Market power on exchanges: linking price impact to trading fees*

Sarah Draus†

Abstract

Recent regulations, aimed at making trading and exchange services more competitive, triggered the proliferation of competing trading venues and resulted in high order flow fragmentation. This paper demonstrates that fragmentation allows liquidity providers and exchanges to retain market power. The results are derived in the framework of a theoretical duopoly competition model, in which a rational trader fragments his order to reduce price impact whenever exchanges are not completely liquid. This lowers the price elasticity of both liquidity provider's asset demand and exchange trading volume, inducing mark-ups on transaction prices and on exchange trading fees. Surprisingly, the theoretical results indicate that less competitive liquidity provision feeds back into higher exchange trading fees in equilibrium. Moreover, considering the cross-section of exchanges, the model predicts that the exchange with better liquidity charges a higher trading fee and attracts larger market shares. These results are consistent with anecdotal evidence and deliver empirical implications for the effect of introducing exchange competition on implicit and explicit trading costs.

JEL Codes: G20, G18, L10

Keywords: stock exchanges, competition, order fragmentation, trading fee, liquidity, regulation

*Comments and suggestions by Dion Bongaerts, Jean-Edouard Colliard, Marta Szymanowska, Mark Van Achter, Wolf Wagner and participants at the 2nd Dauphine Microstructure Workshop are gratefully acknowledged.

†Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands, email: draus@rsm.nl

1 Introduction

This paper studies the interplay between exchange competition and liquidity provision in a fragmented market. Several regulations (e.g. Reg NMS in the US or the European Directive on Markets in Financial Instruments (MiFID) in the European Union (EU), both implemented in 2007) have resulted in the proliferation of trading venues competing for order flow with incumbent dominating exchanges. The case of the EU is striking, since most exchanges enjoyed a monopolistic position prior to the implementation of MiFID.[1] To justify MiFID, European institutions argued that enhanced competition between trading venues would lead to substantially cheaper trading infrastructures for investors[2]. Following MiFID, fragmentation of order flow rose rapidly and fees decreased. Interestingly, the traditional exchanges retained large market shares and maintained fees at a higher level than their competitors. The London Stock Exchange (LSE) and Euronext, for instance, had market shares between 60% and 70% over the year 2017.[3] Their trading fees, ranging between 0.45 bps and 0.20 bps on the LSE and between 0.95 bps and 0.45 bps on Euronext (in its default pricing scheme), were higher than those of their main competitor, CBOE Europe Equities, which charged a total fee of about 0.15 bps on its lit order books.[4] Such dispersion in trading fees is difficult to explain with traditional price competition models, since exchanges use increasingly similar trading technologies, most market participants are large professional traders with easy access to several exchanges, and trading is done in an almost perfectly homogeneous product (shares).

This paper develops a model of exchange competition that accounts for illiquidity in asset trading. It demonstrates that exchanges competing for order flow in identical shares retain market power when a trader benefits from splitting his order between multiple exchanges due to illiquidity. Equilibrium trading fees are then lower than monopoly

[1] Some European exchanges (e.g. Paris Bourse) benefited from the "concentration rule". It provided that trading takes place only on regulated exchanges that admitted the asset, and was abolished by MiFID. In countries without the concentration rule, large exchanges were de facto monopolies, as described in Boneva, Linton, and Vogt (2016) for the London Stock Exchange and in Gomber (2016) for Deutsche Boerse. These exchanges started facing strong competition after MiFID's adoption. In the US, in contrast, alternative trading platforms existed well before the implementation of Reg NMS. The latter boosted competition between different trading venues substantially.

[2] Two major novelties in MiFID triggered competition between exchanges: facilitated cross border transactions and the creation of a new type of trading venues called Multilateral Trading Facilities (MTFs). Official statements are for instance the document MEMO/06/57 from 06/02/2006 or the speech "Preparing for MiFID" (SPEECH/06/430) from 30/06/2006.

[3] Source: http://fragmentation.fidessa.com, visited on 14/12/2017.

[4] Sources: Trading fee guide for cash market members issued by Euronext, Trading services price list issued by the LSE, and Trading price list issued by CBOE Europe Equities (previously ChiX-Europe until 2011 and Bats Europe between 2011 and 2017).

fees, but strictly higher than those implied by perfect competition. Consistent with the aforementioned observations, the results derived from the model imply that exchanges with cheaper liquidity provision charge higher fees. Moreover, those results suggest that exchanges set fees such as to handle only a portion of total trading volume, leaving the remaining volume to their competitor. Hence, non-competitive fees induce order flow fragmentation.[5]

The results are derived in a three-stage model with strategic and profit-maximizing agents. In the first stage, two competing exchanges set trading fees. In the second stage, competing liquidity providers on each exchange set the price schedules at which they are willing to trade. In the third stage, a discretionary liquidity trader determines how much to trade on which exchange and with which liquidity provider, and trading takes place. Liquidity providers bear inventory holding costs and post therefore quantity contingent price schedules. The discretionary trader faces a trade-off between minimizing price impact (by splitting his order among multiple liquidity providers and therefore multiple exchanges) and minimizing exchange fees (by concentrating his order on the exchange with the lowest trading fee).

The main theoretical results are summarized in the following paragraphs, considering first the decisions of liquidity providers, and thereafter the decisions of exchanges. Liquidity providers anticipate the incentive of the trader to split his order, which allows them to obtain market share even with non-competitive price schedules. As a consequence, all price schedules are strictly steeper than competitive ones in equilibrium. Their slope increases the higher inventory holding costs are and the fewer liquidity providers are active on the market. This result generalizes findings in Bernardt and Hughson (1997), Biais, Martimort, and Rochet (2000) and Bondarenko (2001), who demonstrate that competing liquidity providers charge strictly positive mark-ups in trading models with asymmetric information.

Turning to competition between exchanges, this paper shows that illiquidity gives market power to exchanges. Indeed, steeper price schedules strengthen the incentive of the discretionary trader to split his order. This reduces the trader's sensitivity to the difference in trading fees. The trader's utility loss if a fraction of his order is executed on the exchange with the highest fee, is then over-compensated by the reduction of his total price impact. From the point of view of exchanges, illiquidity reduces the fee elasticity

[5] Alternative explanations such as collusion among exchanges, barriers to entry or strong liquidity externalities seem unlikely in this context. Collusion is prohibited under MiFID, barriers to entry for new trading venues have been substantially reduced in the past decades (Macey and O'Hara, 2005), and liquidity externalities would have led to high order flow concentration (Pagano, 1989).

of trading volume, lowering thereby the competitive pressure on trading fees. Therefore, equilibrium trading fees are strictly larger than competitive ones. Moreover, they increase the less competitive liquidity provision is. In equilibrium, it is profitable for an exchange to set high fees at which it satisfies only a portion of the total trading demand, leaving the remaining to its competitor. A high degree of fragmentation is thus not the driver of strong competition, but the result of imperfect competition in fees between exchanges.

These results imply that exchanges achieve their highest profit with a moderate level of liquidity. Indeed, bad liquidity harms the trader twice: directly due to large price impacts and indirectly through high trading fees. This reduces his asset demand and thus trading volume on both exchanges. On the other hand, good liquidity lowers exchange fees and thus trading revenue per traded share. It is questionable whether exchanges can influence their level of liquidity by influencing sufficiently inventory holding costs, as these depend to a large extent on individual attributes of liquidity providers and on the properties of the traded asset. However, exchanges may be able to exert some influence on the size of the liquidity provision sector, and thereby on the competitiveness of price schedules. As an example, the second version of the European directive MiFID (that came into force in January 2018), provides that regulated exchanges determine rules for liquidity provision and sign explicit agreements with all financial firms pursuing market making strategies.[6] Moreover, exchanges can determine the space and technology available for colocation of high frequency traders, who act often as liquidity providers.[7] The complexity of this model prevents any explicit analysis of the optimal size of the liquidity provision sector from the point of view of exchanges. However, a numerical simulation of equilibrium exchange profits (derived from the model assumptions) suggests that exchanges achieve high profits by limiting the number of liquidity providers, in particular when inventory holding costs are small.

New empirical implications can be derived from this model. The first set of implications relates to the effects that shocks on liquidity supply (changes in overall liquidity due to e.g. regulations or new technologies) have on exchange decisions. The model predicts that a positive shock to liquidity supply on one exchange triggers a decrease in trading fees on the competing exchange. Moreover, such a shock decreases strongly the competitor's trading revenue in equilibrium. It induces also exchanges to restrict the number of liquidity providers. The second set of implications is about the relation between different

[6] Directive 2014/65/EU, article 48, paragraphs 2,3

[7] See Menkveld (2016) and Brogaard, Hagströmer, Nordén, and Riordan (2015) for descriptions and analyses of high frequency trading strategies.

types of trading costs. In a time-series perspective, the results show that an oligopolistic exchange industry structure generates a positive correlation between implicit trading costs (illiquidity) and explicit trading costs (exchange fees). This is in contrast to monopolistic exchanges that would raise explicit trading costs the lower implicit costs are. In addition, considering the cross-section of exchanges, the most liquid exchanges are expected to charge the highest total trading fees and to attract the largest market share. More generally, these results indicate that introducing competition in the exchange industry may not have a large impact on exchange fees per se. Rather, market-wide liquidity improvements combined with exchange competition reduce strongly trading costs.

This paper contributes to the literature on the pricing of exchange services. Few papers study pricing decisions of competing exchanges. They mostly analyze fees as tools to cater to different types of traders or firms. Foucault and Parlour (2004), for instance, show that two exchanges competing for listings set different listing fees and trading technologies: the exchange with the most efficient trading technology has the highest listing fee and attracts initial public offerings from the most productive firms. Chao, Yao, and Ye (2017) demonstrate that identical competing limit order books structure their maker/taker fees differently in order to cater to different types of liquidity providers: those with larger gains from trade post limit orders on the exchange with the higher maker fee and the higher execution probability. Recent papers have focussed particularly on rationales explaining the maker/taker fee scheme and its consequences on market quality, as these fee schemes were introduced on alternative trading venues and on some large incumbent exchanges around the introduction of Reg NMS and MiFID (see, e.g., Panayides, Rindi, and Werner, 2017; Foucault, Kadan, and Kandel, 2013; Colliard and Foucault, 2012). In contrast, the focus in this paper is deliberately placed on the role of illiquidity, observed on all trading venues, in mitigating exchange competition. Thereby, this paper highlights that two commonly made assumptions are not innocuous with regard to the pricing of trading services. First, restricting market participants to trade one indivisible unit on one or another exchange (e.g. in Colliard and Foucualt, 2012) omits that order splitting is rational in the presence of illiquidity and that order splitting induces non-competitive fees. Second, the assumption of competitive market makers (e.g. in Kyle, 1985; Chowdhry and Nanda, 1991; Baruch, Karolyi, and Lemmon, 2007) does not correspond to the optimal strategic choice of competing liquidity providers whenever traders benefit from trading with multiple market makers (and on several venues). This assumption artificially strengthens competition in a model that incorporates endogenous exchange trading fees.

In addition, this paper contributes to the literature on the industrial organization of exchanges. It highlights the importance of integrating illiquidity, specific to financial asset prices that are usually contingent on the traded quantity, into traditional models of industrial organization in order to understand exchange competition. Non-competitive equilibrium prices are obtained in traditional competition models assuming differentiated goods and a constant unit price (see, e.g., Shaked and Sutton, 1982, or Sing and Vives, 1984). This model does not rely on product differentiation, in line with the fact that exchanges become increasingly similar (as pointed out in Chao, Yao, and Ye, 2017). Indeed, the traded asset is identical for all traders and equally accessible on both exchanges. Considering the trading service sold by exchanges, understood as the aggregation of exchange-wide liquidity provision, it may differ depending on inventory holding costs and on the number of liquidity providers per exchange. However, these differences are irrelevant for the results derived in this paper. Rather, it is illiquidity alone that triggers non-competitive pricing at both the level of liquidity provision and the level of exchange competition (although exchanges charge a constant fee per traded share similarly to firms in the standard models mentioned above). Moreover, imperfect competition in liquidity provision reinforces imperfect competition in exchange trading services.

Literature on price competition for homogeneous goods (with constant unit prices) concludes that prices are competitive in equilibrium, unless exogenous capacity constraints prevent single firms from satisfying the entire demand (see, e.g., Bertrand, 1883; Levitan and Shubik, 1972; Allen and Hellwig, 1986; Dudey, 1992). In this model, single liquidity providers (and by extension exchanges) can satisfy the entire liquidity demand of the discretionary trader. Although inventory holding costs limit their risk absorption capacity, the quantity contingent price schedules at which they are willing to trade ensure non-negative profits for any traded quantity. Thus, imperfect competition and fragmentation are caused by the trader's unwillingness to trade with a single liquidity provider, rather then by the liquidity provider's inability to satisfy the entire asset demand of the trader.

More generally, differentiation in trading services is often advanced as an explanation for the coexistence of several exchanges. Indeed, liquidity externalities alone would naturally lead to the consolidation of exchanges (Pagano, 1989) and act as a barrier to entry (Economides and Siow, 1988). With differentiation in trading services, trading venues can cater to different trader groups, creating scope for coexistence of exchanges (see, e.g., Babus and Parlatore, 2017; Pagnotta and Philippon, 2011; Chemmanur and Fulghieri, 2006; Foucault and Parlour, 2004). This paper demonstrates that differentiation is not a necessary condition for the coexistence of exchanges. Indeed, easy access to multiple exchanges (made possible by new technology and encouraged by regulations such as MiFID[8]) also creates scope for the profitable coexistence of non-differentiated exchanges.

The remainder of the paper is organized as follows. Section 2 sets out the model. Section 3 studies the quoting decisions of liquidity providers and the trading decision of the discretionary trader. Section 4 analyzes the optimal decisions of exchanges. Section 5 concludes the paper.

2 Model setup

The economy has two stock exchanges, A and B, on which a single risky asset can be traded in a unique trading round. The payoff of the asset has mean μ normalized to zero and variance $\sigma^2 > 0$. A rational liquidity trader (labelled "the trader" from here onwards) can trade simultaneously on both exchanges against price schedules determined strategically by liquidity providers. Each exchange charges a trading fee per traded share, f_j with $j = \{A, B\}$. The game is sequential: first exchanges set their trading fees simultaneously, then liquidity providers post price schedules simultaneously, and eventually the trader determines his asset demand and trading takes place. At the end of the game, the value of the asset is realized.

2.1 Exchanges

Exchanges offer a trading platform to liquidity providers and to the trader. They obtain revenue by charging a volume-based trading fee and do not bear any costs. The profit of exchange $j \in \{A, B\}$ corresponds to its revenue from trading:

$$\Pi_j = V_j f_j \tag{1}$$

where V_j is the traded volume. Exchanges determine their trading fees to maximize their profit. This results in a Nash equilibrium in fees denoted by (f_A^*, f_B^*) at the first stage of the game.

[8] For instance, the provision on best execution in article 27 of the second version of MiFID (Directive 2014/65/EU) states the "obligation to execute orders on terms most favorable to the client", which may imply routing parts of an order to multiple venues.

The model is solved assuming that trading fees are paid only by the trader and not by liquidity providers. This assumption is in line with common practice among large exchanges and trading platforms. For instance, the LSE and CBOE Europe Equities charge no fee for adding liquidity to the lit market and a strictly positive fee for removing it. Literature has demonstrated that the split of the total trading fee into maker and taker fees matters particularly in combination with a discrete price grid (Colliard and Foucault, 2012; Foucault, Kadan, and Kandel, 2013). In that case, liquidity providers cannot pass the entire maker fee on to liquidity takers, and such a fee structure helps enhancing liquidity provision. In this model, however, price schedules are continuous, giving therefore the possibility to liquidity providers to pass all fees their would pay entirely on to the trader. Hence, trading fees can be interpreted as the total fees collected by exchanges.[9]

2.2 Liquidity providers

Liquidity is provided by N_j market makers on exchange j. All liquidity providers post price schedules simultaneously, before trading takes place. In order to keep the model tractable and to ensure closed form solutions at this stage of the game, liquidity supply is assumed to be segmented: market makers on exchange A cannot post price schedules on exchange B and vice versa. However, price schedules are determined strategically, taking into account the actions of competing liquidity providers on both exchanges.[10] Liquidity provision is modeled following the inventory holding costs framework (Stoll, 1978; Ho and Stoll, 1983). All liquidity providers bear a quadratic inventory holding cost that might differ by exchange, but is identical for all market makers on a specific exchange, $\gamma_j > 0$. They start the game with an empty inventory. Hence, the utility of liquidity provider i from executing the portion q_{ji} of volume V_j is:

$$U_{ji} = p_{ji}q_{ji} - \gamma_j q_{ji}^2 \tag{2}$$

[9]Solving this model with distinct maker and taker fees is intractable due to imperfect competition. However, economic intuition suggests that such a fee split is irrelevant for the results. Indeed, liquidity providers would always pass their entire costs on to the trader. Distinct maker fees would eventually only affect the decisions of the trader, in the same way as taker fees. Hence, exchanges would compete in their total trading fee.

[10]The liquidity provision game is intractable if strategic liquidity providers can post price schedules on every exchange and exchanges charge trading fees. In the absence of exchange trading fees, and considering only one liquidity provider per exchange, strategic cross-market quoting leads to multiple equilibria in price schedules. In addition, segmented liquidity provision is a common assumption in the models analyzing multi-market trading (see, e.g., Chowdhry and Nanda, 1991; Baruch, Karolyi, and Lemmon, 2007; Cespa and Colla, 2017).

Inventory holding costs may arise due to the risk that the value of the inventory position changes adversely as a result of asset price movements. In that case, the utility function U_{ji} derives from a setting where liquidity providers are risk averse with a CARA utility and the asset payoff is normally distributed. The parameter γ_j captures then the product of individual risk aversion with the variance of the asset payoff. Inventory holding costs may also arise due to margin constraints or risk management constraints associated with the market making activity. Literature has shown that such constraints induce risk averse behavior (Nagel, 2012).

Following the methodology in Bernhard and Hughson (1997) (BH in what follows), I assume that all price schedules, p_{ji}, are linear with slopes $a_i > 0$ for intermediary i on exchange A and $b_i > 0$ for intermediary i on exchange B, and $p_{ji}(0) = \mu = 0$. Hence, price schedule p_{ji} is upward sloping in q_{ji} if the trader buys ($q_{ji} > 0$) and downward sloping if he sells ($q_{ji} < 0$). Liquidity providers determine the slopes of their price schedules simultaneously to maximize their utility. This results in a Nash equilibrium in price schedule slopes, denoted $\left(a_1^*, ..., a_{N_A}^*, b_1^*, ..., b_{N_B}^*\right)$, at the second stage of the game. This setting features always strictly positive price impact. Indeed, if liquidity providers were to set competitive price schedules, the price impact would be such as to compensate them for the inventory holding costs: $a_i = \gamma_A$ and $b_i = \gamma_B$. BH apply this methodology to solve a trading model with market power among liquidity providers and information asymmetry. I transpose it to an inventory holding model to ensure tractability in the first stage of the game. Indeed, BH's setting, which focusses on the quoting decisions of intermediaries, leads to solutions that are intractable in the current setting where exchange trading fees are determined strategically at an earlier stage. However, the properties of price schedules are similar in both settings.

2.3 Discretionary liquidity trader

The trader is risk averse with mean-variance utility and risk aversion parameter $\rho > 0$. He has a private value of the asset, v, that differs from μ for exogenous reasons. Whenever $v > 0$ ($v < 0$), the trader buys (sells) the asset provided that trading costs (price impact and trading fees) do not offset the gain from trade. The results are derived assuming that $v > 0$, such that the asset demand per liquidity provider is $q_{ji} \geq 0$. The opposite case is symmetric. The trader starts the game without an initial endowment in the asset. If he decides to trade, he can split his order between exchanges and between intermediaries such as to trade against a small portion of many price schedules. Since the trader buys the asset if he trades, he faces upward sloping price schedules on both exchanges, $p_{Ai} = a_i q_{Ai}$ and $p_{Bi} = b_i q_{Bi}$. His utility is then:

$$U_T = \sum_{j \in \{A,B\}} \sum_{i=1}^{N_j} (v - f_j) q_{ji} - \sum_{i=1}^{N_A} a_i (q_{Ai})^2 - \sum_{i=1}^{N_B} b_i (q_{Bi})^2 - \rho \left(\sum_{j \in \{A,B\}} \sum_{i=1}^{N_j} q_{ji} \right)^2 \tag{3}$$

A necessary condition for the trader to participate is $v - f_j > 0$. The trader determines his exchange- and intermediary specific asset demand at the third stage of the game, such as to maximize his utility. His orders are then routed to the chosen exchanges and intermediaries, and executed.

3 Trading equilibrium

3.1 Trader's asset demand

I first analyze trading decisions, taking exchange fees as given. The trader determines how many shares to buy from which liquidity provider, after he observes the exchange trading fees set in the first stage, and the price schedules posted by liquidity providers in the second stage. His optimization problem is:

$$q_{ji}^* = \arg\max_{q_{ji} \geq 0} U_T \tag{4}$$

with U_T detailed in equation 3. Due to risk aversion, the trader's marginal utility decreases in the total quantity of shares, leading to an optimal asset holding that is finite also in the absence of trading costs. Both types of trading costs considered here, illiquidity and trading fees, lower the trader's utility of holding the asset, and hence the quantity of shares he optimally buys. The trader's main concern is then to determine his asset demand per exchange and per liquidity provider such as to minimize trading costs. Therefore, the following discussion focusses exclusively on the relation between q_{ji} and trading costs.

With linear price schedules, the price impact incurred by the trader, when he executes his order with a single liquidity provider, is convex in the order size. If the trader splits his order such as to buy from many liquidity providers, he needs to walk up only a small part of each single price schedule. This lowers his total price impact. Hence, the only strategy that allows to minimize price impact is to split the order between all liquidity providers.

However, reaching all liquidity providers on both exchanges comes at the cost of higher exchange trading fees, if one exchange charges a higher fee than the other. Hence, the more trading fees differ, the less evenly the trader splits his order between exchanges. In equilibrium, the asset demands per liquidity provider (and per exchange) balance savings on the price impact with savings on exchange fees. The following proposition characterizes the trader's optimal decisions.

Proposition 1 *In equilibrium, the trader's asset demand per liquidity provider is on exchange A:*

$$q_{Ai}^{*} = \frac{1}{\Phi}\left(\frac{v - f_A}{2a_i} - \frac{(f_A - f_B)\,\rho}{2a_i}\sum_{n=1}^{N_B}\frac{1}{b_n}\right), \quad \forall i = 1, ..., N_A \tag{5}$$

and on exchange B:

$$q_{Bi}^{*} = \frac{1}{\Phi}\left(\frac{v - f_B}{2b_i} - \frac{(f_B - f_A)\,\rho}{2b_i}\sum_{n=1}^{N_A}\frac{1}{a_n}\right), \quad \forall i = 1, ..., N_B \tag{6}$$

with

$$\Phi = \left[1 + \rho\left(\sum_{n=1}^{N_A}\frac{1}{a_n} + \sum_{n=1}^{N_B}\frac{1}{b_n}\right)\right] \tag{7}$$

Proof. See appendix ■

As expected, q_{Ai}^{*} and q_{Bi}^{*} diminish in risk aversion. Moreover, asset demand directed to liquidity provider i on exchange j decreases in his own price impact (a_i for liquidity providers on exchange A and b_i for those on exchange B) and in the trading fee of the liquidity provider's exchange. Asset demand increases the less competitive all other liquidity providers are (captured by Φ that increases in a_n and b_n).

The consolidated order flow per exchange diminishes in the exchange's own trading fee and increases in the competitor's fee. Interestingly, the responsiveness of exchange specific volume to changes in trading fees depends crucially on the liquidity level on both exchanges. The steeper price schedules are on exchange A or B, the more profitable it is for the trader to split his order between all liquidity providers. This reduces his sensitivity to the difference in exchange trading fees. Hence, exchange volume becomes less responsive to changes in fees. This implies that the strength of competition between exchanges depends crucially on the level of overall market liquidity, and therefore on the strength of competition between liquidity providers or on inventory holding costs. A single exchange can attract the entire order, only if it sets a small enough trading fee and offers a good enough overall liquidity, as specified in the following lemma from the point of view of exchange A.

Lemma 1 *The trader directs his entire order to exchange A if*

- $f_B > v$ *and* $f_A < v$

or

- $f_B \leq v$ *and the two following conditions hold simultaneously:*

$$f_A < f_B \tag{8}$$

and

$$\sum_{n=1}^{N_A} \frac{1}{a_n} > \frac{(v - f_B)}{\rho \left(f_B - f_A\right)} \tag{9}$$

Proof. See appendix ■

3.2 Liquidity provision

Turning to the decision of liquidity providers, the construction of the equilibrium is illustrated from the perspective of a given intermediary k on exchange A. This intermediary conjectures schedules of all competitors on his exchange, $p_{Ai} = a_i q_{Ai}$ with $i \in \{1, ..., N_A\} \setminus \{k\}$, and on the competing exchange, $p_{Bi} = b_i q_{Bi}$ with $i = 1, ..., N_B$. Given these conjectures and given the best response of the trader at the next stage, the intermediary chooses his own schedule to maximize his utility:

$$a_k^* = \arg\max_{a_k} U_{Ak}\left(q_{Ak}^*\right) \tag{10}$$

with U_{Ak} defined in equation 2. The first order derivative of the utility function is equal to zero whenever the following equality holds:

$$2\gamma_A - a_k + \frac{2\rho\left(a_k - \gamma_A\right)}{\Phi a_k} = 0 \tag{11}$$

where the parameter Φ (defined in equation 7) captures the strategic interaction with all other liquidity providers. A symmetric reasoning applies for any intermediary on exchange B. The condition for the first order derivative of the utility function to be zero is than equation 11 replacing γ_A by γ_B and a_k by b_k. Since intermediaries differ between exchanges but are identical on a given exchange, the equilibrium in slopes is symmetric per exchange:

$$a_i = a,\ \forall i = 1, .., N_A \text{ and } b_i = b,\ \forall i = 1, .., N_B \tag{12}$$

With symmetric equilibria, first order conditions are satisfied when the following equalities hold:

$$\frac{2\rho}{\Phi^s} = a\frac{a - 2\gamma_A}{a - \gamma_A} \tag{13}$$

$$\frac{2\rho}{\Phi^s} = b\frac{b - 2\gamma_B}{b - \gamma_B} \tag{14}$$

where $\Phi^s = 1 + \rho(N_A/a + N_B/b)$. Solving equation 13 for a and equation 14 for b yields the best response functions (their expressions are provided in the appendix, in the proof of proposition 2). Equations 13 and 14 reveal some essential properties of equilibrium schedule slopes. First, equilibrium price schedules do not depend on exchange trading fees, as a consequence of the assumption of segmented liquidity provision. Second, the left hand side of both equalities ($2\rho/\Phi^s$) is strictly positive and approaches zero as liquidity providers become infinitely numerous on one or both exchanges. This property must hold for the right hand side of both equations, implying:

$$a^* > 2\gamma_A \text{ and } b^* > 2\gamma_B \tag{15}$$

and

$$\lim_{N_j \to \infty} a^* = 2\gamma_A \text{ and } \lim_{N_j \to \infty} b^* = 2\gamma_B \text{ with } j = A, B \tag{16}$$

Equilibrium price schedules are always steeper than competitive ones. Liquidity providers always obtain a strictly positive utility. Third, the right hand sides of the equalities are increasing in the respective slopes (a in equation 13 and b in equation 14), and they must be equal in equilibrium. Consequently schedule slopes are strategic complements: the steeper schedules are on exchange A, the steeper they are on exchange B and vice versa. This complementarity is caused by the trader's incentive to split his order when he faces price impact, which is always the case due to inventory holding costs. Order splitting, in turn, induces liquidity providers to raise their price schedules above the competitive level. Closed-form solutions to the strategic liquidity provision game are intractable. Therefore, all following results are stated in general terms. Explicit results are given in an illustration with specific parameter assumptions at the end of the analysis (see section 4.2). The following proposition summarizes the main properties of equilibrium price schedule slopes.

Proposition 2 *There is a unique equilibrium in price schedule slopes* (a^*, b^*) *with symmetric strategies per exchange. It has the following properties:*

- *Price schedules are always steeper than those resulting from competitive liquidity supply.*
- *Price schedules flatten as the number of liquidity providers increases.*
- *Price schedules steepen in own inventory holding costs and are independent of the other exchange's inventory holding costs.*
- *Liquidity providers with higher inventory holding costs set steeper price schedules.*

Proof. See appendix ■

The properties of the equilibrium in price schedules are similar to those found in BH and in Biais, Martimort, and Rochet (2000), who analyze strategic liquidity provision in a setting featuring information asymmetry. In both cases, a finite number of liquidity providers face an inelastic asset demand and trade at non-competitive price schedules. However, the origin of asset demand inelasticity is different. In their setting, it is caused by an attribute of traders (possessing private information), while in this setting it is caused by an attribute of liquidity providers (their inventory holding costs). Inventory holding costs do not vanish as liquidity providers become more numerous, therefore equilibrium price schedules do not approach the competitive level in that case. Only when inventory holding costs are inexistent ($\gamma_j = \gamma_{-j} = 0$) do equilibrium price schedules correspond to the competitive ones.

4 Exchange competition

4.1 General analysis

Exchanges set their trading fees similarly to the way liquidity providers set price schedules. Exchange j conjectures his competitor's trading fee, f_{-j}, and sets his own fee to maximize his profit, accounting for equilibrium price schedules at the second stage and for the best response of the liquidity trader at the third stage of the game:

$$f_j^* = \arg\max_{f_j} f_j \sum_{n=1}^{N_j} q_{jn}^* \left(a^*, b^*, f_j, f_{-j}\right) \tag{17}$$

The equilibrium in trading fees is characterized in the following proposition.

Proposition 3 *There is a unique equilibrium in exchange trading fees:*

$$f_A^* = v \frac{2b^*(a^* + N_A) + a^* N_B}{4a^* \left(b^* + N_B\right) + N_A \left(4b^* + 3N_B\right)} \tag{18}$$

$$f_B^* = v \frac{2a^*(b^* + N_B) + b^* N_A}{4a^* \left(b^* + N_B\right) + N_A \left(4b^* + 3N_B\right)} \tag{19}$$

Proof. See appendix ■

In equilibrium, trading fees are strictly positive at any level of illiquidity although exchanges do not incur any costs. Hence, undercutting in fees as in the traditional Bertrand setting does not take place. Both fees are bounded above at $\frac{v}{2}$, the profit maximizing fee of a monopolistic exchange. As with liquidity provision, the explanation for non-competitive fees relates to the incentive of the trader to split his order between liquidity providers to reduce price impact. This implies that he also splits his order between exchanges provided that the difference in fees is not excessively large (see lemma 1). Exchanges are not perfect substitutes from the point of view of the trader. Undercutting in fees is then particularly costly for them as a large decrease in the trading fee relative to the competitor's one leads only to a limited increase in the trading volume. Conversely, raising the fee is little costly since the decrease in the trading volume is limited too. Equilibrium fees are identical if and only if both exchanges are identical (i.e. $\gamma_A = \gamma_B$ and $N_A = N_B$). Otherwise, the exchange offering the best liquidity to the trader charges the highest fee ($f_A^* > f_B^* \Leftrightarrow a^*/N_A < b^*/N_B$).

Moreover, both fees increase the less liquid exchanges are. More specifically, an exchange's fee increases in both the illiquidity of its competitor (e.g. $\partial f_A^*/\partial b^* > 0$) and its own illiquidity (e.g. $\partial f_A^*/\partial a^* > 0$). The former can be explained by the enhanced attractiveness of the exchange. The latter, however, seems counter-intuitive as the exchange's attractiveness diminishes in that case. In addition, more illiquidity (on one or both exchanges) reduces the size of the trader's order, and hurts both exchanges by reducing trading volume. Illiquidity has, however, another effect that dominates in this context: it reduces the trader's sensitivity to the difference in trading fees. Hence, with more illiquidity, raising fees becomes less costly for exchanges as exchange specific trading volume is less responsive to such a move. This induces an increase in fees on both exchanges. In contrast, liquidity improvements trigger a reduction of all trading fees as volume becomes more responsive to the difference in fees, regardless of where the liquidity improvement has taken place. The positive link between the exchange's fee and its own level of illiquidity is crucially linked to the imperfect substitutability of trading services from the point of view of the trader. If, in contrast, exchanges were perfect substitutes, any worsening in own illiquidity would be offset by a lower trading fee in order to be more attractive than the competing exchange. Perfect substitutability may arise in the absence of price impact (if $\gamma_A = \gamma_B = 0$ in this model) or if rules force the execution of the entire order on one single exchange. The latter is often assumed in microstructure models (see, e.g., Colliard and Foucault, 2012). This analysis demonstrates that such an assumption is not innocuous whenever exchange competition in trading fees is considered, as it strengthens artificially the competitive pressure on fees.

Exchanges cannot directly influence price schedules as these do not depend on exchange fees. Rather, the link between equilibrium fees and illiquidity runs exclusively through the trader's asset demand. A major consequence of this linkage is that the size of the liquidity provision sector affects indirectly the level of equilibrium fees: the more liquidity providers are active, the stronger is competition between them which leads to flatter price schedules and consequently to lower exchange trading fees. In addition, the number of liquidity provider has a direct effect on equilibrium fees. A higher number of liquidity providers allows the trader to split his order in smaller parts and to bear therefore a smaller price impact. This raises his sensitivity to the difference in trading fees, which enhances competition between exchanges. In the limiting case of an infinite number of liquidity providers on both exchanges, equilibrium fees converge to the competitive level (i.e. zero in this model).

The characteristics of equilibrium fees lead to two empirical implications on the interplay between liquidity and exchange competition.

Implication 1 *In an industry with competing exchanges, a shock to liquidity provision on one exchange leads to an adjustment of trading fees on all competing exchanges.*

A shock to liquidity provision can be understood as a market-wide event that causes a large change in overall liquidity provision. Such events are for instance the introduction of new technologies that reduce the cost of liquidity provision or trigger the massive entrance of new traders (e.g. high frequency traders), or regulations facilitating or triggering multi-market trading (e.g. Reg NMS or MiFID).

Implication 2 *Regulation that leads to new entrants in a monopolistic exchange industry affects trading fees through two channels: first through exchange competition per se (holding all else constant), and second through the thereby improved liquidity.*

Which of the two channels dominates in the aftermath of an event such as the implementation of MiFID in 2007, is an empirical question. Estimating the importance of each channel is relevant from a competition policy perspective. Indeed, it indicates whether policy should focus on keeping the exchange industry fragmented (if the competition channel dominates) or on ensuring a good liquidity in a oligopolistic but concentrated industry (if the liquidity channel dominates) in order to reduce total trading costs.

In equilibrium, both exchanges attract order flow. When exchanges have the same number of liquidity providers and the same inventory holding costs, they set identical fees and all liquidity providers set identical price schedules. As a consequence, the trader splits his order in equal parts between exchanges and liquidity providers. When the liquidity provision sector differs between exchanges ($N_j \neq N_{-j}$ or $\gamma_j \neq \gamma_{-j}$), volume is never concentrated on a single exchange. The highest volume concentration on any exchange j is reached when this exchange has infinitely many liquidity providers ($N_j \to \infty$) and the competing exchange has a monopolistic liquidity provider ($N_{-j} = 1$). The trader can divide his order between many price schedules on exchange j, thereby going up an infinitely small portion of each price schedule. In contrast, he would trade against one single price schedule on exchange $-j$, resulting in a higher total price impact. Hence, he directs the largest part of his order to exchange j. Although this exchange could attract the entire order flow with a trading fee that is only slightly smaller than the competitor's one (this is implied by equation 9 in Lemma 1 taking the limit $N_A \to \infty$), it optimally charges a fee that is twice as large and induces fragmentation.[11] Its market share tends to one, only in the improbable case of an infinitely large inventory holding cost on the competing exchange. Generally, the exchange with the highest market share is the one charging the highest trading fee. Indeed, this is the exchange with the best liquidity from the point of view of the trader. Hence, the higher trading fee never over-compensates the liquidity advantage of that exchange. The previous results are summarized in the following lemma.

Lemma 2 *The market share on exchange j, $\frac{V_j}{V_j+V_{-j}}$, has the following characteristics:*

- *It is strictly smaller than* 1.
- *It tends to one if $N_j \to \infty$ and $\gamma_{-j} \to \infty$*

The exchange with the largest market share is the one with the highest equilibrium fee:

$$\frac{V_j}{V_j + V_{-j}} > \frac{1}{2} \Leftrightarrow f_j^* > f_{-j}^* \tag{20}$$

Proof. See appendix ■

The findings on volume concentration and trading costs imply that a competitive liquidity provision sector benefits an exchange, relative to its competitor, in terms of market share. However, competitive liquidity provision need not maximize the profit of exchanges due to lower fee levels. Attracting the largest portion of the order flow with a competitive liquidity provision sector comes at the cost of a small trading revenue per unit and possibly a small profit.

In equilibrium, the profits of exchanges are strictly positive:

$$\Pi_A\left(f_A^*, f_B^*\right) = \left(f_A^*\right)^2 \frac{N_A\left(b^* + N_B\right)}{2\left(b^* N_A \rho + a^*\left(b^* + N_B \rho\right)\right)} \tag{21}$$

$$\Pi_B\left(f_A^*, f_B^*\right) = \left(f_B^*\right)^2 \frac{N_B\left(a^* + N_A\right)}{2\left(b^* N_A \rho + a^*\left(b^* + N_B \rho\right)\right)} \tag{22}$$

[11] With $N_j \to \infty$ and $N_{-j} = 1$, the equilibrium fees are: $f_j^* = 4\frac{v\gamma_{-j}}{3+8\gamma_{-j}}$ and $f_{-j}^* = \frac{1}{2} f_j^*$.

They are non-monotonic in both the levels of illiquidity, a^* and b^*, and the numbers of liquidity providers, N_A and N_B. Exchanges' possibilities to influence inventory holding costs are rather limited as these depend to a large extent on the characteristics of liquidity providers (funding constraints and costs, risk aversion etc.) and on the characteristics of assets (in particular price volatility). However, exchanges may influence the number of liquidity providing firms by creating liquidity provision programs to which financial firms must be admitted formally (as requested, e.g., by the second version of MiFID - see footnote 6), or by regulating physical access to the trading venue (e.g. by limiting or broadening colocation space and cable width used by high frequency traders).[12]

To understand how competition in the liquidity provision sector affects the profit of exchanges, I analyze exchange A's profit with respect to N_A. The previous analysis has shown that the equilibrium fee diminishes with increases in N_A. The effect on equilibrium volume, however, is ambiguous. The number of intermediaries affects volume in three ways. First, as a direct effect, holding all else equal including liquidity levels, a higher N_A raises the volume on exchange A since this leads to a smaller total price impact for the trader. The two other effects are indirect as they run through the positive impact of a higher N_A on liquidity levels on both exchanges. As a second effect, an increase in N_A improves liquidity on exchange A, leading also to a higher volume. The ambiguity comes from the third effect related to improved liquidity on exchange B. Indeed, a lower level of b^* makes the competing exchange more attractive (which lowers the volume on exchange A), but it also reduces trading costs on exchange A (both price impact and trading fees), which tends to raise volume on that exchange. The direction of the third effect depends on parameters. Concluding, exchanges in this model do not always benefit from competitive liquidity provision, because high competition in liquidity provision increases the pressure on trading fees, and might also have an adverse effect on trading volume.

4.2 Numerical illustration

The general setting considered so far is not tractable enough to analyze explicitly how the characteristics of liquidity provision, i.e. the level of inventory holding costs and the number of liquidity providers, affect equilibrium fees and exchange profits. To gain more

[12] Some exchanges restrict the use of colocation services or their upgrades to specific firms and charge additional costs. This is for instance the case on OMX, analyzed in Brogaard, Hagströmer, Nordén, and Riordan (2015). On the contrary, the Johannesburg stock exchange introduced colocation services for all interested firms without additional costs in 2014 (source: Traders Magazine, 21/07/2014: "Johannesburg stock exchange delivers democratic colocation").

insights into these relationships, I simplify the setting to obtain tractable closed-form solutions at the liquidity provision stage, and discuss numerical simulations of the profit functions of exchanges. More specifically, I impose equal inventory holding costs for all liquidity providers and a specific degree of risk aversion:

$$\gamma_A = \gamma_B = \gamma \tag{23}$$

and

$$\rho = 1 \tag{24}$$

With these parameter specifications, equilibrium price schedule slopes are:

$$a^* = b^* = \gamma + 1 - \frac{N_A + N_B}{2} + \frac{\sqrt{(N_A + N_B - 2)^2 + 4\gamma (N_A + N_B + \gamma)}}{2} \tag{25}$$

The mathematical expressions of equilibrium fees and of the resulting profit functions are obtained by replacing a^* and b^* with their expression in equations 18 and 19, and equations 21 and 22 respectively.

In equilibrium, trading fees approach their upper bound, $v/2$, when inventory holding costs are infinitely large. With finite inventory holding costs, fees decrease in the number of liquidity providers and are bounded below at a strictly positive level. Considering a realistic scenario in which inventory holding costs are between 0 and 1 (see, e.g., Bollen, Smith, and Whaley, 2004; Stoll, 2000; Affleck-Graves, Hedge, and Miller, 1994) and in which there are multiple liquidity providers on exchanges, the numerical examples in Figure 1 show that equilibrium fees (scaled by the gross gain from trade v) tend to be small (generally below 0.05, and for small enough γ, below 0.01).

Please insert Figure 1 around here.

Turning to the profits of exchanges, they are always positive and converge towards the following level as the number of their own liquidity providers becomes infinitely large:

$$\lim_{N_j \to \infty} \Pi_j \left(f_j^*, f_{-j}^*\right) = 4v^2\gamma \frac{2\gamma + N_{-j}}{(8\gamma + 3N_{-j})^2} \tag{26}$$

Only when both exchanges have infinitely many market makers do profits converge to zero, since competition in fees is then strongest leading to zero fees. A numerical analysis of profit functions shows that the desirability for exchange j of a large liquidity provision sector depends crucially on the size of this sector at the competing exchange as well as on the level of inventory holding costs. Figure 2, that depicts the profit of exchange j for several combinations of (γ, N_{-j}), hints at the existence of four cases:

Case 1 *Very small levels of γ or a small γ combined with a medium to large N_{-j}: Π_j decreases always in N_j. The highest profit is obtained with a single liquidity provider.*

Case 2 *Small levels of γ and N_{-j}, or high levels of γ and N_{-j}: Π_j reaches its highest level with a finite but small number of liquidity providers.*

Case 3 *High levels of γ combined with medium levels of N_{-j}: Π_j has a local maximum at a small N_j, but it reaches its highest level with an infinitely large N_j.*

Case 4 *High levels of γ combined with small levels of N_{-j}: Π_j increases always in N_j. The highest profit is obtained with an infinitely large N_j.*

Please insert Figure 2 around here.

Fixing inventory holding costs and the number of liquidity providers on exchange j, the profit of this exchange varies strongly with small changes in the competitor's number of liquidity providers. Consider for instance the illustrations in Figure 2 with $\gamma = \frac{1}{5}$ and $N_j = 50$. Exchange j obtains a profit of 5.3 when the competing exchange hosts a single intermediary ($N_{-j} = 1$). Its profit drops to 3.3 with $N_{-j} = 2$, to 0.84 with $N_{-j} = 10$, and to 0.1 with $N_{-j} = 100$. This leads to the following implication.

Implication 3 *A positive (negative) chock to the size of the liquidity provision sector on an exchange lowers (increases) strongly the trading revenue of the competing exchange.*

Straightforward calculations using the parameters of the numerical simulation allow to determine the optimal number of liquidity providers per exchange for specific values of inventory holding costs. These calculations are in line with the intuition conveyed by the previous formal results: exchanges optimally keep liquidity at an intermediate level such as to dampen competition in fees and at the same time encourage sufficiently large trading volume. If inventory holding costs are small, both exchanges allow only a small number of liquidity providers. As an illustration, with $\gamma \leq \frac{1}{20}$ the equilibrium in the number of liquidity providers is $(N_A^*, N_B^*) = (1, 1)$, and with $\gamma = \frac{1}{5}$ the equilibrium is

$(N_A^*, N_B^*) = (3, 3)$. The trader is then forced to trade larger portions of his order with each liquidity provider, raising his total price impact and consequently his incentive to split across exchanges. These, in turn, charge high fees and benefit from a high profit. With higher inventory holding costs, equilibria are asymmetric: one exchange has a very large number of liquidity providers while the other has a small number. Thereby, the price impact incurred by the trader is limited due to flat price schedules and the possibility to reach many liquidity providers, but trading fees are strictly positive and away from zero, and both exchanges attract a substantial portion of the volume.

5 Conclusion

This paper studies a competition model in which liquidity providers' price schedules and exchange trading fees are set endogenously. It shows that order flow fragmentation induces non-competitive pricing at the trading level and at the level of exchange competition. A discretionary liquidity trader always splits his order between liquidity providers and exchanges to minimize price impact. Liquidity providers anticipate this behavior, and optimally set price schedules that are steeper than competitive ones. Illiquidity, in turn, gives market power to exchanges. Indeed, steep price schedules strengthen the incentive of the liquidity trader to split his order between exchanges, reducing thereby his sensitivity to the difference in trading fees. From the exchange's point of view, this reduces the fee-elasticity of their trading volume. As a consequence, exchanges face a lower pressure on trading fees and charge fees that are strictly larger than competitive ones. In equilibrium, fees increase the more costly and the less competitive liquidity provision is. Moreover, the exchange with the best liquidity conditions from the point of view of the trader charges the highest fee. Thereby, this exchange forgoes the possibility to attract the entire order flow with a sufficiently small fee. Rather, it benefits from its better liquidity by charging a high fee, satisfying thereby only a portion of the trading demand and leaving the remaining to the less liquid competitor. A numerical analysis of equilibrium exchange profits suggests that exchanges optimally keep liquidity at an intermediate level such as to dampen competition in fees and at the same time encourage sufficiently large trading volume.

The model delivers empirical implications. First, the model predicts that a positive shock to liquidity supply (due e.g. to new regulations or technologies) on one exchange triggers a decrease in trading fees on the competing exchange. Moreover, such a shock decreases strongly the competitor's trading revenue and may induce a reduction of the number of liquidity providers. Second, the results show that an oligopolistic exchange industry structure generates a positive correlation between implicit trading costs (illiquidity) and explicit trading costs (exchange fees). This is in contrast to monopolistic exchanges that would raise explicit trading costs the lower implicit costs are. In addition, the model predicts that the most liquid exchange charges the highest trading fee and attracts the largest market share. More generally, these results indicate that introducing competition in the exchange industry (as was the case for instance with the first version of MiFID in the European Union in 2007) may not have a large impact on exchange fees per se. Rather, market-wide liquidity improvements combined with exchange competition reduce strongly trading costs.

Appendix

Proof of Proposition 1. The utility function U_T, indicated in equation 3, is quadratic and concave in q_{ji}. Therefore, the optimal asset demand per exchange and per liquidity provider is determined by the first order conditions. Taking the first order derivative of U_T with respect to every q_{ji} and equating it to zero results in the following sets of equations:

$$q_{Ai}a_i = \frac{v - f_A}{2} - \rho \sum_{j\in\{A,B\}} \sum_{n=1}^{N_j} q_{jn} \quad \text{for all } i = 1, ..., N_A \tag{A.1}$$

and

$$q_{Bi}b_i = \frac{v - f_B}{2} - \rho \sum_{j\in\{A,B\}} \sum_{n=1}^{N_j} q_{jn} \quad \text{for all } i = 1, ..., N_B \tag{A.2}$$

Since the right hand sides in A.1 are identical and constant for all $i = 1, ..., N_A$, the left hand sides must also be equal. As a consequence, asset demands on exchange A are linked pair-wise through their price impacts: any pair (q_{Aj}, q_{Ak}), with $j \neq k$ and both indicators from the set $\{1, ..., N_A\}$, is related as follows:

$$q_{Aj} = q_{Ak}\frac{a_k}{a_j} \tag{A.3}$$

Similar computations using the set of equations in A.2 specify the relation between any pair (q_{Bj}, q_{Bk}), with $j \neq k$ and both indicators from the set $\{1, ..., N_B\}$:

$$q_{Bj} = q_{Bk}\frac{a_k}{a_j} \tag{A.4}$$

Combining equations in A.1 and in A.2 pair-wise leads to the link between any pair (q_{Aj}, q_{Bk}) with $j \in \{1, ..., N_A\}$ and $k \in \{1, ..., N_B\}$:

$$q_{Aj} = q_{Bk}\frac{b_k}{a_j} + \frac{f_B - f_A}{2a_j} \tag{A.5}$$

Replacing expressions in A.3, A.4 and A.5 in equations A.1 and A.2, and rearranging terms yields the optimal asset demands given in expressions 5 and 6. ■

Proof of Lemma 1. Volume per exchange is:

$$V_A = \sum_{i=1}^{N_A} q^*_{Ai} = \frac{1}{\Phi}\sum_{i=1}^{N_A}\frac{1}{a_i}\left(\frac{v - f_A}{2} - \frac{(f_A - f_B)\,\rho}{2}\sum_{n=1}^{N_B}\frac{1}{b_n}\right) \tag{A.6}$$

and

$$V_B = \sum_{i=1}^{N_B} q^*_{Bi} = \frac{1}{\Phi} \sum_{i=1}^{N_B} \frac{1}{b_i} \left(\frac{v - f_B}{2} - \frac{(f_B - f_A)\,\rho}{2} \sum_{n=1}^{N_A} \frac{1}{a_n} \right) \tag{A.7}$$

Exchange A attracts the entire order if and only if $V_A > 0$ and $V_B = 0$. This is the case whenever the following three conditions hold:

$$\frac{v - f_A}{2} - \frac{(f_A - f_B)\,\rho}{2} \sum_{n=1}^{N_B} \frac{1}{b_n} > 0 \tag{A.8}$$

$$\frac{v - f_B}{2} - \frac{(f_B - f_A)\,\rho}{2} \sum_{i=1}^{N_A} \frac{1}{a_n} < 0 \tag{A.9}$$

$$f_A < v \tag{A.10}$$

If exchange B charges a trading fee that is larger than the gross gain from trade, $f_B > v$, the trader never trades on that exchange. Exchange A attracts the entire order with any trading fee smaller than v. If exchange B charges a smaller trading fee, $f_B < v$, exchange A attracts the entire order if and only if it charges a lower fee, $f_A < f_B$ and its liquidity level is good enough:$\sum_{i=1}^{N_A} 1/a_n \geq (v - f_B)\,/\,(f_B - f_A)$. ■

Proof of Proposition 2. Solving equation 13 for a and equation 14 for b leads to the following best response functions, a^R and b^R:

$$a^R = \gamma_A + \frac{b\rho(2 - N_A)}{2(b + N_B\rho)} + \frac{\sqrt{8b\gamma_A\rho\,(N_A - 1)\,(b + N_B\rho) + (2\gamma_A\,(b + N_B\rho) - b\rho\,(N_A - 2))^2}}{2(b + N_B\rho)} \tag{A.11}$$

$$b^R = \gamma_B + \frac{a\rho(2 - N_B)}{2(a + N_A\rho)} + \frac{\sqrt{8a\gamma_B\rho\,(N_B - 1)\,(a + N_A\rho) + (2\gamma_B\,(a + N_A\rho) - a\rho\,(N_B - 2))^2}}{2(a + N_A\rho)} \tag{A.12}$$

Reaction functions have the following properties:

- They are increasing:$\partial a^R/\partial b > 0$ and $\partial b^R/\partial a > 0$
- They are strictly positive since $a^R(b = 0) = 2\gamma_A$ and $b^R(a = 0) = 2\gamma_B$
- They are strictly concave: $\partial^2 a^R/\partial b^2 < 0$ and $\partial^2 b^R/\partial a^2 < 0$

These properties imply the existence of a single crossing point between best response functions that defines the equilibrium (a^*, b^*). The equilibrium is such that $a^* > 2\gamma_A$ and $b^* > 2\gamma_B$.

The competitive price schedule slopes that lead to zero profits for liquidity providers can be inferred from equation 2: $a = \lambda_A$ and $b = \lambda_B$. Equilibrium price schedules are always steeper than competitive ones.

The best response functions decrease in the number of liquidity providers on both exchanges: $\partial a^R / \partial N_A < 0$, $\partial a^R / \partial N_B < 0$, $\partial b^R / \partial N_A < 0$ and $\partial b^R / \partial N_B < 0$. As a result, the equilibrium slopes decrease in both N_A and N_B.

Best response functions increase in own inventory holding costs: $\partial a^R / \partial \lambda_A > 0$ and $\partial b^R / \partial \lambda_B > 0$. As a result, the equilibrium slopes increase in own inventory holding costs.

Finally, considering equations 13 and 14, both right hand sides must be equal. Since they both decrease in inventory holding costs and increase in price schedule slopes, liquidity providers with the highest cost set the steepest price schedule: $\gamma_A > \gamma_B \Leftrightarrow a^* > b^*$. ■

Proof of Proposition 3. The profit function of exchange j is:

$$\Pi_j = f_j \sum_{n=1}^{N_j} q_{jn}^* \left(a^*, b^*, f_j, f_{-j}\right) \tag{A.13}$$

where the index $-j$ stands for the competing exchange. The profit function is strictly concave in f_j. Hence, the equilibrium fees stated in the proposition are obtained by solving simultaneously the first order conditions, $\partial \Pi_j / \partial f_j = 0$, for $j = \{A, B\}$. Moreover, strict concavity ensures that the solution is unique. ■

Proof of Lemma 2. The proof is provided considering the market share of exchange A. A symmetric reasoning holds for exchange B.

Equilibrium volume on both exchanges is computed replacing a, b, f_A and f_B by their equilibrium values in expressions A.6 and A.7. Market share is then:

$$\frac{V_A}{V_A + V_B} = \frac{N_A \left(b^* + N_B\right)\left(2b^* \left(a^* + N_A\right) + a^* N_B\right)}{2a^{*2} N_B \left(b^* + N_B\right) + b^* N_A^2 \left(2b^* + 3N_B\right) + a^* N_A \left(2b^{*2} + 6b^* N_B + 3N_B^2\right)} \tag{A.14}$$

This expression is strictly smaller than one for any finite N_A, N_B, a^* and b^*. Taking N_A

to infinity results in the following expression:

$$\lim_{N_A \to \infty} \frac{V_A}{V_A + V_B} = \frac{2\left(b^* + N_B\right)}{2b^* + 3N_B} \tag{A.15}$$

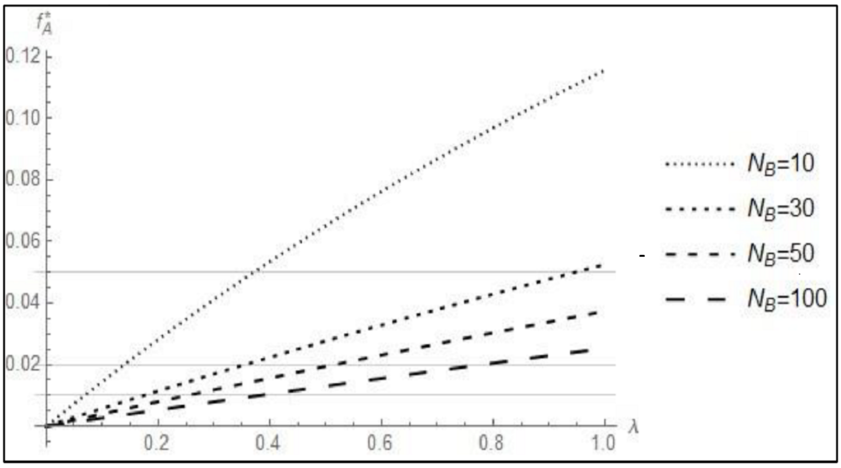
The limit of expression A.15 when b^* tends to infinity is 1.

Expression A.14 is large than 1/2 if and only if

$$\frac{a^*}{N_A} < \frac{b^*}{N_B} \tag{A.16}$$

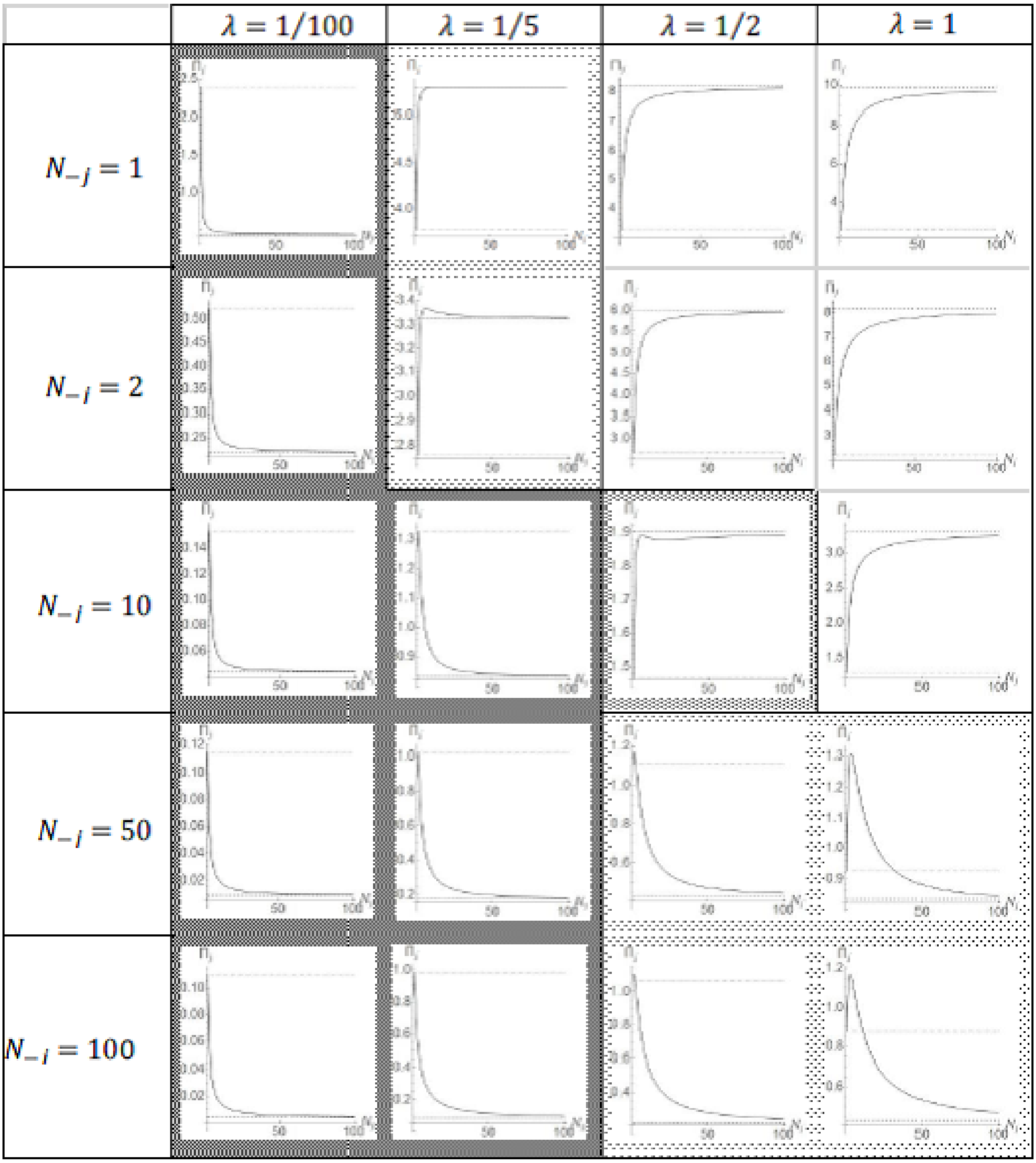
This condition implies that $f_A^* > f_B^*$. The third bullet point in Lemma 2 follows. ■

Figure 1: Equilibrium trading fee



This figure depicts the equilibrium trading fee set by exchange A as a function of λ, for different values of N_B. Other parameters are set as follows: $v = 10$ and $N_A = 50$.

Figure 2: Equilibrium exchange profit



This figure depicts the equilibrium profit of exchange j as a function of N_j, for different combinations of λ and N_{-j}. The parameter v is set to 10. In all graphs, the dotted horizontal line is the profit level with $N_j = 1$, and the dashed horizontal line is the profit with $N_j \to \infty$. The cases described in section 4.2 correspond to the following areas: the heavily dotted area illustrates case 1, the lightly dotted area illustrates case 2, the medium dotted area ($N_{-j} = 10$ and $\lambda = 1/2$) illustrates case 3, and the white area illustrates case 4.

References

Affleck-Graves, J., Hegde, S.P., Miller, R.E.,1994. Trading Mechanisms and the Components of the Bid-Ask Spread. The Journal of Finance 49, 1471-88.

Allen, B., Hellwig, M., 1986. Bertrand-Edgeworth oligopoly in large markets. The Review of Economic Studies 53, 175-204

Babus, A., Parlatore, C., 2017. Strategic Fragmented Markets. Unpublished working paper. NYU Stern.

Baruch, S., Karolyi, G.A., Lemmon, M.L., 2007. Multimarket Trading and Liquidity: Theory and Evidence. The Journal of Finance 62, 2169-2200

Bernhardt, D., Hughson, E., 1997. Splitting orders. Review of Financial Studies 10, 69-101.

Bertrand, J., 1883. Theorie mathematique de la richesse sociale. Journal des Savants 48, 499–508,

Biais, B., Martimort, D., Rochet, J.C., 2000. Competing mechanisms in a common value environment. Econometrica 68, 799-837.

Bollen, N.P., Smith, T., Whaley, R.E., 2004. Modeling the bid/ask spread: measuring the inventory-holding premium. Journal of Financial Economics 72, 97-141.

Bondarenko, O., 2001. Competing market makers, liquidity provision, and bid–ask spreads. Journal of Financial Markets 4, 269–308

Boneva, L., Linton, O., Vogt, M., 2016. The effect of fragmentation in trading on market quality in the UK equity market. Journal of Applied Econometrics 31, 192–213

Brogaard, J., Hagströmer, B., Nordén, L., Riordan, R., 2015. Trading fast and slow: Colocation and liquidity. The Review of Financial Studies 28, 3407-43.

Cespa, G., Colla, P., 2017. Market fragmentation, dissimulation, and the disclosure of insider trades. Unpublished working paper. CEPR Discussion paper.

Chao, Y.,Yao, C., Ye, M., 2017. Why Discrete Price Fragments US Stock Exchanges and Disperses Their Fee Structures. Unpublished working paper. University of Illinois at Urbana-Champaign.

Chemmanur, T.J. and Fulghieri, P., 2006. Competition and cooperation among exchanges: A theory of cross-listing and endogenous listing standards. Journal of Financial Economics 82, 455-489.

Chowdhry, B. and Nanda, V., 1991. Multimarket trading and market liquidity. Review of Financial Studies 4, 483-511.

Colliard, J.E. and Foucault, T., 2012. Trading fees and efficiency in limit order markets. Review of Financial Studies 25, 3389-3421.

Dudey, M., 1992. Dynamic Edgeworth-Bertrand Competition. The Quarterly Journal of Economics 107, 1461-1477

Economides, N., Siow, A., 1988. The division of markets is limited by the extent of liquidity. The American Economic Review 78, 108-121

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

Foucault, T. and Parlour, C.A., 2004. Competition for listings. Rand Journal of Economics 35, 329-355.

Gomber, P., 2016.The German equity trading landscape. Unpublished Working Paper. Goethe University Frankfurt.

Ho, T.S., Stoll, H.R., 1983. The dynamics of dealer markets under competition. The Journal of finance 38, 1053-74.

Kyle, A.S, 1985. Continuous auctions and insider trading. Econometrica 53, 1315-1335.

Levitan, R., Shubik, M., 1972. Price duopoly and capacity constraints. International Economic Review 13, 111-122

Macey, J.R., O'Hara, M., 2005. From markets to venues: securities regulation in an evolving world. Stanford Law Review 58, 563-599.

Menkveld, A.J., 2016. The economics of high-frequency trading: Taking stock. Annual Review of Financial Economics 8, 1-24

Nagel, S., 2012. Evaporating liquidity. The Review of Financial Studies 25, 2005-39.

Pagano, M., 1989. Trading volume and asset liquidity. The Quarterly Journal of Economics 104, 255-274.

Pagnotta, E. and Philippon, T., 2011. Competing on speed. Unpublished working paper. National Bureau of Economic Research.

Panayides, M.A., Rindi, B., Werner, I.M., 2017. Trading Fees and Intermarket Competition. Unpublished working paper. Ohio State University.

Shaked, A., Sutton, J., 1982. Relaxing price competition through product differentiation. The Review of Economic Studies 49, 3-13

Singh, N., Vives, X., 1984. Price and quantity competition in a differentiated duopoly. The RAND Journal of Economics 15, 546-554

Stoll H.R., 1978. The supply of dealer services in securities markets. The Journal of Finance 33, 1133-51.

Stoll H.R., 2000. Presidential address: friction. The Journal of Finance 55, 1479-514.