Trading in Fragmented Markets*

Markus Baldauf† Joshua Mollner‡

November 16, 2017

Abstract

This paper studies fragmentation of equity trading using a model of imperfect competition among exchanges. The addition of an exchange has theoretically ambiguous consequences for market quality. Increased competition places downward pressure on trading fees. However, additional arbitrage opportunities arise in fragmented markets, which intensify adverse selection. These opposing forces imply that the effects of fragmentation are context-dependent. To empirically investigate the ambiguity in a single context, we estimate key parameters of the model with order-level data pertaining to an Australian security. At the estimates, the benefits of increased competition are outweighed by the costs of multi-venue arbitrage. Compared to the prevailing duopoly, we predict that the counterfactual monopoly spread monopoly would be 23 percent lower.

JEL classification: G18, D43, D47

Keywords: exchange competition, fragmentation, high-frequency trading, bid-ask spread

*We are indebted to our advisors Timothy Bresnahan, Gabriel Carroll, Jonathan Levin, Paul Milgrom, and Monika Piazzesi. We would also like to thank Lanier Benkard, Philip Bond, Jean-Edouard Colliard (discussant), Carole Comerton-Forde, Alan Crawford, Darrell Duffie, Liran Einav, Lorenzo Garlappi, Duncan Gilchrist, Terrence Hendershott, Christiaan Hogendorn, Katya Malinova (discussant), Albert Menkveld, Peter Reiss, Alvin Roth, Georgios Skoulakis, Laura Tuttle, Brian Weller, Mao Ye, and seminar participants, as well as various industry experts for valuable comments. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. We acknowledge financial support by the Kohlhagen Fellowship Fund and the Kapnick Fellowship Program through grants to the Stanford Institute for Economic Policy Research.

†Sauder School of Business, University of British Columbia. E-mail: baldauf@mail.ubc.ca.

‡Kellogg School of Management, Northwestern University. E-mail: joshua.mollner@kellogg.northwestern.edu.

1 Introduction

Equity markets have become increasingly fragmented over the past ten to fifteen years. During that period, a surge of new exchanges and alternative trading systems entered and began to attract trading volume that had previously been concentrated on primary venues. Illustrative is the decline in the NYSE's share of the trading volume of NYSE-listed stocks, from 82 percent in June 2004 to 24 percent in October 2017 (Angel, Harris and Spatt, 2015; CBOE, 2017).

This proliferation of trading venues and the accompanying dispersion of trades were actively encouraged by the Securities and Exchange Commission, arguing that "vigorous competition among markets promotes more efficient and innovative trading services" (SEC, 2005, Reg NMS). The intuition that investors benefit from competition should resonate with any economist. However, in public equity markets there may be drawback to spreading out trade across markets: fragmentation may create opportunities for fast traders to engage in high-frequency arbitrage across venues.

We capture this tradeoff in a tractable and estimable model of limit order book trading with imperfect competition among exchanges, which extends the high-frequency trading model of Budish, Cramton and Shim (2015). Traders are either investors with an intrinsic motive to buy or sell, or high-frequency traders who trade for profit in two ways. High-frequency traders either provide liquidity, by offering to intermediate between investors, or they engage in arbitrage, by exploiting differences between quoted prices and the fundamental security value. With more exchanges, competition forces them to charge lower fees in order to attract investors. However, traders who provide liquidity face greater adverse selection, because news about fundamentals enables an arbitrageur to trade against more active quotes in aggregate, with a given amount of investors against whom to offset these losses. We show that depending on factors such as the strength of the private transaction motives of investors, their arrival rate to the market, and their elasticities of substitution among different exchanges, the introduction of new exchanges can either increase or decrease the transaction costs faced by investors.

We then employ the Australian market for an empirical application of the model. Using data from the first half of 2014 to estimate the parameters of our model, we find that investors are worse off under the prevailing duopoly than they would be under a monopoly exchange.

We find support for this conclusion from a natural experiment in which a technical issue forced one of the two Australian exchanges to shut down for a day.

In section 3 we develop a model of exchange competition, and we analyze its equilibrium in section 4. The baseline model features a single security whose shares are traded in limit order books on multiple exchanges. The fundamental value of the security is public information and follows a random walk. There are three categories of strategic decision makers: exchanges, high-frequency traders, and investors. Exchanges operate trading platforms and earn profits from trading fees. High-frequency traders may trade for profit by speculating or by facilitating transactions with other traders. Investors arrive stochastically with private trading motives and are differentiated along two dimensions. First, they differ in the strength of their private motive to transact. Second, they differ in their propensity to substitute among venues for a given price difference. In particular, not every investor trades at the exchange that offers the best price, which may be thought of as a reduced form representation of some market friction.[1]

We use this model primarily to explore the effect of fragmentation on transaction costs. The relevant measure of transaction costs in this paper is the cum-fee spread: the quoted bid-ask spread plus twice the take fee (the take fee is levied by the exchange on the party that initiates the trade). In practice, this quantity is a significant component of the transaction costs of equity trading, and in the model it is a sufficient statistic for welfare. Henceforth, we use "spread" as shorthand for the cum-fee spread unless otherwise specified.

Two forces give rise to a spread in this model: (*i*) the market power of exchanges, and (*ii*) adverse selection stemming from a race to react to information. Regarding the second force, although information is publicly observable, adverse selection arises from a liquidity provider's inability to cancel mispriced quotes before they are exploited by arbitrageurs, as in Budish, Cramton and Shim (2015) and Aït-Sahalia and Sağlam (2017a,b). A change in the number of venues affects the magnitude of each of these two forces, and there are consequently two opposing channels through which an additional exchange affects the equilibrium spread. First, an increase in the number of exchanges reduces the spread through the "competition channel." Intuitively, exchanges have less market power when there are more of them.

[1]For direct evidence that corroborates this modeling choice, see section 5.4. In the data, 35.0 percent of trades that occur when exchanges offer different prices take place at the exchange offering the worse quote, which indicates the presence of significant market frictions.

Consequently, each charges a lower trading fee, which results in a lower spread, other things being equal. Second, an increase in the number of exchanges raises the spread through the "exposure channel." Because it is difficult to forecast the venue at which the next investor will seek to trade, the amount demanded by an investor (in the model, one indivisible share) must be offered at the bid and the ask at each exchange in equilibrium. With more exchanges present, the aggregate book is therefore deeper.[2] More aggregate depth, in turn, implies that for any given change in fundamentals there are more mispriced quotes and thus, more arbitrage opportunities. This creates more adverse selection for liquidity providers, who in turn demand a higher spread, other things being equal. Theory is silent on whether lower spreads will prevail under a monopoly or an oligopoly, since either the competition channel or the exposure channel can dominate. While forces of this nature have been identified by previous work, our contribution is to provide a coherent and empirically suitable model encompassing both.

This theoretical ambiguity is consistent with the diversity of findings that have been reached by an old and extensive empirical literature, of which some papers report a positive association between fragmentation and liquidity while others report the reverse. Virtually all of those studies leverage a variation in market structure to determine the effects of fragmentation in a given market at a given point in time. However, a different approach is needed to ascertain the effects of fragmentation in other settings for which identifying variation is unavailable, and to that end, our approach may be a viable alternative. To illustrate, we utilize an Australian security for an empirical application of the model.

In section 5, we discuss the Australian market and our data. This setting is a natural fit for the model because it is particularly simple and self-contained. In particular, there are just two formal exchanges active in Australia: the Australian Securities Exchange (ASX) and Chi-X Australia (Chi-X). Nevertheless, the Australian market is modern and similar to the American market in most other respects. For the analysis we use order-level data pertaining to the Australian exchange-traded fund SPDR S&P/ASX 200 FUND (STW). The sample covers trading on both ASX and Chi-X, and it comprises 80 trading days from the first half

[2] Fragmentation also leads to greater aggregate depth in the models of Dennert (1993) and van Kervel (2015), as the result of similar forces. Moreover, that aggregate depth is increasing in the number of trading venues is also a stylized fact that has been documented in the empirical literature (Boehmer and Boehmer, 2003; Fink, Fink and Weston, 2006; Foucault and Menkveld, 2008; He, Jarnecic and Liu, 2015; Aitken, Chen and Foley, 2016).

of 2014.

In section 6, we estimate the parameters of the model on the Australian data. The parameters are identified via the response of order flow to variation in prices at ASX and Chi-X, together with the average level of the spread. We then use the estimated model to conduct counterfactual analysis, by comparing the currently observed duopoly outcome to what would prevail under a monopoly. We find that the counterfactual monopoly spread would be 23 percent lower than the duopoly spread of 2.88 cents. In other words, the exposure channel dominates the competition channel in the case of STW. Although these findings pertain only to the trading of one security in Australia, they do emphasize that unbridled competition among trading platforms is not always desirable.

Finally, in section 7, we conduct a separate, out-of-sample analysis of a natural experiment in which Chi-X experienced a technical issue and halted its trading for the remainder of the day, leaving ASX as the only exchange in operation. We find that the STW spread on ASX is significantly lower on the day of the Chi-X shutdown than on the surrounding days, as well as relative to an unaffected control group. This effect is consistent with the exposure channel postulated by the model, and its magnitude is in line with the estimates. Thus, this event provides additional support for the model's prediction that STW spreads would be lower under a monopoly than the prevailing duopoly. Moreover, the model's success in forecasting the result of this natural experiment provides us with some confidence that the model may be a useful methodological tool for predicting the effects of fragmentation.

2 Related Literature

Theory. This paper adds to the theoretical literature on competition among trading platforms in financial markets. In early contributions to this literature, multiple markets are typically modeled as operating in isolation without cross-venue arbitrage. In such settings, network externalities are important: fragmentation can harm liquidity by making it less likely that buyers and sellers find each other. These network externalities are at the heart of several early models, including Cohen, Maier, Schwartz and Whitcomb (1982), Economides and Siow (1988), Pagano (1989), as well as the clearing house model of Mendelson (1987). However, these network externalities are less relevant to modern trading, in which exchanges are electronically linked and information flows quickly from one venue to another. For example, in the United States, Reg NMS stipulates that exchanges must publish their quotes to a consolidated feed.

An early model of electronically-linked limit order book markets is that of Glosten (1994). He demonstrates that in an idealized, frictionless setting in which investors can costlessly send simultaneous orders to separate exchanges in order to complete their trades at the best possible price, the liquidity of the aggregate market is invariant to the degree of fragmentation. In our model, in contrast, investors do not always act to complete their trades at the best possible price so that Glosten's invariance result does not apply. Rather, liquidity does depend on the amount of fragmentation in our model in two ways: (*i*) the competition channel, via which fragmentation leads to more intense competition on trading fees, thereby improving liquidity; and (*ii*) the exposure channel, via which fragmentation increases the number of arbitrage opportunities, thereby amplifying adverse selection and harming liquidity. In consequence, our paper is connected to two branches of the literature on the relationship between liquidity and fragmentation.

First, our paper is related to models that have focused on how fragmentation can improve liquidity by enhancing competition. Some models focus on competition among market makers via quotes. Examples include Dennert (1993), Bernhardt and Hughson (1997), Biais, Martimort and Rochet (2000), as well as the dealer model of Mendelson (1987). These papers tend to show that competition among market makers may be advantageous because it induces them to post tighter quotes. However, since modern markets permit virtually free entry into market making, we are more interested in studying competition among the trading venues themselves, while taking competitive market making as given. More directly connected, therefore, are the models of Colliard and Foucault (2012), Pagnotta and Philippon (2016), and Chao, Yao and Ye (2017), which focus on competition among exchanges. In line with the "competition channel" of our paper, they find that competition may be advantageous because it induces exchanges to lower their trading fees. Finally, some other models have identified ways in which competition among exchanges can change the nature of competition among market makers. For example, Foucault and Menkveld (2008) illustrate how the absence of cross-exchange time priority means that fragmentation can improve liquidity by inducing market makers to quote greater depth in the aggregate.

Second, our paper is closely related to models that have focused on how fragmentation can harm market quality by altering the opportunities available to informed traders. In

Chowdhry and Nanda (1991), when trade is fragmented across several batch auctions (à la Kyle, 1985) running in parallel, an informed trader can more readily camouflage himself among noise traders, and markets become less liquid in response. Similarly, in Dennert (1993), when more market makers actively quote, an informed trader obtains more opportunities to trade. Although we model continuous time limit order books—rather than discrete time batch auctions—and focus on changes in the number of exchanges—rather than in the number of market makers—these forces are similar to the "exposure channel" of our paper. Our primary theoretical contribution lies in developing a framework that combines these two forces from the literature—the competition channel and the exposure channel—into a single tractable and estimable model.

Also connected are several other papers that, although they do not deal directly with fragmentation, focus on trading in limit order books in the presence of adverse selection from privately informed traders. Notable models include Copeland and Galai (1983), Glosten and Milgrom (1985), and Foucault (1999). More recently, Budish, Cramton and Shim (2015) and Aït-Sahalia and Sağlam (2017a,b) have demonstrated that similar forces arise in limit order books even when information is public. Our model is built upon the Budish, Cramton and Shim (2015) framework, but it also enriches that framework in several ways. For instance, exchanges in our model strategically set trading fees in competition for investors. Moreover, that competition may be imperfect, in which case the market power of exchanges constitutes a second source of transaction costs beyond the adverse selection modeled by Budish, Cramton and Shim (2015). Finally, we model the possibility that large spreads can crowd out some investors and some gains from trade.

Empirics. This paper is related to a rich empirical literature on fragmentation of financial markets. The effects of fragmentation have been studied using a diverse set of empirical strategies, including cross-section and panel regression, matched sample analysis, as well as studies of entry events, consolidation events, cross-listing events, and rule changes. Many of these studies find positive associations between fragmentation and liquidity (Branch and Freed, 1977; Hamilton, 1979; Neal, 1987; Cohen and Conroy, 1990; Battalio, 1997; Mayhew, 2002; Weston, 2002; Boehmer and Boehmer, 2003; De Fontnouvelle, Fishe and Harris, 2003; Fink, Fink and Weston, 2006; Nguyen, Van Ness and Van Ness, 2007; Foucault and Menkveld, 2008; Chlistalla and Lutat, 2011; O'Hara and Ye, 2011; Menkveld, 2013). Particularly relevant to our paper are He, Jarnecic and Liu (2015) and Aitken, Chen and Foley (2016), who also study the Australian market and observe that liquidity improves on average in conjunction with the 2011 entry of Chi-X Australia.

Many other studies find negative associations between fragmentation and liquidity (Bessembinder and Kaufman, 1997; Arnold, Hersch, Mulherin and Netter, 1999; Amihud, Lauterbach and Mendelson, 2003; Hendershott and Jones, 2005; Bennett and Wei, 2006; Gajewski and Gresse, 2007; Nielsson, 2009; Bernales, Riarte, Sagade, Valenzuela and Westheide, 2017). Additionally, some other studies find an inverted-U relationship, so that liquidity is maximized under moderate degrees of fragmentation (Boneva, Linton and Vogt, 2016; Degryse, de Jong and van Kervel, 2015). Likewise, Haslag and Ringgenberg (2016) also find a nuanced association: fragmentation benefits liquidity for large stocks, but harms it for small stocks.[3]

The diversity of these findings indicates that the consequences of fragmentation are highly context-dependent. We contribute to this literature by proposing a model to explain the role of context. Moreover, the model may be used as a tool for predicting the consequences of fragmentation in a given context.

3 Model

The model includes a single security, which is traded at one or more exchanges by two categories of traders: investors and high-frequency traders.

3.1 Trading Environment

Security. There is a single security whose fundamental value at time t is v_t, which is public information. Time begins at $t = 0$ and ends at $t = T$. During the interval $[0, T]$, v_t evolves as a compound Poisson jump process with arrival rate $\lambda_j \in \mathbb{R}_+$. Positive and negative jumps occur with equal probability and all have a size of $\sigma \in \mathbb{R}_+$.

For the main analysis, we assume that the evolution of v_t obeys the simple process described above. However, as we argue in appendix F.3, the results of our analysis would continue to hold if a diffusive component were added to the process or if there were a

[3] See appendix B.1 for a more detailed description of these members of the empirical literature on fragmentation.

degree of heterogeneity in the sizes of the discrete jumps. Furthermore, in appendix F.5 we demonstrate that the main conclusions of the model are robust to the addition of short-lived private information.

Limit order book. The limit order book is the mechanism that governs trading in the model. In what follows, we refer to four types of orders. A *limit order* consists of (*i*) a buy/sell designation, (*ii*) a limit price, where we allow for continuous prices, (*iii*) a quantity, measured in number of shares, and (*iv*) a time until when the order stays in force. Limit orders, unless otherwise specified, are assumed to be "good 'til cancelled." An *immediate-or-cancel* order is a limit order with a time in force of zero. A *market order* can be thought of as an immediate-or-cancel order with a limit price of positive or negative infinity. A *cancellation order* instructs the exchange to remove a previously placed limit order from the book.

The *quoted bid* is the highest price at which there exists an offer to buy. The *quoted ask* is the lowest price at which there exists an offer to sell. The *mid price* is the average of the bid and ask. The *quoted spread* is the difference between the bid and ask.

Orders are processed sequentially, in the order they are received. In the event that two orders are received simultaneously, ties are broken uniformly at random (i.e. by random latency). An incoming limit order is processed as follows. If it is a buy order specifying a price at or above the ask, then a trade occurs at the ask price. Likewise, if it is a sell order specifying a price at or below the bid, then a trade occurs at the bid price. In either case, the incoming order is referred to as the *aggressive* order, and the matching order is referred to as the *passive* order. In other cases, the incoming order does not trigger a trade, but is instead added to the book.

3.2 Decision Makers

There are two categories of traders: investors and high-frequency traders. In addition, exchanges are also strategic decision makers. All agents are risk-neutral and do not discount the future.

Exchanges. There are X exchanges, each of which is a platform on which shares of the security can be traded throughout the interval $[0, T]$. In the baseline model, each exchange operates a separate limit order book. Exchanges are horizontally differentiated. We model this by assuming that exchanges are metaphorically located at equally spaced points around a circle of unit length, as in Salop (1979).[4] The location of exchange x is denoted l_x. Exchange x sets make and take fees $\tau_{x,\text{make}}$ and $\tau_{x,\text{take}}$, which are collected, respectively, from the passive and aggressive parties of each trade that occurs on the exchange. Trading fees are chosen once and for all before trading commences at time zero.

For the main analysis, we assume that exchanges do not face a cost of operation. However, as we argue in appendix F.2, the results of our analysis would continue to hold unchanged in the presence of such a cost, provided it is sufficiently small.

Investors. Investors arrive at a Poisson rate $\lambda_i \in \mathbb{R}_+$ with a motive to trade one share of the security. An investor has a two-dimensional type $(\tilde{l}, \tilde{\theta})$. The first component, $\tilde{l}$, is drawn independently and identically distributed from $U[0,1]$ and denotes a position on the aforementioned circle. The second component, $\tilde{\theta}$, is drawn independently and identically distributed from $U[-\theta, \theta]$ and denotes a private value for trading a share of the security.

Let $b_{x,t}$ and $a_{x,t}$ denote, respectively, the prevailing cum-fee bid and cum-fee ask at exchange x at time t.[5] An investor who arrives at time t can act only at time t and is restricted to market orders.[6] By submitting a market order for $y \in \{-1, 0, 1\}$ shares to exchange $x \in \{1, \ldots, X\}$, such an investor obtains utility

$$u_t(y, x|\tilde{\theta}) = \begin{cases} v_t + \tilde{\theta} - a_{x,t} & \text{if } y = 1 \\ b_{x,t} - v_t - \tilde{\theta} & \text{if } y = -1 \\ 0 & \text{if } y = 0 \end{cases}$$

However, investors do not necessarily act to maximize their utility—although the model will

[4]An exchange's location on the unit circle represents the non-price characteristics that it offers. To the extent that modern stock exchanges are differentiated in practice, the primary mode of differentiation is not geographic location, but rather characteristics such as available order types, processing speed, etc. Consequently, we intend location to be interpreted in this metaphorical sense (as in, e.g. Hotelling, 1929). Note that we model neither the entry game of exchanges nor their location game. Rather, our approach is to solve for the equilibrium under a fixed number of equally spaced exchanges.

[5]That is, if $\hat{b}_{x,t}$ and $\hat{a}_{x,t}$ denote, respectively, the prevailing quoted bid and quoted ask at exchange x at time t, then $b_{x,t} = \hat{b}_{x,t} - \tau_{x,\text{take}}$ and $a_{x,t} = \hat{a}_{x,t} + \tau_{x,\text{take}}$.

[6]The restriction of investors to market orders prevents them from providing liquidity and is fairly standard in the literature, for example as in Glosten and Milgrom (1985) and Budish, Cramton and Shim (2015).

permit that case. Instead, an investor who arrives at time t chooses a quantity $y \in \{-1, 0, 1\}$ and an exchange $x \in \{1, \ldots, X\}$ to maximize

$$\hat{u}_t(y, x|\tilde{l}, \tilde{\theta}) = u_t(y, x|\tilde{\theta}) - 2\alpha \cdot d(\tilde{l}, l_x)^2,$$

where the function d yields the distance between two points on the unit circle. Formally, $d(l_1, l_2) = \min(|l_1 - l_2|, 1 - |l_1 - l_2|)$.[7] That investors do not always act to maximize their utility could be thought of as the result of an unmodeled market friction, such as difficulties associated with monitoring prices in real time or the agency problem between an investor and the broker who routes his orders to an exchange.[8] The parameter α governs the extent of this market friction. One possibility is $\alpha = 0$, in which case there is no friction and investors are utility maximizers. In the other extreme, as α grows large, investors become increasingly likely to choose the exchange closest to them on the unit circle, without regard for the terms of trade.[9]

Under the interpretation that the market friction is an investor/broker agency problem, the location parameter $\tilde{l}$ could be interpreted as a property of the broker, which influences his actions in such a way that they are not always in the investor's best interest. Several forces could give rise to such a conflict of interest in practice. For example, a broker might be tempted to route a market order to an exchange that offers an inferior price if by doing so he could collect a rebate from that exchange.[10] This explanation is less applicable to this paper's eventual empirical application in Australia, where there are no such rebates. More germane to that setting are the following possibilities: (i) a broker might not possess a

[7] Our framework bears a similarity to models with travel costs, in which a quadratic functional form is a standard choice (d'Aspremont, Gabszewicz and Thisse, 1979; Economides, 1989). The primary difference is that the "travel cost" entering $\hat{u}$ applies even if the investor does not ultimately trade (i.e. chooses $y = 0$). This feature of our model would be slightly unnatural if we were to interpret the "travel costs" literally, as a feature of the underlying preferences of investors (*cf.* footnote 9). However, it is not incompatible with our preferred interpretation of the "travel cost," which is as reduced form for a market friction.

[8] In appendix F.6, we explain in more detail how an imperfect ability to monitor prices in real time might lead to investor behavior consistent with maximizing an objective with the functional form of $\hat{u}$.

[9] An alternative interpretation of the model, although not our preferred one, is that in which the utility function of an investor is represented by $\hat{u}$ itself, rather than by u. Under that interpretation, exchanges are horizontally differentiated from the perspective of an investor, who may have an intrinsic preference for trading at one exchange over another.

[10] Indeed, Battalio, Corwin and Jennings (2016)—although focusing on limit orders—document empirical evidence of brokers deviating from their best execution obligation to focus instead on collecting rebates.

smart order router that automatically checks prices at all exchanges, in which case checking prices at all exchanges might require costly effort;[11] (*ii*) a broker might have an ownership stake in a certain exchange, which could introduce a financial incentive to route orders there; or (*iii*) a broker might have a non-pecuniary preference for trading at a certain exchange, perhaps as a result of physical location, order types offered, the speed and infrastructure of the exchange, or relationship capital. These latter explanations appear especially plausible in the context of Australia where the rules describing a broker's best execution duty are not cast strictly in terms of routing to the best price, as is the case in the United States.[12]

For the main analysis, we assume that investors do not split their orders across exchanges, with each investor choosing a single exchange to maximize $\hat{u}$ as described above. However, in appendix F.4 we demonstrate that the equilibrium we study remains intact even if order splitting is allowed.

High-frequency traders.[13] There is an infinite number of high-frequency traders.[14] Each is a speculator whose utility is determined solely by its trading profits. The action space of a high-frequency trader at any time t includes whether to submit any limit orders or cancellations.

For the main analysis, we assume that high-frequency traders face neither inventory constraints nor costs of operation. However, as we argue in appendix F.1, the results of our analysis would continue to hold unchanged in the face of inventory constraints, provided that they do not bind at sufficiently low levels. Moreover, as we argue in appendix F.2, the results of our analysis would continue to hold unchanged even if there were a cost to monitoring active limit orders, provided that cost is sufficiently small.

[11] In a similar spirit, certain traders in the model of Foucault and Menkveld (2008) check prices only at the incumbent exchange—in other words, it is infinitely costly for them to check prices at the entrant exchange.

[12] Details regarding Australia's Market Integrity Rules, the ownership structures of the two Australian exchanges, and the order types they offer are provided in appendix B.2.

[13] We label these speculators "high-frequency traders" because their equilibrium behavior in the model resembles strategies often associated with modern high-frequency trading. Nevertheless, this label does also connote a number of other characteristics and strategies, not all of which are embodied in the model.

[14] In practice, the number of high-frequency traders is quite large, and the Australian market is a representative case. According to a study by the Australian Securities and Investments Commission, there were 550 separate high-frequency traders active in that market in 2012 (ASIC, 2013).

3.3 Parametric Assumptions

This section describes the restrictions on the parameter space that are required for the subsequent analysis. These restrictions guarantee, among other things, the existence of an equilibrium in which the market does not break down due to adverse selection.

Assumption 1, below, requires the arrival rate of jumps in the value of the security, λ_j, to be sufficiently small relative to the arrival rate of investors, λ_i. Thus, the role of this assumption is to ensure that episodes of adverse selection are sufficiently infrequent.

Assumption 1. $X\lambda_j \leq \lambda_i$

Assumption 2, below, requires that the size of a jump in the value of the security exceeds each investor's private value for trading a share. The implication of this restriction is that any increase in the spread will crowd out more liquidity-based trades from investors than information-based trades from high-frequency traders.[15]

Assumption 2. $\sigma > \theta$.

The final assumption is the following.

Assumption 3. $\lambda_i \left(1 - \frac{1}{\theta}\frac{\Sigma}{2}\right)\frac{\Sigma}{2} \geq \lambda_j X \left(\sigma - \frac{\Sigma}{2}\right)$, where Σ is defined in terms of the underlying parameters as follows:

$$\Sigma \equiv \begin{cases} \theta\left(1 + \dfrac{\lambda_j}{\lambda_i}\right) & \text{if } X = 1 \\ \theta + \dfrac{4\alpha}{X^2} - \sqrt{\theta^2 + \dfrac{16\alpha^2}{X^4} - \dfrac{8\alpha\theta\lambda_j}{X\lambda_i}} & \text{if } X \geq 2 \end{cases}$$

Assumption 3 requires that if Σ is the cum-fee spread prevailing at all exchanges, then the resulting payments by investors would exceed the profits of high-frequency traders.[16] This restriction is needed to guarantee the existence of a symmetric equilibrium in pure strategies. To understand this assumption, a helpful analogy is to oligopolistic price competition in an industry with fixed costs. In that analogous setting, a symmetric equilibrium in pure

[15] Similar assumptions also appear in Glosten (1994) (his Assumption 2) and Biais, Martimort and Rochet (2000) (their assumption that $v'(\theta) \geq 0$). Likewise, Back and Baruch (2013) argue that assumptions of this nature are required to guarantee the existence of an equilibrium in the model they study.

[16] As lemma 1(*i*) establishes, assumption 1 implies $\Sigma \in \mathbb{R}$, so that this is a well-defined inequality.

strategies can exist only if (*i*) the fixed cost is not too large, (*ii*) price competition is not too intense, and (*iii*) the market is not too small. In the current setting, adverse selection plays the role of fixed costs. The parameters σ, λ_j, and X determine the magnitude of adverse selection, and the assumption prevents them from being too large. Likewise, α determines the intensity of competition among exchanges, and the assumption prevents it from being too small. Finally, λ_i and θ determine the available gains from trade, and the assumption prevents them from being too small.

4 Equilibrium

In this section, we describe the subgame perfect Nash equilibrium of the model. We then discuss how equilibrium outcomes depend on the parameters of the model.

Throughout, the focus of the analysis is the *cum-fee spread*: the spread plus twice the take fee. This is the appropriate quantity because it measures the transaction costs borne by investors. Moreover, because gains from trade are derived only from the private values of investors, the cum-fee spread also serves as a sufficient statistic for the welfare implications of the model. Specifically, an increase in the spread affects welfare in two ways: (*i*) gains from trade are reduced since some marginal investors may cease to trade, and (*ii*) there are transfers away from the inframarginal investors who continue to trade.

4.1 Equilibrium Description

Recall that the timing of events in the model is as follows. First, each exchange sets its make and take fees. Then, after observing these fees, high-frequency traders and investors submit orders. We refer to the latter sets of events as the "trading subgames," and we solve the model by backward induction: first characterizing equilibrium behavior in each trading subgame, then taking those outcomes as given to identify equilibrium fee choices. We provide an intuitive description of equilibrium strategies in the paragraphs that follow, and we defer to the proofs of propositions 1 and 2 in appendix A for a complete treatment.

We begin by discussing the case of a monopoly exchange, and we extend the analysis to the oligopoly case later in the section. In equilibrium, high-frequency traders sort into two roles as in Budish, Cramton and Shim (2015). One high-frequency trader plays the role of "liquidity provider" and establishes quotes of one share at the bid and one share at the ask, which are maintained so that the mid price tracks the value of the security. The remaining high-frequency traders play the role of "stale-quote sniper," attempting to trade whenever a jump in the value of the security generates a mispricing in the quotes of the liquidity provider. In addition, these stale-quote snipers enforce a zero-profit condition for the liquidity provider, as in Bertrand competition over the spread.[17]

After each jump, the high-frequency traders race to react: the liquidity provider to cancel her mispriced quotes, and the stale-quote snipers to exploit them. Each race results in a tie, which is broken uniformly at random (i.e. by random latency). Since an infinite number of high-frequency traders assume the sniper role in equilibrium, the liquidity provider loses each race. Thus, one cost in the liquidity provider's zero-profit condition is due to the stale-quote snipers. A second cost to the liquidity provider is the make fee set by the exchange. On the other hand, the liquidity provider's revenue is derived from the investors. Investors, upon arrival, decide whether to buy, sell, or hold. Investors hold if the magnitude of their private value for trading does not exceed half of the cum-fee spread, and they otherwise trade, buying or selling in correspondence with their types. From every investor who trades, the liquidity provider earns the half spread as a commission.

This zero-profit condition determines the spread, and since the spread influences an investor's decision to trade, it also affects trading volumes. Inducting backwards, the monopolist exchange sets make and take fees to maximize its expected revenue. The key tradeoff is that while higher fees generate more revenue per trade, they are also passed on as larger spreads, which crowd out some investor trades and reduce volume.

Proposition 1 asserts existence of an equilibrium for the case of a monopolist exchange and also characterizes the equilibrium spread. In appendix F we show that this result persists in more general settings, as does the subsequent result for the oligopoly case.

Proposition 1 (monopoly)**.** *With a single exchange ($X = 1$), under assumptions 1, 2, and 3, there exists a subgame perfect Nash equilibrium with cum-fee spread*

$$s^* = \theta \left(1 + \frac{\lambda_j}{\lambda_i} \right).$$

[17] Although excessive profits may have accrued in the early days of high-frequency trading, they were short-lived (The Wall Street Journal, 2017). Moreover, much of the rest of the literature takes a similar approach (e.g. Glosten and Milgrom, 1985; Kyle, 1985; Budish, Cramton and Shim, 2015).

The expression for the spread in proposition 1 illustrates the two sources of the spread in this model: market power of the exchange and adverse selection. First, liquidity providers use the spread to offset the costs of being adversely selected. As a result, the spread is a function of the relative arrival rates of information and investors, λ_j/λ_i, which governs the degree of adverse selection. Second, a monopolist exchange takes into account that fewer investors will trade at a higher price. Thus, the spread is also a function of θ, which governs the price elasticity of demand. Indeed, absent adverse selection (i.e. if $\lambda_j = 0$), the pricing equation reduces to the classic Lerner condition, which equates the markup over marginal cost to the inverse of the demand elasticity.[18]

This analysis can also be extended to the case of multiple exchanges. The oligopoly case is similar to the monopoly case, but with two straightforward modifications. First, exchanges set their trading fees in a simultaneous move game. Second, investors not only choose whether to trade, but also choose where to trade. As before, high-frequency traders sort into two roles in equilibrium. One high-frequency trader per exchange plays the role of liquidity provider, maintaining quotes of one share at the bid and one share at the ask, while the remainder play the role of stale-quote sniper. This behavior is optimal for the liquidity providers because of the structure of investor demand: each investor demands a single share and attempts to fill their entire demand by trading at a single exchange. Thus, quoting less than one share would mean foregoing profitable trades with investors. On the other hand, quoting more than one share would expose liquidity providers to additional adverse selection losses without offsetting benefits from increasing the investor trading volume.

While a monopolist exchange is constrained only by the own-price elasticity of investors (the probability that raising trading fees would cause an investor to refuse to trade), an oligopolist exchange must also consider cross-price elasticities (the probability that raising trading fees would cause an investor to switch to a different exchange). Taking these tradeoffs into account, exchanges set their trading fees in a simultaneous move game. We focus our analysis on the symmetric equilibrium of this game.

Proposition 2 asserts existence of a symmetric equilibrium for the oligopoly case.

Proposition 2 (oligopoly)**.** *With multiple exchanges ($X \geq 2$), under assumptions 1, 2 and*

[18]The parameter α, which governs the strength of market frictions affecting the exchange choices of investors, does not enter the monopoly spread for the reason that with a single exchange, there are no exchange choices to be made.

3, there exists a subgame perfect Nash equilibrium with cum-fee spread

$$s^* = \theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}.$$

As in the monopoly case, the oligopoly spread is influenced by θ and λ_j/λ_i. Yet in addition there is now also a role for X, the number of exchanges, as well as α, which governs the strength of market frictions affecting the exchange choices of investors. One reason for the presence of these two parameters is that they interact with the prevailing spreads to determine the number of investors who choose to trade at a given exchange: (*i*) if X increases, then each exchange obtains a smaller share of investor trades at equal spreads, and (*ii*) if α increases, then each exchange's share of investor trades becomes less responsive to its spread. As a result, these parameters affect the tradeoffs exchanges face when they set trading fees, and in turn, the equilibrium spread as well.

For the remainder of this section we focus on the mathematical derivation of the expressions for the monopoly and oligopoly spreads via backward induction from the zero-profit condition of the liquidity provider. The following discussion focuses on the the ask side of the book. The version for the bid side is analogous. At any instant, one of two things may affect the ask-side profits of a liquidity provider: an investor may arrive with a desire to buy, or the value of the security may jump upward. We begin by considering investor arrivals. Investors with a desire to buy arrive at the rate $\lambda_i/2$. In the case of an oligopoly, when the cum-fee ask is a_x on an exchange x and a_{-x} on the other exchanges, exchange x is the preferred exchange of such an investor with probability $\max\left[0, \min\left(1, \frac{X}{2\alpha}(a_{-x} - a_x) + \frac{1}{X}\right)\right]$. In the case of a monopoly, the monopolist exchange is always the preferred exchange. Conditional on exchange x being the chosen exchange of such an investor, that investor trades with probability $\max\left(0, 1 - \frac{a_x - v}{\theta}\right)$. Thus, investors buy at exchange x at the rate $\frac{\lambda_i}{2}\max\left(0, 1 - \frac{a_x - v}{\theta}\right)$ in the case of a monopoly and at the rate

$$\frac{\lambda_i}{2}\max\left[0, \min\left(1, \frac{X}{2\alpha}(a_{-x} - a_x) + \frac{1}{X}\right)\right]\max\left(0, 1 - \frac{a_x - v}{\theta}\right) \tag{1}$$

in the case of an oligopoly. From each of these trades, the liquidity provider earns $a_x - v$, and must pay the make fee $\tau_{x,\text{make}}$ to the exchange. Next, we consider jumps in the value of the security. Upward jumps arrive at the rate $\lambda_j/2$. Conditional on such a jump occurring, the liquidity provider at exchange x will lose $\sigma + v - a_x$ to a stale-quote sniper and must also pay the make fee $\tau_{x,\text{make}}$ to the exchange.[19] Combining all this, the zero-profit condition that determines the liquidity provider's ask is then

$$\frac{\lambda_i}{2} \max\left(0, 1 - \frac{a_x - v}{\theta}\right)(a_x - v - \tau_{x,\text{make}}) = \frac{\lambda_j}{2}(\sigma + v - a_x + \tau_{x,\text{make}})$$

in the case of a monopoly, and is

$$\begin{aligned}\frac{\lambda_i}{2} \max\left[0, \min\left(1, \frac{X}{2\alpha}(a_{-x} - a_x) + \frac{1}{X}\right)\right] \max\left(0, 1 - \frac{a_x - v}{\theta}\right)(a_x - v - \tau_{x,\text{make}}) \\ = \frac{\lambda_j}{2}(\sigma + v - a_x + \tau_{x,\text{make}})\end{aligned}$$

in the case of an oligopoly. Conditional on exchange x setting the make and take fees $\tau_{x,\text{make}}$ and $\tau_{x,\text{take}}$, the liquidity provider on exchange x quotes so that the ask a_x satisfies the appropriate zero-profit condition. That is, if a_x is the zero-profit cum-fee ask, then the liquidity provider sets the quoted ask $\hat{a}_x = a_x - \tau_{x,\text{take}}$.

Taking the above behavior of liquidity providers as given, each exchange x sets fees to maximize its profits, which are the product of $\tau_{x,\text{make}} + \tau_{x,\text{take}}$ with the volume traded on its platform. The resulting equilibrium spread is as described in proposition 1 for the case of a monopoly. For the case of an oligopoly, we focus on symmetric equilibria, in which the same total fee is set by each exchange, and the resulting equilibrium spread is as described in proposition 2.

4.2 The Effect of Fragmentation

A key insight formalized by this model is that it is theoretically ambiguous how the spread (and hence welfare) depends upon the number of exchanges. The ambiguity is caused by two opposing channels. On one hand, the addition of another exchange may reduce the spread through the "competition channel." Intuitively, exchanges have less market power when

[19] In the model, the liquidity provider races against an infinite number of other high-frequency traders, and therefore loses the race with probability one. Of course, there are not an infinite number of high-frequency traders in practice, but only a large finite number of them. If, in the model, there were N high-frequency traders, then the liquidity provider would lose the race with probability $\frac{N-1}{N} \approx 1$, and the liquidity provider's zero-profit condition would be approximately the same.

they have more competitors and must therefore reduce their trading fees to retain investors. All else equal, lower fees induce a lower spread. However, the addition of another exchange may raise the spread through the "exposure channel." Intuitively, more shares are quoted in aggregate. Therefore, whenever the fundamental value of the security moves away from the current posted prices, more shares are exposed to that mispricing, which creates larger losses for liquidity providers. All else equal, the spread rises as liquidity providers quote wider to compensate.

The ambiguity is illustrated by two limiting cases of the model. First, consider the limiting case as α diverges to infinity, which is to say that investors do not condition their choice of exchange on the price difference. In that case, the expression for the oligopoly spread converges to $s^* = \theta\left(1 + X\lambda_j/\lambda_i\right)$. Thus, every additional exchange raises the spread by $\theta\lambda_j/\lambda_i$. The reason is that if investors do not respond to prices, then multiple exchanges are a collection of isolated monopolists. Yet additional exchanges provide snipers with more opportunities to trade on a given piece of information, which increases adverse selection. Intuitively, the competition channel is shut down, so that the exposure channel dominates.

Second, consider the case in which $\lambda_j = 0$, which is to say that the fundamental security value is constant. In that case, the monopoly spread is θ, which exceeds the duopoly spread of $\theta + \alpha - \sqrt{\theta^2 + \alpha^2}$, and the spread decreases still further as X increases beyond two. The reason is that adverse selection does not increase with the number of exchanges, as with a constant security value there is no adverse selection. Yet additional exchanges intensify price competition, resulting in smaller trading fees and hence smaller spreads. Intuitively, the exposure channel is shut down, so that the competition channel dominates.

The primary focus of our empirical application is upon the comparison between monopoly and duopoly. For that comparison, we obtain the following corollary of propositions 1 and 2.

Corollary 1. *Under assumptions 1, 2, and 3,* $s^*_{monopoly} \geq s^*_{duopoly}$ *if and only if*

$$\frac{\alpha}{\theta} \leq \min\left(\frac{\lambda_j}{\lambda_i}, \frac{\lambda_i}{2\lambda_j} - \frac{\lambda_j}{2\lambda_i}\right).$$

An implication of corollary 1 is that a given shock to fragmentation might produce differing implications in the cross-section, leading to larger spreads for one subset of securities and smaller spreads for another subset (as in Haslag and Ringgenberg, 2016). To extend this insight one step further, the model suggests that there may be scope for security-specific policies on fragmentation rather than a one-size-fits-all approach.

4.3 Comparative Statics

In this section we use the characterizations of the equilibrium spread from propositions 1 and 2 to study how it varies with respect to the remaining parameters. There exist monotone comparative statics for three of those primitives: α, the strength of market frictions that distort the exchange choice of investors, λ_i, the arrival rate of investors, and λ_j, the arrival rate of information. These results are standard and in line with what intuition would dictate. There is, however, no monotone comparative static with respect to θ, which determines the strength of the private transaction motive of investors.

Proposition 3 (comparative statics)**.** *Within the set of parameters that satisfy assumptions 1, 2, and 3, the equilibrium spread is*

(i) weakly increasing in α,

(ii) weakly decreasing in λ_i, and

(iii) weakly increasing in λ_j.

The parameter α determines the strength of the market frictions that distort the exchange choice of investors. It therefore controls the magnitude of the "competition channel." When α is large, market frictions are high, which mutes competition on trading fees and results in a large spread. On the other hand, when α is small, fee competition is strong and the spread is small.

The arrival rates of investors and of information, λ_i and λ_j, affect transaction costs by controlling the amount of adverse selection that liquidity providers face. If investors arrive more frequently (an increase in λ_i), then the liquidity provider faces less adverse selection, since she trades with relatively more investors and relatively fewer stale-quote snipers. She therefore demands a smaller spread. The reverse is true if trades based on changes in the fundamental value occur more frequently (an increase in λ_j).

5 Empirical Application

We employ the Australian market for an empirical application of the model.[20] This section lays the groundwork for that exercise. We begin by discussing the relevant details of the Australian market, then describe our datasets, and finally define the variables that we construct from the data and eventually use for estimation.

5.1 Industry Background

Two formal exchanges are currently active in Australia: the Australian Securities Exchange (ASX) and Chi-X Australia (Chi-X). Over the 80 trading days in our main sample, the average daily value of trades in the Australian cash market (i.e. equity, warrant and interest-rate market transactions) amounted to $4.8 billion (ASX, 2014), or roughly fifty times smaller than the United States market at the time.

The baseline make and take fees charged by ASX are each 0.15 basis points of the value of the trade (ASX, 2016). Fees exceeding the baseline are applied to certain advanced order types. For Chi-X, the make and take fees are, respectively, 0.06 basis points and 0.12 basis points of the value of the trade (Chi-X, 2011).

The Australian market is a natural fit for the model because it is particularly simple and self-contained. In particular, there are just two formal exchanges, with independent ownership. Moreover, there are limited overlaps with foreign markets with respect to trading hours and securities. In contrast, at the beginning of our main sample, there were thirteen formal exchanges in the United States, possessing overlapping ownership structures, and there is also substantial overlap with European and Canadian markets. Furthermore, off-exchange trading is more prevalent in the United States than it is in Australia (largely due to a minimum price improvement rule in Australia, as well as a prohibition against payment for order flow).

Despite these differences, Australia is similar to the United States in many respects. In particular, many of the same large trading firms are active in Australia as in the United States. In addition, the technical protocol that is used by Australian exchanges is owned by

[20]Other recent studies of the Australian market include the aforementioned He, Jarnecic and Liu (2015) and Aitken, Chen and Foley (2016). In addition, Foley and Putniņš (2016) and Comerton-Forde and Putniņš (2015) also study Australia, although they focus primarily upon dark trading.

Nasdaq OMX Group and is effectively the same as that used on Nasdaq.

Additional details pertaining to the Australian market are discussed in appendix B.2.

5.2 Data

We have order-level data from both ASX and Chi-X. In each case, the data are a historical record of the messages that are broadcast by the exchange, which market participants can access in real time for a fee. These messages are sufficient to construct the lit book at each exchange at any point in time and also to identify all trades that take place in the lit book. Appendix B.3 documents details of the order-level data and the specific steps required to process it.

Our analysis focuses on a single security: the exchange-traded fund SPDR S&P/ASX 200 FUND (STW), which tracks the Australian market index S&P/ASX 200.[21] This security is a natural object of focus for three reasons. First, it is an extremely significant security, not only because it is Australia's largest ETF but also because it tracks the benchmark index for the Australian market. Second, STW's broad exposure parallels the model's assumption that information is purely public. While certain traders may be quite likely to possess private information about an individual stock, private information is a less significant feature of the trading of a broad composite such as STW. Third, STW's relatively large average spread parallels the model's assumption that the price space is continuous. For many other thickly-traded securities, the bid-ask spread is often constrained by the minimum tick size (typically one cent), in which case the discreteness of prices is highly significant. However, discrete prices are less salient for securities whose spread typically exceeds one tick, as is the case for STW. Fourth, STW is not internationally cross-listed. To the extent that foreign and domestic fragmentation generate different effects, international cross-listings would complicate the interpretation of our findings.

The ASX data cover the trading days in the months of February through June 2014. The Chi-X data cover only February through May. From the sample, we drop two days that were affected by data issues.[22] Our main analysis requires data from both exchanges, and

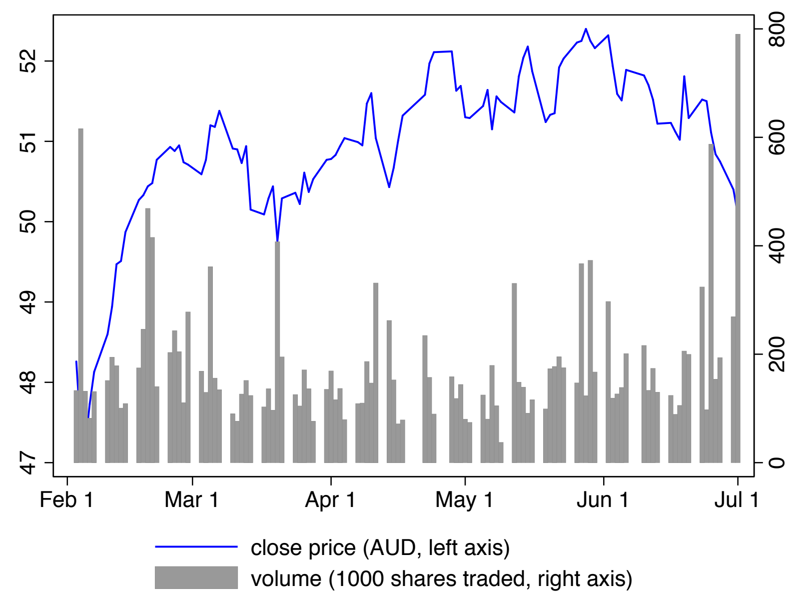
[21]As of June 2014, STW had AUD 2.3 billion under management, with 45 millions units on offer. The fund consisted of 205 constituents, with a weighted average market capitalization of AUD 54 million (State Street, 2014).

[22]One of those days is February 11, on which there were known issues with the ASX data feed (Chi-X,

therefore uses data on the 80 remaining days in February through May. A supplemental, out-of-sample analysis in section 7 uses ASX data on the 20 trading days in the month of June.

Figure 1 plots the close price and traded volume of STW for each trading day in the months of February through June of 2014. Aside from a rally over the first part of February, the price remains fairly stable in the neighborhood of $51, and there are no drastic fluctuations. Traded volumes are somewhat more volatile.

Figure 1: Price and traded volume (STW, 2014)



Price and traded volume of SPDR S&P/ASX 200 FUND (STW) for all trading days between Feb 3, 2014 and June 30, 2014. Price is the close price as announced by ASX, measured in Australian dollars. Volume includes all trading in the Australian market, measured in thousands of contracts. Data are from Bloomberg.

Table 1 displays summary statistics pertaining to the trading of STW. During our main sample, 169,930 contracts per day were traded on average across the Australian market. Our analysis focuses on lit book volume. Formally, we say that a trade takes place on the lit

2014a). The other day is May 2, for which our record of the ASX data feed is incomplete. Any missing orders render the entire remainder of the trading day unusable, which is due to the fact that processing order-level data requires us to replay the trading day from the beginning (for details, see appendix B.3).

book if it is recorded as an "E message" in the data feed of the exchange. This includes all on-exchange trades during the continuous session in which the passive order is visible, and typically also includes trades against the visible portion of iceberg orders. See the market data specification files ASX (2012b) and Chi-X (2012). Of total STW volume, 72.3 percent is traded in the lit book of ASX, and 16.5 percent in the lit book of Chi-X. The remaining 11.2 percent of traded volume includes (*i*) trading in the ASX opening and closing crosses, (*ii*) off-exchange trading (e.g. trading in crossing systems, block trades, and internalization), which must be reported to a trade reporting facility managed by either ASX or Chi-X, and (*iii*) dark on-exchange trading, including trades in which the passive order is hidden, as well as trades against the hidden portion of iceberg orders.

Table 1 also displays statistics on the messages in the ASX and Chi-X data feeds that pertain to STW. The two exchanges are quite comparable in terms of the total number of messages. On an average day, trading in STW generates 19,350 messages at ASX compared to 17,711 messages at Chi-X. However, owing to differences in the amount of volume traded, the ratio of total messages to trade messages tends to be much higher at Chi-X than at ASX. Finally, table 1 also displays statistics pertaining to the volatility of STW (computed using the standard deviation of one-second returns) and price movement (computed using the absolute value of the daily return).

Table 1: Summary statistics for trading of STW ($N = 80$ days)

	mean	st. dev.	quartile 1	median	quartile 3
traded volume (1,000 contracts)					
ASX lit book	122.90	84.92	72.06	96.11	143.22
ASX other	10.52	19.20	3.12	4.81	10.96
Chi-X lit book	28.00	20.85	15.66	22.32	33.19
Chi-X other	8.51	18.33	0.00	0.05	1.77
Total	169.93	100.41	109.26	139.58	178.87
number of messages					
ASX	19,350	5,010	16,000	19,053	21,537
Chi-X	17,711	6,175	13,040	16,400	19,962
message to trade					
ASX	66.97	27.21	49.12	62.01	77.33
Chi-X	320.77	209.47	181.48	268.42	376.94
volatility (bps)	0.32	0.06	0.28	0.31	0.35
price movement (bps)	32.13	26.84	11.86	26.24	41.90

The table reports summary statistics for trading of STW throughout our sample of 80 trading days. For each day and exchange, "lit book" volume is the volume accounted for by E messages in the data feed of the exchange, and "other" volume is calculated by subtracting that from the total volume as obtained from Bloomberg. For each day and exchange, the number of messages is obtained from our record of the data feed of the exchange and the message to trade ratio is calculated by dividing the number of messages by the number of trade messages. Volatility is the standard deviation of the one-second returns of STW between 10:30 and 16:00, where returns are computed using the average of all four cum-fee quotes and are expressed in basis points. Price movement is the absolute value of the daily return of STW, computed using ASX open and close prices and expressed in basis points.

5.3 Variables Used in Estimation

For the estimation we discretize time into intervals of one second in length. For each second between 10:30 and 16:00 in each of the 80 trading days in our main sample, we construct variables pertaining to the prices that prevailed and the trades that took place.[23]

[23]On ASX, the continuous trading session for STW begins at a random point in the interval [10:08:45, 10:09:15] and ends at 16:00. On Chi-X, the continuous trading session begins at 10:00 and ends at 16:12. However, we limit attention to continuous trading between 10:30 and 16:00 in order to ensure a balanced panel and to avoid contamination from the opening and closing auctions at ASX.

Prices. Cum-fee bids, asks, and spreads constitute one set of variables used in the estimation. As an intermediate step, for each exchange $x \in \{\text{ASX}, \text{Chi-X}\}$ and each second t, we extract the quoted bid and ask that prevail on the lit order book at the beginning of the second, which we denote $\hat{b}_{x,t}$ and $\hat{a}_{x,t}$, respectively. The ASX take fee is 0.15 basis points of the value of the trade. Thus, for each second t, the ASX cum-fee bid is $b_{\text{ASX},t} = 0.999985\hat{b}_{\text{ASX},t}$, and the ASX cum-fee ask is $a_{\text{ASX},t} = 1.000015\hat{a}_{\text{ASX},t}$. Similarly, the Chi-X take fee is 0.12 basis points of the value of the trade. Thus, for each second t, the Chi-X cum-fee bid is $b_{\text{Chi-X},t} = 0.999988\hat{b}_{\text{Chi-X},t}$, and the Chi-X cum-fee ask is $a_{\text{Chi-X},t} = 1.000012\hat{a}_{\text{Chi-X},t}$. For each exchange $x \in \{\text{ASX}, \text{Chi-X}\}$ and each second t, we also use $s_{x,t} = a_{x,t} - b_{x,t}$ to denote the cum-fee spread.

Trades. A second set of variables used in the estimation pertains to different categories of trades. For our analysis, we consider only lit book trades, which as described in section 5.2 are those trades that give rise to an E message in the data feed of the exchange. As mentioned, this includes all on-exchange trades during the continuous session in which the passive order is visible, and typically also includes trades against the visible portion of iceberg orders. However, it excludes both off-exchange trading and dark on-exchange trading.

In the model, investor trades occur in isolation from other trades. In contrast, sniper trades occur in clusters: taking place on both exchanges simultaneously and in the same direction. Leveraging this distinction, we use isolated trades and clustered trades as empirical proxies for the investor and sniper trades of the model, respectively. Specifically, we classify a trade as isolated if no other trade in the same direction occurs within a certain cutoff on either exchange. In the baseline, we set this cutoff to one second.[24] For each exchange $x \in \{\text{ASX}, \text{Chi-X}\}$ and each second t, we define the indicators $buy_{x,t}$ and $sell_{x,t}$, which evaluate to unity if, respectively, an isolated trade that is an aggressive buy or an isolated trade that is an aggressive sell occurs on exchange x in second t. The remaining trades are classified as clustered. Finally, for each second t, we define the indicator $clustered_t$, which evaluates to unity if a clustered trade occurs in second t.[25]

[24] Note that it is possible for a single marketable limit order to lead to multiple E messages if it executes against multiple resting orders (one message for each match). For the purposes of this analysis, it is appropriate to treat all such executions as a single trade. This can be done because the corresponding execution messages appear in consecutive order with the same timestamp.

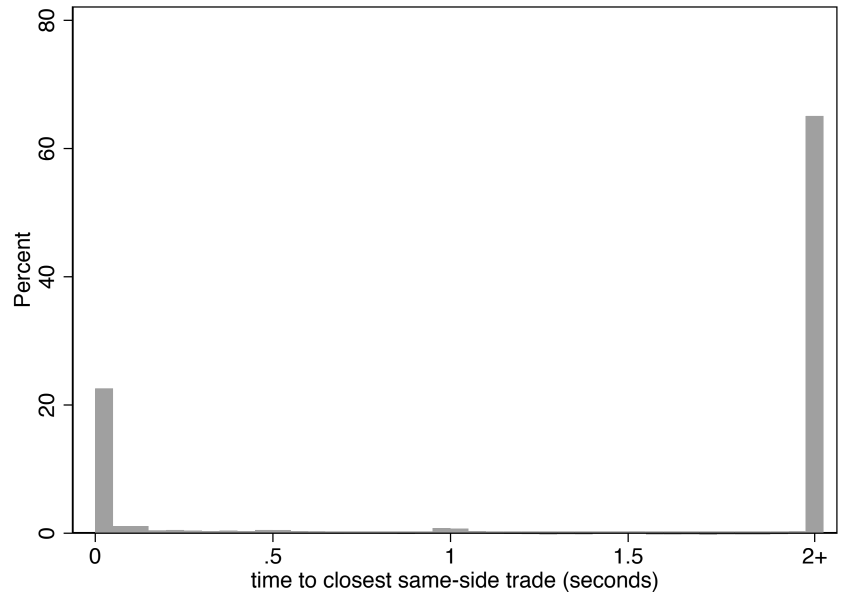
[25] van Kervel (2015) uses essentially the same classification scheme, proxying for the fraction of fast traders

A potential concern is classification error, which could arise if the distinction between isolated and clustered trades is less clean in practice than the sharp dichotomy predicted by the model. We alleviate this concern in two ways. First, in figure 2, we illustrate that the predicted dichotomy is, in fact, not far from the truth. For each trade, we compute the length of time to the nearest trade on the same side of the book on either exchange; the figure plots the empirical distribution of this variable. A large fraction of these trades are within 50 milliseconds of another trade in the same direction, and after that there is a long tail. Therefore, the extent of type I and II errors that stem from classification error is relatively limited because the mass between any two cutoff points is small compared to the entire distribution. In particular, for only 11.2 percent of trades is the nearest trade in the same direction between 0.1 and 2 seconds away. Second, in appendix D.1, we demonstrate that our results are robust to alternative choices of the cutoff used to classify trades as isolated or clustered.

Additionally, in appendix E.4, we demonstrate that clustered trades are better predictors of subsequent price movements than isolated trades. This evidence further supports the use of isolation as a proxy for the liquidity-motivated investor trades of the model, and conversely for using clustering as a proxy for the information-motivated sniper trades of the model.

with the percentage of market orders that occur at approximately the same time. One difference is that he uses 0.1 seconds as his cutoff, whereas we use 1 second as the cutoff in our baseline specification. Nevertheless, we demonstrate in appendix D.1 that using a 0.1 second cutoff leads to similar results.

Figure 2: Distribution of time gaps between trades



To construct this figure, we isolate all lit book trades (i.e. trades that give rise to an E message) of STW that take place on either ASX or Chi-X between 10:30 and 16:00 on one of the 80 trading days in the sample. For each trade, we compute the difference in timestamps to the nearest trade on the same side of the book on either exchange. The figure summarizes the distribution of those differences. The width of a bar is 50 milliseconds.

Summary statistics. Table 2 presents summary statistics for seven key variables used in the estimation: the cum-fee spreads at ASX and Chi-X, indicators for isolated buys and sells at ASX and Chi-X, and an indicator for clustered trades.[26] For each variable we report the mean and standard deviation over the sample, which comprises the seconds between 10:30 and 16:00 in each of the 80 trading days in the sample. The spread at ASX tends to be slightly lower and more volatile than its counterpart at Chi-X. Reflecting the fact that ASX accounts for most of the volume traded in Australia, isolated buys and sells are more frequent there than at Chi-X. Finally, trades that we classify as isolated are somewhat more frequent, in aggregate, than trades classified as clustered.

In addition, figure 3 plots the joint density of quoted spreads at ASX and Chi-X. Although

[26]We do not report summary statistics for the remaining four variables used in estimation (the cum-fee bids and asks at ASX and Chi-X), because those statistics have little economic meaning in and of themselves.

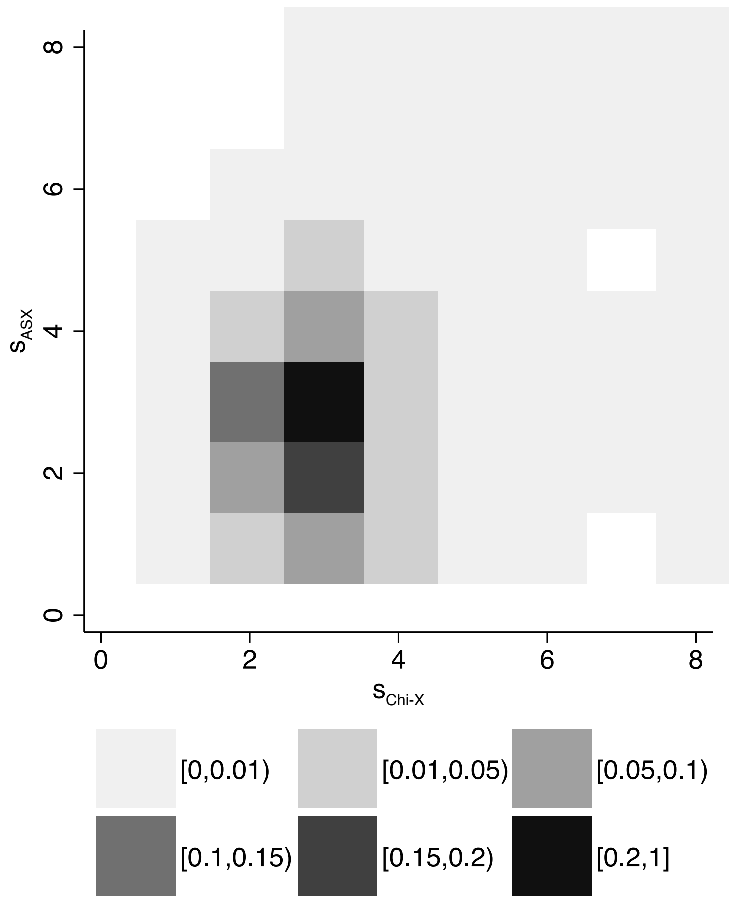
it is cum-fee spreads that are used in estimation, quoted spreads are more natural to plot because they lie on a grid. The unique mode is given by symmetric quoted spreads of three cents, which occurs 22.7 percent of the time. Quoted spreads are equal 37.1 percent of the time and within one cent 81.7 percent of the time.

Table 2: Summary statistics of variables used in estimation
($N = 1,584,000$ seconds)

	Mean	Std. Dev.
s_{ASX}	2.80865	1.02835
$s_{\text{Chi-X}}$	2.94904	0.74820
buy_{ASX}	0.00380	0.06151
$buy_{\text{Chi-X}}$	0.00097	0.03117
$sell_{\text{ASX}}$	0.00356	0.05955
$sell_{\text{Chi-X}}$	0.00100	0.03160
clustered	0.00200	0.04466

An observation is one second between 10:30 and 16:00 in one of the 80 trading days in the sample. For each exchange x, s_x refers to the cum-fee spread at exchange x; it is measured in AUD cents and is evaluated at the start of each second. For each exchange x, the indicators buy_x and $sell_x$ evaluate to unity for seconds during which, respectively, an isolated buy or an isolated sell occurs; a trade is classified as isolated if it is a lit book trade and if no other lit book trade in the same direction occurs on either exchange within one second before or after. The indicator *clustered* evaluates to unity for seconds during which a clustered trade occurs; a trade is classified as clustered if it is a lit book trade and if another lit book trade in the same direction occurs on either exchange within one second before or after.

Figure 3: Joint density of Chi-X and ASX quoted spreads



An observation is one second between 10:30 and 16:00 in one of the 80 trading days in the sample. The quoted spreads of Chi-X and ASX prevailing at the beginning of the second are represented on the horizontal and vertical axes, respectively, and are measured in AUD cents. The shading refers to the fraction of seconds in which a pair of quoted spreads $(s_{\text{Chi-X}}, s_{\text{ASX}})$ is observed in the sample.

5.4 Direct Evidence of Market Frictions

A key feature of the model is that it permits the existence of frictions affecting the exchange choices of investors. If these frictions are strong, then investors would often fail to trade at the exchange with the best prevailing price. Conversely, if investors often fail to trade at the best prevailing price, then these frictions must be strong. The latter statement provides a direct and model-free way to assess the strength of these frictions. In this section, we demonstrate that, for trades taking place when quotes differ at Chi-X and ASX (so that a single "best price" exists), a significant fraction of them occur at the exchange offering the worse price. This evidence indicates that these frictions are indeed strong.

To conduct this analysis, we focus on isolated trades, which are empirical proxies for the investor trades of the model.[27] For every isolated trade, we determine which exchange or exchanges features the best quoted price at the time (i.e. the lowest ask in the case of an isolated buy, or the highest bid in the case of an isolated sell), and we compare that to the exchange on which the trade actually occurs. The results of this analysis are tabulated in table 3. Motivated by order protection rules that prevail in the United States and other jurisdictions—although not in Australia—this analysis defines "best price" and "inferior price" with respect to the marginal price for the first traded share, or the price "at the top of the book." In appendix E.1, we conduct similar analyses using the total price of the entire order to measure execution quality, and we find similar results.

Table 3: Exchange choice for isolated trades, number of trades

	different quotes			identical quotes	total
	best price	inferior price	subtotal		
	(1)	(2)	(3)	(4)	(5)
BUY	1,949	1,131	3,080	4,466	7,546
SELL	2,089	1,040	3,129	4,092	7,221
BUY or SELL	4,038	2,171	6,209	8,558	14,767

The table classifies isolated trades that occur at ASX and at Chi-X by prevailing quotes at the two exchanges. The columns show the number of isolated trades in each of the following cases: (1) the trade occurs at the exchange with strictly better price; (2) the trade occurs at the exchange with strictly inferior price; and (4) the trade occurs when both exchanges quoted identical prices. Best price is defined as the lowest quoted ask in the case of an isolated buy, or the highest quoted bid in the case of an isolated sell.

[27] A second reason for focusing on isolated trades is the following. A trader who wishes to trade a large volume at one point in time might use an intermarket sweep order to access liquidity essentially simultaneously at multiple exchanges. In such cases, that investor might trade at an exchange featuring a worse price, and would be doing so for reasons unrelated to the frictions we posit. However, that investor would also trade at the exchange offering the better price at essentially the same time. Thus, our classification algorithm would label such a trade as clustered, and it would be omitted from the analysis in this section. In summary, given our focus on isolated trades, intermarket sweep orders and similar trading strategies cannot serve as a competing explanation for the frequency with which trades take place at inferior prices.

As the table indicates, a substantial number of isolated trades take place on exchanges offering inferior quotes. Aggregating across buys and sells, 6,209 isolated trades occur when one exchange offers a strictly worse quote. Of those, 35.0 percent trade through a better price and in so doing occur on the exchange offering the worse quote. The magnitude of this frequency suggests that frictions are an empirically important determinant of exchange choice, and it motivates our development of a model that allows for such frictions. Moreover, since the strength of these frictions is parametrized in the model by α, this evidence also indicates that we should expect to obtain a relatively high estimate of α when we estimate the model in the following section.

While this analysis points to a considerable friction in the exchange choice of investors, it is silent as to its micro-foundations. As discussed in section 3, there exist several potential sources.

6 Estimation and Counterfactual Analysis

In this section we first describe how we exploit variation in the data to identify and estimate key parameters of the model. We then discuss the estimates and use them to predict the counterfactual monopoly spread.

6.1 Empirical Strategy

Four parameters require estimation.[28] Three of these parameters govern the demand system of investors: λ_i, their arrival rate; θ, the strength of their private transaction motive; and α, the strength of market frictions that distort their choice of exchange. The fourth parameter is λ_j, the arrival rate of jumps in the value of the security, which affects the amount of adverse selection that a liquidity provider faces. We estimate these parameters with nonlinear least squares.

Identification. In what follows we argue that the parameters are uniquely identified by variation in the data. The identification argument relies on two parts. First, α, θ, and λ_i

[28]The fifth parameter, σ, we do not estimate because it is not needed to derive a prediction of the spread that would prevail in a counterfactual monopoly, or for any of the other counterfactuals that we consider.

are identified by how the arrival rates of isolated buys and sells at ASX and Chi-X fluctuate with variation in prices. Intuitively, α is identified by the cross-price elasticity, θ is identified by the own-price elasticity; and λ_i is identified by the frequency of isolated trades.[29] Second, given values for these three parameters, the duopoly spread is a monotone function of λ_j, where monotonicity follows from proposition 3. Thus, the second step of the argument is to use the level of the spread to identify λ_j.

Estimating equations. Equation (1) in section 4.1 states the arrival rate of investor buys as a function of the prevailing asks. In the duopoly case, its empirical analogue for each exchange $x \in \{\text{ASX}, \text{Chi-X}\}$ and each second t is

$$buy_{x,t} = \frac{\lambda_i}{2} \max\left[0, \min\left(1, \frac{1}{2} + \frac{a_{-x,t} - a_{x,t}}{\alpha}\right)\right] \max\left[0, 1 - \frac{a_{x,t} - v_t}{\theta}\right] + \varepsilon_{x,t}^{\text{buy}} \tag{2}$$

where $\varepsilon_{x,t}^{\text{buy}}$ is an error term whose mean, conditional on the quotes, is assumed to be zero. In the model, investor arrivals follow a Poisson process. The error term captures deviations of this random process from its mean, as well as any unmodeled determinants of trade. Likewise, we obtain the following equation for investor sells

$$sell_{x,t} = \frac{\lambda_i}{2} \max\left[0, \min\left(1, \frac{1}{2} + \frac{b_{x,t} - b_{-x,t}}{\alpha}\right)\right] \max\left[0, 1 - \frac{v_t - b_{x,t}}{\theta}\right] + \varepsilon_{x,t}^{\text{sell}} \tag{3}$$

where $\varepsilon_{x,t}^{\text{sell}}$ is an error term with zero conditional mean. Since we do not observe v_t, we proxy with the average mid price $(b_{\text{ASX},t} + b_{\text{Chi-X},t} + a_{\text{ASX},t} + a_{\text{Chi-X},t})/4$ in both (2) and (3).

Finally, proposition 2 provides an expression for the duopoly spread. Its empirical analogue for each exchange $x \in \{\text{ASX}, \text{Chi-X}\}$ and each second t is

$$s_{x,t} = \theta + \alpha - \sqrt{\theta^2 + \alpha^2 - \frac{4\alpha\theta\lambda_j}{\lambda_i}} + \varepsilon_{x,t}^{\text{spread}} \tag{4}$$

where $\varepsilon_{x,t}^{\text{spread}}$ is an error term with zero expectation. Note that the model does not suggest a reason for why the spread should vary around its mean. Nevertheless, spreads do vary in the data, yet are most of the time either 2 or 3 cents. There are a number of potential explanations for short-term deviations from this long run pricing equation. First, we have assumed

[29] We present reduced-form evidence on these elasticities in appendix E.3.

that the parameters of the model are constant over time. While this seems accurate as a first approximation, small deviations may arise in practice, which would give rise to different values of the spread. For example, days on which macroeconomic news announcements are expected would be associated with larger values of λ_j, and on such days liquidity providers might set a larger spread. Second, in the model it is assumed that all agents are risk-neutral. In practice, it is likely that liquidity providers are either risk averse or constrained in their ability to take on inventory, in which case they may condition their quotes on their net positions.

We assume further that the data generation process is stationary and weakly dependent.

Estimation procedure. We estimate the parameters of the model using systems nonlinear least squares on equations (2), (3), and (4). Specifically, we minimize the objective

$$Q_T(\alpha, \theta, \lambda_i, \lambda_j) = \frac{1}{T} \sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \left(\varepsilon_{x,t}^{\text{sell}}\right)^2 + \left(\varepsilon_{x,t}^{\text{spread}}\right)^2 \tag{5}$$

where t indexes all seconds between 10:30 and 16:00 in the 80 trading days in our main sample.

Given the aforementioned assumptions on the structure of the error terms, it can be shown that this procedure is consistent. The consistency result is formally stated and proven in appendix B.4. The estimation was implemented using SNOPT (Gill, Wong, Murray and Saunders, 2015). In addition, we compute standard errors using a non-overlapping block bootstrap procedure. Appendix B.5 contains further details on the implementation of estimation and the computation of standard errors.

Discussion of estimation approach. GMM is a popular class of estimators, which has been used to estimate other models of limit order book trading (e.g. Sandås, 2001; Biais, Bisiere and Spatt, 2010). Our estimation procedure is also encompassed by the GMM framework, with the sample moment conditions being the first order conditions of equation (5) with respect to the parameters. Establishing that uses the fact that the error terms in equations (2), (3), and (4) enter additively (Cameron and Trivedi, 2005, p. 168).

In addition, others have used MLE to estimate models of limit order book trading (e.g. Easley, Kiefer, O'Hara and Paperman, 1996; Easley, Engle, O'Hara and Wu, 2008). In principle, we could have done the same, but this would require parametrizing the error terms in equations (2), (3), and (4). Since our sample is large, it seems unlikely that the efficiency gains brought by MLE would be sufficient to justify additional distributional assumptions.

6.2 Parameter Estimates

In this section we report the results of our estimation procedure. Table 4 contains the parameter estimates. The point estimate of α of 7.34 cents implies that an exchange attracts every investor in the market only if it offers prices at least 3.67 cents better than its competitor. This considerably exceeds the average half-spread of 1.44 cents, which indicates that there are significant market frictions distorting the exchange choices of investors. The point estimate of θ of 1.53 cents means that the average magnitude of private transaction motives among investors is 53 percent of the average half-spread, which highlights that transaction costs crowd out a substantial number of potential investor trades. Finally, the estimates of λ_i and λ_j indicate that investor arrivals are 2.2 times more frequent than arrivals of information. All estimated parameters are highly significant based on bootstrapped standard errors. In particular, the null hypothesis of frictionless routing (i.e. $\alpha = 0$) is rejected at all common significance levels.

Table 4: Parameter estimates

parameter	point estimate	standard error
α	7.33504	1.06129
θ	1.52817	0.15127
λ_i	0.00172	0.00022
λ_j	0.00078	0.00012

The point estimates are computed using the nonlinear least squares procedure described in section 6.1. An observation is one second between 10:30 and 16:00 in one of the 80 trading days in the sample. The estimation was performed using SNOPT (Gill, Wong, Murray and Saunders, 2015). Standard errors are based on 200 block bootstrap replications.

To illustrate the elasticities implied by the model, note that if, in the model, the liquidity provider at each exchange quotes symmetrically about the value of the security, then the arrival rate of investors at an exchange $x \in \{\text{ASX}, \text{Chi-X}\}$ is

$$\lambda_i \max\left[0, \min\left(1, \frac{s_{-x} - s_x}{\alpha} + \frac{1}{2}\right)\right] \max\left(0, 1 - \frac{1}{\theta}\frac{s_x}{2}\right).$$

Evaluating at the equilibrium spread and the estimated parameters, the elasticity with respect to s_x is -17.1, while the elasticity with respect to s_{-x} is 0.8. The relatively high own-price elasticity is due to the spread falling on the highly elastic portion of the linear demand schedule, and the relatively low cross-price elasticity reflects the presence of considerable market frictions that govern exchange choice.

One test of the extent to which the estimated model is a good fit for the data is to see if it can match a moment in the data that was not leveraged in estimation. In the model, sniper trades arrive at the rate λ_j, which we estimate to be 0.00078 per second. Clustered trades constitute an empirical proxy for sniper trades, which from table 2 arrive at the rate 0.00200 per second. Thus, although our model does not perfectly match this moment, it does generate a prediction that is of the same order of magnitude. In addition, we also conduct a formal test of overidentifying restrictions in appendix D.2. Lastly, we establish that the estimates possess various forms of robustness elsewhere in appendix D.

6.3 The Effect of Fragmentation

In our model it is theoretically ambiguous whether competition benefits investors. The reason for this ambiguity is that is that the competition channel (i.e. additional exchanges induce greater price competition) and the exposure channel (i.e. additional exchanges lead to greater adverse selection against liquidity providers) act in opposite directions. Nevertheless, the estimates from the previous section can be used to resolve this theoretical ambiguity for the case of Australia and STW.

At the moment, the market structure in Australia is given by an exchange duopoly, in which each exchange operates a separate order book. The estimates can be used to consider the effect of a counterfactual reduction in the number of exchanges. In the first row of table 9, we compare the prevailing duopoly to the counterfactual of a monopoly. The average spread observed in the data is 2.88 cents. However, our estimates imply that in the counterfactual of a monopoly exchange, the average STW spread would be 2.22 cents, or 22.9 percent lower. Moreover, the monopoly spread is lower than the prevailing duopoly spread in each of the 200 bootstrap replications. Thus, our findings show that the competition channel is dominated by the exposure channel.

In principle, the estimates could also be used to consider the effects of counterfactual increases in the number of exchanges. According to the estimated model, it is not possible for three or more exchanges to operate simultaneously without market breakdown, because assumption 3 would be violated with $X \geq 3$. Intuitively, our estimates imply that the exposure channel is so strong that the Australian market cannot support more than two exchanges. However, one reason for interpreting this counterfactual prediction with caution is that exchanges in practice have revenue sources outside trading fees. Although such additional revenue streams may be strategically orthogonal to the trading fee decision, and hence to our previous conclusions, they would imply that assumption 3 might be more than is needed to ensure simultaneous profitability of all X exchanges.

7 Natural Experiment: Chi-X Shutdown

Our estimates imply that the spread of STW would be substantially lower if, instead of two exchanges, the Australian market were to feature only one. To provide additional support for this conclusion, we study a natural experiment in which Chi-X shut down for a day, leaving ASX as the only operating exchange. Because of the short and unanticipated nature of the shutdown, ASX did not strategically alter its fees in response. Thus, this episode constitutes an isolated test of the exposure channel of the model. Consistent with the predictions of the model, the STW spread is substantially smaller on the "monopoly day" than on the surrounding "duopoly days," as well as relative to an unaffected control group. And this remains the case even after we attempt to account for the competition channel with a back-of-the-envelope calculation.

The Chi-X shutdown has several advantages as a natural experiment with which to test the predictions of the model. First, the shutdown occurred within weeks of the end of our estimation sample. Consequently, the underlying parameters are more likely to resemble those prevailing during the estimation sample than if we had used a natural experiment several years prior or hence. Second, the shutdown was unanticipated, abrupt, and clean.[30]

[30] Another potential candidate for testing the model is the natural experiment generated by the Chi-X entry in 2011. However, we have been unable to obtain data from that year. More significantly, it is probable that

7.1 Event Description

On the morning of June 16, 2014, Chi-X experienced a technical issue which led them to halt their trading at 11:08 and remain closed for the rest of the trading day (Chi-X, 2014b). Trading resumed as usual the following morning. For the duration of the Chi-X shutdown, ASX was the only exchange in operation. The variation in market structure created by this episode gives rise to a natural experiment on the effects of fragmentation in the Australian market.

The technical issue arose because Chi-X made a small change to the way in which it handled certain incoming orders. That change was implemented with an error, which, when noticed, prompted Chi-X to shut down. Hence, the shutdown was exogenous to trading activity on that day. Indeed, June 14 (the "monopoly day") appears similar to the other trading days in June 2014 (which we term "duopoly days") in terms of volume, message flow, volatility, and price movement. Table 5 presents the results of these comparisons. These comparisons give no indication that the monopoly day was an extreme outlier in any respect. For each statistic, the monopoly day is within one standard deviation of the mean.[31]

the parameters governing Australian trading would have changed during the more than two years separating the entry from our estimation sample. Thus, if the entry episode led to findings that were at odds with our counterfactual analysis, it would be unclear whether to interpret that as a rejection of the model or as evidence that the parameters had changed.

[31] While not an extreme outlier, the monopoly day does feature relatively low volume. Nevertheless, this does not seem to be driving our findings. In appendix E.2, we show that our main results survive even if we add volume as a control variable or if we estimate only on low-volume days.

Table 5: Summary statistics for trading of STW, June 2014

	monopoly day	duopoly days (June 2014, $N = 19$ days)				
		mean	st. dev.	quartile 1	median	quartile 3
volume (1,000 contracts)						
ASX	121.65	160.18	97.10	90.33	145.32	198.16
Total	123.84	197.81	115.36	126.62	173.73	215.93
number of messages						
ASX	20,941	17,222	4,531	13,179	16,192	20,537
message to trade						
ASX	85.82	76.19	36.89	48.49	68.10	88.83
volatility (bps)	.298	.354	.162	.278	.332	.35
price movement (bps)	17.60	41.99	31.26	13.69	37.30	69.28

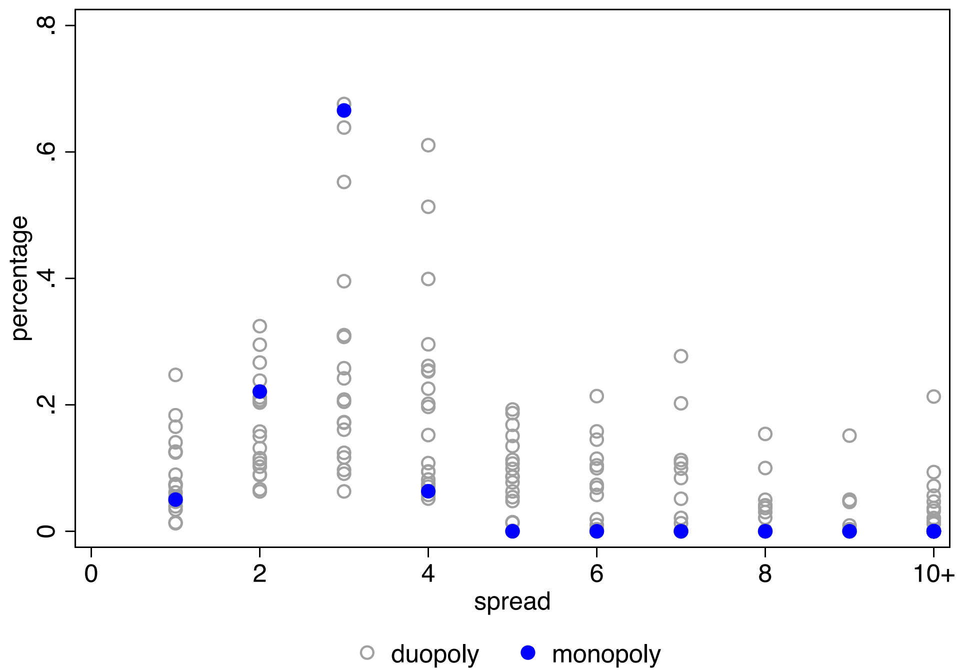
The sample comprises all 20 trading days in June 2014. Volume is the number of contracts of STW traded in Australia as obtained from Bloomberg. For each day, the number of messages is obtained from the ASX data feed, and the message to trade ratio is calculated by dividing the number of messages by the number of trade messages. Volatility is the standard deviation of the one-second returns of STW between 10:30 and 16:00, where returns are computed using the average of the ASX cum-fee quotes and are expressed in basis points. Price movement is the absolute value of the daily return of STW, computed using ASX open and close prices and expressed in basis points.

7.2 Analysis

To corroborate the conclusions of our structural estimation, we investigate how the spread of STW is affected by the change in market structure. To that end, we compute both the quoted spread and cum-fee spread on ASX that prevailed at the beginning of every second between 11:08 and 16:00 of every trading day in June 2014. Note that we begin at 11:08 because that was the time of the Chi-X technical issue.

Figure 4 displays the probability mass function of the empirical distribution of the quoted spread on ASX for each of the trading days of June 2014. In the figure, the day of the Chi-X shutdown is labeled “monopoly,” and remaining days are labeled “duopoly.” The figure suggests that quoted spreads are on average lower on the monopoly day than on the surrounding duopoly days. In particular, the right tail of the spread distribution is much thinner on the monopoly day, without a single second in which the spread exceeded four cents.

Figure 4: Distribution of ASX quoted spread of STW, June 2014



Probability mass functions of the empirical distribution of quoted bid-ask spreads (cents) of STW on ASX for each of the 20 trading days of June 2014. For each day, the sample comprises all seconds between 11:08 and 16:00.

To test formally whether duopoly spreads are larger on average, we regress the STW spread on a dummy variable that evaluates to unity for seconds on June 16, 2014. In our baseline specification, the sample consists of all seconds between 11:08 and 16:00 in all twenty trading days in June 2014. The results, reported in column (1) of table 6, indicate that the monopoly day is associated with a substantial reduction in the spread. Depending on the measure of spread, the reduction is 29.1 percent (spread in cents), 28.0 percent (cum-fee spread in cents), or 28.8 percent (spread in bps). All results are highly significant.

A potential concern is that this reduction in the spread is attributable to forces other than the shock to fragmentation. For instance, a lower spread might also be caused by a reduction in the rate of information arrival or an increase in retail trader activity. To the extent that such forces are correlated with day of the week or with time of the year, we can control for them by repeating the above analysis with different subsets of trading days. In column (2) of table 6, we rule out day-of-the-week effects by focusing only on Mondays. In columns (3) and (4), we focus on shorter event windows around the monopoly day, of 5 and 3 trading days, respectively. None of the subsamples leads to qualitatively different results, which suggests that the finding is indeed driven by the change in market structure.

As a second means of ruling out alternative explanations, we introduce a control group of eight securities against which to compare the response of STW to the Chi-X shutdown. Appendix B.6 provides details about the selection of this control group. Like STW, these are ETFs with exposure to Australian equities that were traded on ASX in June 2014, but unlike STW, they were not also traded on Chi-X. Thus, this control group enables us to pursue a difference-in-difference approach, thereby isolating the effects of the change in the number of venues on which STW was traded from those of any market-wide shocks that may have influenced trading conditions on the day of the shutdown. The results of this analysis are reported in column (5) of table 6, and we again find results that are qualitatively similar to those of the baseline specification.

Table 6: ASX spreads, Australian equity ETFs, June 2014

	(1) All	(2) Mondays	(3) ± 5 days	(4) ± 3 days	(5) All
A. Dependent Variable: quoted spread (cents)					
STW × Monopoly	-1.126^{++}	-0.980^{++}	-1.068^{++}	-0.881^{++}	-1.037^{++}
	(0.0567)	(0.0901)	(0.0673)	(0.0818)	(0.0643)
STW Duopoly Mean	3.869	3.722	3.810	3.623	3.869
Change (Percent)	-29.12	-26.33	-28.04	-24.31	-26.79
B. Dependent Variable: cum-fee spread (cents)					
STW × Monopoly	-1.127^{++}	-0.981^{++}	-1.069^{++}	-0.882^{++}	-1.036^{++}
	(0.0567)	(0.0901)	(0.0673)	(0.0818)	(0.0643)
STW Duopoly Mean	4.023	3.877	3.965	3.777	4.023
Change (Percent)	-28.02	-25.31	-26.97	-23.34	-25.76
C. Dependent Variable: spread (bps)					
STW × Monopoly	-2.167^{++}	-1.863^{++}	-2.034^{++}	-1.691^{++}	-3.091^{++}
	(0.111)	(0.174)	(0.131)	(0.159)	(0.186)
STW Duopoly Mean	7.526	7.223	7.394	7.050	7.526
Change (Percent)	-28.79	-25.80	-27.51	-23.98	-41.07
Fixed Effects					
Day × Hour	NO	NO	NO	NO	YES
Security	YES	YES	YES	YES	YES
Control Group	NO	NO	NO	NO	YES
Observations	350400	70080	192720	122640	3153600

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.005$, $^{++}$ $p < 0.001$

In panels A, B and C, the dependent variable is, respectively, the quoted spread in cents, the cum-fee spread in cents, and the spread in bps prevailing on ASX at the beginning of the second. An observation is a second between 11:08 and 16:00 for a security traded on ASX during June 2014. The variable STW × Monopoly is an indicator for June 16, 2014 interacted with STW. In column (1), the sample includes all trading days in June 2014 and the security STW. In column (2), the sample is restricted to all Mondays in June 2014 and the security STW. In column (3), the sample is restricted to all trading days from June 6 until June 23 and the security STW. In column (4), the sample is restricted to all trading days from June 11 until June 19 and the security STW. In column (5), the sample includes all trading days in June 2014 and the securities STW, IOZ, ISO, MVW, QOZ, SSO, VAS, VLC, and VSO. The latter eight securities are a control group that consists of all ETFs with exposure to Australian equity which were traded on ASX but not on Chi-X. Coefficients are estimated by ordinary least squares. Standard errors are clustered by 120 second blocks on each trading day. STW Duopoly Mean refers to the average spread pertaining to STW on non-monopoly days. Change (Percent) refers to the coefficient estimate relative to the STW duopoly mean.

7.3 Discussion

In the model, the full equilibrium effect of subtracting an exchange is determined by the magnitudes of both the competition and the exposure channels. The fact that spreads are on average 29.1 percent smaller on the monopoly day is consistent with the large exposure channel predicted by the model. However, the Chi-X shutdown does not provide any insight into the competition channel: because the shutdown was both unexpected and short-lived, ASX did not respond to it by strategically altering its fees. Therefore, this figure can be interpreted only as an upper bound on the net effect.

A back-of-the-envelope estimate of the magnitude of the competition channel might be derived from the July 2010 fee change that ASX implemented in response to the announcement that Chi-X would enter the following year. ASX reduced its total trading fee from 0.56 to 0.30 basis points. Multiplying this 0.26 basis point change by the closing price of STW on the monopoly day ($51.23) amounts to a change in trading fees of 0.133 cents. Interpreting this figure as the magnitude of the competition channel and adding it to the -1.126 cents estimate from column (1) of table 6, which we interpret as the magnitude of the exposure channel, suggests a net effect of -0.992 cents. This would equate to a 25.7 percent decrease from the average spread on duopoly days, which is very much in line with the 22.9 percent reduction predicted by the model.

Even though the identification strategy underlying our analysis is strong, a potential concern is that the behavior we observe on the day of the shutdown might not reflect long-run behavior in a monopoly environment. For instance, certain traders, knowing that the shutdown was temporary, could have responded by withdrawing from trading on that day, whereas they would have adapted over the long run to a permanent change in the number of exchanges. Consistent with this particular concern is that volume was below average on the day of the shutdown (*cf.* table 5). Nevertheless, several factors mitigate this worry. First, the only agent in the model whose strategy is markedly different in the monopoly and duopoly equilibria is the liquidity provider who must set a different spread; the other agents can maintain essentially the same strategy and so have no reason to withdraw. Moreover, since we do observe a change in the spread on the monopoly day, it would seem that the real-world analogues of the liquidity provider do in fact adjust their strategies rather than withdraw. Second, private conversations with industry participants in Australia and the United States support the view that traders are well-equipped to deal with such contingencies and have no need to withdraw from trading. Consistent with this, a similar episode that occurred in the US one year later had almost no impact on trading (The Wall Street Journal, 2015), despite a primary venue, NYSE, suffering an unexpected shutdown in that case. Third, total trading volume for STW on the day after the shutdown was only 88,953 contracts, the lowest value for the entire month of June. This is somewhat at odds with the theory that certain traders withdrew from trading on the day of the shutdown because, in that case, we would expect the displaced trades materialize on the following day and contribute to a relatively large volume.

Nevertheless, we concede that whatever effects we find are necessarily short-run effects, and it is theoretically unclear whether the long-run effects are greater or smaller in comparison. Still, this event does provide some additional support for the model's prediction that STW spreads would be lower in a monopoly than in the prevailing duopoly. Moreover, these results can be interpreted as a proof of concept of the model, indicating that the approach we take in this paper might also be a valid way of ascertaining the effects of fragmentation in other contexts.

8 Potential Concerns

The model is quite stylized in several respects, omitting a number of important features possessed by financial markets in practice. Some of these omissions prove to be without consequence. As we argue in appendix F, our main findings persist in extensions involving (*i*) inventory constraints, (*ii*) costs of operation for liquidity providers and exchanges, (*iii*) Brownian components of the process governing the evolution of the security value, as well as discrete jumps of heterogeneous sizes, (*iv*) across-exchange order splitting by investors, and (*v*) short-lived private information about the value of the security. Nevertheless, other omissions could have import, and might therefore represent potential concerns.

One potential concern is our focus on symmetric equilibria, which could limit the extent to which our model is a good fit for our chosen empirical application as well as for other potential applications. While ASX and Chi-X are reasonably symmetric with respect to some key variables (e.g. quoted spreads and message traffic), symmetry fails to hold in other respects. For example, ASX handles more volume than Chi-X, and ASX also charges higher fees than Chi-X. As a result, it is possible an alternative model that allows for certain asymmetries might be a better fit for the Australian market. One possibility would be to relax the assumption that investors are uniformly distributed on the circle.[32]

A second limitation is that our model and estimation focus on only a single security. A key takeaway from our analysis is that the welfare consequences of fragmentation are a function of deep parameters—which need not be constant across securities or across time. Thus, while our estimates indicate that fragmentation is on net detrimental for the trading of STW, that finding need not extrapolate to other securities, in Australia or elsewhere. For instance, He, Jarnecic and Liu (2015) and Aitken, Chen and Foley (2016) use data from an earlier point in time and aggregate across a broader group of securities than do we, so there is little reason to expect that they would arrive at the same conclusions. Indeed, their analysis suggests that the 2011 entry of Chi-X was beneficial for the Australian market on the whole. Related is the fact that a model with only a single security is incapable of capturing any cross-security substitution patterns. An interesting direction for future work would be to expand this framework to a multi-security setting.

Third, the model makes quite stark parametric assumptions. In particular, we assume a linear demand framework. In addition, we assume that both investors and information arrive at constant and exogenous rates. While these assumptions are restrictive, they can be thought of as first-order approximations, which also improve tractability and allow for the derivation of clean, closed-form expressions. Nevertheless, it might be interesting to consider more flexible models, and possibly to endogenize the information and investor arrival processes.

Fourth, while the model allows for some forms of heterogeneity among investors, it does not permit differences in either the volume that they seek to trade or in their patience for spreading trades over time. With such heterogeneity, the spread would cease to be a sufficient welfare statistic as it is in our model. Moreover, such heterogeneity might also extend to concerns with the empirical classification of clustered trades as information-motivated, since a large but uninformed investor might initiate simultaneous trades on multiple exchanges. It is always difficult to speculate about the equilibrium outcomes of an alternative model, and in this case it is not clear how the existence of such investors might interact with our key

[32]Technically, a second possibility would be to allow for a mass of investors who lack access to the smaller exchange, as in Foucault and Menkveld (2008). Although this may have been an appropriate modeling choice for their 2004–2005 data, it seems less suitable for capturing modern trading, wherein essentially all traders now have access to all exchanges.

result on the negative welfare implications of fragmentation for traders of STW in Australia.

Finally, the model necessarily falls short of incorporating every potential channel through which fragmentation might affect market quality. Some channels that are absent from our model, despite having received attention in the earlier literature, seem less relevant than the channels we have considered. For instance, while network externalities were the focus of many early models on fragmentation, they are less significant in the electronically-linked trading environments of today. However, other channels that are absent from our model could remain important. For example, the competitive benefits of fragmentation might not be limited to the price dimension. Rather, fragmentation could induce some exchanges to differentiate themselves by improving their speed (as in Pagnotta and Philippon, 2016) or other aspects of their infrastructure. Further, fragmentation together with maker/taker pricing, allows liquidity providers to compete on a finer price grid, which could reduce the spread of the aggregate book (as in Chao, Yao and Ye, 2017). On the other hand, fragmentation would increase the costs associated with communication (as in Mendelson, 1987), and fixed costs are also required to establish new venues. While our focus on two channels has the advantage of yielding a parsimonious and tractable model, it would also be valuable to develop a richer model that incorporates some of these other elements.

9 Conclusion

This paper provides a tractable and estimable model that assembles several first-order insights about the interaction between fragmentation and liquidity. In the model, fragmentation has ambiguous consequences for market quality as a result of two countervailing forces: (*i*) the competition channel, whereby adding more exchanges induces lower fees and therefore smaller spreads; and (*ii*) the exposure channel, whereby adding more exchanges increases the costs of liquidity providers and therefore induces larger spreads. This theoretical ambiguity is consistent with the empirical literature on fragmentation, which collectively demonstrates that the effects of fragmentation depend upon the context. Moreover, by estimating the model, it can be used as a tool for predicting the effects of a change in fragmentation in a given context, and as such, might be useful for guiding policy.

The parameters of the model are identified from data by the average spread, together with how the incidence of certain types of trades depends upon the prevailing prices. We also demonstrate how these parameters can be estimated via nonlinear least squares. For an empirical application of the model, we use data pertaining to the trading of STW, which is the largest ETF in the Australian market. Our estimates imply the existence of significant market frictions, which suggests that the competition channel is limited in magnitude. Indeed, at the estimates, we find that the competition channel is outweighed by the exposure channel, so that traders of STW would fare better under a monopoly than under Australia's prevailing duopoly. This prediction is corroborated by a separate analysis based on exogenous variation in the market structure, which supports the validity of our approach. We believe that our approach can be useful when rule changes affecting fragmentation have to be decided, especially in the absence of sufficient identifying variation in market structure.

References

Aït-Sahalia, Yacine and Mehmet Sağlam, "High Frequency Market Making: Optimal Quoting," *Working Paper*, 2017. http://ssrn.com/abstract=2331613.

_ **and** _ , "High Frequency Market Making: Implications for Liquidity," *Working Paper*, 2017. http://ssrn.com/abstract=2908438.

Aitken, Michael, Haoming Chen, and Sean Foley, "The Impact of Fragmentation, Exchange Fees and Liquidity Provision on Market Quality," *Working Paper*, 2016. http://ssrn.com/abstract=2284635.

Amihud, Yakov, Beni Lauterbach, and Haim Mendelson, "The Value of Trading Consolidation: Evidence From the Exercise of Warrants," *Journal of Financial and Quantitative Analysis*, 2003, *38* (4), 829–846.

Angel, James J., Lawrence E. Harris, and Chester S. Spatt, "Equity Trading in the 21st Century: An Update," *The Quarterly Journal of Finance*, 2015, *5* (1), 1–39.

Arnold, Tom, Philip Hersch, J. Harold Mulherin, and Jeffry Netter, "Merging Markets," *The Journal of Finance*, 1999, *54* (3), 1083–1107.

ASX Group, "ITCH - Glimpse Message Specification," ASX Market Information 2012.

_ , "Past Editions of Australian Cash Market Reports," 2014. http://www.asx.com.au/services/trading-services/past-editions-australian-cash-market-reports.htm.

_ , "ASX Trade: Markets Participant and Trading Schedule of Fees," 2016. http://www.asx.com.au/about/asx-fees.htm.

Australian Securities and Investments Commission, "Report 331: Dark Liquidity and High-Frequency Trading," 2013. http://download.asic.gov.au/media/1344182/

rep331-published-18-March-2013.pdf.

Back, Kerry and Shmuel Baruch, "Strategic Liquidity Provision in Limit Order Markets," *Econometrica*, 2013, *81* (1), 363–392.

Battalio, Robert H., "Third Market Broker-Dealers: Cost Competitors or Cream Skimmers?," *The Journal of Finance*, 1997, *52* (1), 341–352.

Battalio, Robert, Shane A. Corwin, and Robert Jennings, "Can Brokers Have It All? On the Relation between Make-Take Fees and Limit Order Execution Quality," *The Journal of Finance*, 2016, *71* (5), 2193–2238.

Bennett, Paul and Li Wei, "Market Structure, Fragmentation, and Market Quality," *Journal of Financial Markets*, 2006, *9* (1), 49–78.

Bernales, Alejandro, Italo Riarte, Satchit Sagade, Marcela Valenzuela, and Christian Westheide, "A Tale of One Exchange and Two Order Books: Effects of Fragmentation in the Absence of Competition," *Working Paper*, 2017.

Bernhardt, Dan and Eric Hughson, "Splitting Orders," *Review of Financial Studies*, 1997, *10* (1), 69–101.

Bessembinder, Hendrik and Herbert M. Kaufman, "A Cross-Exchange Comparison of Execution Costs and Information Flow for NYSE-Listed Stocks," *Journal of Financial Economics*, 1997, *46* (3), 293–319.

Biais, Bruno, Christophe Bisiere, and Chester Spatt, "Imperfect Competition in Financial Markets: An Empirical Study of Island and Nasdaq," *Management Science*, 2010, *56* (12), 2237–2250.

_ , David Martimort, and Jean-Charles Rochet, "Competing Mechanisms in a Common Value Environment," *Econometrica*, 2000, *68* (4), 799–837.

Boehmer, Beatrice and Ekkehart Boehmer, "Trading Your Neighbor's ETFs: Competition or Fragmentation?," *Journal of Banking & Finance*, 2003, *27* (9), 1667–1703.

Boneva, Lena, Oliver Linton, and Michael Vogt, "The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market," *Journal of Applied Econometrics*, 2016, *31* (1), 192–213.

Branch, Ben and Walter Freed, "Bid-Asked Spreads on The AMEX and The Big Board," *The Journal of Finance*, 1977, *32* (1), 159–163.

Budish, Eric, Peter Cramton, and John Shim, "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, 2015, *130* (4), 1547–1621.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

Cboe Global Markets, Inc., "Historical Market Volume Data," 2017. http://markets.cboe.com/us/equities/market_statistics/historical_market_volume/.

Chao, Yong, Chen Yao, and Mao Ye, "Why Discrete Price Fragments U.S. Stock Exchanges and Disperses Their Fee Structures," *Working Paper*, 2017. http://ssrn.com/abstract=2530572.

Chi-X Australia, "Market Operations Notice (Reference Number: 0019/11)," 2011. http://au.chi-x.com/resources/au/file/Market%20Operations%20Notice%200019-11.pdf.

— , "Market Data Feed Specification," Chi-X Global Inc. 2012.

— , "Market Operations Notice (Reference Number: 0002/14)," 2014. http://cmsau.chi-x.com/Portals/15/Docs/Market%20Operations%20Notice%200002-14.pdf.

— , "Technical Notice (Reference Number: 0012/14)," 2014. http://cmsau.chi-x.com/Portals/15/Docs/Chi-X%20Technical%20Notice%200012-14.pdf.

Chlistalla, Michael and Marco Lutat, "Competition in Securities Markets: The Impact on Liquidity," *Financial Markets and Portfolio Management*, 2011, *25* (2), 149–172.

Chowdhry, Bhagwan and Vikram Nanda, "Multimarket Trading and Market Liquidity," *Review of Financial Studies*, 1991, *4* (3), 483–511.

Cohen, Kalman J. and Robert M. Conroy, "An Empirical Study of the Effect of Rule 19c-3," *Journal of Law and Economics*, 1990, *33* (1), 277–305.

— , Steven F. Maier, Robert A. Schwartz, and David K. Whitcomb, "An Analysis of the Economic Justification for Consolidation in a Secondary Security Market," *Journal of Banking & Finance*, 1982, *6* (1), 117–136.

Colliard, Jean-Edouard and Thierry Foucault, "Trading Fees and Efficiency in Limit Order Markets," *Review of Financial Studies*, 2012, *25* (11), 3389–3421.

Comerton-Forde, Carole and Tālis J. Putniņš, "Dark Trading and Price Discovery," *Journal of Financial Economics*, 2015, *118* (1), 70–92.

Copeland, Thomas E. and Dan Galai, "Information Effects on the Bid-Ask Spread," *The Journal of Finance*, 1983, *38* (5), 1457–1469.

d'Aspremont, Claude, J. Jaskold Gabszewicz, and J-F Thisse, "On Hotelling's 'Stability in Competition'," *Econometrica*, 1979, *47* (5), 1145–1150.

Degryse, Hans, Frank de Jong, and Vincent van Kervel, "The Impact of Dark Trading and Visible Fragmentation on Market Quality," *Review of Finance*, 2015, *19* (4), 1587–1622.

Dennert, Jürgen, "Price Competition Between Market Makers," *The Review of Economic Studies*, 1993, *60* (3), 735–751.

Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, "Liquidity, Information, and Infrequently Traded Stocks," *The Journal of Finance*, 1996, *51* (4), 1405–1436.

— , Robert F. Engle, Maureen O'Hara, and Liuren Wu, "Time-Varying Arrival Rates of Informed and Uninformed Trades," *Journal of Financial Econometrics*, 2008, *6* (2), 171–207.

Economides, Nicholas, "Symmetric Equilibrium Existence and Optimality in Differentiated Product Markets," *Journal of Economic Theory*, 1989, *47* (1), 178–194.

_ **and Aloysius Siow**, "The Division of Markets is Limited by the Extent of Liquidity (Spatial Competition with Externalities)," *The American Economic Review*, 1988, *78* (1), 108–121.

Fink, Jason, Kristin E. Fink, and James P. Weston, "Competition on the Nasdaq and the Growth of Electronic Communication Networks," *Journal of Banking & Finance*, 2006, *30* (9), 2537–2559.

Foley, Sean and Tālis J. Putniņš, "Should We be Afraid of the Dark? Dark Trading and Market Quality," *Journal of Financial Economics*, 2016, *122* (3), 456–481.

Fontnouvelle, Patrick De, Raymond P.H. Fishe, and Jeffrey H. Harris, "The Behavior of Bid-Ask Spreads and Volume in Options Markets During the Competition for Listings in 1999," *The Journal of Finance*, 2003, *58* (6), 2437–2464.

Foucault, Thierry, "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market," *Journal of Financial Markets*, 1999, *2* (2), 99–134.

_ **and Albert J. Menkveld**, "Competition for Order Flow and Smart Order Routing Systems," *The Journal of Finance*, 2008, *63* (1), 119–158.

Gajewski, Jean-François and Carole Gresse, "Centralised Order Books Versus Hybrid Order Books: A Paired Comparison of Trading Costs on NSC (Euronext Paris) and SETS (London Stock Exchange)," *Journal of Banking & Finance*, 2007, *31* (9), 2906–2924.

Gill, Philip E., Elizabeth Wong, Walter Murray, and Michael A. Saunders, "User's Guide for SNOPT Version 7.5: Software for Large-Scale Nonlinear Programming," 2015. http://www.cam.ucsd.edu/~peg/papers/sndoc7.pdf.

Glosten, Lawrence R., "Is the Electronic Open Limit Order Book Inevitable?," *The Journal of Finance*, 1994, *49* (4), 1127–1161.

_ **and Paul R. Milgrom**, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 1985, *14* (1), 71–100.

Hamilton, James L., "Marketplace Fragmentation, Competition, and the Efficiency of the Stock Exchange," *The Journal of Finance*, 1979, *34* (1), 171–187.

Haslag, Peter H. and Matthew Ringgenberg, "The Causal Impact of Market Fragmentation on Liquidity," *Working Paper*, 2016. http://ssrn.com/abstract=2591715.

He, Peng William, Elvis Jarnecic, and Yubo Liu, "The Determinants of Alternative Trading Venue Market Share: Global Evidence from the Introduction of Chi-X," *Journal of Financial Markets*, 2015, *22*, 27–49.

Hendershott, Terrence and Charles M. Jones, "Island Goes Dark: Transparency, Fragmentation, and Regulation," *Review of Financial Studies*, 2005, *18* (3), 743–793.

Hotelling, Harold, "Stability in Competition," *The Economic Journal*, 1929, *39* (153), 41–57.

Kyle, Albert S., "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53* (6), 1315–1335.

Mayhew, Stewart, "Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets," *The Journal of Finance*, 2002, *57* (2), 931–958.

Mendelson, Haim, "Consolidation, Fragmentation, and Market Performance," *Journal of Financial and Quantitative Analysis*, 1987, *22* (2), 189–207.

Menkveld, Albert J., "High Frequency Trading and the *New Market* Makers," *Journal of Financial Markets*, 2013, *16* (4), 712–740.

Neal, Robert, "Potential Competition and Actual Competition in Equity Options," *The Journal of Finance*, 1987, *42* (3), 511–531.

Nguyen, Vanthuan, Bonnie F. Van Ness, and Robert A. Van Ness, "Short-and Long-Term Effects of Multimarket Trading," *Financial Review*, 2007, *42* (3), 349–372.

Nielsson, Ulf, "Stock Exchange Merger and Liquidity: The Case of Euronext," *Journal of Financial Markets*, 2009, *12* (2), 229–267.

O'Hara, Maureen and Mao Ye, "Is Market Fragmentation Harming Market Quality?," *Journal of Financial Economics*, 2011, *100* (3), 459–474.

Pagano, Marco, "Trading Volume and Asset Liquidity," *The Quarterly Journal of Economics*, 1989, *104* (2), 255–274.

Pagnotta, Emiliano and Thomas Philippon, "Competing on Speed," *Working Paper*, 2016. http://ssrn.com/abstract=1967156.

Salop, Steven C., "Monopolistic Competition with Outside Goods," *The Bell Journal of Economics*, 1979, *10* (1), 141–156.

Sandås, Patrik, "Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market," *The Review of Financial Studies*, 2001, *14* (3), 705–734.

State Street Global Advisors, "SPDR S&P/ASX 200 Fund," 2014. http://spdrs.com.au/etf/fund/fund_detail_STW.html.

U.S. Securities and Exchange Commission, "Regulation NMS: Final Rules and Amendments to Joint Industry Plans," *Federal Register*, June 29 2005, *70* (124).

van Kervel, Vincent, "Competition for Order Flow with Fast and Slow Traders," *Review of Financial Studies*, 2015, *28* (7), 2094–2127.

The Wall Street Journal, "Ho-Hum, Just Another Day for Traders," July 2015. https://www.wsj.com/articles/why-the-nyse-outage-may-not-matter-to-investors-1436374396.

— , "High-Frequency Traders Fall on Hard Times," March 2017. https://www.wsj.com/articles/high-frequency-traders-fall-on-hard-times-1490092200.

Weston, James P., "Electronic Communication Networks and Liquidity on the Nasdaq," *Journal of Financial Services Research*, 2002, *22* (1/2), 125–139.

Internet Appendix

A Proofs

A.1 Lemma 1

Lemma 1. *As before, let*

$$\Sigma \equiv \begin{cases} \theta\left(1+\dfrac{\lambda_j}{\lambda_i}\right) & \text{if } X = 1 \\ \theta + \dfrac{4\alpha}{X^2} - \sqrt{\theta^2 + \dfrac{16\alpha^2}{X^4} - \dfrac{8\alpha\theta\lambda_j}{X\lambda_i}} & \text{if } X \geq 2 \end{cases}$$

Under assumptions 1, 2 and 3:

(i) $\Sigma \in \mathbb{R}$,

(ii) $\Sigma/2 \leq \theta$, *and*

(iii) if $X \geq 2$, *then* $\theta^2 + \dfrac{16\alpha^2}{X^4} - \dfrac{8\alpha\theta\lambda_j}{X\lambda_i} > 0$.

Proof. To prove the first claim, notice that by assumption 1,

$$\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i} \geq \theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta}{X^2} = \left(\theta - \frac{4\alpha}{X^2}\right)^2 \geq 0,$$

which therefore implies that $\Sigma \in \mathbb{R}$.

The proof of the second claim has two parts. First, if $X = 1$, then the result follows directly from assumption 1. Second, if $X \geq 2$, then by assumption 1,

$$\theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}} \leq \theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta}{X^2}} = \theta + \frac{4\alpha}{X^2} - \left|\theta - \frac{4\alpha}{X^2}\right| \leq 2\theta.$$

In light of the proof of the first claim, proving the third claim reduces to deriving a contradiction from the combination of $X \geq 2$ and $\theta = \frac{4\alpha}{X^2}$. If this is the case, then $\Sigma = 2\theta$. Plugging that into assumption 3 yields $0 \geq \lambda_j X(\sigma - \theta)$, which contradicts assumption 2. □

A.2 Proposition 1 (Monopoly)

Proof. The proof proceeds in two parts. First we describe equilibrium strategies, and second we show that no player has a profitable deviation.

Part One (Description): Let s^* be defined as in the statement of the proposition:

$$s^* = \theta\left(1 + \frac{\lambda_j}{\lambda_i}\right).$$

The strategy of the exchange is to set a make fee $\tau^*_{\text{make}} \geq 0$ and a take fee $\tau^*_{\text{take}} \geq 0$ such that

$$\tau^*_{\text{make}} + \tau^*_{\text{take}} = \frac{\lambda_i\left(1 - \frac{1}{\theta}\frac{s^*}{2}\right)\frac{s^*}{2} - \lambda_j\left(\sigma - \frac{s^*}{2}\right)}{\lambda_i\left(1 - \frac{1}{\theta}\frac{s^*}{2}\right) + \lambda_j}.$$

Note that assumption 3 ensures that the expression on the right hand side is nonnegative, so that this is feasible. Only the total fee will be relevant.

One high-frequency trader plays the role of a "liquidity provider." A second high-frequency trader plays the role of an "enforcer." The remaining high-frequency traders (infinitely many) play the role of "stale-quote snipers."

To define the strategy of the liquidity provider, we consider the following polynomial in s, where we use $\tau = (\tau_{\text{make}}, \tau_{\text{take}})$ to denote the fees set by the exchange:

$$\pi(s|\tau) := \lambda_i\left(1 - \frac{1}{\theta}\frac{s}{2}\right)\left(\frac{s}{2} - \tau_{\text{make}} - \tau_{\text{take}}\right) + \lambda_j\left(\frac{s}{2} - \tau_{\text{make}} - \tau_{\text{take}} - \sigma\right).$$

The polynomial $\pi(s|\tau)$ represents, roughly speaking, the profits that would accrue to the liquidity provider if the spread were s and trading fees were given by τ. There are two cases. First, if τ is such that there is no value of $s \in [0, 2\theta]$ that is a root of $\pi(s|\tau)$, then the liquidity provider never quotes. Otherwise, let $s(\tau)$ be defined implicitly as the smallest such root.[33] In that case, the liquidity provider then acts as follows. At time zero, she submits to the exchange a limit order to buy one share at $\hat{b}_0 = v_0 - \frac{s(\tau)}{2} + \tau_{\text{take}}$ and a limit order to sell one share at $\hat{a}_0 = v_0 + \frac{s(\tau)}{2} - \tau_{\text{take}}$.[34] If one of her standing limit orders is filled by an investor, then she immediately submits an identical order to replace it. If v_t jumps, then she immediately submits to the exchange the following orders: (i) cancellations for her limit orders, (ii) a limit order to buy one share at $\hat{b}_{t^+} = v_{t^+} - \frac{s(\tau)}{2} + \tau_{\text{take}}$, and (iii) a limit order to sell one share at $\hat{a}_{t^+} = v_{t^+} + \frac{\hat{s}(\tau)}{2} - \tau_{\text{take}}$.[35]

The strategy of the enforcer is as follows. She never submits any orders unless the liquidity provider is observed to have deviated, in which case the enforcer begins to take the

[33] Specifically,

$$s(\tau) = \tau_{\text{make}} + \tau_{\text{take}} + \theta\left(1 + \frac{\lambda_j}{\lambda_i}\right) - \sqrt{(\tau_{\text{make}} + \tau_{\text{take}})^2 - 2\theta(\tau_{\text{make}} + \tau_{\text{take}})\left(1 + \frac{\lambda_j}{\lambda_i}\right) + 2\theta\frac{\lambda_j}{\lambda_i}(\theta - 2\sigma) + \theta^2\left(1 + \frac{\lambda_j^2}{\lambda_i^2}\right)}.$$

Note that $s(\tau^*) = s^*$.

[34] The resulting cum-fee spread is $s(\tau)$, and note that it depends upon the make and take fees only through their sum.

[35] For any continuous time variable X_t, we use the shorthand X_{t^+} to denote $\lim_{s \to t^+} X_s$ and X_{t^-} to denote $\lim_{s \to t^-} X_s$.

actions that were prescribed for the liquidity provider.

The strategy of a stale-quote sniper is as follows. If v_t jumps upward (downward), then she immediately submits to the exchange an IOC order to buy (sell) at the price $v_{t^-} + \sigma - \tau_{\text{take}}$ $(v_{t^-} - \sigma + \tau_{\text{take}})$.

An investor who arrives at time t with type $(\tilde{l}, \tilde{\theta})$ chooses a quantity $y \in \{-1, 0, 1\}$ to maximize $\hat{u}_t(y|\tilde{l}, \tilde{\theta})$.

Part Two (Verification): It can be shown that $s(\tau^*) = s^*$. We now argue that if all other players behave as specified, then the liquidity provider has no profitable deviations. The arguments are similar to those in Budish, Cramton and Shim (2015, proof of Proposition 1). By lemma 1(*ii*), $s^*/2 \leq \theta$, and by assumption 2, $s^*/2 < \sigma$. Thus, if the liquidity provider sets a spread $s \leq s^*$, then her flow profits are $\pi(s|\tau^*)$. These profits are zero at s^*, since $s(\tau^*)$ is defined as a root of $\pi(s|\tau^*)$, in particular the smallest root. Moreover, since $\pi(s|\tau^*)$ is a concave, second-order polynomial, profits must be negative at spreads $s < s^*$. Thus, it is not profitable to deviate by setting a smaller spread. It is also not profitable to deviate by setting a larger spread, since the enforcer would then undercut her, and she would receive none of the benefits (from investor orders), but might receive adverse selection costs (from stale-quote sniper orders). Finally, it is also not profitable to deviate by quoting more than a single unit at either the bid or the ask, since her benefits would be the same (only one unit at each is needed to satisfy investor demand) but her costs would increase (since more units are exposed to adverse selection from stale-quote snipers).

We now argue that the stale-quote snipers and the enforcer have no profitable deviations. The arguments are also similar to those in Budish, Cramton and Shim (2015, proof of Proposition 1). They also earn zero profits in the equilibrium, and it therefore remains to show that none of them possesses a deviation that would yield positive profits. It is also not profitable to attempt to provide liquidity at a smaller spread than the liquidity provider, since that would result in negative expected profits for the same reason as above. It is also not profitable to attempt to provide liquidity at the same spread as the liquidity provider, since these quotes have the same adverse selection costs (from stale-quote sniper orders) that the liquidity provider faces in equilibrium but only half the benefits (from investor orders), and would therefore result in negative expected profits. Finally, it is not profitable to attempt to provide liquidity at a larger spread than the liquidity provider, since these orders would receive none of the benefits (from investor orders), but might receive adverse selection costs (from stale-quote sniper orders).

We finally argue that the exchange has no profitable deviations. Given the behavior of the traders, the profits of the exchange are zero for the case in which the liquidity provider does not quote. In the other case, the profits of the exchange are

$$(\tau_{\text{make}} + \tau_{\text{take}}) \left[\lambda_j + \lambda_i \left(1 - \frac{1}{\theta} \frac{s(\tau)}{2} \right) \right],$$

which, using the fact that $s(\tau)$ is a root of $\pi(s|\tau)$, can be shown to equal

$$\lambda_i \left(1 - \frac{1}{\theta} \frac{s(\tau)}{2} \right) \frac{s(\tau)}{2} + \lambda_j \left(\frac{s(\tau)}{2} - \sigma \right).$$

The above expression is a concave function of the spread, $s(\tau)$, which is maximized when $s(\tau) = s^*$. Since $s(\tau^*) = s^*$, the exchange has no profitable deviations to other fee structures τ under which the liquidity provider quotes. Furthermore, by assumption 3, this yields nonnegative profits for the exchange, so the exchange also has no profitable deviations to fee structures τ for which the liquidity provider does not quote. □

A.3 Proposition 2 (Oligopoly)

Proof. The proof proceeds in two parts. First we describe equilibrium strategies, and second we show that no player has a profitable deviation.

Part One (Description): Let s^* be defined as in the statement of the proposition:

$$s^* = \theta + \frac{4\alpha}{X^2} - \sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}.$$

The strategy of each exchange x is to set a make fee $\tau^*_{x,\text{make}} \geq 0$ and a take fee $\tau^*_{x,\text{take}} \geq 0$ such that

$$\tau^*_{x,\text{make}} + \tau^*_{x,\text{take}} = \frac{\lambda_i \left(1 - \frac{1}{\theta}\frac{s^*}{2}\right)\frac{s^*}{2} - \lambda_j X \left(\sigma - \frac{s^*}{2}\right)}{\lambda_i \left(1 - \frac{1}{\theta}\frac{s^*}{2}\right) + \lambda_j X}.$$

Note that assumption 3 ensures that the expression on the right hand side is nonnegative, so that this is feasible. Only the total fee will be relevant.

One high-frequency trader per exchange plays the role of a "liquidity provider." A second high-frequency trader plays the role of an "enforcer." The remaining high-frequency traders (infinitely many) play the role of "stale-quote snipers."

To define the strategy of the liquidity provider for exchange x, we consider the following polynomial in s_x, where we use $\tau_x = (\tau_{x,\text{make}}, \tau_{x,\text{take}})$ to denote the fees set by exchange x.

$$\begin{aligned}\pi(s_x|\tau_x) :=& \lambda_i \left[\frac{X}{2\alpha}\left(\frac{s^*}{2} - \frac{s_x}{2}\right) + \frac{1}{X}\right]\left(1 - \frac{1}{\theta}\frac{s_x}{2}\right)\left(\frac{s_x}{2} - \tau_{x,\text{make}} - \tau_{x,\text{take}}\right) \\ &+ \lambda_j \left(\frac{s_x}{2} - \tau_{x,\text{make}} - \tau_{x,\text{take}} - \sigma\right).\end{aligned}$$

The polynomial $\pi(s_x|\tau_x)$ represents, roughly speaking, the profits that would accrue to the liquidity provider on exchange x if the spread on that exchange were s_x, trading fees on that exchange were given by τ_x, and if the spreads on other exchanges were s^*. There are two cases. First, if τ_x is such that there is no value of $s_x \in [0, 2\theta]$ that is a root of $\pi(s_x|\tau_x)$, then the liquidity provider never quotes. Otherwise, let $s(\tau_x)$ be defined implicitly as the smallest such root. In that case, the liquidity provider for exchange x then acts as follows. At time zero, she submits to the exchange a limit order to buy one share at $\hat{b}_{x,0} = v_0 - \frac{s(\tau_x)}{2} + \tau_{x,\text{take}}$ and a limit order to sell one share at $\hat{a}_{x,0} = v_0 + \frac{s(\tau_x)}{2} - \tau_{x,\text{take}}$.[36] If one of her standing limit orders

[36]The resulting cum-fee spread for exchange x is $s(\tau_x)$, and note that it depends upon the make and take

is filled by an investor, then she immediately submits an identical order to replace it. If v_t jumps, then she immediately submits to exchange x the following orders: (i) cancellations for her limit orders, (ii) a limit order to buy one share at $\hat{b}_{x,t^+} = v_{t^+} - \frac{s(\tau_x)}{2} + \tau_{x,\text{take}}$, and (iii) a limit order to sell one share at $\hat{a}_{x,t^+} = v_{t^+} + \frac{s(\tau_x)}{2} - \tau_{x,\text{take}}$.

The strategy of the enforcer is as follows. She never submits any orders unless a liquidity provider is observed to have deviated, in which case the enforcer begins to take the actions that were prescribed for the deviating liquidity provider.

The strategy of a stale-quote sniper is as follows. If v_t jumps upward (downward), then she immediately submits to each exchange an IOC order to buy (sell) at the price $v_{t^-} + \sigma - \tau_{\text{take}}$ $(v_{t^-} - \sigma + \tau_{\text{take}})$.

An investor who arrives at time t with type $(\tilde{l}, \tilde{\theta})$ chooses an exchange $x \in \{1, \dots, X\}$ and a quantity $y \in \{-1, 0, 1\}$ to maximize $\hat{u}_t(y, x|\tilde{l}, \tilde{\theta})$.

Part Two (Verification): We claim that $s(\tau_x^*) = s^*$. It can be shown that s^* is a root of $\pi(s_x|\tau_x^*)$, so it remains only to be shown that it is the smallest root. To begin, note that since $\tau_{x,\text{make}}^* + \tau_{x,\text{take}}^* \geq 0$ and $\sigma > \theta$, it follows that $\pi(2\theta|\tau_x^*) < 0$. In addition, by lemma 1(ii), $s^* \leq 2\theta$. Because $\pi(s_x|\tau_x^*)$ is a third-order polynomial in s_x with a positive leading coefficient, we conclude from these facts that it has three roots and that s^* is not the largest of those. In addition, it can be shown that $\pi'(s^*|\tau_x^*) \geq 0$, which means that s^* cannot be the middle root. This establishes the claim.

We now argue that if all other players behave as specified, then the liquidity provider at exchange x has no profitable deviations. The arguments are again similar to those in Budish, Cramton and Shim (2015, proof of Proposition 1). By lemma 1(ii), $s^*/2 \leq \theta$, and by assumption 2, $s^*/2 < \sigma$. Thus, if the liquidity provider at exchange x sets a spread $s_x \leq s^*$, then her flow profits are $\pi(s_x|\tau_x^*)$. These profits are zero at s^*, since $s(\tau^*)$ is defined as a root of $\pi(s_x|\tau_x^*)$, in particular the smallest root. Moreover, since $\pi(s_x|\tau_x^*)$ is a third-order polynomial with a positive leading coefficient, profits must be negative at spreads $s_x < s^*$. Thus, it is not profitable to deviate by setting a smaller spread. It is also not profitable to deviate by setting a larger spread, since the enforcer would then undercut her, and she would receive none of the benefits (from investor orders), but might receive adverse selection costs (from stale-quote sniper orders). Finally, it is also not profitable to deviate by quoting more than a single unit at either the bid or the ask, since her benefits would be the same (only one unit at each is needed to satisfy investor demand) but her costs would increase (since more units are exposed to adverse selection from stale-quote snipers).

That the other high-frequency traders also have no profitable deviations can be argued as in the proof of proposition 1.

We now argue that the exchange has no profitable deviations. Given the behavior of the traders and other exchanges, the profits of exchange x are zero for the case in which the liquidity provider does not quote. For the case in which the liquidity provider does quote, the profits of exchange x are

$$(\tau_{x,\text{make}} + \tau_{x,\text{take}}) \left(\lambda_j + \lambda_i \left[\frac{X}{2\alpha} \left(\frac{s^*}{2} - \frac{s(\tau_x)}{2} \right) + \frac{1}{X} \right] \left(1 - \frac{1}{\theta} \frac{s(\tau_x)}{2} \right) \right),$$

fees of the exchange only through their sum.

which, using the fact that $s(\tau_x)$ is a root of $\pi(s_x|\tau_x)$, can be shown to equal

$$\lambda_i \left[\frac{X}{2\alpha} \left(\frac{s^*}{2} - \frac{s(\tau_x)}{2} \right) + \frac{1}{X} \right] \left(1 - \frac{1}{\theta} \frac{s(\tau_x)}{2} \right) \frac{s(\tau_x)}{2} + \lambda_j \left(\frac{s(\tau_x)}{2} - \sigma \right).$$

Note that when the liquidity provider does quote, she sets a spread in the domain $[0, 2\theta]$. We claim that the spread $s(\tau_x) = s^*$ maximizes the above expression on this domain. It can be shown that s^* is a critical point of the expression. Note also that by assumption 3 the expression is nonnegative at s^*. Moreover, the expression is negative at 2θ (the first term in the last factor is zero at 2θ, and the second term is negative, since $\sigma > \theta$ by assumption 2). In addition, by lemma 1(*ii*), $s^* \leq 2\theta$. Because the expression is a third-order polynomial in $s(\tau_x)$ with a positive leading coefficient, we conclude from these facts that s^* must be the unique local maximum. It therefore remains only to show that the expression is not larger at either endpoint. We have already argued that the expression is nonnegative at s^* and negative at 2θ. It is also negative at 0, which establishes the claim.

Since $s(\tau_x^*) = s^*$, the exchange has no profitable deviations to other fee structures τ_x under which the liquidity provider quotes. Furthermore, by assumption 3, this yields nonnegative profits for the exchange, so the exchange also has no profitable deviations to fee structures τ_x for which the liquidity provider does not quote. □

A.4 Proposition 3 (Comparative Statics)

Proof. We consider separately the case of monopoly and the case of oligopoly.

Case One ($X = 1$): In this case, the claims follow straightforwardly from the derivatives of the expression for s^* given in proposition 1 with respect to those parameters.

Case Two ($X \geq 2$): In this case, the claims follow from the derivatives of the expression for s^* given in proposition 2 with respect to those parameters. To establish this, we first compute these derivatives:

$$\frac{\partial s^*}{\partial \alpha} = \frac{\frac{4\theta\lambda_j}{X\lambda_i} - \frac{16\alpha}{X^4} + \frac{4}{X^2}\sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}}{\sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}}$$

$$\frac{\partial s^*}{\partial \lambda_i} = \frac{-4\alpha\theta\lambda_j}{\lambda_i^2 X\sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}}$$

$$\frac{\partial s^*}{\partial \lambda_j} = \frac{4\alpha\theta}{\lambda_i X\sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}}}$$

The derivatives with respect to λ_i and λ_j have the desired sign. To sign the derivative with respect to α, we reason from assumption 1:

$$\begin{aligned}
&\lambda_i \geq X\lambda_j \\
\Longrightarrow\ &\frac{16\theta^2}{X^4\lambda_i^2}(\lambda_i^2 - X^2\lambda_j^2) \geq 0 \\
\Longleftrightarrow\ &\frac{16}{X^4}\left(\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}\right) \geq \left(\frac{16\alpha}{X^4} - \frac{4\theta\lambda_j}{X\lambda_i}\right)^2 \\
\Longrightarrow\ &\frac{4}{X^2}\sqrt{\theta^2 + \frac{16\alpha^2}{X^4} - \frac{8\alpha\theta\lambda_j}{X\lambda_i}} \geq \frac{16\alpha}{X^4} - \frac{4\theta\lambda_j}{X\lambda_i} \\
\Longrightarrow\ &\frac{\partial s^*}{\partial\alpha} \geq 0.
\end{aligned}$$

□

B Additional Details

In appendix B.1, we summarize in greater detail the empirical literature on fragmentation that was mentioned in section 2. In appendix B.2, we describe the structure of the Australian market in greater detail. In appendix B.3, we describe our ASX and Chi-X data in greater detail and explain the steps required to process it. In appendix B.4, we establish consistency of our estimation procedure. In appendix B.5, we describe our implementation of the estimation procedure. In appendix B.6, we provide details about our selection of the control group used for our analysis of the natural experiment described in section 7.

B.1 Empirical Literature on Fragmentation

Paper	Data	Source of Variation
A. Positive association between fragmentation and liquidity		
Branch and Freed (1977)	NYSE and AMEX securities	cross section
Hamilton (1979)	NYSE equities	cross section
Neal (1987)	AMEX options	cross section
Cohen and Conroy (1990)	NYSE equities	Rule 19c-3
Battalio (1997)	NYSE securities	Madoff entry
Mayhew (2002)	CBOE options	panel
Weston (2002)	Nasdaq equities	panel
Boehmer and Boehmer (2003)	US ETFs	NYSE entry
De Fontnouvelle, Fishe and Harris (2003)	US options	cross-listing events
Fink, Fink and Weston (2006)	Nasdaq equities	panel
Nguyen, Van Ness and Van Ness (2007)	US options	panel
Foucault and Menkveld (2008)	Dutch equities	LSE entry
Chlistalla and Lutat (2011)	French equities	Chi-X Europe entry
O'Hara and Ye (2011)	US equities	cross section
Menkveld (2013)	Dutch equities	Chi-X Europe entry
He, Jarnecic and Liu (2015)	Global equities	Chi-X entries (various)
Aitken, Chen and Foley (2016)	Australian equities	Chi-X Australia entry
B. Negative association between fragmentation and liquidity		
Bessembinder and Kaufman (1997)	US equities	Rule 19c-3
Arnold, Hersch, Mulherin and Netter (1999)	US securities	exchange mergers
Amihud, Lauterbach and Mendelson (2003)	Tel-Aviv Stock Exchange warrants	warrant exercises
Hendershott and Jones (2005)	US ETFs	Island goes dark
Bennett and Wei (2006)	NYSE & NASDAQ equities	listing switches
Gajewski and Gresse (2007)	LSE & Euronext Paris equities	cross section
Nielsson (2009)	European equities	Euronext mergers
Bernales, Riarte, Sagade, Valenzuela and Westheide (2017)	European equities	Euronext order book consolidation
C. Mixed association between fragmentation and liquidity		
Boneva, Linton and Vogt (2016)	UK equities	panel
Degryse, de Jong and van Kervel (2015)	Dutch equities	panel
Haslag and Ringgenberg (2016)	US securities	Reg NMS

B.2 Industry Background

This appendix describes additional details of the Australian market, including the regulatory environment, the microstructure of ASX and Chi-X, and the off-exchange trading environment.

Market Integrity Rules. The Australian Securities and Investments Commission (ASIC) regulates the Australian market under the Market Integrity Rules (ASIC, 2011). In most respects, these rules are similar to Reg NMS and Reg ATS, their counterparts in the United States. However, one notable difference lies in the definition of best execution. For a retail client, ASIC's guidance is that best execution is based on total consideration, typically interpreted as the best average price.[37] (This is in contrast to the United States, where the order protection rule requires the best marginal price, at least for the top of the book.) For wholesale clients, ASIC's guidance is that best execution can also include factors such as speed. Another notable difference is that payment for order flow is not allowed in Australia.[38] There is also a minimum price improvement rule: trades that take place outside of pre-trade transparent order books must receive price improvement (although there are exceptions for block trades, large portfolio trades, and times outside of trading hours).[39] A consequence of both of these aforementioned facts is that a greater proportion of retail order flow executes on exchanges in Australia than in the United States.

Other relevant aspects of the Market Integrity Rules include a requirement that markets synchronize their clocks to within 20 milliseconds of UTC,[40] a requirement that visible orders have priority over hidden orders at the same price,[41] and a mandated minimum tick size (which, for equities priced above two dollars, is 1 cent).[42]

To cover the costs of market supervision, ASIC imposes fees on market participants. Some of these fees are activity-based, which are calculated on the basis of the market share of trading activity and messaging activity at ASX and Chi-X.

ASX. ASX is the larger and older of Australia's two extant exchanges, having been formed in 1987. It is operated by ASX Limited, which is a publicly traded company. ASX conducts opening and closing auctions. However, the majority of trading takes place in the intervening continuous session. During this session, ASX operates a transparent limit order book called TradeMatch. ASX TradeMatch features pre-trade anonymity for equities, although not for ETFs. In 2011, ASX introduced a second book, PureMatch, which offers fewer functionalities but faster speeds. However, PureMatch failed to attract any significant volume.

Beyond standard limit orders, ASX TradeMatch also offers the following advanced order types: (*i*) iceberg orders, where at least 500 shares must be displayed, (*ii*) undisclosed orders, in which the precise quantity is not disclosed, provided that the value of the order exceeds $0.5 million, and (*iii*) tailor-made combination orders, which can be used for multi-leg transactions.[43]

Chi-X. Chi-X is the smaller and newer of Australia's two exchanges. Like ASX, it is located in Sydney. Chi-X entered the Australian market in 2011. During the sample period, the exchange was operated by a subsidiary of Chi-X Global Holdings LLC, which was privately owned by a consortium of major financial institutions including BofA Merrill Lynch, GETCO LLC, Goldman Sachs, Morgan Stanley, Nomura Group, Quantlab Group LP, and UBS. Chi-X

[37] Rule 3.1.1.

[38] Rule 7.5.1.

[39] Rule 4.1.1.

[40] Rule 6.3.1.

[41] Rule 4.1.7.

[42] Rule 6.4.1.

[43] Details on these order types, as well as other aspects of the ASX operating rules are available at ASX (2016).

Global has made similar entries into other markets, including Europe, Canada, and Japan.[44] Fewer securities are traded on Chi-X than on ASX, and in addition, Chi-X does not perform a listing function. Chi-X offers neither an opening auction nor a closing auction, but just a single limit order book. There is pre-trade anonymity for all securities.

Like ASX, Chi-X offers standard limit orders, as well as iceberg orders and undisclosed orders (but does not offer combination orders). Unlike ASX, Chi-X offers pegged orders, which can reference the bid, ask, or mid price of the NBBO. Also unlike ASX, Chi-X allows the placement of completely hidden orders, which interact with visible orders in the book. Chi-X permits broker preferencing for these hidden orders, which allows brokers to cross with their own orders regardless of time priority (yet with regard to price and visibility priority). Chi-X also allows brokers to specify a minimum executable quantity for their orders.[45]

In addition, Chi-X also offers market on close orders, which are fully hidden and trade continuously with each other throughout the day at the ASX closing price (both before and after that price is determined).

Other trading. While the ASX and Chi-X books serve as the main trading venues in Australia, there also exist several other modes of trading, including crossing systems, block trading, and internalization. The largest crossing system is CentrePoint, which is operated by ASX. CentrePoint trades take place at the prevailing TradeMatch mid price, and the venue features full pre-trade anonymity (ASX, 2012a). Other functionalities include (*i*) the ability to specify a minimum executable quantity, (*ii*) broker preferencing, and (*iii*) sweep orders, which allow for simultaneous access of CentrePoint and TradeMatch.

Excluding CentrePoint, twenty other crossing systems, with sixteen separate operators, were active at the beginning of our sample (ASIC, 2016). By the end of our sample, those numbers had fallen to eighteen crossing systems and fifteen operators. The largest of these crossing systems are those operated by Credit Suisse, Goldman Sachs, and Citigroup (ASIC, 2015). These crossing systems account for just 2.4 percent of total equity market turnover (ASIC, 2015). Additionally, they are not required to provide fair access, and many are accessible by only a small number of traders.

Block trades comprise the majority of off-exchange trading. The minimum size requirement for a block trade is \$0.2 million, \$0.5 million, or \$1 million, depending upon the category of the security in question. Unlike smaller off-exchange trades, which, under the Market Integrity Rules, must receive price improvement relative to the NBBO, block trades can be negotiated at any price.

B.3 Data

The starting point of our empirical investigation are the message feeds from ASX and Chi-X, which are marketed under the names "ITCH – Glimpse" and "Chi-X MD Feed," respectively. The outbound data feeds of both exchanges are based on NASDAQ's proprietary ITCH pro-

[44] Chi-X Global has sold all their exchanges since then, although they continue to operate. In particular, Chi-X Europe was sold to Bats Global Markets in 2011, Chi-X Canada was sold to Nasdaq in 2015, and both Chi-X Australia and Chi-X Japan were sold to private equity firm JC Flowers in 2016.

[45] Details on the operating rules of Chi-X are available at Chi-X (2013).

tocol. These data are a complete historical record of the information that market participants observe in real-time for a fee. Every trading day is recorded in a separate file, within which messages are recorded chronologically.

These messages are sufficient to construct the lit book at each exchange at any point in time and also to identify all trades that take place in the lit book. Two steps of processing are necessary to obtain that information: message parsing and order book reconstruction. We implement routines to do both using the high-performance computing system *Blacklight* at the Pittsburgh Supercomputing Center, as part of an allocation at XSEDE (Extreme Science and Engineering Discovery Environment).

Message parsing. Every message is binary encoded using MoldUDP64, a networking protocol that allows efficient and scaleable transmission of data. A message is read in as a message block, which consists of the message length and the message data. The length is given by the first two bytes, which contain the number of message data bytes. Depending on the type of message, the message data contain a different amount of information. Table 7 contains examples of the type of information that is contained in common messages.

Every second, a timestamp message is broadcast. Other message types include add orders, cancellations, and executions of existing orders. Those messages specify the time, in nanoseconds, relative to the previous timestamp message, as well as any incremental changes to the lit book.[46]

[46] Rather than transmit, for example, the current bid and ask prices, only incremental changes to the book are broadcast. This is to ensure high-performance for latency-sensitive traders.

Table 7: Examples of ASX message data formats

	length	value
Timestamp Message		
Message Type	1	"T"
Second	4	Numeric
Add Order Message		
Message Type	1	"A"
Timestamp – Nanoseconds	4	Numeric
Order ID	8	Numeric
Order Book ID	4	Numeric
Side (Buy or Sell)	1	Alpha
Order Book Position	4	Numeric
Quantity	8	Numeric
Price	4	Numeric
Order Delete Message		
Message Type	1	"D"
Timestamp – Nanoseconds	4	Numeric
Order ID	8	Numeric
Order Book ID	4	Numeric
Side (Buy or Sell)	1	Alpha
Side (Buy or Sell)	1	Alpha

The table contains a sample of ASX message specifications (ASX, 2012b). The length of a field is measured in number of bytes.

Order book reconstruction. Each message conveys only an incremental change. Therefore, reconstructing the lit order book for a given day and security requires re-running the message broadcast in chronological order beginning at market open and replicating the matching process used by the exchange. When an add order arrives, it is added to the book at the limit price that it specifies. In case of a cancellation, the active order in question is removed. Finally, in the event of an execution, the affected order is removed or its quantity is adjusted. The time series of inside quotes can then be computed by reading off the bid and ask that prevail after the processing of each message. The pseudocode in figure 5 describes the algorithm used for this process.

Figure 5: Algorithm for computing inside quotes from order messages

```
def getBidAsk(LOB):
    S   = sellOrders(LOB)
    ask = minPrice(S)
    B   = buyOrders(LOB)
    bid = maxPrice(B)
    return [bid, ask]

LOB = {}
for m in M:
    if isAddOrder(m):
        addMessage(LOB,m)
    if isCancelOrder(m):
        removeMessage(LOB,m)
    if isTrade(m):
        removeMessage(m)
    print getBidAsk(LOB)
```

Note: M denotes a chronologically sorted list of messages pertaining to a security. The functions *isAddOrder*, *isCancelOrder*, and *isTrade* evaluate to true or false depending on the type of message m. The functions *addMessage* and *removeMessage* modify the current order book *LOB* by adding or removing the quantity at the limit price specified by message m.

B.4 Consistency

This appendix establishes consistency of the nonlinear least squares (NLS) estimation procedure described in section 6.1. While we assume that, conditional on the quotes, both $\varepsilon_{x,t}^{\text{buy}}$ and $\varepsilon_{x,t}^{\text{sell}}$ have mean zero, it would not be correct to make the same assumption for $\varepsilon_{x,t}^{\text{spread}}$. Therefore, consistency of our estimation procedure does not follow immediately from the standard arguments that typically establish consistency of NLS. Nevertheless, consistency is restored by two special features of the model. First, λ_j is excluded from equations (2) and (3). Second, in a neighborhood of $(\alpha, \theta, \lambda_i)$, the right-hand side of equation (4), as a function of λ_j, is injective and contains a neighborhood of the expected spread $\theta + \alpha - \sqrt{\theta^2 + \alpha^2 - 4\alpha\theta\lambda_j/\lambda_i}$ in its range.

Proposition 4. *Assuming that the data generation process is as described in section 6.1, where the parameters satisfy assumptions 1, 2 and 3, the NLS estimation procedure described in section 6.1 is consistent for those parameters.*

Proof. Define

$$Q_T(\alpha, \theta, \lambda_i, \lambda_j) = \frac{1}{T}\sum_{t=1}^{T} \sum_{x\in\{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \left(\varepsilon_{x,t}^{\text{sell}}\right)^2 + \left(\varepsilon_{x,t}^{\text{spread}}\right)^2$$

and

$$\tilde{Q}_T(\alpha, \theta, \lambda_i) = \frac{1}{T} \sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \left(\varepsilon_{x,t}^{\text{sell}}\right)^2,$$

where $\varepsilon_{x,t}^{\text{buy}}$, $\varepsilon_{x,t}^{\text{sell}}$, and $\varepsilon_{x,t}^{\text{spread}}$ are as defined implicitly by equations (2), (3), and (4), respectively. The NLS estimates are then a selection

$$(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j) \in \operatorname*{arg\,min}_{\alpha,\theta,\lambda_i,\lambda_j} Q_T(\alpha, \theta, \lambda_i, \lambda_j).$$

Define also

$$\bar{s} = \frac{1}{2T} \sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} s_{x,t}$$

Finally, define

$$\tilde{Q}_T^* = \left\{ (\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i, \tilde{\lambda}_j) \;\middle|\; \begin{array}{l} (\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i) \in \operatorname{arg\,min}_{\alpha,\theta,\lambda_i} \tilde{Q}_T(\alpha, \theta, \lambda_i), \\ \text{and } \tilde{\lambda}_j = \frac{\tilde{\lambda}_i}{4\tilde{\alpha}\tilde{\theta}} \left(2\tilde{\alpha}\bar{s} + 2\tilde{\theta}\bar{s} - \bar{s}^2 - 2\tilde{\alpha}\tilde{\theta}\right) \end{array} \right\}.$$

With these definitions in hand, we complete the proof by establishing two claims.

Claim 1: Any selection $(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i, \tilde{\lambda}_j) \in \tilde{Q}_T^*$ is consistent for $(\alpha, \theta, \lambda_i, \lambda_j)$.

Proof of claim: By the assumptions imposed upon $\varepsilon_{x,t}^{\text{buy}}$ and $\varepsilon_{x,t}^{\text{sell}}$, the usual arguments for consistency of NLS establish that

$$(\tilde{\alpha}, \tilde{\theta}, \tilde{\lambda}_i) \xrightarrow{p} (\alpha, \theta, \lambda_i).$$

Standard arguments also imply

$$\bar{s} \xrightarrow{p} \theta + \alpha - \sqrt{\theta^2 + \alpha^2 - \frac{4\alpha\theta\lambda_j}{\lambda_i}}.$$

Thus, the continuous mapping theorem yields

$$\tilde{\lambda}_j \xrightarrow{p} \lambda_j.$$

Claim 2: The NLS estimates $(\hat{\alpha}, \hat{\theta}, \hat{\lambda}_i, \hat{\lambda}_j)$ converge almost surely, and hence in probability, to $\tilde{Q}_T^*$.

Proof of claim: First note that

$$Q_T(\hat{\alpha},\hat{\theta},\hat{\lambda}_i,\hat{\lambda}_j) = \tilde{Q}_T(\hat{\alpha},\hat{\theta},\hat{\lambda}_i) + \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX},\text{Chi-X}\}}\left(s_{x,t} - \hat{\theta} - \hat{\alpha} + \sqrt{\hat{\theta}^2 + \hat{\alpha}^2 - \frac{4\hat{\alpha}\hat{\theta}\hat{\lambda}_j}{\hat{\lambda}_i}}\right)^2$$
$$\geq \min_{\alpha,\theta,\lambda_i}\tilde{Q}_T(\alpha,\theta,\lambda_i) + \frac{1}{T}\sum_{t=1}^{T}\sum_{x\in\{\text{ASX},\text{Chi-X}\}}(s_{x,t} - \bar{s})^2. \tag{6}$$

A parameter vector $(\hat{\alpha},\hat{\theta},\hat{\lambda}_i,\hat{\lambda}_j)$ achieves the lower bound (6) only if both of the following conditions hold:

$$(\hat{\alpha},\hat{\theta},\hat{\lambda}_i) \in \arg\min_{\alpha,\theta,\lambda_i}\tilde{Q}_T(\alpha,\theta,\lambda_i)$$
$$\hat{\theta} + \hat{\alpha} - \sqrt{\hat{\theta}^2 + \hat{\alpha}^2 - \frac{4\hat{\alpha}\hat{\theta}\hat{\lambda}_j}{\hat{\lambda}_i}} = \bar{s}$$

Note that the second of these conditions holds only if

$$\hat{\lambda}_j = \frac{\hat{\lambda}_i}{4\hat{\alpha}\hat{\theta}}\left(2\hat{\alpha}\bar{s} + 2\hat{\theta}\bar{s} - \bar{s}^2 - 2\hat{\alpha}\hat{\theta}\right).$$

Thus, the lower bound (6) is achieved only if $(\hat{\alpha},\hat{\theta},\hat{\lambda}_i,\hat{\lambda}_j) \in \tilde{Q}_T^*$.

Let $\delta > 0$ and take a selection $(\tilde{\alpha},\tilde{\theta},\tilde{\lambda}_i,\tilde{\lambda}_j) \in \tilde{Q}_T^*$. We have seen that $\bar{s} \xrightarrow{p} \theta + \alpha - \sqrt{\theta^2 + \alpha^2 - 4\alpha\theta\lambda_j/\lambda_i}$ and that $\tilde{\theta} + \tilde{\alpha} \xrightarrow{p} \theta + \alpha$. Thus, $\tilde{\theta} + \tilde{\alpha} - \bar{s} \xrightarrow{p} \sqrt{\theta^2 + \alpha^2 - 4\alpha\theta\lambda_j/\lambda_i}$. By lemma 1(*iii*), this limit is strictly positive. There thus exists some T' such that if $T \geq T'$, then $\Pr(\bar{s} \leq \tilde{\theta} + \tilde{\alpha}) > 1 - \delta$. In such circumstances,

$$\tilde{\theta} + \tilde{\alpha} - \sqrt{\tilde{\theta}^2 + \tilde{\alpha}^2 - \frac{4\tilde{\alpha}\tilde{\theta}\tilde{\lambda}_j}{\tilde{\lambda}_i}} = \bar{s}.$$

Since we also have

$$(\tilde{\alpha},\tilde{\theta},\tilde{\lambda}_i) \in \arg\min_{\alpha,\theta,\lambda_i}\tilde{Q}_T(\alpha,\theta,\lambda_i),$$

$(\tilde{\alpha},\tilde{\theta},\tilde{\lambda}_i,\tilde{\lambda}_j)$ achieves the lower bound (6). To summarize: in such circumstances, the lower bound (6) is achievable, and so any selection $(\hat{\alpha},\hat{\theta},\hat{\lambda}_i,\hat{\lambda}_j) \in \arg\min_{\alpha,\theta,\lambda_i,\lambda_j} Q_T(\alpha,\theta,\lambda_i,\lambda_j)$ must achieve it and must therefore be contained in $\tilde{Q}_T^*$. □

B.5 Estimation

Estimation is conducted by minimizing the NLS objective function (5), as described in section 6.1. This minimization was performed using SNOPT (Gill, Wong, Murray and Saunders, 2015). To evade the possibility of identifying a local minimum that is not the global minimum, we repeat the optimization for 200 different randomly chosen starting values. During estimation we impose non-negativity constraints on the parameters.

We use a bootstrap procedure to compute standard errors. To allow for temporal dependence, we use a block bootstrap. The asymptotically optimal block length grows with the sample size T at a rate proportional to $T^{1/3}$ (Hall, Horowitz and Jing, 1995). In our case, $T = 1,584,000$, which gives $T^{1/3} \approx 117$. So as to avoid blocks that span days, which would not correctly capture the dependence in the data, we use non-overlapping blocks and round to a block length of 120 seconds. Thus, each of the 80 trading days in the sample is divided into 165 blocks. For each bootstrap repetition we draw 165×80 blocks with replacement. Standard errors are computed based on 200 bootstrap replications.

B.6 Control Group Selection

In this appendix we provide details about the selection of the control group that we have used in the analysis of the Chi-X shutdown in section 7, where the results of that analysis are reported in column (5) of table 6. The control group consists of eight securities that are similar to STW in that they are also ETFs with exposure to Australian equities. However, unlike STW, they were not traded on Chi-X in June 2014. To construct the control group, we start from a list of all 17 ETFs traded on ASX with exposure to Australian equity securities as of February 2017 (ASX, 2017). One of these ETFs is STW, the focus of our analysis. Six of these ETFs were admitted only after June 2014, and are therefore discarded. Another two were also traded on Chi-X as of June 2014, so they are discarded as well, which leaves us with the remaining eight ETFs that constitute the control group for our analysis. Table 8 summarizes these ETFs.

Table 8: Australian equity ETFs, June 2014

Code	Benchmark	Admission Date
A. Control Group		
IOZ	S&P/ASX 200	Dec 2010
ISO	S&P/ASX Small Ordinaries	Dec 2010
MVW	MVIS Australia Equal Weight Index	Mar 2014
QOZ	FTSE RAFI Australia 200	Jul 2013
SSO	S&P/ASX Small Ordinaries	Apr 2011
VAS	S&P/ASX 300	May 2009
VLC	MSCI Large Cap Index	May 2011
VSO	MSCI Small Cap Index	May 2011
B. Treatment Group		
STW	S&P/ASX 200	Aug 2001

The control group was selected from the set of ETFs with with exposure to Australian equities provided by ASX (2017). Such an ETF was selected into the control group if (*i*) it existed in June 2014; and (*ii*) it was traded on ASX but not on Chi-X as of May 30, 2014. An ETF's admission date refers to when it was first admitted for trading at ASX.

C Additional Counterfactuals

In appendix C.1, we investigate the consequences of replacing the limit order book with either of two counterfactual trading mechanisms: frequent batch auctions or non-cancellation delay. In appendix C.2, we investigate the counterfactual in which an order protection rule is applied to the Australian market. And in appendix C.3, we investigate the counterfactual of a regulated monopolist exchange.

C.1 Alternative Trading Mechanisms

A current debate among policy makers, industry participants, and researchers concerns whether alternative trading mechanisms can improve upon the prevailing limit order book. In this section, we consider the counterfactual of the model under two of those alternatives: frequent batch auctions and non-cancellation delay.

The frequent batch auction mechanism would replace the limit order book with sealed-bid, uniform price double auctions conducted at discrete intervals. Frequent batch auctions are the focus of Budish, Cramton and Shim (2015), who show that they improve upon the limit order book by eliminating stale-quote sniping on the basis of public news, and the same remains true in our extension of their framework. The intuition is as follows. Frequent batch auctions delay the processing of all orders received during a batch interval until the end of that interval, which ensures that orders submitted at the same time are processed together. This effectively allows liquidity providers to update their stale quotes before they can be sniped, thereby eliminating adverse selection.

The non-cancellation delay mechanism would modify the limit order book by adding a small, possibly random, delay to all orders except cancellations. It is considered by Baldauf and Mollner (2017), who show that it also eliminates what they refer to as "aggressive-side order anticipation," but it also eliminates stale-quote sniping on the basis of public news of the form modeled by Budish, Cramton and Shim (2015) and by this paper. The intuition is as follows. When the liquidity provider submits an order to cancel a mispriced quote at the same time that a stale-quote sniper submits an order to trade against that mispriced quote, either could be processed first under the limit order book mechanism. In contrast, under non-cancellation delay, the cancellation is guaranteed to be processed first.

In this model, stale-quote sniping is the only source of adverse selection. Thus, the consequences of eliminating it are mathematically equivalent to what would transpire if there were never any jumps in the fundamental value of the security (i.e. $\lambda_j = 0$). While this can be established formally, we omit the derivation in the interest of brevity. Given this, immediate corollaries of propositions 1 and 2 are the following characterizations of the spread that prevails under either frequent batch auctions or non-cancellation delay.

Corollary 2 (monopoly)**.** *With a single exchange ($X = 1$) that uses either frequent batch auctions or non-cancellation delay, there exists a subgame perfect Nash equilibrium with cum-fee spread*

$$s^*_{FBA} = s^*_{NCD} = \theta.$$

Corollary 3 (oligopoly)**.** *With multiple exchanges ($X \geq 2$) that use either frequent batch auctions or non-cancellation delay, there exists a subgame perfect Nash equilibrium with cum-fee spread*

$$s^*_{FBA} = s^*_{NCD} = \theta + \frac{2\alpha}{X} - \sqrt{\theta^2 + \frac{4\alpha^2}{X^2}}.$$

In addition, an immediate corollary of proposition 3 is that frequent batch auctions and non-cancellation delay result in a spread that is guaranteed to be smaller that which prevails under the limit order book. The intuition is that, by eliminating stale-quote sniping, these alternative trading mechanisms eliminate the portion of the spread stemming from adverse selection, leaving only the portion stemming from the market power of exchanges.

Corollary 4. *Under assumptions 1, 2 and 3, $s^*_{FBA} = s^*_{NCD} \leq s^*$.*

Theory therefore dictates that either frequent batch auctions or non-cancellation delay would improve outcomes by reducing transaction costs. What is more, our empirical approach can quantify this reduction. Evaluating the expressions from corollaries 2 and 3 at the estimated parameters, we find that in the counterfactual of frequent batch auctions or non-cancellation delay, the duopoly spread is 52.4 percent lower relative to the spread in the prevailing limit order book duopoly (*cf.* column 2 in table 9).

Moreover, since frequent batch auctions and non-cancellation delay eliminate all adverse selection from the model, they also shut down the exposure channel. Thus, fragmentation unambiguously lowers spreads under the regime of frequent batch auctions or non-cancellation delay. For example, moving from duopoly to monopoly raises spreads by 11.5 percent, from 1.37 to 1.53 cents. Furthermore, a triopoly is not only feasible, but also would feature spreads that are still lower.

Table 9: Cum-fee spreads under various counterfactuals (cents)

	number of exchanges		
	1	2	3
limit order book	2.22	2.88	
frequent batch auctions/non-cancellation delay	1.53	1.37	1.19

The counterfactual spreads are based on the parameter estimates in table 4. Rows refer to trading mechanisms, and columns refer to the number of exchanges. At the parameter estimates, there does not exist an equilibrium with three active exchanges operating limit order books.

C.2 Order Protection Rule

One of the stipulations of Reg NMS in the United States is the order protection rule (also known as Rule 611, or the trade-through rule). It requires trading venues to maintain and enforce procedures that limit the possibility of a trade occurring on that venue when a better price is available elsewhere. In contrast, the legal framework governing trading in Australia does not currently incorporate an order protection rule. In this appendix, we use the estimated model to shed some light on the counterfactual in which Australia were to adopt such a rule.

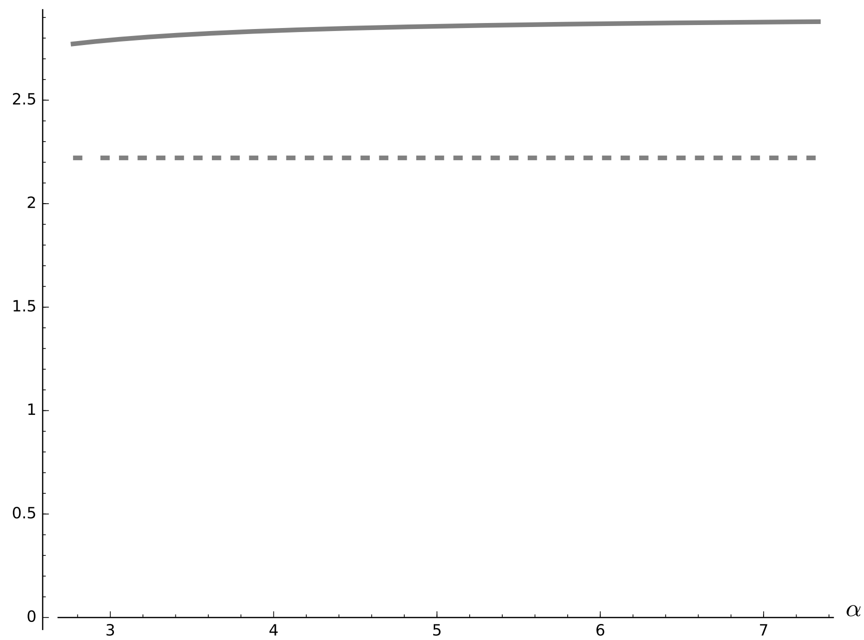
The most natural way to interpret an order protection rule within the model would be as a reduction in α, the magnitude of the frictions that prevent investors from filling their orders at the best price. Proposition 3 implies that such a reduction in α would lead to a reduction of the spread. The remainder of this appendix is dedicated to quantifying the extent of this reduction.

It is unclear what the precise magnitude of the reduction in α would be. In particular, we should not expect a complete eradication of market frictions (i.e. a full reduction of α to zero). Rather, some residual frictions would likely remain, due, for example, to (*i*) difficulties with monitoring prices in real time, (*ii*) imperfect enforcement of the rule, or (*iii*) discrete prices.

Moreover, the value of α cannot be reduced arbitrarily without violating assumption 3, evaluated in the case of a duopoly. If the effect of the rule were to reduce α below this threshold, then competition among exchanges would be intensified to such an extent that it would be impossible for both ASX and Chi-X to operate profitably, and we might expect a reduction in the number of exchanges in the long run. In such cases, the relevant counterfactual would be the monopoly case. On the other hand, if α remains above this threshold, then exchange profitability would be reduced but might not be completely eliminated. The precise value of this threshold depends upon σ. While our procedure does not produce an estimate of σ, a lower bound for it is given by assumption 2. Evaluating at that lower bound, we obtain a lower bound for the threshold: $\bar{\alpha} = 2.772$ is the smallest value of α that is consistent with the model assumptions at the parameter estimates, such that a duopoly remains feasible.

Figure 6 illustrates the spread that would prevail in a duopoly for values of α below that estimated in the data, yet above $\bar{\alpha}$, implicitly assuming that σ is small enough to satisfy assumption 3. The dashed line depicts the monopoly spread and is included for purposes of comparison. Interestingly, this analysis suggests that the effects of an order protection rule would be fairly minimal, except for the case in which the rule induces the exit of an exchange. The reason for this is that the estimated model indicates that ASX and Chi-X are already competing quite intensely with each other. There is therefore not much scope for intensifying this competition without eliminating exchange profitability altogether.

Figure 6: Counterfactual spreads with order protection



The figure plots the counterfactual spreads for monopoly (dashed line) and duopoly (solid line) for values of α below the estimate of the parameter (*cf.* table 4), yet large enough to be consistent with the estimates of the other parameters and the assumptions of the model (*viz.* assumptions 2 and 3).

C.3 Regulated Monopoly

Returning to an analogy made in section 3.3, adverse selection in our model bears a resemblance to fixed costs in oligopolistic price competition. Moreover, our estimates indicate that adverse selection—and therefore the exposure channel—is relatively large. A common regulatory approach in industries with large fixed costs is to grant a monopoly while constraining prices at break-even levels. In this appendix we investigate whether a similar approach would be wise in our setting.

We say that the regulated monopoly outcome for our setting is what would transpire with a single exchange that sets a revenue-neutral fee schedule $\tau_{\text{make}} + \tau_{\text{take}} = 0$. The spread that would prevail in this setting is derived by plugging these fees into the expression for $s(\tau)$ derived in the proof of proposition 1:

$$s^{regulated} = \theta\left(1 + \frac{\lambda_j}{\lambda_i}\right) - \sqrt{2\theta\frac{\lambda_j}{\lambda_i}(\theta - 2\sigma) + \theta^2\left(1 + \frac{\lambda_j^2}{\lambda_i^2}\right)}.$$

While we have estimates of the parameters $(\theta, \lambda_i, \lambda_j)$, we do not possess an estimate of σ. Nevertheless, upper and lower bounds on σ can be derived from assumptions 2 and 3, respectively, implying that $\hat{\sigma} \in [1.528, 1.532]$. These bounds on $\hat{\sigma}$ likewise imply bounds on the counterfactual regulated monopoly spread: $s^{regulated} \in [1.386, 1.392]$.

In summary, this analysis indicates that the spread in the counterfactual of a regulated monopoly would be approximately 37 percent lower than the spread in the counterfactual of an unregulated monopoly (and approximately 52 percent lower than in the prevailing unregulated duopoly). This suggests a potential case for regulating stock exchanges in a way that resembles the approach taken for public utilities and other natural monopolies. Nevertheless, a significant portion of the spread remains even if the transaction fee is set to zero, which reflects the presence of a fair amount of adverse selection.

D Robustness of Estimates

In appendix D.1, we demonstrate that the results of our estimation procedure are robust to alternative choices of the cutoff that is used to distinguish between isolated and clustered trades. In appendix D.2, we demonstrate that our results are also robust to using only buys or only sells as the basis for estimation, and we also show that we fail to reject an overidentifying restriction. Finally, in appendix D.3, we also demonstrate robustness with respect to the timeframe of the sample.

D.1 Robustness to Classification Error

For the empirical analysis in the main text, we use a one second cutoff for distinguishing between isolated and clustered trades: a lit book trade is classified as isolated if no other such trade occurs in the same direction within one second on either exchange, and it is classified as clustered otherwise. This appendix demonstrates that the main results are robust to changes in this cutoff.

Column (3) of table 10 contains the baseline results reported in the main text. To construct the remaining columns, we repeat our estimation procedure for four alternative choices of this cutoff, ranging from 0.1 seconds to 5 seconds. The table reveals that the precise definition of what separates an isolated trade from a clustered trade—at least within this range—has very little impact on the parameter estimates and, consequently, does not change the results qualitatively. Over these different robustness checks, the counterfactual monopoly spread is never smaller than 2.19 cents and never larger than 2.28 cents.

Table 10: Estimates for different definitions of isolated/clustered trades

	value of cutoff (seconds)				
	0.1	0.5	1	2	5
A. parameter estimates					
α	9.64338	4.92955	7.33504	12.10799	5.48684
	(1.45972)	(0.71493)	(1.06129)	(1.77839)	(0.76835)
θ	1.6213	1.57209	1.52817	1.54993	1.48682
	(0.16318)	(0.17743)	(0.15127)	(0.14898)	(0.16159)
λ_i	0.00184	0.00176	0.00172	0.00166	0.00164
	(0.00024)	(0.00024)	(0.00022)	(0.00022)	(0.00023)
λ_j	0.00074	0.00077	0.00078	0.00072	0.00078
	(0.00012)	(0.00012)	(0.00012)	(0.00012)	(0.00011)
B. counterfactual spreads					
monopoly	2.27722	2.26421	2.22093	2.22753	2.19527
	(0.1065)	(0.14064)	(0.10647)	(0.09287)	(0.12471)
duopoly	2.87885	2.87885	2.87885	2.87885	2.87885
	(0.00476)	(0.00457)	(0.0048)	(0.00497)	(0.0046)

The point estimates are computed using the NLS procedure described in section 6.1. An observation is one second between 10:30 and 16:00 in one of the 80 trading days in the sample. The columns refer to five different definitions of what constitutes an isolated trade. The estimation was performed using SNOPT (Gill, Wong, Murray and Saunders, 2015). Standard errors are based on 200 block bootstrap replications.

D.2 Robustness: Side of the Book

In the main text, we report parameter estimates based on an estimation procedure that leverages both buys and sells. In this appendix, we first demonstrate that we obtain similar results from estimation procedures that leverage only buys or only sells. We then use this framework as the basis for testing an overidentifying restriction implied by the model.

Column (3) of table 11 contains the baseline results reported in the main text. Columns (1) and (2) report the results obtained from analogous estimation procedures that are based on only buys or only sells, respectively. To describe those estimation procedures in more detail, consider the following estimating equations:

$$buy_{x,t} = \frac{\lambda_i^B}{2} \max\left[0, \min\left(1, \frac{1}{2} + \frac{a_{-x,t} - a_{x,t}}{\alpha^B}\right)\right] \max\left[0, 1 - \frac{a_{x,t} - v_t}{\theta^B}\right] + \varepsilon_{x,t}^{\text{buy}} \tag{7}$$

$$sell_{x,t} = \frac{\lambda_i^S}{2} \max\left[0, \min\left(1, \frac{1}{2} + \frac{b_{x,t} - b_{-x,t}}{\alpha^S}\right)\right] \max\left[0, 1 - \frac{v_t - b_{x,t}}{\theta^S}\right] + \varepsilon_{x,t}^{\text{sell}} \tag{8}$$

$$s_{x,t} = \theta^B + \alpha^B - \sqrt{(\theta^B)^2 + (\alpha^B)^2 - \frac{4\alpha^B \theta^B \lambda_j^B}{\lambda_i^B}} + \varepsilon_{x,t}^{\text{spread,buy}}$$

$$s_{x,t} = \theta^S + \alpha^S - \sqrt{(\theta^S)^2 + (\alpha^S)^2 - \frac{4\alpha^S \theta^S \lambda_j^S}{\lambda_i^S}} + \varepsilon_{x,t}^{\text{spread,sell}}$$

As before, t indexes the seconds in the sample and x indexes the exchanges $\{\text{ASX}, \text{Chi-X}\}$. And as before, we proxy for v_t with the average mid price $(b_{\text{ASX},t} + b_{\text{Chi-X},t} + a_{\text{ASX},t} + a_{\text{Chi-X},t})/4$ in both (7) and (8).

The estimation procedure for buys is to minimize the objective

$$Q_T^B(\alpha^B, \theta^B, \lambda_i^B, \lambda_j^B) = \frac{1}{T} \sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{buy}}\right)^2 + \frac{1}{2} \left(\varepsilon_{x,t}^{\text{spread,buy}}\right)^2$$

Likewise, the estimation procedure for sells is to minimize the objective

$$Q_T^S(\alpha^S, \theta^S, \lambda_i^S, \lambda_j^S) = \frac{1}{T} \sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \left(\varepsilon_{x,t}^{\text{sell}}\right)^2 + \frac{1}{2} \left(\varepsilon_{x,t}^{\text{spread,sell}}\right)^2$$

Comparing across the columns of table 11, the point estimates do differ somewhat depending on the estimation procedure. However, the counterfactual monopoly spreads are remarkably similar.

Note that we can perform the buy and sell estimation procedures jointly by minimizing the objective

$$Q_T^J(\alpha^B, \theta^B, \lambda_i^B, \lambda_j^B, \alpha^S, \theta^S, \lambda_i^S, \lambda_j^S) = \frac{1}{T} \sum_{t=1}^{T} \sum_{x \in \{\text{ASX,Chi-X}\}} \sum_{i \in \{\text{buy,sell}\}} \left(\varepsilon_{x,t}^{i}\right)^2 + \frac{1}{2} \left(\varepsilon_{x,t}^{\text{spread},i}\right)^2$$

Doing so allows us to test the restriction that the parameters in the buy and sell equations are equal: $\alpha^B = \alpha^S$, $\theta^B = \theta^S$, $\lambda_i^B = \lambda_i^S$, $\lambda_j^B = \lambda_j^S$. To do so, we perform a standard Wald test using the bootstrapped variance-covariance matrix. We compute a test statistic of 6.75, based on which we conclude that we fail to reject the Null of parameter equality, even at the ten percent level.

Finally, note that the baseline estimation is equivalent to performing the above joint estimation with the restrictions in place. As a result, the aforementioned test can also be interpreted as a test of an overidentifying restriction for our estimation procedure.

Table 11: Estimates based on buys and sells

	Buy	Sell	Joint
A. parameter estimates			
α	8.73269	4.92870	7.33504
	(1.25134)	(0.70999)	(1.06129)
θ	1.52424	1.56523	1.52817
	(0.15689)	(0.18383)	(0.15127)
λ_i	0.00162	0.00173	0.00172
	(0.00023)	(0.00022)	(0.00022)
λ_j	0.00073	0.00077	0.00078
	(0.00011)	(0.00013)	(0.00012)
B. counterfactual spreads			
monopoly	2.21552	2.25878	2.22093
	(0.10531)	(0.14853)	(0.10647)
duopoly	2.87885	2.87885	2.87885
	(0.00478)	(0.00480)	(0.00480)

An observation is one second between 10:30 and 16:00 in one of the trading days in the sample. The estimation was performed using SNOPT (Gill, Wong, Murray and Saunders, 2015). Standard errors are based on 200 block bootstrap replications. Column (3) is based on minimizing the objective Q_T, as described in section 6.1. Columns (1) and (2) are based on minimizing the objectives Q_T^B and Q_T^S, respectively, as described in this appendix. We perform a Wald test of the restriction $H_0 : \alpha^B = \alpha^S, \theta^B = \theta^S, \lambda_i^B = \lambda_i^S, \lambda_j^B = \lambda_j^S$. The test statistic based on the bootstrapped variance-covariance matrix is 6.7539 and the critical values of the χ^2 distribution with four degrees of freedom are 13.2767, 9.4877, and 7.7794, for a 1 percent, 5 percent, or 10 percent test, respectively.

D.3 Robustness: Timeframe

In table 12 we investigate whether the parameter estimates are constant over time. Column (5) of the table contains the baseline results reported in the main text. To construct the remaining columns, we repeat our estimation procedure for subsamples of trading days pertaining to February, March, April, and May of 2014. Although the some of the point estimates differ, the monopoly spread, which is our key counterfactual, remains quite stable across subsamples.

Table 12: Estimates for different months

	February	March	April	May	Full
A. parameter estimates					
α	7.33507	9.64332	4.92913	12.108	7.33504
	(1.05744)	(1.429)	(0.69578)	(1.73488)	(1.06129)
θ	1.52989	1.62702	1.56615	1.55160	1.52817
	(0.15613)	(0.16883)	(0.16855)	(0.16920)	(0.15127)
λ_i	0.00172	0.00184	0.00177	0.00166	0.00172
	(0.00023)	(0.00025)	(0.00023)	(0.00023)	(0.00022)
λ_j	0.00078	0.00075	0.00076	0.00073	0.00078
	(0.00013)	(0.00013)	(0.00012)	(0.00012)	(0.00012)
B. counterfactual spreads					
monopoly	2.22558	2.29256	2.24185	2.23249	2.22092
	(0.11048)	(0.11138)	(0.13515)	(0.10489)	(0.10647)
duopoly	2.88732	2.9060	2.83104	2.88767	2.87884
	(0.01248)	(0.00932)	(0.00911)	(0.01032)	(0.00480)

The point estimates are computed using the NLS procedure described in section 6.1. An observation is one second between 10:30 and 16:00 in one of the trading days in the sample. The estimates of column (5) are computed using the entire sample. The estimates of each column (1)–(4) are computed using only the trading days of a single month in the sample. The estimation was performed using SNOPT (Gill, Wong, Murray and Saunders, 2015). Standard errors are based on 200 block bootstrap replications.

E Additional Evidence

In appendix E.1, we argue the exchange choices of investors continue to appear to be distorted even under a more charitable benchmark for execution quality than was applied in the main text (*cf.* section 5.4). In appendix E.2, we demonstrate that the results we obtain in our analysis of the Chi-X shutdown (*cf.* section 7.2) survive even if we control for traded volume in various ways.

In appendix E.3, we present reduced form evidence of the own-price and cross-price elasticities, which are at the heart of the demand system for investors that is postulated by the model. Finally, in appendix E.4, we demonstrate that clustered trades predict subsequent price movements better than do isolated trades, which supports our use of isolation as a proxy for trades that are precipitated by liquidity-motivated investors and clustering as a proxy for trades that are precipitated by information-motived snipers.

E.1 Direct Evidence of Market Frictions: Additional Details

Section 5.4 documents that a significant number of isolated trades take place on exchanges offering inferior quotes. Specifically, 35 percent of isolated trades that occur when quotes differ take place on the exchange offering the worst quote. This is strong evidence of the existence of market frictions that distort the exchange choice of investors.

Moreover, it might be argued that this analysis *understates* these distortions by failing to fully consider the possibilities of order splitting. A more stringent benchmark for evaluating execution quality would be to compare the total price obtained in the trade against the total price that could have been obtained by splitting orders across exchanges in an optimal way. Every trade highlighted by the previous analysis fails to meet this benchmark, but additional trades would fail as well.

In contrast, a much less stringent benchmark would be to evaluate execution quality assuming an artificial constraint against order splitting. For every isolated trade, we determine which exchange or exchanges offered the best total price for a trade of that size at the time (i.e. the lowest total price in the case of an isolated buy, or the highest total price in the case of an isolated sell), and we compare that to the exchange on which the trade actually occurs.[47] The results of this analysis are tabulated in table 13. This table is completely analogous to table 3 in the main text, with the exception that the relevant price is the total price for a trade of that size rather than the quoted price (i.e. the price for the first traded share).

As the table indicates, a substantial number of isolated trades take place on exchanges offering inferior total prices. Aggregating across buys and sells, 9,698 isolated trades occur when the two exchange offer different total prices. Of those, 8.9 percent occur on the exchange offering the worse price. The magnitude of the frequency with which execution quality fails to meet even this extremely generous benchmark reinforces our conclusion that frictions are an empirically important determinant of exchange choice.

Table 13: Exchange choice for isolated trades, number of trades

	different price			identical price	total
	best price	inferior price	subtotal		
	(1)	(2)	(3)	(4)	(5)
BUY	4,533	441	4,974	2,572	7,546
SELL	4,299	425	4,724	2,497	7,221
BUY or SELL	8,832	866	9,698	5,069	14,767

The table classifies isolated trades that occur at ASX and at Chi-X by prevailing quotes at the two exchanges. The columns show the number of isolated trades in each of the following cases: (1) the trade occurs at the exchange with strictly better price; (2) the trade occurs at the exchange with strictly inferior price; and (4) the trade occurs when both exchanges quoted identical prices. Best price is defined as the lowest *total* price in the case of an isolated buy, or the highest *total* price in the case of an isolated sell.

[47] It is sometimes the case that it would have been infeasible to trade the same volume at the counterfactual exchange, due to insufficient depth. In all such cases, we classify the counterfactual exchange as offering an inferior total price.

E.2 Natural Experiment: Controlling for Volume

In terms of total volume, the monopoly day is in the bottom quartile of the trading days in June 2014. A potential concern is therefore that our findings are driven not by the Chi-X shutdown but by alternative factors that are correlated with low volume. Table 14 addresses this concern in two ways. First, we show that our results survive even if we restrict our analysis to samples of low-volume days. Columns (1) and (2) replicate columns (1) and (5) of table 6 for days in the bottom quartile with respect to total volume. Columns (3) and (4) do the same for days in the bottom half. Second, we show that our results survive even if volume is added as a control variable. Column (5) adds volume as a control but otherwise replicates column (1) of table 6.

Table 14: ASX spreads, Australian equity ETFs, June 2014

	(1)	(2)	(3)	(4)	(5)
	Volume: bottom quartile		Volume: bottom half		Full sample
A. Dependent Variable: quoted spread (cents)					
STW × Monopoly	-0.481^{++}	-0.376^{++}	-0.694^{++}	-0.561^{++}	-0.823^{++}
	(0.0642)	(0.0725)	(0.0607)	(0.0671)	(0.0601)
Volume					0.00410^{++}
					(0.000487)
STW Duopoly Mean	3.223	3.223	3.436	3.436	3.869
Change (Percent)	-14.93	-11.65	-20.20	-16.33	-21.27
B. Dependent Variable: cum-fee spread (cents)					
STW × Monopoly	-0.482^{++}	-0.376^{++}	-0.695^{++}	-0.561^{++}	-0.824^{++}
	(0.0642)	(0.0725)	(0.0607)	(0.0671)	(0.0601)
Volume					0.00410^{++}
					(0.000487)
STW Duopoly Mean	3.378	3.378	3.590	3.590	4.023
Change (Percent)	-14.27	-11.12	-19.35	-15.63	-20.48
C. Dependent Variable: spread (bps)					
STW × Monopoly	-0.909^{++}	-1.938^{++}	-1.336^{++}	-1.860^{++}	-1.575^{++}
	(0.125)	(0.209)	(0.119)	(0.194)	(0.117)
Volume					0.00800^{++}
					(0.000951)
STW Duopoly Mean	6.269	6.269	6.696	6.696	7.526
Change (Percent)	-14.50	-30.92	-19.95	-27.78	-20.93
Fixed Effects					
Day × Hour	NO	YES	NO	YES	NO
Security	YES	YES	YES	YES	YES
Control Group	NO	YES	NO	YES	NO
Observations	87600	788400	175200	1576800	350400

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.005$, $^{++}$ $p < 0.001$

In panels A, B and C, the dependent variable is, respectively, the quoted spread in cents, the cum-fee spread in cents, and the spread in bps prevailing on ASX at the beginning of the second. An observation is a second between 11:08 and 16:00 for a security traded on ASX during June 2014. The variable STW × Monopoly is an indicator for June 16, 2014 interacted with STW. In columns (1) and (2), the sample is restricted to trading days on which total STW volume was in the bottom quartile of days in June 2014. In columns (3) and (4), the sample is restricted to trading days on which total STW volume was in the bottom half of days in June 2014. Columns (2) and (4) include a control group consisting of the securities IOZ, ISO, MVW, QOZ, SSO, VAS, VLC, and VSO. In column (5), the sample includes all trading days in June 2014, and total STW volume is added as a control variable. Coefficients are estimated by ordinary least squares. Standard errors are clustered by 120 second blocks on each trading day. STW Duopoly Mean refers to the average spread pertaining to STW on non-monopoly days. Change (Percent) refers to the coefficient estimate relative to the STW duopoly mean.

E.3 Reduced Form Evidence of Own-Price and Cross-Price Elasticities

The driving force of the model is a demand system that governs the exchange choice of investors. In particular, the model predicts an own-price elasticity (i.e. fewer investors trade at an exchange when its prices become less favorable) as well as a cross-price elasticity (i.e. more investors trade at an exchange when prices at other exchanges become less favorable). In this appendix, we present reduced form evidence for the existence of these elasticities in the data.

Table 15 displays the coefficients of regressions that explain variation in the occurrence of isolated buys and sells at ASX and Chi-X in terms of the prices prevailing at the two exchanges. For column (2), we perform the regression

$$buy_{x,t} = \beta_0 + \beta_1(a_{-x,t} - a_{x,t}) + \beta_2(a_{x,t} - v_t) + \varepsilon_{x,t},$$

and similarly, for column (3), we perform the regression

$$sell_{x,t} = \beta_0 + \beta_1(b_{x,t} - b_{-x,t}) + \beta_2(v_t - b_{x,t}) + \varepsilon_{x,t}.$$

In both cases, the sample consists of all exchanges $x \in \{\text{ASX}, \text{Chi-X}\}$ and all seconds t between 10:30 and 16:00 in the 80 trading days of the sample. Moreover, because we do not observe v_t, we proxy with the average cum-fee mid price $(b_{\text{ASX},t} + b_{\text{Chi-X},t} + a_{\text{ASX},t} + a_{\text{Chi-X},t})/4$ in both cases. In the table, we refer to the difference in the own and other cum-fee asks (or the own and other cum-fee bids) as the "price difference," and we refer to the difference between the own cum-fee ask and v_t (or the own cum-fee bid and v_t) as the "half spread."

In addition, column (1) of the table reports the results of combining both regressions and estimating them together. Formally, we index the sides of the trade by $i \in \{\text{buy}, \text{sell}\}$. We then define $iso_{\text{buy},x,t} = buy_{x,t}$ and $iso_{\text{sell},x,t} = sell_{x,t}$. Likewise, we define $price\_difference_{\text{buy},x,t} = a_{-x,t} - a_{x,t}$ and $price\_difference_{\text{sell},x,t} = b_{x,t} - b_{-x,t}$. Finally, we define $half\_spread_{\text{buy},x,t} = a_{x,t} - v_t$ and $half\_spread_{\text{sell},x,t} = v_t - b_{x,t}$. Given these definitions, column (1) reports the results of the regression

$$iso_{i,x,t} = \beta_0 + \beta_1 price\_difference_{i,x,t} + \beta_2 half\_spread_{i,x,t} + \varepsilon_{i,x,t},$$

where the sample consists of both sides $i \in \{\text{buy}, \text{sell}\}$, all exchanges $x \in \{\text{ASX}, \text{Chi-X}\}$, and all seconds t between 10:30 and 16:00 in the 80 trading days of the sample. Moreover, we proxy for v_t with the average cum-fee mid price, as before.

Focusing on the estimates in column (1), we find that a one cent increase (decrease) in the ask (bid) on an exchange is associated with a decrease in the number of isolated buys (sells) on that exchange of approximately 1.5 per hour, as well as an increase in the number of isolated buys (sells) at the other exchange of approximately 2.3 per hour. Qualitatively similar results are found in column (2), where the focus is only on isolated buys, and in column (3), where the focus is only on isolated sells.

Table 15: Isolated trades as function of quotes

	(1) BUY or SELL	(2) BUY	(3) SELL
	(1)	(2)	(3)
price difference	0.000638^{++}	0.000548^{++}	0.000737^{++}
	(0.0000566)	(0.0000657)	(0.0000772)
half spread	-0.000405^{++}	-0.000379^{++}	-0.000431^{+}
	(0.000103)	(0.000111)	(0.000133)
Constant	0.00291^{++}	0.00293^{++}	0.00290^{++}
	(0.000149)	(0.000163)	(0.000193)
Observations	6336000	3168000	3168000

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, $^{+}$ $p < 0.005$, $^{++}$ $p < 0.001$

For columns (2) and (3), an observation is an exchange-second pair, where the exchange is either ASX or Chi-X, and the second is between 10:30 and 16:00 in one of the 80 trading days in the sample. The dependent variables are, respectively, indicators for the occurrence of an isolated buy or an isolated sell on that exchange in that second. For column (1), the estimation combines buys and sells. The independent variable "price difference" is defined as the difference between the own and other cum-fee bids or asks (see appendix E.3). The independent variable "half spread" is defined as the difference between the own cum-fee bid or ask and the mid price (see appendix E.3). Coefficients are estimated by ordinary least squares. Standard errors are clustered by 120 second blocks on each trading day.

E.4 Trade Clustering and Price Changes

Our estimation strategy relies upon using isolated trades as a proxy for the liquidity-motivated investor trades of the model, and conversely upon using clustered trades as a proxy for the information-motivated sniper trades of the model. If this approach is valid, then we should expect clustered trades to be better predictors of subsequent price movements in the same direction than their isolated counterparts. In this appendix, we present evidence to show that this relationship is indeed borne out in the data.

To that end, we define $R(\Delta)_t = (m_{t+\Delta} - m_t)/m_t$ to be the return over a time interval of length Δ, where m_t is the NBBO mid price. We define indicators $clusterBuy_t$ and $clusterSell_t$ for, respectively, whether a clustered buy or a clustered sell occurs on either exchange in second t. We also define $B_t = buy_{\text{ASX},t} + buy_{\text{Chi-X},t} + clusterBuy_t$ and $S_t = sell_{\text{ASX},t} + sell_{\text{Chi-X},t} + clusterSell_t$ to be indicators for, respectively, whether a buy or a sell occurs on either exchange in second t.

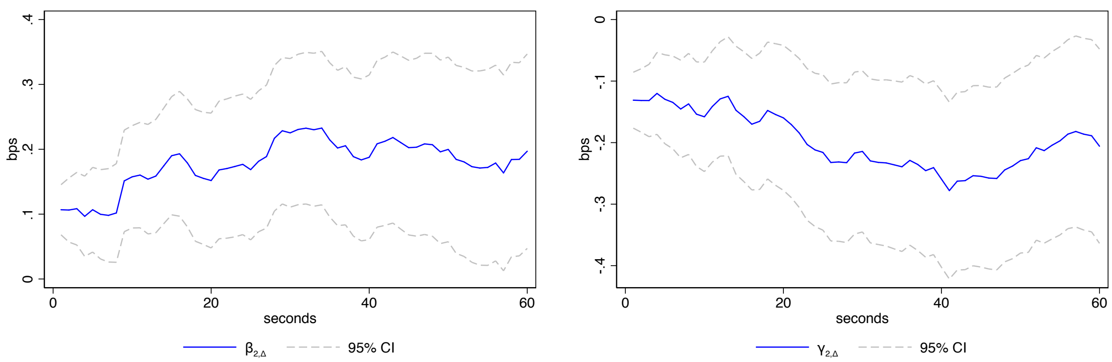
Then, for each value of Δ, we run the regression

$$R(\Delta)_t = \alpha_\Delta + \beta_{1,\Delta} B_t + \beta_{2,\Delta}\, clusterBuy_t + \gamma_{1,\Delta} S_t + \gamma_{2,\Delta}\, clusterSell_t + \varepsilon_t.$$

In each case, the sample consists of all seconds t in the 80 trading days of the sample that are between (i) 10:30 and (ii) Δ seconds before 16:00.

In Figure 7, we plot point estimates and 95 percent confidence intervals for $\beta_{2,\Delta}$ and $\gamma_{2,\Delta}$ for values of Δ up to one minute. As expected, we find that clustered trades are more predictive of price movements. The occurrence of a clustered buy (sell) in second t is associated with a 30-second return of 0.23 bps (−0.21 bps) more than an isolated buy (sell). Qualitatively, the result is robust to the choice of time horizon.

Figure 7: Predicting price changes using trade clustering



(a) Aggressive BUY (b) Aggressive SELL

The graphs plot point estimates for $\beta_{2,\Delta}$ and $\gamma_{2,\Delta}$ from the regression defined in the text for Δ ranging from 1 to 60 seconds. An observation is one second between (*i*) 10:30 and (*ii*) Δ seconds before 16:00 in one of the 80 trading days in the sample. 95 percent confidence intervals as based on standard errors clustered by 120 second blocks on each trading day.

F Extensions of the Model

This appendix discusses three extensions of the model. In appendix F.1, we allow for the possibility that each high-frequency trader faces a constraint on its net inventory. In appendix F.2, we add costs of operation for exchanges and liquidity providers. In appendix F.3, we consider an extension of the model in which the process governing the evolution of the value of the security is enriched to include heterogeneous jump sizes and a Brownian component. In appendix F.4, we relax the indivisibility constraint on investor demand, allowing them to split orders across exchanges. Finally, in appendix F.5, we allow for some agents who acquire and trade on short-lived private information. None of these extensions affects the expressions for the equilibrium spread. Consequently, our empirical findings remain valid under each.

Lastly, in appendix F.6, we explain in more detail how an imperfect ability to monitor prices in real time might produce market frictions of the nature assumed in the baseline model.

F.1 Inventory Constraints

In the baseline model, high-frequency traders face no limits in the amount of inventory that they are able to take on. This is admittedly somewhat unrealistic. In contrast, due to capital controls and a desire to limit their risk, high-frequency traders typically limit the extent to which they allow themselves to build up large inventories.

While there is a rich literature in market microstructure theory about market maker inventory management, most of those papers feature a monopolistic market maker (Stoll, 1978; Amihud and Mendelson, 1980; Ho and Stoll, 1981). In contrast, our model features an infinite number of high-frequency traders, any one of which can provide liquidity. As we argue here, inventory management is a less important determinant of the equilibrium spread when market making is competitive in this way. Under the additional assumption that any high-frequency trader can take on a small amount of inventory without cost, the equilibrium spread remains unchanged.

Suppose that each high-frequency trader faces inventory constraints that prohibit its net position from exceeding $K \geq 1$ shares long or short at any point in time.[48] Given this modification of the model, it is relatively straightforward to adapt the proofs of propositions 1 and 2 to show that the equilibrium spread remains unchanged. The strategies of high-frequency traders adjust so that they become inactive if their net inventory ever reaches $\pm(K-1)$ shares long or short. And if the liquidity provider ever becomes inactive, then one of the stale-quote snipers abandons its original strategy to assume the liquidity provider's role.

F.2 Cost of Operation

In the baseline model, neither exchanges nor traders must pay a cost of operation. In this extension, we add per-time costs of operation for exchanges and liquidity providers. Although one might have suspected that these operating costs would be an additional source of a spread, they affect neither the equilibrium spread nor therefore our empirical findings.

Formally, suppose that a high-frequency trader who has active limit orders on x exchanges must pay a monitoring cost of xc_1. Suppose also that an operation cost of c_2 must be paid by any active exchange. Also, amend assumption 3 to read:

Assumption 3′. $\lambda_i \left(1 - \frac{1}{\theta}\frac{\Sigma}{2}\right)\frac{\Sigma}{2} \geq \lambda_j X \left(\sigma - \frac{\Sigma}{2}\right) + X(c_1 + c_2).$

Given the modified assumption, it is relatively straightforward to extend the proofs of propositions 1 and 2 to show that the equilibrium spread remains unchanged. The strategies of the liquidity providers adjust to accommodate the inclusion of c_1 in their zero-profit conditions. Working backwards from this, it can be shown that the profits of each exchange, as a function of the cum-fee spreads, are $c_1 + c_2$ less than before. These additive constants fall out of the optimization problem, and the same spreads arise in equilibrium.

There are other costs (e.g. order processing costs and inventory costs) for which the analysis is less clean. Nevertheless, the findings discussed above provide us with some confidence that our empirical findings would not necessarily be diluted or reversed by enriching

[48] This specification can be thought of as a special case of a model in which high-frequency traders face a cost of inventory that is a function of their net position (which in itself can be thought of as reduced form for a model in which high-frequency traders are risk averse). Specifically, this corresponds to the special case in which the inventory cost is zero on $[-K, K]$ and prohibitively high outside that interval.

the model to include various costs, even those that are commonly interpreted as components of the spread.

F.3 Enriched Information Arrival Processes

In the baseline model, v_t is affected by jumps of only one size, σ, which arrive at the rate λ_j. While this is a quite simple model of price movement, the main conclusions remain unchanged even if the process is enriched in various ways.

In particular, the findings remain intact if the baseline jump process is augmented with a Brownian motion, possibly with time-varying volatility. Equilibrium strategies change only slightly under this extension: liquidity providers update their stale quotes not only after jumps, but also in a continuous fashion that tracks the Brownian motion. Stale-quote snipers, on the other hand, attempt to trade only upon jumps. Given this, it is relatively straightforward to establish that the equilibrium spread continues to be as described by propositions 1 and 2.

Furthermore, the findings would also remain unchanged if the size of each discrete jump were drawn independently from a distribution, F that satisfies the following modifications of assumptions 2 and 3:

Assumption 2″. $\theta < \min \operatorname{supp}(F)$.

Assumption 3″. $\lambda_i \left(1 - \frac{1}{\theta}\frac{\Sigma}{2}\right)\frac{\Sigma}{2} \geq \lambda_j X \left(\int \sigma dF(\sigma) - \frac{\Sigma}{2}\right)$, where Σ, as before, is defined in terms of the underlying parameters as follows:

$$\Sigma \equiv \begin{cases} \theta \left(1 + \dfrac{\lambda_j}{\lambda_i}\right) & \text{if } X = 1 \\ \theta + \dfrac{4\alpha}{X^2} - \sqrt{\theta^2 + \dfrac{16\alpha^2}{X^4} - \dfrac{8\alpha\theta\lambda_j}{X\lambda_i}} & \text{if } X \geq 2 \end{cases}$$

In words, assumption 2 is modified to apply to the minimum jump size, and assumption 3 is modified to apply to the expected jump size.

F.4 Order Splitting

In the baseline model, investor demand is indivisible: each investor is limited to buying or selling a single share at a single exchange. In this extension, we allow investors to split orders across exchanges in such a way that they buy or sell one share in total. The baseline equilibrium survives unchanged even if this form of order splitting is allowed.[49]

Suppose that an investor arriving at time t has the following available actions: (i) submit market orders for $y_x \geq 0$ shares at each exchange x, where $\sum_{x=1}^{X} y_x = 1$; (ii) submit market orders for $y_x \leq 0$ shares at each exchange x, where $\sum_{x=1}^{X} y_x = -1$; or (iii) submit market orders for $y_x = 0$ shares at each exchange x.

[49]We have not, however, ruled out the possibility that new equilibria would be introduced by this modification.

Supposing that liquidity providers continue to quote one share at the bid and one share at the ask, then an investor who arrives at time t and submits the orders $\boldsymbol{y} = (y_1, \dots, y_X)$ obtains utility

$$u_t(\boldsymbol{y}|\tilde{\theta}) = \begin{cases} v_t + \tilde{\theta} - \sum_{x=1}^{X} y_x a_{x,t} & \text{if } \sum_{x=1}^{X} y_x = 1 \\ \sum_{x=1}^{X} |y_x| b_{x,t} - v_t - \tilde{\theta} & \text{if } \sum_{x=1}^{X} y_x = -1 \\ 0 & \text{if } \sum_{x=1}^{X} y_x = 0 \end{cases}$$

As in the baseline model, investors do not necessarily act to maximize their utility. Instead, an investor who arrives at time t chooses orders $\boldsymbol{y} = (y_1, \dots, y_X)$ to maximize

$$\hat{u}_t(\boldsymbol{y}|\tilde{l}, \tilde{\theta}) = u_t(\boldsymbol{y}|\tilde{\theta}) - 2\alpha \cdot \sum_{x=1}^{X} |y_x| d(\tilde{l}, l_x)^2.$$

Note that conditional on choosing $\boldsymbol{y}$ so that $\sum_{x=1}^{X} y_x = 1$, it is optimal under $\hat{u}$ to trade the entire quantity at the exchange x that maximizes $-a_{x,t} - 2\alpha d(\tilde{l}, l_x)^2$. Likewise, conditional on choosing $\boldsymbol{y}$ so that $\sum_{x=1}^{X} y_x = -1$, it is optimal under $\hat{u}$ to trade the entire quantity at the exchange x that maximizes $b_{x,t} - 2\alpha d(\tilde{l}, l_x)^2$. In other words, the investor does not avail himself of the opportunity to split orders, as a result of the linear way in which $|y_x|$ enters $\hat{u}$, and his behavior remains as in the baseline model.

To show that the baseline equilibrium remains intact under this modification, it remains only to verify that no liquidity provider can profitably deviate by quoting less than one share of depth. Reducing depth in this way would reduce the volume of trade with stale-quote snipers, but it would also induce investors who would have traded at that exchange to reduce their demand by a proportionate amount. This leaves the liquidity provider's zero-profit condition unchanged, and so the deviation would not be profitable.

F.5 Private Information

In the baseline model, information is purely public. Nevertheless, the main conclusions remain unchanged even if short-lived private information is incorporated in a particular way.

In this extension, we augment the model by adding a trader ("the analyst") who may acquire private information. Jumps in the value of the asset continue to be of size σ and arrive at the rate λ_j. A fraction η of the jumps are, as in the baseline model, revealed publicly when they occur. The remaining $1 - \eta$ fraction of the jumps are revealed publicly only after an infinitesimal delay. But the analyst observes *every* jump when it occurs, providing him with short-lived private information in the latter cases. The analyst may trade at as many exchanges as he wishes, but is restricted to immediate-or-cancel orders.

In the equilibrium of this extension, the analyst submits immediate-or-cancel orders to each exchange to buy (sell) one share immediately after observing an upward (downward) jump. Equilibrium strategies of the liquidity providers change only slightly in this extension: after a trade against one of their orders, they wait for an infinitesimal length of time before replenishing the limit order book, updating the price if information arrives in the interim. Equilibrium strategies of investors and stale-quote snipers remain unchanged. Given this, it is relatively straightforward to show that the expressions for the equilibrium spreads prevailing under the limit order book remain as before.

There would, however, be changes to the expressions for the spreads prevailing under frequent batch auctions and non-cancellation delay (*cf.* appendix C.1). Intuitively, these alternative trading mechanisms remove the adverse selection due to stale-quote sniping on the basis of public information, but not the adverse selection due to the analyst's private information. Thus, while these mechanisms would continue to improve upon the limit order book in the presence of private information of this nature, the reduction in transaction costs would be less dramatic.

F.6 Market Friction Microfoundation

A key aspect of the model is that it allows for market frictions that distort the exchange choices of investors. In the main text, we are deliberately agnostic as to the precise source of these frictions. Nevertheless, we do suggest a number of potential sources, one of which is that it is inherently difficult to monitor prices in real time with perfect accuracy. In this appendix, we develop this explanation in greater detail.

For the purposes of this appendix, we focus on the case in which $X = 2$. The approach can be extended to accommodate larger values of X. As in the main text, denote the private transaction motive of an investor by $\tilde{\theta}$. Ignore, for the purposes of this appendix, the other component of the investor's type, $\tilde{l}$, which had designated the investor's location on the unit circle. For expositional ease, we focus the following discussion on investors with an inclination to buy (i.e. for whom $\tilde{\theta} \geq 0$) and, correspondingly, the ask-side of the book. Denote the asks at the two exchanges at time t by $a_{1,t}$ and $a_{2,t}$. Suppose, however, that investors are limited in their capacity to monitor these prices. This could be the case if their information comes from some lagged data feed (such as the SIP), or if so-called "fleeting orders" prevent investors from obtaining a clear view of the market.

As a stylized model of these limitations, suppose that rather than observing the two asks, an investor who arrives at time t observes only the noisy signals

$$\hat{a}_{1,t} = a_{1,t} + \varepsilon_t$$
$$\hat{a}_{2,t} = a_{2,t} - \varepsilon_t,$$

where $\varepsilon_t \sim U[-\alpha, \alpha]$. Suppose that the investor routes an immediate-or-cancel order with limit price $v_t + \tilde{\theta}$ to the exchange with best (i.e. lowest) signal.[50]

This behavior induces the same probability distribution over choices as would be obtained if investors of every type $(\tilde{l}, \tilde{\theta})$ were to optimize $\hat{u}_t(x, y|\tilde{l}, \tilde{\theta})$ as described in the main text, and we were then to integrate over $\tilde{l}$. In particular, conditioning on the true quotes $(a_{1,t}, a_{2,t})$, the investor routes to exchange 1 with probability $\max\left[0, \min\left(1, \frac{1}{2} + \frac{a_{2,t} - a_{1,t}}{\alpha}\right)\right]$ and to exchange 2 otherwise.

[50]Such behavior is optimal for the investor if he is constrained to send orders only at time t and if he possesses a prior distribution that is symmetric about $a_{1,t}$ and $a_{2,t}$.

G Notation and Variables

Table 16: List of Mathematical Notation

Name	Description
Parameters	
α	strength of frictions distorting exchange choice of investors
θ	maximum investor private transaction motive
λ_i	Poisson arrival rate of investors
λ_j	Poisson arrival rate of jumps
σ	jump size
X	number of exchanges
Other Notation	
v	fundamental value of security
s	cum-fee spread*
a	cum-fee ask*
b	cum-fee bid*
τ_{make}	make fee
τ_{take}	take fee
l	location on unit circle

* We also use $\hat{s}$, $\hat{a}$, and $\hat{b}$ for the quoted spread, ask, and bid, respectively.

Table 17: Variables Used in Main Estimation

Name	Description
$s_{x,t}$	cum-fee spread of STW on exchange x in second t (cents)
$a_{x,t}$	cum-fee ask of STW on exchange x in second t (cents)
$b_{x,t}$	cum-fee bid of STW on exchange x in second t (cents)
v_t	fundamental value of security*
$buy_{x,t}$	indicator for an isolated buy of STW on exchange x in second t
$sell_{x,t}$	indicator for an isolated sell of STW on exchange x in second t
$clustered_{x,t}$	indicator for a clustered trade of STW on exchange x in second t

* Proxied by $(a_{\text{ASX,t}} + b_{\text{ASX,t}} + a_{\text{Chi-X,t}} + b_{\text{Chi-X,t}})/4$.

References

Aitken, Michael, Haoming Chen, and Sean Foley, "The Impact of Fragmentation, Exchange Fees and Liquidity Provision on Market Quality," *Working Paper*, 2016. http://ssrn.com/abstract=2284635.

Amihud, Yakov and Haim Mendelson, "Dealership Market: Market-Making with Inventory," *Journal of Financial Economics*, 1980, *8* (1), 31–53.

_ , **Beni Lauterbach, and Haim Mendelson**, "The Value of Trading Consolidation: Evidence From the Exercise of Warrants," *Journal of Financial and Quantitative Analysis*, 2003, *38* (4), 829–846.

Arnold, Tom, Philip Hersch, J. Harold Mulherin, and Jeffry Netter, "Merging Markets," *The Journal of Finance*, 1999, *54* (3), 1083–1107.

ASX Group, "ASX Centre Point Block," 2012. http://www.asx.com.au/documents/products/asx_centre_point_factsheet.pdf.

_ , "ITCH - Glimpse Message Specification," ASX Market Information 2012.

_ , "ASX Operating Rules Procedures," 2016. http://www.asx.com.au/documents/rules/asx_or_procedures.pdf.

_ , "Managed Funds and ETP Product List," 2017. http://www.asx.com.au/products/etf/managed-funds-etp-product-list.htm.

Australian Securities and Investments Commission, "ASIC Market Integrity Rules (Competition in Exchange Markets)," 2011. http://asic.gov.au/regulatory-resources/markets/market-integrity-rules.

_ , "Report 452: Review of High-Frequency Trading and Dark Liquidity," 2015. http://download.asic.gov.au/media/3444836/rep452-published-26-october-2015.pdf.

_ , "List of Crossing Systems Registered with ASIC," 2016. http://www.asic.gov.au/crossing-systems.

Baldauf, Markus and Joshua Mollner, "High-Frequency Trading and Market Performance," *Working Paper*, 2017. http://ssrn.com/abstract=2674767.

Battalio, Robert H., "Third Market Broker-Dealers: Cost Competitors or Cream Skimmers?," *The Journal of Finance*, 1997, *52* (1), 341–352.

Bennett, Paul and Li Wei, "Market Structure, Fragmentation, and Market Quality," *Journal of Financial Markets*, 2006, *9* (1), 49–78.

Bernales, Alejandro, Italo Riarte, Satchit Sagade, Marcela Valenzuela, and Christian Westheide, "A Tale of One Exchange and Two Order Books: Effects of Fragmentation in the Absence of Competition," *Working Paper*, 2017.

Bessembinder, Hendrik and Herbert M. Kaufman, "A Cross-Exchange Comparison of Execution Costs and Information Flow for NYSE-Listed Stocks," *Journal of Financial Economics*, 1997, *46* (3), 293–319.

Boehmer, Beatrice and Ekkehart Boehmer, "Trading Your Neighbor's ETFs: Competition or Fragmentation?," *Journal of Banking & Finance*, 2003, *27* (9), 1667–1703.

Boneva, Lena, Oliver Linton, and Michael Vogt, "The Effect of Fragmentation in Trading on Market Quality in the UK Equity Market," *Journal of Applied Econometrics*, 2016, *31* (1), 192–213.

Branch, Ben and Walter Freed, "Bid-Asked Spreads on The AMEX and The Big Board," *The Journal of Finance*, 1977, *32* (1), 159–163.

Budish, Eric, Peter Cramton, and John Shim, "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, 2015, *130* (4), 1547–1621.

Chi-X Australia, "Operating Rules: Procedures," 2013. http://cmsau.chi-x.com/Portals/15/Compliance%20Page/CXA%20Operating%20Rules%20Procedures%20v1.5.pdf.

Chlistalla, Michael and Marco Lutat, "Competition in Securities Markets: The Impact on Liquidity," *Financial Markets and Portfolio Management*, 2011, *25* (2), 149–172.

Cohen, Kalman J. and Robert M. Conroy, "An Empirical Study of the Effect of Rule 19c-3," *Journal of Law and Economics*, 1990, *33* (1), 277–305.

Degryse, Hans, Frank de Jong, and Vincent van Kervel, "The Impact of Dark Trading and Visible Fragmentation on Market Quality," *Review of Finance*, 2015, *19* (4), 1587–1622.

Fink, Jason, Kristin E. Fink, and James P. Weston, "Competition on the Nasdaq and the Growth of Electronic Communication Networks," *Journal of Banking & Finance*, 2006, *30* (9), 2537–2559.

Fontnouvelle, Patrick De, Raymond P.H. Fishe, and Jeffrey H. Harris, "The Behavior of Bid-Ask Spreads and Volume in Options Markets During the Competition for Listings in 1999," *The Journal of Finance*, 2003, *58* (6), 2437–2464.

Foucault, Thierry and Albert J. Menkveld, "Competition for Order Flow and Smart Order Routing Systems," *The Journal of Finance*, 2008, *63* (1), 119–158.

Gajewski, Jean-François and Carole Gresse, "Centralised Order Books Versus Hybrid Order Books: A Paired Comparison of Trading Costs on NSC (Euronext Paris) and SETS (London Stock Exchange)," *Journal of Banking & Finance*, 2007, *31* (9), 2906–2924.

Gill, Philip E., Elizabeth Wong, Walter Murray, and Michael A. Saunders, "User's Guide for SNOPT Version 7.5: Software for Large-Scale Nonlinear Programming," 2015. http://www.cam.ucsd.edu/~peg/papers/sndoc7.pdf.

Hall, Peter, Joel L. Horowitz, and Bing-Yi Jing, "On Blocking Rules For the Bootstrap with Dependent Data," *Biometrika*, 1995, *82* (3), 561–574.

Hamilton, James L., "Marketplace Fragmentation, Competition, and the Efficiency of the Stock Exchange," *The Journal of Finance*, 1979, *34* (1), 171–187.

Haslag, Peter H. and Matthew Ringgenberg, "The Causal Impact of Market Fragmentation on Liquidity," *Working Paper*, 2016. http://ssrn.com/abstract=2591715.

He, Peng William, Elvis Jarnecic, and Yubo Liu, "The Determinants of Alternative Trading Venue Market Share: Global Evidence from the Introduction of Chi-X," *Journal of Financial Markets*, 2015, *22*, 27–49.

Hendershott, Terrence and Charles M. Jones, "Island Goes Dark: Transparency, Fragmentation, and Regulation," *Review of Financial Studies*, 2005, *18* (3), 743–793.

Ho, Thomas and Hans R. Stoll, "Optimal Dealer Pricing Under Transactions and Return Uncertainty," *Journal of Financial Economics*, 1981, *9* (1), 47–73.

Mayhew, Stewart, "Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets," *The Journal of Finance*, 2002, *57* (2), 931–958.

Menkveld, Albert J., "High Frequency Trading and the *New Market* Makers," *Journal of Financial Markets*, 2013, *16* (4), 712–740.

Neal, Robert, "Potential Competition and Actual Competition in Equity Options," *The*

Journal of Finance, 1987, *42* (3), 511–531.

Nguyen, Vanthuan, Bonnie F. Van Ness, and Robert A. Van Ness, "Short-and Long-Term Effects of Multimarket Trading," *Financial Review*, 2007, *42* (3), 349–372.

Nielsson, Ulf, "Stock Exchange Merger and Liquidity: The Case of Euronext," *Journal of Financial Markets*, 2009, *12* (2), 229–267.

O'Hara, Maureen and Mao Ye, "Is Market Fragmentation Harming Market Quality?," *Journal of Financial Economics*, 2011, *100* (3), 459–474.

Stoll, Hans R., "The Supply of Dealer Services in Securities Markets," *The Journal of Finance*, 1978, *33* (4), 1133–1151.

Weston, James P., "Electronic Communication Networks and Liquidity on the Nasdaq," *Journal of Financial Services Research*, 2002, *22* (1/2), 125–139.