemma 3 then characterizes the feasible set of fee structures to ensure that trading occurs and guaranteeing the implied trading price and participation:

**Lemma 3 (Fee Structure, Trading Price and Participation with One Exchange):** With tick size requirement (7), given make-take fees \((f_m, f_t)\), the following results hold.

(i) 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,

\[ f_m \cdot f_t < 0. \quad (8) \]

and

\[ f_m + f_t \leq d. \quad (9) \]

(ii) Trading prices under alternate fee structures: The buyer will propose a trading price of

\[ P = \begin{cases} 0 & \text{when } f_m > 0 \text{ (so that } f_t < 0) \\ d & \text{when } f_m < 0 \text{ (so that } f_t > 0) \end{cases}. \quad (10) \]

(iii) Participation: A buyer with \(v_b \geq \max\{P + f_m, p_0 + d/2\}\) will trade with a seller with \(v_s \leq \min\{P - f_t, d/2\}\).

**Proof:** See the appendix.

Lemma 3 states that the make-take fees have to carry opposite signs, which fits perfectly with the stylized facts in the real world. Cardella, Hao and Kalcheva (2013) show that the fee structures changed 108 times during 2008~2010, including the switch from subsidizing takers to
subsidizing makers, but no major exchange ever adopts the halfway case of charging both sides.\textsuperscript{13} To the best of our knowledge, we are the first to explain this.

Compared with charging zero fees on both sides, charging positive fees on both sides makes the bid-ask spread even wider than one tick, which aggravates the trade-hindering effect of the tick size constraints. By charging one side and in the meantime subsidizing the other side, the exchange essentially creates effective buy and effective sell prices within the same tick. In particular, under $F = (f_m, f_t)$, the effective buy price is $P + f_m$, and the effective sell price is $P - f_t$.

Hereafter, we denote the effective buy and sell prices $(p_b, p_s)$ as

$$
p_b \equiv P + f_m = \begin{cases} 
  f_m & \text{when } f_m > 0 \\
  d + f_m & \text{when } f_m < 0 
\end{cases}
$$

$$
p_s \equiv P - f_t = \begin{cases} 
  f_t & \text{when } f_m > 0 \\
  d - f_t & \text{when } f_m < 0 
\end{cases}
$$

Figure 3 shows the effective buy and sell prices. The exchange will set an effective buy price below some buyers’ valuations and an effective sell price above certain sellers’ valuations, thus enabling trading to occur. As will be shown in Section V.A., a profit-maximizing exchange will not set effective buy and sell prices that facilitate all possible trades, as this would leave the exchange with zero profit. Yet a necessary condition for profit maximization is that make-take fees are set in a way that allows some trades to occur.

\textbf{B. Competition between the Maker/Taker Market and the Taker/Maker Market}

In this subsection we consider the competition between the maker/taker market and the taker/maker market with exactly inverted fee structures.

\textsuperscript{13} Direct Edge A switched in August 1, 2011 and switched back to subsidizing takers on September 1, 2012. A new platform IEX currently charge both sides based on the argument of fairness, but does not achieve significant volume till now.
In our game, the liquidity taker seems to play a passive role: the liquidity taker can trade only in markets selected by the liquidity maker, because the unchosen exchange has an empty limit-order book. It thus seems that the priority of an exchange is attracting liquidity makers, and a natural way to do so is to subsidize them. This intuition, however, is incorrect because a liquidity maker profits only when her limit order is accepted by a liquidity taker, and hence she will take into account the exchange at which her order is more likely to be accepted when selecting an exchange. It turns out that a liquidity maker has to post a more aggressive limit order in the market that subsidizes him. According to Lemma 3, in a market that subsides liquidly makers, the buyer must post a limit order at price $P = d$, because no seller would trade with a limit buy order at 0 when she has to pay a positive take fee. In contrast, in a market that charges liquidity makers, the buyer will post a limit buy order at a price $P = 0$, because some sellers can accept the limit order of 0 after the subsidy. Hence, the subsidy actually imposes a cost on the liquidity maker to quote a less favorable nominal price to herself. Proposition 1 shows that makers prefer a market that charges them over a market that subsidizes them when the tick size is large relative to the sum of the fees’ absolute values, and vice versa.

**Proposition 2 (An Example in which All Buyers May Prefer Being Charged):** With tick size requirement (1), suppose exchange 1 adopts fee structure $(f_m, f_t)$, and exchange 2 adopts fee structure $(f_t, f_m)$, where $d > f_m > -f_t > 0$. All buyers prefer exchange 1 (or 2) when $|f_m| + |f_t| < d$ (or $|f_m| + |f_t| > d$), and they are indifferent between the two exchanges only when $|f_m| + |f_t| = d$.

**Proof:** See the appendix.

Proposition 2 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 her and subsidizes the liquidity taker.
Fixing the level of make-take fees, as the tick size decreases the liquidity maker gradually shifts her preference to the market that subsidizes her and charges the liquidity taker.

Proposition 2 provides a plausible justification for the existence of the taker/maker market. The emergence of markets which charge liquidity makers is a puzzle, particularly when regulations can put taker/maker markets at a disadvantage. One such policy is the trade-through rule.\textsuperscript{14} 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. Our paper fills this gap. The result is also consistent with the empirical evidence in Yao and Ye (2014) that taker/maker markets attract volume for securities with large relative tick size and maker/taker markets attract volume for securities with low relative tick size.

\textbf{IV Product Differentiation and Buyer’s Segmentation}

This section demonstrates that the non-neutrality of the fee structure, led by the tick size constraints, allows operators to create vertical differentiation for otherwise identical exchanges. The previous section shows that the nature of fee competition is to choose the effective buy and sell prices \((p_b, p_s)\) within the tick. In the following analysis, we consider each exchange’s decision variables as the effective buy and sell prices to avoid a tedious discussion of infinite many fee structures which achieve the same equilibrium outcome.

\textbf{A. Vertical Product Differentiation}

\textsuperscript{14} 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 the competition with the maker/taker market because the latter has a better nominal price, ceteris paribus.
Given effective sell price $p_s$, the marginal seller’s valuations are given by

$$\bar{v}_s \equiv \min \left\{ p_s, \frac{d}{2} \right\}. \quad (12)$$

So the probability that a seller accepts the limit order is given by

$$q_s = \Pr(\nu_s \leq \bar{v}_s) = \frac{2}{d} \cdot \bar{v}_s \quad (13)$$

Therefore, a higher effective sell price implies a higher probability of execution. From a buyer’s point of view, an exchange with a higher effective sell price is of a higher quality because it offers a higher probability of realizing the gains from the trade. All buyers prefer an exchange with a higher effective sell price to another with a lower effective sell price if both offer identical effective buy prices. As a result, two exchanges with differentiated effective sell prices are vertically differentiated.

More importantly, two exchanges have another instrument—the effective buy price—they can use to charge buyers prices that reflect variations in execution probability, so that exploiting such vertical differentiation becomes possible. It is worth noting that the space needed for two intrinsically identical exchanges to do so stems from the two-sidedness of the markets, or equivalently, the non-neutrality of the fee structure. Such non-neutrality is a direct consequence of the tick size constraints. When the tick size is zero, the total fee is neutral, and thus the exchange cannot use the breakdown of the total fee to price-discriminate based on vertical differentiation. This discussion again shows that the market outcome depends crucially on the ability of end users to neutralize the fee. We illustrate the result using the stock exchange industry, but we believe the intuition should hold in other contexts as well. If end users can neutralize the fee that is set by the exchange, the competition among exchanges is only one-dimensional. If end users cannot neutralize the fee, then the exchange has more power and room for manipulating the fee structure.
The non-neutrality of the fee structure under two-sidedness allows platforms to create product differentiation that is not available in a one-sided market.

Next, we show that such product differentiation leads to market fragmentation. This provides a formal justification of the observation that “(i)t is relatively uncommon for industries based on two-sided platforms to be monopolies or near monopolies” (Evans and Schmalensee, 2007, p. 166). The fragmentation of two-sided markets is a puzzle, because the overwhelming majority of work in the two-sided market literature is based on network effects (Rysman, 2009), and network effects in general tend to induce consolidation. To the best of our knowledge, our paper is the first to show that two-sidedness may add an extra dimension that facilitates product differentiation and hence market fragmentation.

B. Buyer Segmentation under Two Exchanges

In this subsection, we look closely at buyers’ choices between two vertically differentiated exchanges.

Given the effective buy and sell prices from exchanges 1 and 2, say \((p_b^1, p_s^1)\) and \((p_b^2, p_s^2)\), the buyer’s surpluses when choosing exchange 1 and exchange 2 are

\[
BS^1 = \frac{d}{2} \cdot (v_b - p_b^1) \cdot \tilde{v}_s^1 = \frac{d}{2} \cdot (v_b - p_b^1) \cdot p_s^1 \\
BS^2 = \frac{d}{2} \cdot (v_b - p_b^2) \cdot \tilde{v}_s^2 = \frac{d}{2} \cdot (v_b - p_b^2) \cdot p_s^2
\]

(14)

Here the equalities above follow from \(\tilde{v}_s^i = p_s^i (i = 1,2)\). This is because neither exchange would set \(p_s^i > \frac{d}{2}\) so that \(\tilde{v}_s^i = \frac{d}{2} (i = 1,2)\), as doing so only hurts its per trade profit while not gaining any trading volume.
When \( p_s^1 = p_s^2 \), \( BS^1 \geq BS^2 \) if and only if \( p_b^1 \leq p_b^2 \). Without loss of generality, suppose that \( p_s^1 < p_s^2 \). Following the previous section’s discussion, exchange 1 becomes a low execution probability exchange while exchange 2 becomes a high execution probability exchange for buyers. The buyer’s surpluses under the two exchanges are shown in Figure 4.

**Insert Figure 4 about Here**

When \( p_b^1 \geq p_b^2 \), as shown in panel (a) of Figure 4, \( BS^1 \leq BS^2 \) for any \( v_b \geq p_b^2 \). So all buyers will choose exchange 2. This result is intuitive, because exchange 2 offers higher execution probability together with a lower effective buy price, so all traders prefer exchange 2.

When \( p_b^1 < p_b^2 \), as shown in panels (b) and (c) of Figure 4, there exists a unique intersection

\[
\varphi = p_b^1 + (p_b^2 - p_b^1) \cdot \frac{p_s^2}{p_s^2 - p_s^1}.
\]  \hspace{1cm} (15)

And \( BS^1 \leq BS^2 \) for any \( v_b \geq \varphi \). Recall that \( v_b \in \left[ \frac{d}{2}, d \right] \), and we can show that \( \varphi > \frac{d}{2} \), so all that remains is to check is whether \( \varphi \geq d \). The boundary of \( \varphi = d \) in \((p_b^2, p_s^2)\)-plane is given by

\[
p_s^2 = \rho(p_b^2) = p_s^1 \cdot \frac{d - p_b^1}{d - p_b^2}.
\]  \hspace{1cm} (16)

or equivalently,

\[
p_b^2 = \omega(p_s^2) = d - p_s^1 \cdot \frac{d - p_b^1}{p_s^2}.
\]  \hspace{1cm} (17)

When \( \varphi > d \), or \( p_b^2 > d - (d - p_b^1) \cdot \frac{p_b^1}{p_s^2} \), all the buyers will go to the exchange with low execution probability, as shown in panel (c) of Figure 4, because the price of the high-quality exchange with is too high to justify its higher execution probability.
Panel (b) of Figure 4 demonstrates the most interesting case in which the exchange with high execution probability and the exchange with low execution probability co-exist. This happens when the exchange with high execution probability charges a price that is higher than the exchange with low execution probability charges, but not too high to drive even the buyer with the highest valuation to the exchange with low execution probability. The split of the market comes from the heterogeneity of buyers’ valuations. Ceteris paribus, all traders prefer a higher execution probability. Yet, the degrees of their favors over the higher execution probability differ. Buyers with relatively high valuations have larger gains from trading, so they care more about the execution probability than do buyers with relatively low valuations. The heterogeneity of valuations across traders gives the vertically differentiated exchanges some room to charge different prices for different execution probabilities. The key trade-offs for traders in this model is between execution probability and gains from trading conditional on execution. Engle, Ferstenberg, and Russell (2008) consider such execution risk and “reward” as the fundamental trade-offs for the execution desk, just as the mean and the variance of return are the fundamental trade-offs for the portfolio manager.

The buyers’ choices given \((p_b^1, p_s^1)\) and \((p_b^2, p_s^2)\) are summarized in the lemma below and Figure 5.

**Lemma 4 (Buyer Segmentation under Two Exchanges)** For any given \((p_b^1, p_s^1) \in \left[\frac{d}{2}, d\right] \times [0, \frac{d}{2}]\), the square \(\left[\frac{d}{2}, d\right] \times [0, \frac{d}{2}]\) in the \((p_b^2, p_s^2)\)-plane can be divided into the following four areas:

(i) \(s_1 \equiv \{(p_b^2, p_s^2) \mid \frac{d}{2} \leq p_b^2 \leq p_b^1, p_s^1 \leq p_s^2 \leq \frac{d}{2}\} : \text{no buyer chooses exchange 1, and all buyers with } p_b^2 \leq v_b \leq d \text{ choose exchange 2};\)
\( s_2 \equiv \{ (p_b^2, p_s^2) | p_b^1 \leq p_b^2 \leq \omega(p_s^2), p_s^1 \leq p_s^2 \leq \frac{d}{2} \} : \) buyers with \( p_b^1 \leq v_b \leq \varphi \) choose exchange 1, and buyers with \( \varphi \leq v_b \leq d \) choose exchange 2;

\( s_3 \equiv \{ (p_b^2, p_s^2) | p_b^1 \leq p_b^2 \leq d, p_s^1 \leq p_s^2 \leq \min\{\rho(p_b^2), \frac{d}{2}\} \} : \) all buyers with \( p_b^1 \leq v_b \leq d \) choose exchange 1, and no buyer chooses exchange 2;

\( s_4 \equiv \{ (p_b^2, p_s^2) | p_b^1 \leq p_b^2 \leq d, 0 \leq p_s^2 \leq p_s^1 \} : \) all buyers with \( p_b^1 \leq v_b \leq d \) choose exchange 1, and no buyer chooses exchange 2;

\( s_5 \equiv \{ (p_b^2, p_s^2) | \frac{d}{2} \leq p_b^2 \leq p_b^1, 0 \leq p_s^2 \leq \rho(p_b^2) \} : \) buyers with \( \varphi \leq v_b \leq d \) choose exchange 1, and buyers with \( p_b^2 \leq v_b \leq \varphi \) choose exchange 2;

\( s_6 \equiv \{ (p_b^2, p_s^2) | \frac{d}{2} \leq p_b^2 \leq p_b^1, \rho(p_b^2) \leq p_s^2 \leq p_s^1 \} : \) no buyer chooses exchange 1, and all buyers with \( p_b^2 \leq v_b \leq d \) choose exchange 2.

Here \( \varphi, \rho(p_b^2), \) and \( \omega(p_s^2) \) are given by (15), (16), and (17), respectively.

**Insert Figure 5 About Here**

For any given \((p_b^1, p_s^1)\), the square \( \left[ \frac{d}{2}, d \right] \times \left[ 0, \frac{d}{2} \right] \) in the \((p_b^2, p_s^2)\)-plane can be divided into six areas.

In area \( s_1 \), exchange 2 offers a higher execution probability with a lower buyer price because \( p_s^2 > p_s^1 \) and \( p_b^2 < p_b^1 \). This attracts all traders to exchange 2. This area corresponds to panel (a) of Figure 5. In area \( s_2 \), both exchanges co-exist and split the markets. Exchange 2 attracts buyers with high valuations because its execution probability is higher, and exchange 1 appeals to buyers with low valuations. This corresponds to panel (b) of Figure 5. The curve \( p_s^2 = \rho(p_b^2) \) describes the case in which the buyer with the highest valuation \( d \) is indifferent between the two exchanges.

Area \( s_3 \) corresponds to panel (c) of Figure 5. In this case, the effective buy price of exchange 1 is so low compared with \( p_b^2 \) that even the buyer with the highest valuation prefers it. The areas \( s_4, s_5 \) and \( s_6 \) follow similar logic of \( s_1, s_2 \) and \( s_3 \).
Our results pertaining to buyer segmentation explain the client puzzle. Skjeltorp, Sojli, and Tham (2011) find evidence that NASDAQ and NASDAQ BX, two exchanges operated by the NASDAQ OMX group, have different clients.\textsuperscript{15} Foucault (2011) suggests that the co-existence of various make/take fees should serve to screen investors by type. However, Foucault (2011) also mentions that “it is not clear however how the differentiation of make/take fees suffices to screen different types of investors since, in contrast to payments for order flow, liquidity rebates are usually not contingent on investors’ characteristics (e.g., whether the investor is a retail investor or an institution).” We explain this puzzle: when end users cannot neutralize the breakdown of the fee, and the markets therefore become two-sided, the operators can screen one side of the market based on the other side. This explains, as we show in the next session, why operators have an incentive to open multiple exchanges to price-discriminate against traders.

V Price Discrimination

In this section, we demonstrate the second degree price discrimination facilitated by product differentiation. To simplify our analysis, we consider the decision of a monopoly operator with the option to establish up to two exchanges. Recall from Section II that, under a continuous tick size, a monopoly operator has no incentive to operate more than one exchange, because only the total fee matters and all traders will trade on the exchange with the lowest total fee. However, we show that a discrete tick size allows an exchange with a higher total fee to co-exist with another involving a lower total fee because of product differentiation, as described in the previous section. Section A first considers the benchmark case in which the monopoly operator establishes one

\textsuperscript{15} They find that NASDAQ BX might be used by algorithmic investors who use algorithms to minimize execution costs (agency algorithms) rather than to quickly exploit private information.
exchange, and Section B then analyzes the economic incentive for the monopoly operator to open another exchange on which to practice price discrimination.

A. Benchmark: One Operator with One Exchange

The monopoly operator with one exchange chooses \((p_b, p_s)\) to maximize its profit

\[
\pi = (p_b - p_s) \cdot q_b \cdot q_s
\]

\[
= \frac{4}{d^2} \cdot (p_b - p_s) \cdot (d - \tilde{v}_b) \cdot \tilde{v}_s,
\]

\[
= \frac{4}{d^2} \cdot (p_b - p_s) \cdot (d - p_b) \cdot p_s
\]

where the first equality follows from (13). Since \(\pi\) increases with \(p_b\) (or decreases with \(p_s\)) whenever \(\tilde{v}_b = \max\left\{p_b, \frac{d}{2}\right\} = d/2\) (or \(\tilde{v}_s = \min\left\{p_s, \frac{d}{2}\right\} = d/2\)), it is easy to see that the monopoly exchange will always set \((p_b, p_s)\) such that \(\max\left\{p_b, \frac{d}{2}\right\} = p_b\) and \(\min\left\{p_s, \frac{d}{2}\right\} = p_s\). So the second equality follows.

The key trade-offs for the exchange are the profit conditional on execution and both sides’ participation probabilities. A difference between the buyer’s optimization and the exchange’s is that the buyer optimizes along only two dimensions: her profit depends on execution and sellers’ participation probability, while the operator needs to balance both buyers’ and sellers’ participation and profit conditional on participation. The equilibrium outcomes are summarized in the proposition below.

**Proposition 3 (Optimal Monopoly Fees and Equilibrium Surplus Divisions):** With tick size requirement (1), the monopoly stock exchange will set its effective buy and sell prices as

\[
p_b^M = \frac{2}{3} \cdot d, \quad p_s^M = \frac{1}{3} \cdot d.
\]
The corresponding optimal make-take fees are given by

$$\begin{align*}
  f^M_m &= \frac{2}{3} \cdot d \\
  f^M_t &= -\frac{1}{3} \cdot d'
\end{align*}$$

and

$$\begin{align*}
  \tilde{f}^M_m &= -\frac{1}{3} \cdot d \\
  \tilde{f}^M_t &= \frac{2}{3} \cdot d'.
\end{align*}$$

(19)

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

$$\begin{align*}
  BS^M &= \frac{2}{27} \cdot d, \\
  SS^M &= \frac{2}{27} \cdot d, \\
  \pi^M &= \frac{4}{27} \cdot d.
\end{align*}$$

(20)

The two monopoly fee structures impose an effective buy price of $\frac{2}{3}d$ and an effective sell price of $\frac{1}{3}d$. The implied bid-ask spread is $\frac{1}{3}d$ under the monopoly fee. The exchange obtains $\frac{1}{3}d$ once a trade happens, and 0 otherwise. This fee structure excludes buyers with low valuations (i.e., $[\frac{1}{2}d, \frac{2}{3}d]$) and sellers with high valuations (i.e., $[\frac{1}{3}d, \frac{1}{2}d]$), each of which comprises one-third of the buyer or seller population. These buyers and sellers are those with relatively lower incentive to trade and relatively lower trading surplus to realize when trading among them occurs. By excluding buyers and sellers from whom the exchange profits less, the exchange attracts only high-valuation buyers and low-valuation sellers, and enjoys monopoly profits. Indeed, the effective buy price under monopoly is $\frac{1}{6}d$ higher than the effective buy price under the social optimum, and the effective sell price is $\frac{1}{6}d$ lower than the effective sell price under the social optimum. However, the implied bid-ask spread of $\frac{1}{6}d$ is still lower here than it is in the case with no make-take fees, which is exactly one tick. As a result, compared with the results obtainable with zero fees on both sides, the monopoly fees achieve a Pareto improvement: the maker, the taker, and the exchange all benefit.

Another way to understand Proposition 3 is that the operator’s fee choice entails two dimensions. The first one is the total fee. The second one is the breakdown of the fee. This two-
dimensional optimization differentiates ours from two existing studies on make-take fees. Colliard and Foucault (2012) study the total fee when the breakdown is neutral, and Foucault, Kadan, and Kandel (2013) focus on the breakdown of the fees but with a fixed total fee. As we shall see, such two-dimensional optimization will generate drastically different predictions from those to be found in the existing literature.

B. Price Discrimination with Two Exchanges

In this subsection, we explain the incentives that induces an operator to run multiple exchanges. We introduce a new mechanism to the price-discrimination literature: when end users cannot neutralize the fee structure, the operator can use one side of the market to price-discriminate against the other side. In our model, the price discrimination for buyers comes from differentiated prices for sellers. With differentiated execution probability implied by heterogeneous sell prices, makers self-select based on their expected surpluses from exchanges with differentiated execution probability. This mechanism corresponds to second-degree price discrimination, which is different from payments for order flow, a common third-degree price discrimination under which traders are charged differentially based on their identities (retail or institutional).

Proposition 4 (A Monopoly Opens Two Exchanges): With tick size requirement (1), a monopoly operator always has an incentive to run two exchanges. It will set its effective buy and sell prices as

\[ p_b^1 = \frac{7}{10} \cdot d, \quad p_s^1 = \frac{2}{5} \cdot d, \quad p_b^2 = \frac{3}{5} \cdot d, \quad p_s^2 = \frac{1}{5} \cdot d. \]  

(21)

The corresponding optimal make-take fees for exchange 1 are given by

\[
\begin{align*}
    f_m^1 &= \frac{7}{10} \cdot d, \\
    f_t^1 &= -\frac{2}{5} \cdot d,
\end{align*}
\]

and

\[
\begin{align*}
    \tilde{f}_m^1 &= -\frac{3}{10} \cdot d, \\
    \tilde{f}_t^1 &= \frac{3}{5} \cdot d.
\end{align*}
\]

(22)
The corresponding optimal make-take fees for exchange 2 are given by

\[
\begin{aligned}
    f_m^2 &= \frac{3}{5} \cdot d, \\
    f_t^2 &= -\frac{1}{5} \cdot d.
\end{aligned}
\]  

The buyer surplus, seller surplus, and the monopoly firm’s profit are

\[
\begin{aligned}
    BS_{M2E} &= \frac{2}{25} \cdot d, \\
    SS_{M2E} &= \frac{2}{25} \cdot d, \\
    \pi_{M2E} &= \frac{4}{25} \cdot d.
\end{aligned}
\]  

(24)

Proposition 4 shows that operators can engage in second-degree price discrimination for buyers, as illustrated in Figure 6. Such discrimination is implemented by offering the buyer two exchanges with different effective buy prices and execution probabilities (or equivalently, qualities). Compared with the single exchange we analyzed in the previous subsection, exchange 2 are with low execution probability but it charges a lower effective buy price, and exchange 1 are with higher execution probability but it charges a higher effective buy price. Figure 6 illustrates that such a menu of two price–execution probability packages induces buyers to self-select: buyers with valuations between \([\frac{3}{5}d, \frac{4}{5}d]\) select exchange 2 and buyers with valuations between \([\frac{4}{5}d, d]\) select exchange 1.

Another interesting fact is that whether an exchange is expensive or cheap is not well defined when the markets have two sides. From the buyer’s perspective, exchange 2 is cheaper because its effective buy price is lower. However, exchange 2 is more costly for a seller because its effective sell price is lower as well. More interestingly, the low-quality exchange (exchange 2) with lower execution probability actually charges a higher total fee than the high-quality exchange
(exchange 1) does. By contrast, under a continuous tick size, exchange 1 would dominate exchange 2 because of its lower total fee.

Moreover, the proposition shows that such second-degree price discrimination improves social welfare. More surprisingly, buyers, sellers and the operator all benefit from price discrimination. One source of this enhanced efficiency comes from larger-scale participation. For example, buyers with valuations between $\left[ \frac{3}{5} d, \frac{2}{3} d \right]$ now can post limit orders on exchange 2, whereas they are excluded under a monopoly exchange. As the inefficiency arises mainly from the discrete tick size, the creation of new price levels by setting up new exchanges reduces this inefficiency and generates gains from trading for all parties.

VI. Competing Operators and Non-existence of Pure-strategy Equilibrium

Now we consider the case with two competing operators, each of which establishes one exchange. Even with this simplification, the model yields complex and interesting outcomes. Section VI.A. shows the non-existence of pure-strategy equilibrium under the tick size constraints. This result contrasts sharply with the case involving zero tick size in Colliard and Foucault (2012), who predict pure-strategy equilibrium with zero fees and profit. Section VI.B. then shows that symmetric mixed-strategy equilibria, in which both exchanges earn strictly positive profits, always exists. As an illustration, we also characterize one set of mixed-strategy equilibria.

A. No Pure-strategy Equilibrium

The fact that the tick size constraints destroy Bertrand equilibrium can be understood intuitively based on product differentiation. Operators can create exchanges with various execution probabilities, which relieves the competitive pressure on otherwise identical exchanges. What is
more interesting is that the tick size constraints also destroy any pure-strategy equilibrium, which is summarized in Proposition 5.

**Proposition 5 (No Pure-strategy Equilibrium):** There is no pure-strategy equilibrium when two exchanges compete under tick size $d$.

**Proof:** See the appendix.

The detailed proof of the proposition is in the appendix. Here we offer a sketch of the proof and the corresponding intuitions. We first prove the non-existence of pure-strategy equilibrium with any exchange earning a positive profit, which follows from the Bertrand argument in Colliard and Foucault (2012). Without loss of generality, suppose exchange 1 earns a strictly positive profit and exchange 2 begins by earning a lower profit or having the same profit as exchange 1. In the former case, exchange 2 can always increase its profit by undercutting exchange 1’s effective buy price by $\epsilon$ and mimicking its effective sell price. By doing so, exchange 2 offers the same execution probability with a lower effective buy price. So all buyers will choose exchange 2, and exchange 2 will earn what exchange 1 earned before. In the latter case, by the same undercutting strategy, exchange 2 can corner the entire market, rather than sharing the market with exchange 1, with only $\epsilon$ concession per trade. Therefore, there is no pure strategy equilibrium with any exchange earning positive profits.

The standard Bertrand argument seems to suggest that both exchanges should end up with zero profits and zero fees. We find however that, due to the two-sidedness of the markets as well as the heterogeneity of buyer/seller valuations, one exchange can always find a profitable deviation strategy given the other exchange maintain a pure strategy.

In particular, there are two possibilities for zero-profit outcomes: 1) at least one side of the market does not participate; 2) the effective buy and sell prices are equal. It is easy to see that the
first case cannot be sustained in equilibrium, because one of the exchanges would have incentives to facilitate some trading and profit from it. Now we look at what happens when the effective buy price is equal to the effective sell price. There are three possible cases to be considered.

Suppose exchange 1 charges \( p_b^1 = p_s^1 > \frac{1}{2} \cdot d \). Then exchange 2 can easily deviate by setting \( p_b^2 = p_b^1 - \epsilon, p_s^2 = \frac{1}{2} \cdot d \). In this case exchange 2 reduces the effective sell price but it does not reduce the execution probability for the buyer because all sellers accept \( p_s^2 = \frac{1}{2} \cdot d \). Then all buyers choose exchange 2 because it offers the same quality but cheaper trading.

The most interesting result is that a fee structure of \( p_b^1 = p_s^1 = \frac{1}{2} \cdot d \) cannot be sustained in any pure-strategy equilibrium. Panel (a) of Figure 7 demonstrates the intuitive deviation, \( p_b^2 = \frac{1}{2} d - \mu \epsilon, p_s^2 = \frac{1}{2} d - \epsilon \), with \( \epsilon > 0 \) and \( 0 < \mu < 1 \). This deviation reduces the execution probability, which leads to a flatter buyer surplus function on exchange 2. However, the intersection of \( BS_2 \) with the horizontal axis falls to the left of the intersection of \( BS_1 \) with the horizontal axis, which implies that \( BS_2 \) crosses \( BS_1 \) at the point where \( v_b^1 > \frac{1}{2} \cdot d \). Buyers with valuations between \( \left( \frac{1}{2} \cdot d, v_b \right) \) then prefer exchange 2 to exchange 1, and thus exchange 2 can enjoy strictly positive profits. The intuition is as follows: even when exchange 1 provides the maximum execution quality and charges the lowest price for sustaining that quality, exchange 2 can deviate by charging an even lower price while making a small sacrifice in execution probability. This strategy caters to buyers with low valuations, such as those with valuations close to \( \frac{1}{2} \cdot d \).

Insert Figure 7 about Here

When \( p_b^1 = p_s^1 < \frac{1}{2} \cdot d \), the execution probability is less than 1. Then one exchange can deviate by setting \( p_b^2 = p_b^1 + \epsilon \) and \( p_s^2 = p_s^1 + \mu \cdot \epsilon \), with \( \epsilon > 0 \) and \( 0 < \mu < 1 \). Such a deviation
involves a higher effective price for buyers but they are compensated with higher probability of execution. Panel (b) of Figure 7 illustrates that this deviation makes $BS_2$ steeper than $BS_1$, although the intersection of $BS_2$ with the horizontal axis falls to the right of $BS_1$. It is easy to see that there exists a point $\overline{v}_b < d$ such that buyers with valuations higher than $\overline{v}_b$ will go to exchange 2.

Such a profitable deviation always exists because, in this two-sided market, exchanges compete along two dimensions to attract order flow—the buyer’s surplus conditional on trading and the trade execution probability—via two price instruments. Thus, the effective buy and sell prices do not have to be equally appealing to all buyers and sellers. As indicated in Lemma 4, when exchange 1’s pricing strategy is fixed, exchange 2 can always find some price range that is attractive to some buyers in a two-dimensional plane while at the same time ensuring that $p^2_b > p^2_s$. In the formal proof in the appendix, we have constructed profitable deviation strategies for all possible cases.

One implication of Proposition 5 is that there exists no pricing strategy on which competing exchanges converge. 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. Our result explains the total fee fluctuations displayed in Figure 2, which further motivates us to investigate, in the next sub-section, the random nature of competing fee structures.

B. Mixed-strategy Equilibrium

This subsection focuses on characterizing symmetric mixed-strategy equilibria, in which both exchanges follow the same randomization when deciding their effective buy and sell prices $(p_b, p_s)$.

**Proposition 6 (Mixed-strategy Equilibrium):**
(i) There exist symmetric mixed-strategy equilibria, in which \((p_b, p_s)\) has a convex support on \([\frac{d}{2}, d] \times [0, \frac{d}{2}]\):

(ii) In equilibrium, both exchanges earn strictly positive profits, and their profits increase linearly with \(d\).

(iii) In any of the symmetric equilibria,

\[
\frac{p_s}{d} = g\left(\frac{p_b}{d}\right),
\]

where \(g(\cdot)\) is an increasing function.

Proposition 6 proves the existence and describes some properties of mixed-strategy equilibria. The mixed strategies have a convex support, which implies that there is a connected range of effective buy and sell price pairs in which no specific pair 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 with respect to probability. At first glance, buyers and sellers 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 buyers nor the highest rebate for sellers. Proposition 6 provides a plausible explanation of the diverse fee structures across exchanges.

Part (ii) of Proposition 6 states that the profits under mixed strategies equilibrium are strictly positive, which could induce new entries and market fragmentation. This result arises from the two-sidedness of the markets caused by the tick size regulation. When the tick size is zero, as shown in Section III.C., the markets are one-sided. Hence, the competition between two exchanges can drive profits 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. When the tick size is positive, the markets become two-sided. So the competition between exchanges does not lead to zero profit for the exchanges,
which encourages new entries. Regulators are often concerned that the entry of new trading platforms generates greater market fragmentation (O’Hara and Ye, 2011), but the literature has achieved only a limited understanding of why the market becomes more and more fragmented. We show that one force causing such fragmentation is the existing tick size regulation.

Part (ii) also shows that a large tick size increases operators’ profits, which should create greater incentives for new exchange entry. On April 5, 2012 Congress passed the Jumpstart Our Business Startups (JOBS) Act, which encourages the SEC to examine the possibility of increasing tick size. A pilot program to increase the tick size for 1200 small stocks has been proposed for comments. The rationale for increasing the tick size is that a wider tick size increases market-making revenue and supports sell-side equity research and, finally, increases the number of IPOs (Grant Thornton, 2012). Yet we have not seen any theoretical or empirical work showing that an increase in tick size can increase IPOs. Our paper shows, however, that an increase in the tick size can lead to higher profits for stock exchanges and encourage new entries.

Part (iii) of Proposition 6 says that the support of any symmetric mixed-strategy equilibria must be an upward-sloping curve that pushes the effective buy price higher than the effective sell price. This also confirms that competing exchanges must earn strictly positive profits when they randomize their fees.

In the appendix, we show that mixed-strategy equilibria are in general characterized by two partial differentiation equations along with some boundary conditions. It is a daunting task to find analytical solutions for all possible mixed-strategy equilibria. One set of randomized pricing strategies is given in Corollary 1. Note that there might be some other symmetric or asymmetric mixed-strategy equilibria.

Corollary 2: One set of symmetric mixed-strategy equilibrium is as follows.

(a) \( p_b - p_s = \frac{1}{6} \cdot d \);

(b) \( p_b \) is randomized over \([L, U] \subset \left[ \frac{1}{2}, \frac{7}{12} \right]\) with distribution function \( F(x) \), where

\[
F(x) = \frac{C_1}{\left( \frac{9}{2} \right)^{\frac{1}{3}} \cdot \left( x - \frac{4}{9} \right)^{\frac{1}{3}}} + \frac{C_2}{90 \cdot \left( x - \frac{1}{6} \right)^{\frac{1}{3}}} \cdot \left[ 1 - \frac{H \left( \frac{x - \frac{1}{6}}{\frac{4}{9}} \right)^{\frac{1}{3}}}{\frac{1}{3}} \right]
\] (25)

Here \( H \) is a Hypergeometric function \( _2F_1 \left( \frac{2}{3}, \frac{5}{3}, \frac{5}{18(x - \frac{1}{6})} \right) \), and \((C_1, C_2, L, U)\) satisfy:

\[
\int_x^U t \cdot dF(t) - \int_x^L t \cdot dF(t) + \left( \frac{3}{2} - 2x \right) \cdot F(x) + \left( x - \frac{1}{6} \right) \left( \frac{4}{3} - 3x \right) \cdot F'(x) = \frac{1}{6}
\] (26)

\[
F(L) = 0
\] (27)

\[
F(U) = 1
\] (28)

\[
F'(x) > 0 \text{ for any } x \in [L, U]
\] (29)

Proof: see the appendix.

VII. Empirical Results

One implication of our model is that a larger tick size cause greater market fragmentation. Although securities with prices above one dollar have a uniform one-penny tick size in the U.S. equity market, the relative tick size varies with the price of a security: low-priced securities have larger relative tick sizes and vice versa. This section empirically tests the prediction that a larger relative tick size causes greater market fragmentation. Because our test is based on the cross-sectional variation in the relative tick size, the results need to be carefully interpreted. When an operator establishes an exchange, it is hard for the operator to consider the revenue stock by stock. Therefore, although we believe that the tick size drives market fragmentation, our cross-sectional choice should be more closely related to the choices made by liquidity makers and takers in our
model. When the relative tick size is large, liquidity makers and takers find it hard to neutralize the breakdown of the fees, which generates fragmentation for low-priced stocks. A small relative tick size facilitates neutralization of the fee breakdown and generates consolidation. Section VII.A. describes the data used in testing this prediction; Section VII.B. presents the test results using multivariate regression analysis, and Section VII.C. tests this hypothesis using difference-in-differences analysis following the identification strategy proposed by Yao and Ye (2015).

A. Data and Sample

The empirical analyses use two securities samples from January 2010 through November 2011. The multivariate regression in Section VII.B. uses a sample of stocks selected by Hendershott and Riordan. The original sample includes 60 NYSE–listed and 60 NASDAQ–listed stocks. The stratified 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. During our sample period, three of the 120 stocks were delisted (BARE, CHTT and KTII), our sample is thus consists of 117 stocks.

Section VII.C. tests the causal impact of the relative tick size on market fragmentation using a difference-in-differences approach. The identification follows Yao and Ye (2015), which use the split/reverse splits of leveraged ETFs as shocks to the relative tick size. The test uses leveraged ETFs that have undergone splits/reverse splits as the pilot group, and uses leveraged ETFs that track the same indexes and undergo no splits/reverse splits in our sample period as the control group. Leveraged ETFs amplify the return on the underlying index, and they often appear in pairs that track the same index but in opposite directions. For example, if the leverage ratio is 2:1 and if on one day the underlying index returns 1%, one ETF in the pair will return 2%, and the
other one in the pair will return -2%.\(^{17}\) Although twin leveraged ETFs often have similar nominal prices when launched for IPOs, the return amplification often diverges from their nominal prices after issuance. As the ETFs are commonly issued by the same issuer, the issuers often use splits/reverse splits to keep their nominal prices aligned with each other. We use the Bloomberg Database to collect information on leveraged ETF pairs, and select the pairs that track the same index with an identical multiplier. The data are then merged with the CRSP to identify their reverse splitting events.

The variable of interest, market fragmentation, is 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, 2014). We use the Herfindahl index as a measure of market fragmentation, which is calculated as follows:

\[
Herfindahl \text{ Index}_{i,t} = \sum_{j=1}^{13} \left( \frac{ExchVol_{i,j,t}}{TotalVol_{i,t}} \right)^2
\]

where \(i\) indexes stock and \(t\) indexes time. \(ExchVol_{i,t,j}\) is the trading volume on exchange \(j\), while \(TotalVol_{i,t}\) is the total trading volume on all 13 stock exchanges.\(^{18}\)

The market fragmentation measure is then merged with the sample of 117 stocks and the leveraged ETF sample, respectively. The final sample used in the multivariate regression consists of 117 stocks in 51,950 stock-day observations. The final sample used in the difference-in-differences analysis consists of 5 splits and 23 reverse splits of leveraged ETFs from January 2010 through November 2011.\(^{19}\) The sample window is 5 days before the reverse split event and

\(^{17}\) The actual return will be slightly different, as management fees and transaction are yet to be taken into account.

\(^{18}\) We exclude the volume in TRFs because they have different trading mechanism.

\(^{19}\) To ensure sufficient trading volume in these ETFs, we use leveraged ETFs that experience at least 10,000 trading volume trades each day in the sample period.
5 days after the reverse split event for the treatment and control groups. The summary statistics for the stock and leveraged ETF samples are presented in Panel A and Panel B of Table 1, respectively.

**Insert Table 1 about Here**

**B. Regression Analysis**

Identifying the causal effects of the relative tick size on market fragmentation challenges us. One type of endogeneity arises from omitted variables. The estimation coefficient of the relative tick size would be biased and inconsistent if we did not control for variables that are correlated with both the nominal price and market fragmentation. A necessary condition for the occurrence of omitted variable bias is, therefore, that the omitted variable needs to correlate with nominal price. This criterion narrows our search for four streams of literature on the cross-sectional variation of nominal prices. A more comprehensive discussion of these four lines of literature is in Yao and Ye (2015).

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. We control for this average using an industry fixed effect, whereby industries are classified using the Fama–French classification of 48 industries. The optimal tick size hypothesis argues that firms choose the optimal tick size through splits/reverse splits (Angel, 1997). We include idiosyncratic risk, age, and the number of analysts that may affect the choice of the optimal tick size, from this study. The marketability hypothesis argues that a lower price appeals to individual traders. We include the measure of small investor ownership

---

20 Angel (1997) argues that the relative tick size also depends on whether a firm is in a regulated industry, and we controlled for this effect by including an industry-by-time fixed effect using firm book value as a control for size, which is similar to the market cap for which we have also controlled. When book value is included as an additional control, the results are similar.
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, so we use the probability of informed trading (PIN) proposed by Easley, Kiefer, O’Hara and Paperman (1996) to control for information asymmetry.

The regression takes the following form:

\[
H_i,t = u_{j,t} + \beta \times tick_{relative,i,t} + \Gamma \times X_{i,t} + \epsilon_{i,t}
\] (31)

where \(Herfindahl \ Index_{i,t}\) is the Herfindahl Index, a measure of market fragmentation for stock \(i\) on date \(t\). \(u_{j,t}\) is the industry-by-time fixed effect. The key variable of interest, \(tick_{relative,i,t}\), is the daily inverse of the stock price for stock \(i\). \(X_{i,t}\) are the control variables that include idiosyncratic risk, age, the number of analysts, small investor ownership, and the probability of informed trading. The results are presented in Table 2.

Insert Table 2 about Here

Table 2 shows that the coefficient estimates of the relative tick size are positive and significant at the 1% level. This finding suggests that market fragmentation increases with the relative tick size. Table 2 shows that trading of large firms is more fragmented, as is trading of firms with longer histories. Variables other than the relative tick size, market cap, and firm age do not have a significant impact on market fragmentation.

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

This section establishes the causal relationship between the relative tick size and market fragmentation. The difference-in-differences test follows the identification proposed by Yao and Ye (2015), and uses leveraged ETF splits/reverse splits as a clean means of identification. Leveraged ETFs that have experienced splits/reverse splits are used as the pilot group, and
leveraged ETFs that track the same index but have not undergone reverse splits are used as the control group.

Specifically, we estimate the following model:

\[ H_{i,j,t} = u_{i,t} + \gamma_{i,j} + \rho \times D_{trt_{i,t},j} + \theta \times return_{i,t,j} + \epsilon_{i,t,j} \]  

where \( i \) indexes the underlying index, \( j \) indexes ETFs, and \( t \) indexes time. The dependent variable in the equation is the Herfindahl Index, a measure of market fragmentation. We include index-by-time fixed effects, \( u_{i,t} \), which controls for the time trend that may affect each index. \( \gamma_{i,j} \) capture the ETF fixed effects that absorb the time-invariant differences between two leveraged ETFs that track the same index \( i \). The regression also controls for their returns, \( return_{i,t,j} \), in each period, which is the only main difference left between the ETFs tracking the same index after we control for index-by-time and ETF fixed effects. \( D_{trt_{i,t},j} \) is 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. The coefficient estimate \( \rho \), for the dummy variable, captures the treatment effect and is the main focus of this regression.

To derive an unbiased estimate of the treatment effect, the actual split/reverse split must be uncorrelated with the error term. This does not mean that the actual split/reverse split must be exogenous. As we control for both index-by-time fixed effects and ETF fixed effects, the estimation will be biased only if the actual split/reverse split is somehow related to the contemporaneous idiosyncratic shocks to the dependent variables (Hendershott, Jones and Menkveld, 2011). Two stylized facts are then important for establishing the unbiasedness of the coefficient estimate.
First, the schedule for executing a reverse split is predetermined and announced well before the actual reverse split. Thus it seems highly unlikely that the reverse split schedule could be correlated with idiosyncratic shocks to HFT liquidity provision in the future. In addition, fund companies often conduct multiple splits/reverse splits on the same day for ETFs tracking diversified underlying assets. Such a diversified sample further mitigates the concern that the reverse-split decision is correlated with ETF-specific idiosyncratic shocks. Second, the motivation for ETF reverse splits is transparent. The issuers of ETFs conduct splits/reverse splits when their nominal prices differ dramatically from their pairs. Such differences in price can be captured by the ETF fixed effect, and the estimate of coefficient $\rho$ remains unbiased.

Table 3 displays the regression result for the impacts of splits on the Herfindahl Index for Leveraged ETF splits and reverse splits. Column (1) indicates that the Herfindahl Index decreases after splits, implying that trading becomes more fragmented after an increase in the relative tick size. Column (2) indicates that the Herfindahl Index increases after reverse splits, implying that trading becomes more consolidated after a decrease in the relative tick size.

Insert Table 3 about Here

VIII. Conclusion

We examine the competition between stock exchanges over proposed make-take fees. When traders can quote a continuous price, the breakdown of the make-take fees is neutralized and order flow consolidates to the exchange with the lowest total fee. Under tick size constraints,

---

21 Using a longer time window creates overlap between the pre-announcement and pre-split periods, but we find similar results.
22 For example, the announcement made on April 9, 2010 involves splits ranging from oil, gas, gold, real estate, financial stocks, and basic materials to Chinese indices.
fee breakdowns are no longer neutral, and such non-neutrality of fee structure explains a number of anomalies in price competition between stock exchanges. We first show that one necessary condition for trade to occur is that the make fee and the take fee must carry opposite signs. Second, the two-sidedness of the market allows operators to establish multiple exchanges with heterogeneous fee structures for second-degree price discrimination. Third, we demonstrate the non-existence of pure-strategy equilibrium in the fee game under tick size constraints, which explains the diversity and frequent fluctuations in fee structures.

The price competition under tick size constraints then explains the market fragmentation among nearly homogeneous stock exchanges. First, the same operator has an incentive to operate multiple exchanges to implement second-degree price discrimination. Second, mixed-strategy equilibria entail positive profits for all competing operators, and the equilibrium profits increase with the tick size. This result explains the entry of platforms with new fee structures. Our model predicts that a larger relative tick size leads to greater market fragmentation. We find empirical support for this prediction using a panel regression and a difference-in-differences regression.

Our paper contributes to the literature on make-take fees and tick size, providing direct policy implications for the debate over the two topics. As make-take fees constitute transaction costs imposed on liquidity makers and liquidity takers, they have attracted regulators’ attention. A regulation capping the take fee at 30 cents per one hundred shares has already been implemented, and more aggressive initiatives, such as banning the fee completely, are under discussion among regulators. One argument for banning the fees is based on fairness, because the fees lead to wealth transfer from one side of the market to the other. We show, however, that in the presence of tick size constraints, fees imposed by the exchanges Pareto improve the market relative to a market without fees. The second argument for banning the fee cites its complexity and frequent
fluctuations, which are explained by the mixed-strategy equilibria in the model. The last argument for banning the fee involves agency concerns. 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 fees (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 fees might hinder the would-be efficiency of trading.

Encouraged by the JOBS Act, the SEC is proposing a pilot program to increase the tick size. 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, in which liquidity providers are charged and liquidity takers are subsidized. Another effect of increased tick size is more intense fee competition and potentially a more fragmented market.

Our paper is subject to several limitations. For example, we do not model the competition between liquidity makers, we do not allow limit order queuing, and we do not allow traders to choose their order types. It would be interesting to examine whether more richly structured models would lead to new insights into exchange competition. Also, makers and takers in our model directly execute their own orders, but the execution is often delegated to agents such as brokers or execution desks in current market. It would be interesting to build agency issues into our model along the line of O’Donoghue (2014) and Brolley and Malinova (2013).

Our model and empirical results are based on the exchange industry, but the intuition from the model can be applied to other two-sided markets. The central idea of the paper is that platform operators can exploit the lack of ability from two sides to neutralize the fee structure, and create differentiation for otherwise homogeneous products. We believe one fruitful line of research would be to apply this intuition to other markets. For example, the non-surcharge provision in the credit card industry prevents merchants from charging differentiated prices for individual credit cards, and thus leads to non-neutrality of fee structure. We conjecture that such non-neutrality might be a force leading to the proliferation of credit cards.
REFERENCES


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

Carrillo, Juan and Guofu Tan, 2015, Platform Competition with Complementary Products, Working paper.


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.


Skjeltorp, Johannes, Elvira Sojli, and Wing Wah Tham, 2011, Make take fees and liquidity externalities, Working paper, Erasmus University.


Yao, Chen and Mao Ye, 2015, Tick Size Constraints, High Frequency Trading, and Liquidity, Working paper, University of Illinois.

50
Appendix

Proof of Proposition 1: From profit function expression (3), we discuss two possible cases when choosing the optimal \( T \).

(i) When \( T \geq \frac{d}{2} \), we have \( \max \{ T, \frac{d}{2} \} = T \). Then

\[
\pi^c = \frac{4}{d^3} \cdot T \cdot (d - T) \cdot \int_{T}^{d} (x - T) dx
\]

\[
= \frac{2}{d^3} \cdot T \cdot (d - T)^3.
\]

It is easy to solve for the optimal \( T \) in this case is \( \hat{T} = \frac{d}{2} \), which yields \( \hat{\pi}^c = \frac{d}{8} \).

(ii) When \( T < \frac{d}{2} \), we have \( \max \{ T, \frac{d}{2} \} = \frac{d}{2} \). Then

\[
\pi^c = \frac{4}{d^3} \cdot T \cdot \frac{d}{2} \cdot \int_{\frac{d}{2}}^{d} (x - T) dx
\]

\[
= \frac{1}{d} \cdot T \cdot (\frac{3}{4} d - T).
\]

It is easy to solve for the optimal \( T \), which in this case is \( T^* = \frac{3}{8} \cdot d \), which yields \( \pi^* = \frac{9}{64} \cdot d \).

Thus, the proposition follows. ■

Proof of Lemma 2: First, efficient trades are guaranteed if and only if the effective buy and sell prices coincide at \( \frac{d}{2} \), so that all buyers with \( v_b \geq \frac{d}{2} \) would like to buy, and all sellers with \( v_s \leq \frac{d}{2} \) would like to sell.

Next, we explain how the make-take fees given in the lemma can achieve efficient trading all the time. Given \( f_m^{SO} = (\frac{1}{2} + k) \cdot d \), a buyer must propose a trading price \( P = -k \cdot d \), because otherwise (6) can never hold and there won’t be any trade. Thus, the effective buy price after the fee will be \( P + f_m^{SO} = \frac{d}{2} \). So all buyers will trade. Given \( f_t^{SO} = -(\frac{1}{2} + k) \cdot d \) and the proposed
trading price $P = -k \cdot d$, the effective sell price will be $P - f^S_t = \frac{d}{2}$. Hence, all sellers are willing to trade, too. ■

**Proof of Lemma 3:** If $f_m > 0$, then (6) implies that the buyer must propose $P = 0$, because otherwise $P + f_m > d$. Then from (6) again, it follows that $f_t < 0$, because otherwise we would have $P - f_t = -f_t < 0$. Similarly, if $f_m < 0$, then (6) implies that the buyer must propose $P = d$ and $f_t > 0$. Thus, a necessary condition for the occurrence of a trade is charging one side while subsidizing the other side.

Meanwhile, (6) leads to

$$\begin{cases} 0 \leq P - f_t \\ P + f_m \leq d \end{cases},$$

which is equivalent to

$$\begin{cases} f_t \leq P \\ f_m \leq d - P \end{cases}.$$ 

Thus, for (6) to hold, we must have:

$$f_m + f_t \leq d.$$ 

Other parts of the lemma follows directly. ■

**Proof of Proposition 2:** Under exchange 1’s fee structure $(f_m, f_t)$, according to parts (ii) and (iii) of Lemma 3, the buyer will propose a trading price of $P = 0$, and trade with the seller with $v_s \leq \min\{-f_t, d/2\}$. So the buyer’s surplus when joining exchange 1 is

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

$$= \begin{cases} \frac{2}{d} \cdot (v_b - f_m) \cdot (-f_t) & \text{if } -f_t < d/2 \\ v_b - f_m & \text{if } -f_t \geq d/2. \end{cases}$$
Under exchange 2’s fee structure \((f_t, f_m)\), the buyer will similarly propose a trading price \(P = d\), and trade with the seller with \(v_s \leq \min\{d - f_m, d/2\}\). So the buyer’s surplus when joining exchange 2 is

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

\[
= \begin{cases} 
\frac{2}{d} \cdot (v_b - d - f_t) \cdot (d - f_m) & \text{if } f_m > d/2 \\
\frac{v_b - d - f_t}{2} & \text{if } f_m \leq d/2.
\end{cases}
\]

We must consider the following three possible cases.

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

\[
BS^1 - BS^2 = v_b - f_m - \frac{2}{d} \cdot (v_b - d - f_t) \cdot (d - f_m)
\]

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

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

\[
BS^1 - BS^2 \leq [BS^1 - BS^2]|_{v_b = d}
\]

\[
= \frac{2}{d} \cdot (d - f_m) \cdot \left( \frac{d}{2} + f_t \right) \leq 0.
\]

The buyer prefers exchange 2.

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

\[
BS^1 - BS^2 = \frac{2}{d} \cdot (v_b - f_m) \cdot (-f_t) - \frac{2}{d} \cdot (v_b - d - f_t) \cdot (d - f_m)
\]

\[
= \frac{2}{d} \cdot (f_m - f_t - d) \cdot (v_b - d).
\]

So

\[
BS^1 \geq BS^2 \text{ if and only if } f_m - f_t \leq d.
\]

Case (iii): \(d/2 \geq f_m > -f_t\)
\[ BS^1 - BS^2 = \frac{2}{d} \cdot (v_b - f_m) \cdot (-f_t) - v_b - d - f_t \]

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

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

\[ BS^1 - BS^2 \geq [BS^1 - BS^2] \text{ when } v_b = d \]

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

The buyer prefers exchange 1.

Combining the above three cases, the proposition follows. ■

**Proof of Lemma 4:** We consider the case when \( p^1_s < p^2_s \) as we do in the main text. That is, we focus on the rectangular area \((p^2_b, p^2_s) \in [\frac{d}{2}, d] \times [p^1_s, \frac{d}{2}]\). The analysis of the case when \( p^1_s > p^2_s \) is parallel.

When \( p^1_b \geq p^2_b \), we have \( \varphi \leq p^2_b \leq p^1_b \). So for any \( v_b \geq p^2_b \), we have \( v_b \geq \varphi \). Hence, \( BS^2 - BS^1 \geq 0 \) for any \( v_b \geq p^2_b \).

When \( p^1_b < p^2_b \), we have \( p^1_b < p^2_b < \varphi \). In this case, we can show that \( \varphi > \frac{d}{2} \). Note that

\[ \varphi = p^1_b + (p^2_b - p^1_b) \cdot \frac{p^2_s}{p^2_s - p^1_s} > p^1_b > \frac{d}{2}, \]

where the first inequality follows from \( p^2_b > p^1_b \) and \( p^2_s > p^1_s \). Thus, there are two possible cases to be considered. If \( \frac{d}{2} < \varphi < d \), then for any \( p^1_b \leq v_b \leq \varphi \), \( BS^1 - BS^2 \geq 0 \); for any \( \varphi < v_b \leq d, BS^1 - BS^2 < 0 \). If \( d \leq \varphi \), then for any \( v_b \leq d \), we have \( v_b \leq \varphi \). So \( BS^1 - BS^2 \geq 0 \) for any \( v_b \leq d \). ■

**Proof of Proposition 4:** With \((p^1_b, p^1_s)\) and \((p^2_b, p^2_s)\) from two exchanges, according to Lemma 4 and Figure 5, only one exchange can survive in area \( s_1, s_3, s_4, \text{ or } s_6 \). So the monopoly firm, with only one active exchange, will earn \( \pi = \frac{4}{27} \cdot d \) as in Proposition 3.
Now we look at possible profitable improvement area $s_2$ or $s_5$. Without loss of generality, suppose $p^1_s \geq p^2_s$. In particular, we consider region $s_5$. By Lemma 4, we know buyers with $\varphi \leq v_b \leq d$ choose exchange 1, and buyers with $p^2_b \leq v_b \leq \varphi$ choose exchange 2. Hence, the monopoly firm’s profit is

$$\pi^{M2E} = \frac{4}{d^2} \cdot \left[ (p^1_b - p^1_s) \cdot (d - \varphi) \cdot p^1_s + (p^2_b - p^2_s) \cdot (\varphi - p^2_b) \cdot p^2_s \right].$$

It is straightforward to solve the above maximization with constraints $p^1_s \geq p^2_s, p^1_b \geq p^2_b, and 0 \leq p^2_s \leq \rho(p^2_b)$. The solutions are given in the proposition.

Proof of Proposition 5 (by contradiction) Without loss of generality, suppose the pure-strategy equilibrium exists, and in equilibrium $\pi^1 \geq \pi^2$. There are two possible cases: (i) $\pi^1 > 0$; (ii) $\pi^1 = \pi^2 = 0$.

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

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

$$p^2_b = p^1_b - \varepsilon \text{ and } p^2_s = p^1_s,$$

where $\varepsilon > 0$. Then, we are in area $s_4$. By Lemma 4, no one goes to exchange 1 anymore, and exchange 2’s profit becomes

$$\hat{\pi}^2 = (p^1_b - p^1_s - \varepsilon) \cdot \Pr(v_b \geq \bar{v}^2_b) \cdot \Pr(v_s \leq \bar{v}^2_s)$$

$$\geq (p^1_b - p^1_s - \varepsilon) \cdot \Pr(v_b \geq \bar{v}^1_b) \cdot \Pr(v_s \leq \bar{v}^1_s)$$

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

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

(i-b) $B$ can deviate to (A.1), so that no one goes to exchange 1 anymore, and exchange 2’s profit becomes
\[ \hat{\pi}^2 = (p_b^1 - p_s^1 - \epsilon) \cdot \Pr(v_b \geq \bar{v}_b^2) \cdot \Pr(v_s \leq \bar{v}_s^2) \]
\[ \geq (p_b^1 - p_s^1 - \epsilon) \cdot \Pr(v_b \geq \bar{v}_b^1) \cdot \Pr(v_s \leq \bar{v}_s^1) \]
\[ > (p_b^1 - p_s^1 - \epsilon) \cdot \text{Some of } \Pr(v_b \geq \bar{v}_b^1) \cdot \Pr(v_s \leq \bar{v}_s^1) \]
\[ = \pi^1 - \epsilon \cdot \text{Some of } \Pr(v_b \geq \bar{v}_b^1) \cdot \Pr(v_s \leq \bar{v}_s^1) \]
where the first inequality follows from (A.1), and the second inequality is derived from the fact that exchange 1 does not get all the buyers with \( \Pr(v_b \geq \bar{v}_b^1) \) in this case, because some buyers go to exchange 2. Thus, exchange 2 can always strictly increase its profit by deviation (A.2) with a sufficiently small \( \epsilon \).

(ii) There are two subcases: (ii-a) No trading; (ii-b) Trading with \( p_b^i = p_s^i \in (0, d), i = 1, 2 \).

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

(ii-b) Denote \( p_b^1 = p_s^1 = a \). There are three sub-subcases: (ii-b-I): \( 0 < a < \frac{d}{2} \); (ii-b-II): \( a = \frac{d}{2} \); (ii-b-III): \( \frac{d}{2} < a < d \).

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

\[ p_b^2 = a + \epsilon \text{ and } p_s^2 = a + \mu \cdot \epsilon, \quad \text{(A.2)} \]

where \( \epsilon > 0 \) and \( 0 < \mu < 1 \). For sufficiently small \( \epsilon \), we will have \( \bar{v}_s^2 = p_s^2 > a = \bar{v}_s^1, p_b^2 > p_b^1 \), and

\[ \tilde{\phi} = p_b^2 + (p_b^2 - p_b^1) \cdot \frac{\bar{v}_s^2}{\bar{v}_s^2 - \bar{v}_s^1} \]
\[ = a \left( 1 + \frac{1}{\mu} \right) + \epsilon. \]
Clearly, \( \tilde{\varphi} \) decreases with \( \mu \), \( \lim_{\mu \to 1} \tilde{\varphi} = 2a + \varepsilon < d \), where the inequality follows from \( a < \frac{d}{2} \) and \( \varepsilon \) is sufficiently small. Hence, for sufficiently small \( \varepsilon \), we can always have \( \tilde{\varphi} < d \). And from the buyer segmentation analysis, we know that the buyer with \( v_b \geq \tilde{\varphi} \) will go to exchange 2. So relatively high-valuation buyers will go to exchange 2 and exchange 2 will make a 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

\[
p_b^2 = a - \mu \cdot \varepsilon \quad \text{and} \quad p_s^2 = a - \varepsilon, \tag{A.3}
\]

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

\[
\varphi = p_b^1 + (p_b^1 - p_b^2) \cdot \frac{\tilde{v}_s^2}{\tilde{v}_s^1 - \tilde{v}_s^2}. \\
= a(1 + \mu) - \mu \cdot \varepsilon.
\]

Clearly, \( \varphi \) increases with \( \mu \), and \( \lim_{\mu \to 1} \varphi = d - \varepsilon < d \), where the inequality follows from \( a = \frac{d}{2} \) and \( \varepsilon > 0 \). Hence, for sufficiently small \( \varepsilon \), we can always have \( \varphi < d \). And from the buyer-segmentation analysis, we know that the buyer with \( v_b \leq \varphi \) 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

\[
p_b^2 = a - \mu \cdot \varepsilon \quad \text{and} \quad p_s^2 = a - \varepsilon, \tag{A.4}
\]

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

Proof of Proposition 6:
We establish this proposition in the following 4 steps.

**Step 1:** $0 \leq p_s \leq p_b \leq d$.

Suppose that $p_b > d$ or $p_s < 0$ 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 $0 \leq \tilde{v}_s < \tilde{v}_b \leq d$, so that it will earn strictly positive profit with at 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.1). In case (b), a profitable deviation is given by (A.2), (A.3), and (A.4), respectively, for $0 < p_b = p_s < \frac{d}{2}$, $p_b = p_s = \frac{d}{2}$ and $\frac{d}{2} < p_b = p_s < d$.

**Step 3:** $(p_b, p_s)$ has a convex support on $\left[\frac{d}{2}, d\right] \times \left[0, \frac{d}{2}\right]$.

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

(ii) Given \((p^1_b, p^1_s; p^2_b, p^2_s)\) in the possible support of the symmetric mixed-strategy equilibria

\[
R = \left[ \frac{d}{2}, d \right] \times \left[ 0, \frac{d}{2} \right],
\]

the buyer segmentation is given by Lemma 4, as shown in Figure 5.

If we standardize all terms above by dividing by \( d \), e.g. \( p^i = \frac{p^i}{d} \) \((i = 1,2)\), then we establish a one-to-one mapping between region \( R \) and a half-unit square \( \bar{R} = \left[ \frac{1}{2}, 1 \right] \times \left[ 0, \frac{1}{2} \right] \).

Define \( P(p^2_b) \equiv \frac{p^1_b(1-p^1_b)}{1-p^2_b}, \Omega(p^2_s) \equiv 1 - \frac{p^1_b(1-p^1_b)}{1-p^2_s} \), which corresponds to \( \rho(p^2_b) \) and \( \omega(p^2_s) \) in non-standardized variables, respectively. Areas \( s_1 \) to \( s_6 \), after standardization, correspond to

\[
S_1 = \left\{ (p^2_b, p^2_s) \left| \frac{1}{2} \leq p^2_b \leq p^1_b, p^1_s \leq p^2_s \leq \frac{1}{2} \right. \right\},
\]

\[
S_2 = \left\{ (p^2_b, p^2_s) \left| p^1_b \leq p^2_b \leq \Omega(p^2_s), p^1_s \leq p^2_s \leq \frac{1}{2} \right. \right\},
\]

\[
S_3 = \left\{ (p^2_b, p^2_s) \left| p^1_b \leq p^2_b \leq 1, p^1_s \leq p^2_s \leq \min\{P(p^2_b), \frac{1}{2}\} \right. \right\},
\]

\[
S_4 = \left\{ (p^2_b, p^2_s) \left| p^1_b \leq p^2_b \leq 1, 0 \leq p^2_s \leq p^1_s \right. \right\},
\]

\[
S_5 = \left\{ (p^2_b, p^2_s) \left| \frac{1}{2} \leq p^2_b \leq p^1_b, 0 \leq p^2_s \leq P(p^2_b) \right. \right\},
\]

\[
S_6 = \left\{ (p^2_b, p^2_s) \left| \frac{1}{2} \leq p^2_b \leq p^1_b, P(p^2_b) \leq p^2_s \leq p^1_s \right. \right\}
\]

59
And the buyer segmentation in $\tilde{R}$ is shown in Figure A-1.

![Figure A-1. Buyer segmentation under two exchanges after standardization.](image)

This figure shows the buyer segmentation given $(P_b^1, P_s^1) \in \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$ when $(P_b^2, P_s^2)$ can vary in the square $\left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right]$. Corresponding to areas in Lemma 4, in area $S_1$ and $S_6$, no buyer chooses exchange 1, and all buyers choose exchange 2; in area $S_2$, buyers with relatively low valuations choose exchange 1, while those with relatively high valuations choose exchange 2; in area $S_3$ and $S_4$, all buyers choose exchange 1, and no buyer would choose exchange 2; in area $S_5$, buyers with relatively high valuations choose exchange 1, while those with relatively low valuations choose exchange 2.

Recall that, with standardization,

$$V_j = \frac{v_j}{d}, (j = b, s)$$

Now $V_b \sim U\left[\frac{1}{2}, 1\right], V_s \sim U\left[0, \frac{1}{2}\right]$, an

$$\Phi = \frac{\varphi}{d} = \frac{P_b^1 \cdot P_s^1 - P_b^2 \cdot P_s^2}{P_s^1 - P_s^2}.$$  

Denote the support of the symmetric mixed-strategy equilibrium in $\tilde{R}$ as $M$.

When exchange 2 plays a mixed strategy $(P_b^2, P_s^2)$ with a distribution function $F(x, y)$ over $M$, exchange 1’s expected profit when choosing $(P_b^1, P_s^1)$ is
\[\pi(p^1_b, p^1_s) = (p^1_b - p^1_s) \cdot \Pr(v_s \leq p^1_s) \cdot \left[ \iint_{S_2 \cap M} \Pr(P^1_b \leq V_b \leq \Phi) \, dF(P^2_b, P^2_s) \right] + \iint_{(S_3 \cup S_4) \cap M} \Pr(P^1_b \leq V_b \leq 1) \, dF(P^2_b, P^2_s) + \iint_{S_5 \cap M} \Pr(\Phi \leq V_b \leq 1) \, dF(P^2_b, P^2_s)\]

where \(\Pi(P^2_b, P^2_s) \equiv (P^1_b - P^1_s) \cdot P^1_s \cdot \left[ \iint_{S_2 \cap M} (\Phi - P^1_b) \, dF(P^2_b, P^2_s) + \iint_{(S_3 \cup S_4) \cap M} (1 - P^1_b) \, dF(P^2_b, P^2_s) + \iint_{S_5 \cap M} (1 - \Phi) \, dF(P^2_b, P^2_s) \right].\)

Clearly, \(\Pi(P^1_b, P^1_s)\) is independent of \(d\), and part (ii) follows.

(iii) From (ii), focus without loss of generality on \(\Pi(P^2_b, P^2_s)\) and \(\bar{R}\).

First, we show that, the support of the symmetric mixed-strategy equilibria cannot be a two-dimensional area.

(By contradiction) Suppose \(M\) is an area. From (i), we know that \(M\) must be convex.

However, for any possible convex area, we can always find two points A and N such that,

- at A, the curve \(P^2_s = P(P^2_b)\) is tangent and above \(M\). So according to Figure A-1, all buyers go to exchange 1, and thus \(II > 0\) at A.

- at N, the curve \(P^2_s = P(P^2_b)\) is tangent and below \(M\). So, according to Figure A-1, no buyer goes to exchange 1, and thus \(II = 0\) at A.

So it is impossible to have equal profits at these two points. Thus, \(M\) cannot be an area, as illustrated in panel (a) of Figure A-2. In other words, \(M\) must be a one-dimensional curve.
Figure A-2. The support of the mixed-strategy equilibrium cannot be (a) a two-dimensional area, or (b) any downward-sloping curves. Panel (a) shows that for any two-dimensional area, we always can find two points A and N that are such that, at point A all buyers will choose exchange 1, while at point N no buyer chooses exchange 1, so M cannot be a support of the mixed-strategy equilibrium. Panel (b) shows a similar logic for any downward-sloping curve.

By the same argument, we can show that $M$ curve cannot have any downward-sloping part, as illustrated in panel (b) of Figure A-2. Hence, part (iii) follows.

Proof of Corollary 2: For simplicity of notation, we here denote exchange 1’s effective buy and sell prices as $(a, b)$, and exchange 2’s as $(x, y)$. We take a guess and verify the approach. We guess that $y = x - \frac{1}{6}$ is indeed in equilibrium with distribution $F(x)$ on $[L, U]$, where $[L, U] \subset \left[\frac{1}{2}, \frac{7}{12}\right]$.

The borderline $y = P(x) = \frac{(1-a)b}{1-x} = \frac{(1-a)(a/6)}{1-x}$. $P'(a) = \frac{a-1}{1-a} \geq 1$, if $a \geq \frac{7}{12}$. Thus, for $a \in [L, U] \subset \left[\frac{1}{2}, \frac{7}{12}\right]$, the crossing of $y = P(x)$ is shown in Figure A-3 below.
Figure A-3. One set of the mixed-strategy equilibrium. This figure shows the support of one set of mixed-strategy equilibria \([L, U]\) (in Green) on a straight line \(y = x - \frac{1}{6}\).

Clearly, only \(S_2\) and \(S_5\) are possible.

Note that \(\Phi(x, y) = \frac{ab - xy}{b - y} = a + x - \frac{1}{6}\). Thus, exchange 1’s expected profit is

\[
\Pi(a) = (a - b) \cdot b \left[ \int_a^U (x - \frac{1}{6}) \cdot dF(x) + \int_U^a (1 - a - x + \frac{1}{6}) \cdot dF(x) \right]
\]

\[
= \frac{1}{6} \cdot (a - \frac{1}{6}) \left[ \int_a^U x \cdot dF(x) + \int_L^a x \cdot dF(x) + \left(\frac{4}{3} - a\right) \cdot F(a) - \frac{1}{6} \right]
\]

\[
\Pi'(a) = \int_a^U x \cdot dF(x) - \int_L^a x \cdot dF(x) + \left(\frac{3}{2} - 2a\right) \cdot F(a) - \frac{1}{6} + (a - \frac{1}{6})(\frac{4}{3} - 3a) \cdot f(a)
\]

\[
\Pi''(a) = \left(a - \frac{1}{6}\right) \cdot \left(\frac{4}{3} - 3a\right) \cdot F''(a) + 10 \cdot \left(\frac{1}{3} - a\right) \cdot F'(a) - 2F(a)
\]

Mixed-strategy equilibria requires \(\Pi'(a) = 0\) for any \(a \in [L, U]\), which implies
\[ \Pi''(a) = 0 \text{ for any } a \in [L, U]. \]

Solving it yields (25).

To pin down the equilibrium, we need 4 parameters \((C_1, C_2, L, U)\) that satisfy (26)~(29), where (26) is essentially \( \Pi'(a) = 0 \) for any \( a \in [L, U] \), (27)~(29) are requirements for \( F(x) \) to be a legitimate distribution function.

Because we have 4 parameters and essentially 3 equations to be satisfied, there are a set of parameters, and thus a set of equilibria. ■
Figure 1: Structure of the U.S. Stock Markets and Fee Structure in October 2010
Figure 2: Total Fee of Thirteen Competing Exchanges in the U.S.
Figure 3. Effective Buy and Sell Prices under Fee Structure $(f_m, f_t)$. This figure shows that under fee structure $(f_m, f_t)$, 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 4. Buyer Surpluses under Two Exchanges. This figure shows the buyer’s surpluses when choosing exchanges. Without loss of generality, we assume $p_s^1 < p_s^2$. The buyer will choose the exchange that offers her a higher buyer surplus. Thus, the buyer’s choice is depicted as the upper envelope (in Red) of the two surplus curves. Note that buyers’ valuations have an upper bound $d$. 

(a) $p_b^2 \leq p_b^1$
EX 2 Only

(b) $p_b^1 > p_b^2, \phi < d$
EX 1 and 2 Split

(c) $p_b^1 > p_b^2, \phi \geq d$
EX 1 Only
Figure 5. Buyer Segmentation under Two Exchanges. This figure shows the buyer segmentation given $(p^1_b, p^1_s) \in \left[ \frac{d}{2}, d \right] \times \left[ 0, \frac{d}{2} \right]$ when $(p^2_b, p^2_s)$ can vary in the square $\left[ \frac{d}{2}, d \right] \times \left[ 0, \frac{d}{2} \right]$. As demonstrated in Lemma 4, in area $s_1$ and $s_6$, no buyer chooses exchange 1, and all buyers choose exchange 2; in area $s_2$, buyers with relative low valuations choose exchange 1, while those with relatively high valuations choose exchange 2; in area $s_3$ and $s_4$, all buyers choose exchange 1, and no buyer would choose exchange 2; in area $s_5$, buyers with relatively high valuations choose exchange 1, while those with relatively low valuations choose exchange 2.
Figure 6. Buyer Segmentation under Price Discrimination when One Monopoly Operator Runs Two Exchanges. This figure shows the optimal effective buy and sell prices offered by two exchanges run by one monopoly operator, and the corresponding buyer segmentation.

Execution probability under EX1

Execution probability under EX2

Buyers go to EX1

Buyers go to EX2
Figure 7: Two Types of Deviations from Bertrand Equilibrium. This figure demonstrates two types of profitable deviations for exchange 2 when both exchange 1 and 2 start from a zero total fee. Panel (a) demonstrates the deviation when exchange 1 set \( p^1_b = p^1_s = \frac{1}{2} \cdot d \). The deviation for exchange 2 in panel (a) involves decreasing the effective sell price to \( p^2_b = \frac{1}{2} d - \mu \varepsilon \) and effective buy price to \( p^2_s = \frac{1}{2} d - \varepsilon \), which attracts low valuation buyer to exchange 2. Panel (b) considers the case when exchange 1 set \( p^1_b = p^1_s < \frac{1}{2} \cdot d \). The deviation of exchange 2 involves increase the effective buy price to \( p^2_b = p^1_b + \varepsilon \) and sell price to \( p^2_s = p^1_s + \mu \varepsilon \), which attracts high valuation buyer. \( \varepsilon > 0 \) and \( 0 < \mu < 1 \) in both panels.
Table 1. Summary Statistics

This table reports summary statistics for the two samples used in the empirical test. Panel A presents the summary statistics for the same 117-stock sample as the NASDAQ HFT dataset from January 2010 through November 2011. Panel B provides the summary statistics for the leveraged ETF sample used in the difference-in-differences test, in which the split/reverse split event happens between January 2010 and November 2011. Herfindahl Index is the Herfindahl Index of the security, used as a measure of market fragmentation. tick\textsubscript{relative} is the reciprocal of price. logmcap stands for the log value of market capitalization. logbv\textsubscript{average} is the logarithm of the average book value of equity per shareholder at the end of the previous year (December 2009). idiorisk is a measure of idiosyncratic risk pertaining to the security, calculated as 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. PIN is the probability of informed trading. Return stands for the daily return on the security.

<table>
<thead>
<tr>
<th>Panel A. NASDAQ HFT Sample</th>
<th>Mean</th>
<th>Median</th>
<th>Std.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herfindahl Index</td>
<td>0.317</td>
<td>0.304</td>
<td>0.089</td>
<td>0.123</td>
<td>0.921</td>
</tr>
<tr>
<td>tick\textsubscript{relative}</td>
<td>0.048</td>
<td>0.033</td>
<td>0.039</td>
<td>0.002</td>
<td>0.364</td>
</tr>
<tr>
<td>logmcap</td>
<td>22.02</td>
<td>21.45</td>
<td>1.90</td>
<td>18.69</td>
<td>26.70</td>
</tr>
<tr>
<td>logbv\textsubscript{average}</td>
<td>13.19</td>
<td>13.22</td>
<td>2.15</td>
<td>4.94</td>
<td>18.11</td>
</tr>
<tr>
<td>idiorisk</td>
<td>0.014</td>
<td>0.008</td>
<td>0.020</td>
<td>0.001</td>
<td>0.149</td>
</tr>
<tr>
<td>age (in 1k days)</td>
<td>9.77</td>
<td>7.56</td>
<td>7.86</td>
<td>0.68</td>
<td>31.38</td>
</tr>
<tr>
<td>numAnalyst</td>
<td>14.08</td>
<td>11.00</td>
<td>10.51</td>
<td>1.00</td>
<td>54.00</td>
</tr>
<tr>
<td>PIN</td>
<td>0.118</td>
<td>0.112</td>
<td>0.052</td>
<td>0.021</td>
<td>0.275</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Leveraged ETF Sample</th>
<th>Mean</th>
<th>Median</th>
<th>Std.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herfindahl Index</td>
<td>0.281</td>
<td>0.270</td>
<td>0.076</td>
<td>0.163</td>
<td>0.844</td>
</tr>
<tr>
<td>Log(HerfindahlIndex)</td>
<td>-1.300</td>
<td>-1.310</td>
<td>0.244</td>
<td>-1.812</td>
<td>-0.170</td>
</tr>
<tr>
<td>return</td>
<td>-0.000</td>
<td>-0.001</td>
<td>0.041</td>
<td>-0.245</td>
<td>0.249</td>
</tr>
</tbody>
</table>
Table 2. The Impact of the Relative Tick Size on Market Fragmentation

This table presents the results of the regression of market fragmentation on the relative tick size. The regression uses the same 117-stock sample as the NASDAQ HFT dataset from January 2010 through November 2011. The regression specification is:

\[ \text{Herfindahl Index}_{i,t} = u_{j,t} + \beta \times \text{tick}_{relative,i,t} + \Gamma \times X_{i,t} + \epsilon_{i,t} \]

where \( MktFragmentation_{i,t} \) is measured using Herfindahl Index for stock \( i \) on date \( t \). \( \text{tick}_{relative,i,t} \) is the inverse of the stock price for stock \( i \) on day \( t \). \( u_{j,t} \) represents industry-by-time fixed effects. The definitions of the control variables \( X_{i,t} \) are presented in Table 1. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>Herfindahl Index</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>( \text{tick}_{relative} )</td>
<td>-0.383***</td>
<td>-0.385***</td>
<td>-0.312**</td>
<td>-0.362***</td>
<td>-0.289**</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.11)</td>
<td>(0.12)</td>
<td>(0.11)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>( \text{logmcap} )</td>
<td>-0.023***</td>
<td>-0.023***</td>
<td>-0.020***</td>
<td>-0.021***</td>
<td>-0.017***</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>( \text{logbvaverage} )</td>
<td>-0.000</td>
<td></td>
<td>-0.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>( \text{idiorisk} )</td>
<td>-0.210</td>
<td></td>
<td>-0.232</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td></td>
<td>(0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{age} )</td>
<td>-0.002***</td>
<td></td>
<td>-0.002***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
<td></td>
<td>(0.00)</td>
</tr>
<tr>
<td>( \text{numAnalyst} )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>( \text{PIN} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.139</td>
<td>0.145</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.250</td>
<td>0.250</td>
<td>0.264</td>
<td>0.253</td>
<td>0.270</td>
</tr>
<tr>
<td>( N )</td>
<td>51950</td>
<td>51950</td>
<td>51950</td>
<td>51950</td>
<td>51950</td>
</tr>
</tbody>
</table>
Table 3. Difference-in-differences Test Using Leveraged ETF Splits (Reverse Splits)

This table presents the results of difference-in-differences tests using leveraged ETF split (and reverse split), in which the event window is 5 days before splits/reverse splits from January 2010 through November 2011. The regression specification is:

\[ \text{Herfindahl Index}_{i,t,j} = u_{i,t} + \gamma_{i,j} + \rho \times D_{trt_{i,t,j}} + \theta \times \text{return}_{i,t,j} + \epsilon_{i,t,j} \]

where \( i \) indexes the underlying index, \( j \) indexes ETF and \( t \) indexes time. The dependent variable in is the Herfindahl Index. \( u_{i,t} \) is the index-by-time fixed effects and \( \gamma_{i,j} \) is the ETF fixed effects for ETF \( j \) of index \( i \). 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 splits/reverse splits. \( \text{Return}_{i,t,j} \) is the return on ETF \( j \) of index \( i \) on day \( t \). ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>Herfindahl Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Split</td>
</tr>
<tr>
<td>( D_{trt} )</td>
<td>-0.047**</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>( \text{Return} )</td>
<td>-0.180</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
</tr>
<tr>
<td>( \text{Constant} )</td>
<td>0.272***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.734</td>
</tr>
<tr>
<td>( N )</td>
<td>100</td>
</tr>
</tbody>
</table>