Nº 211
Marzo 2002

Documento de Trabajo
ISSN (edición impresa) 0716-7334
ISSN (edición electrónica) 0717-7593

Liquidity as an Industrial Organization of Stock Exchanges.

Felipe Zurita
Alexander Galetovic

www.economia.puc.cl

Liquidity and the simple IO of stock exchanges*

Alexander Galetovic[†] Felipe Zurita[‡]

March 2002

Abstract

It is usually thought that network externalities, which are inherent to liquidity, make it desirable to concentrate transactions in one stock exchange. This paper shows that when the value of liquidity stems from the ability of potentially reach as many traders as possible, the market is *integrated* when every broker meets every other broker in at least one exchange. Thus, fragmentation is not about trades being executed in different exchanges but of connectedness among brokers.

An implication of this distinction is that in an integrated market the network externality created by liquidity becomes pecuniary and the optimal number of exchanges depends only on the shape of the (physical) technology to execute trades—whether it exhibits increasing, constant or decreasing returns to scale—as in any standard industry.

We characterize the planner's allocation and compare it with that reached by a monopoly. It is shown that when exchanges are natural monopolies a particular ownership structure of the exchange and allocation of voting rights over the exchange fee achieve the planner's optimum.

With decreasing returns to scale the Walrasian allocation is efficient, provided that the market is integrated. Nevertheless, with few exchanges the price-taking assumption is suspect. If exchanges are not price takers, there are many other equilibria, all of them inefficient. Moreover, there are reasons to doubt that the market will become integrated. Fragmentation softens price competition between exchanges and may help a monopolist exchange to erect a barrier to entry even when he has no cost advantage.

Key words: brokerage, exchange fee, fragmentation, liquidity, network externality
JEL classification: G24, L1

* We gratefully acknowledge the financial support from Fondecyt, grant #1000871. For comments we are grateful to Sergio de la Cuadra, Salvador Valdés, seminar participants at UCLA, the IMF Institute, Pontificia Universidad Católica de Chile, LACEA meetings (Montevideo, 2001) Jornadas de Economía del Banco Central del Uruguay (Montevideo, 2001), and LAMES (Buenos Aires, 2001).

† Centro de Economía Aplicada (CEA), Universidad de Chile. Av. República 701, Santiago Chile. Tel: ++56-2-678 4065; Fax: +56-2-689 7895; email: agaleto@dii.uchile.cl.

‡ Instituto de Economía, Pontificia Universidad Católica de Chile. Av. Vicuña Mackenna 4860, Santiago, Chile. Tel. +56-2-686 4318; Fax: +56-2-553 2377; email: fzurita@faceapuc.cl.

1. Introduction

It is usually thought that network externalities, which are inherent to liquidity, make it desirable to concentrate transactions in one stock exchange.[1] But, on the other hand, concentration may hurt competition. Thus some fragmentation of trades in more than one exchange may be a price worth paying to foster competition. What is the adequate balance of this trade off? This paper shows that, in fact, a trade off is not necessary.

It may seem that network externalities imply that all transactions should be centralized in one exchange to take full advantage of it. Nevertheless, we start from the observation that the value of liquidity stems from the ability of potentially reach as many traders as possible. Thus, if every broker meets every other broker in at least one exchange, thereby giving complete reachability to any pair of potential traders, then the market is *integrated* regardless of the number of exchanges where transactions take place. Hence, fragmentation is not about trades being executed in different exchanges but of connectedness among brokers. This distinction turns out to have a somewhat surprising implication: in an integrated market the network externality created by liquidity becomes pecuniary and the IO of the industry is "simple": the optimal number of exchanges depends only on the shape of the (physical) technology to execute trades—whether it exhibits increasing, constant or decreasing returns to scale—as in any standard industry.

To study the IO of stock exchanges we present a model where a large number of investors, half of them buyers and half of them sellers, want to trade at most one unit of an asset. By assumption investors can trade only through a broker. In turn, to execute an order to buy or sell the security brokers must meet pairwise in an exchange where both are members. Brokers are heterogeneous but competitive and have standard, decreasing-returns-to-scale cost functions; exchanges are few in number and have a standard cost function which allows to distinguish the cases of decreasing, constant and increasing returns to scale in volume traded. Thus, we distinguish between brokerage and exchange services, the two vertically-related activities that are needed in fixed proportions to produce a trade[2]. By "brokerage" we mean services typically provided by brokers to investors, like looking for a seller or buyer, executing a trade, back office activities, research, advice, or custody. By "exchange services" we mean those services that stock exchanges provide to brokers, mainly procedures and protocols that reduce the cost of finding a seller or buyer. The price that investors pay for executing a transaction is the sum of a brokerage commission and an exchange fee.

A key feature of the model is that we explicitly consider the role of liquidity. The utility of an investor who posts a *sell* order with broker b increases with the number of *buy* orders handled by

[1] For example, *The Economist* recently stated that "[...] liquidity comes with size, one of the reasons why stock exchanges are often described as natural monopolies." (See "The hunt for liquidity", July 28th, 2001).

[2] This is usually mandated by law. For example, in Chile brokers must execute all trades in an exchange.

brokers who can be reached by b in at least one exchange (and conversely for a buyer-investor). It follows that investors create network externalities: when a buyer-investor posts an order it reduces trading costs of at least some sellers. When every broker is connected to every other broker through at least one exchange all brokers offer exactly the same liquidity, and we say that the market is *integrated*.

Fragmentation in this sense is the lack of integration or connectedness, and not merely the fact that traded volume is split into many exchanges to handle it, as is the general meaning of the term in the literature (see, for example, Lo [1996]).

As a benchmark, we characterize the solution to the planner's problem. We show that the social planner always integrates the market and allows free entry. Then, the optimal number of exchanges and the optimal exchange fee are determined exclusively by the technology to execute trades. Hence, if there are increasing returns to scale in the technology to execute trades then exchanges are a natural monopoly. As in the standard case of any monopoly, the optimal number of exchanges is one, and the optimal exchange fee (subject to the standard self-financing constraint) equals the average cost of the exchange. On the other hand, if the technology to execute trades exhibits decreasing returns to scale, then it is optimal to spread trades in many exchanges and set the exchange fee equal to marginal cost. Thus, as in the textbook case, an exchange is a natural monopoly only if its technology exhibits increasing returns to scale—liquidity does *not* have any consequence. It is for this reason that one can call the IO of stock exchanges "simple".

How can it be that the market is integrated when trades are spread in many exchanges? The key point is that liquidity is in fact an attribute of brokers, not of exchanges. If broker b meets pairwise with all other brokers in at least one exchange, then any sell order taken by broker b can be matched by any other broker with a buy order, and the transaction can be executed in *some* exchange. Thus, all brokers offer the market liquidity, regardless of the identity of the exchange where a particular transaction is executed. It is also noteworthy that in an integrated market the network externality created by investors is of the pecuniary kind. We show that in the optimum the volume of transactions matches the cost of transacting with its value—liquidity included—in the margin.

Next we analyze pricing when exchange services are monopolized. We show that the equilibrium fee will crucially depend on the exchange's ownership structure. A monopolistic exchange owned by shareholders who are not brokers (call it *independent* monopoly) will exploit both its monopoly power over investors and its monopsony power over brokers, but will allow free entry into brokerage. By contrast, when brokers are shareholders and each owns one share (call this a *broker-owned* exchange), all brokers who are active in equilibrium would like to set a lower fee than an independent monopoly[3]. The reason is that while all brokers participate in the exchange's profit, each also internalizes the negative effect on brokerage profits of a higher exchange fee. Nevertheless, brokers differ in their preferred fee.

Thus, as has been pointed out by Hart and Moore (1996) and Pirrong (2000), there is a conflict of interest between members, which is a central feature of broker-owned exchanges. It occurs because the exchange fee plays two roles, exploitation of market power and, given the one member-one share rule, redistribution of profits from relatively efficient broker-members to relatively less efficient ones. A broker who is more efficient than the average receives a smaller share of exchange profits than his contribution to them, and would like to set a low exchange fee, in some cases even equal to the exchange's average cost. At the other extreme, those who are less efficient than the average would like to set a higher exchange fee because by so doing they redistribute profits to them. The extreme case of an "inefficient" broker is a shareholder that chooses not to become an active broker which, not surprisingly, would like to set the same fee as the independent monopoly. We also show that irrespective of their relative efficiency, brokers would vote unanimously to restrict entry.

Understanding the relation between ownership structure and pricing behavior leads to a somewhat unexpected result: one can combine a particular ownership structure and distribution of ownership rights such that the exchange fee set by the broker-owned monopoly equals average cost, which is optimal subject to the self-financing constraint when exchanges are natural monopolies. This structure and distribution of voting rights is as follows. First, the number of shares must be large so that many shareholders choose not to become active brokers in equilibrium. Second, only active brokers may vote to choose the exchange fee. A large number of shareholders, most of them inactive, implies that almost all profits generated by active brokers are redistributed to pure shareholders. Thus, all active brokers would like to set the exchange fee as low as possible—i.e. the exchange fee will be equal to average cost if only active brokers can vote to set it.

Recent developments suggest the possibility that the same security may be traded in several exchanges, each controlled by different owners. It is therefore appropriate to inquire whether competition between exchanges leads to efficient outcomes.

The answer to this question is subtle. It is shown that Walrasian equilibrium (WE) is efficient, provided the market is integrated. However, the existence of a small number of exchanges casts doubts over the price-taking assumption. If exchanges are not price-takers, we find that there are many other equilibria besides the WE, all of them inefficient.

Most importantly, there are reasons to doubt that the market will be integrated. Even though brokers have an incentive to participate in as many exchanges as there are, in order to offer

[3]We allow for shareholders that choose not to become active brokers in equilibrium.

higher liquidity to their clients, exchanges' incentives point in the opposite direction. Fragmentation softens price competition, possibly to the point of allowing to maintain higher profit levels than would result in an integrated market. Similarly, a monopolist may use restrictions to cross memberships to convert liquidity into a barrier to entry.

The literature on network externalities is vast (see, for example, Shy [2001]). Although our modeling is similar to the standard one, this literature has touched on liquidity only few times. Economides and Siow (1988), for instance, find in liquidity a network externality for a specific reason: risk averse traders prefer to trade where others do because larger markets produce lower price variances. They focus, however, on the equilibrium level of variety rather than in the organization of the markets for trading. Pagano (1989) studies a coordination game between traders who have to choose where to trade. Since liquidity depends on the entry decisions of all potential participants, each trader assesses them according to conjectures about entry by others. If trade is equally costly across markets, the network externality leads to the concentration of trade on one market. If not, it can produce multiple conjectural equilibria, some where trade concentrates on one market and others where large traders resort to a separate market or to search for a trading partner. In his model fragmentation is welfare-reducing in the two-market case, but no such ranking is possible if it involves off-exchange search. By contrast, in our model fragmentation is always welfare decreasing.

Di Noia (2001), on the other hand, studies competition among exchanges in the listing of firms' securities. The externalities refer to the value of listing several securities in the same exchange. Even considering that we restrict attention to a unique asset, the flavor of some of his results are present also in this paper. Particularly, the optimality of integration —implicit merger in his context—. Similarly, Foucault and Parlour (1999) study exchange competition for listings, yet deprived of network externalities. By contrast, we study the trading side of exchanges and the interaction between brokerage and the services provided by exchanges in a model with only one security. In our model network externalities emerge because investors' utility is increasing in the number of orders in the market.

Pirrong (1999) also analyzes the organization of financial exchanges, characterizing the endogenous determination of active exchanges and membership. In the one-asset case, he concludes that the exchange is a natural monopoly, and that membership is strategically restricted to an inefficient level. He also investigates competition among exchanges when there are many assets. Our paper departs from his in many respects. The most important difference is that Pirrong's notion of liquidity is price stabilization, performed by heterogeneous, risk-averse market makers, and as such involves no network externality –the market is automatically integrated. Hence, there are no spreads or costs to brokerage: profits are implicit in the expected price changes. By contrast, the network externality is central in our analysis.

Our paper is also related to Hart and Moore (1996), who studied pricing by cooperative exchanges with heterogeneous members. In their model, pricing is decided by majority voting. The median-voter theorem implies that the exchange will choose a high price if the distribution of broker sizes is skewed towards high-cost brokers, but low prices will result if the distribution is skewed towards low-cost brokers. We show how this conflict of interest can be exploited to make a monopoly exchange price exchange services at average cost—the social optimum if exchanges are natural monopolies.

Pirrong (2000) studied how the conflict of interest between heterogeneous brokers affects the choice between a for-profit and not-for-profit exchange. He shows that high-cost brokers may accept to fix low exchange fees if relatively efficient brokers can credibly threaten to form a separate exchange. In his model, high-cost brokers accept low exchange fees because having only one exchange enables brokers to fix rules that facilitate collusion and all brokers earn rents. We go beyond Pirrong's by explicitly modeling the role of liquidity. Doing so enables us to show that competition does not necessarily lead to efficiency, because exchanges can use prohibitions of multiple memberships to fragment the market and soften competition. Liquidity also enables us to give a foundation for an exchange's market power, a fact which has been pointed out by Pirrong (2000). While liquidity may be used to create a barrier to entry, we suggest that by itself it is not sufficient—the exchange must combine it with restrictions to brokers to become members of multiple exchanges. Liquidity is not a barrier to entry in an integrated market.

Finally, Santos and Scheinkman (2001) analyze whether competition among exchanges leads them to set lower guarantee against traders' defaults. By contrast, we assume that brokers never default and concentrate in the determination of the fees that investors pay to trade.

The rest of this paper proceeds as follows. In section 2 we set up the model and study the determinants of liquidity. Section 3 solves the planner's problem. Section 4 studies the case of a monopolistic exchange. Section 5 studies competition among exchanges. Section 6 concludes.

2. The model

We study a secondary market with three types of agent: investors, brokers and exchanges. Each investor buys or sells one unit of a single security. Brokers execute investor's orders in an exchange and charge brokerage commissions for it. Exchanges provide transaction services to brokers and charge fees for them.[4]

[4] See Banner (1998) for an account of the services provided by exchanges.

2.1. Agents: preferences and technology

Investors There is a continuum of investors of mass 2, indexed by $\theta \in [-1, 1]$. A type θ investor for whom $\theta \geq 0$ is a would-be buyer of the security, and her utility function is given by

$$u^{bu}(\theta, \ell^{se}, p^{bu}) \equiv \begin{cases} \theta + \tau\ell^{se} - p^{bu} & \text{if she buys} \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Hence, θ is not only an index but also a taste parameter that determines the investor's valuation. Her valuation is also affected by ℓ^{se}, the mass of the population of sellers with whom the buyer can potentially trade—an indicator of the liquidity of the market. $\ell^{se} \in [0, 1]$ since the mass of all potential sellers equals 1. $\tau \geq 0$ parametrizes the intensity of the preference for liquidity, and p^{bu} is the price of the security paid by the buyer, including any brokerage commissions and fees for using the exchange.

The utility function (2.1) says that buyer θ will buy one unit of the security only if $\theta + \tau\ell^{se} \geq p^{bu}$, that is, when the price inclusive of commissions and fees is at most equal to the utility of having the security. Note that the more "liquid" the market, the higher is the buyer's willingness to pay for the security. That is, we assume that bringing together a large mass of buyers and sellers creates wealth, presumably because it reduces (unmodelled) transaction costs.

Similarly, a type θ investor for whom $\theta \leq 0$ is a potential seller, whose utility function is given by

$$u^{se}(\theta, \ell^{bu}, p^{se}) \equiv \begin{cases} p^{se} + \tau\ell^{bu} + \theta & \text{if he sells} \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

where $\theta \sim \mathsf{U}[-1, 0]$, ℓ^{bu} is the mass of the population of buyers with whom the seller can potentially trade ($\ell^{bu} \in [0, 1]$ since the mass of all buyers equals 1) and p^{se} is the price received by the seller net of commissions and fees. It will be useful to assume $\tau \in [0, 1)$; as will be seen below, $\tau \geq 1$ would imply that sellers may be willing to pay for selling their security.

The inclusion of liquidity in trader's utility function deserves some comments. Liquidity in this model is a market attribute which traders attach value to. The (unmodelled) reasons for that are the associations between trading volume and cost of providing immediacy, as documented in Grossman and Miller (1988), and between expected time to sell an asset and number of quotes received, as in Lippman and McCall (1986). Uncertainty and time are two dimensions at the heart of the very concept of liquidity; yet, including them explicitly in a model of the industrial organization of exchange services was deemed unnecessary. As O'Hara (1995) puts it, "liquidity is not so easily defined as it is recognized." The notion that will be kept in the background is that larger markets are more liquid, and that liquidity is a valuable attribute to traders either because it reduces search costs or because it improves bargaining positions. While liquidity is in essence a dynamic phenomenon, in this paper we follow Santos and Scheinkman (2001) and concentrate in one aspect of it, namely that ceteris paribus, investors prefer to trade in a market with higher volume.

Brokers There is a continuum of possibly heterogenous and perfectly competitive brokers of mass $B \in \mathbb{R}_{++}$. If broker $b \in [0, B]$ executes x transactions, he incurs in total brokerage costs (excluding any fees charged by the exchange fees) of

$$\alpha_b x + \tfrac{1}{2}\beta_b x^2 + \eta,$$

with $\alpha_b \geq 0$, $\beta_b > 0$. This cost function reflects diminishing returns to scale in brokerage services.

We assume that both, α_b and β_b, are non-decreasing and twice continuously differentiable functions of b in $[0, B]$. This implies that for all b, $b' \in [0, B]$ such that $b \leq b'$, broker b can intermediate a given volume x at least as cheaply as broker b'. We also assume that there is a fixed cost $\eta \geq 0$ to become an active broker.

All brokers are required to be members of at least one exchange in order to process any transaction. Those affiliated brokers we call "active." We denote by $\mathcal{B}$ the set of all active brokers.

Exchanges There are $\overline{E} \in \mathbb{N}$ identical exchanges of positive mass, indexed by $e \in \{1, 2, ..., \overline{E}\}$. If an exchange executes a mass y of trades, it incurs a total cost of

$$\gamma y + \tfrac{1}{2}\delta y^2 + \mu,$$

where μ is the sunk cost of establishing an exchange, and $\gamma > 0$. To simplify the exposition, in what follows we will consider two particular cases:

- $\delta = 0$ and $\mu > 0$. That is, the trading technology exhibits constant marginal cost but fixed costs, hence returns to scale are increasing.
- $\delta \geq 0$ and $\mu = 0$. That is, there are non-increasing returns to exchange services.

$\mathcal{E} \subseteq \{1, 2, ..., \overline{E}\}$ is the set of active exchanges (i.e. those that have paid the entry cost μ) and $E \leq \overline{E}$ is the number of active exchanges.

It may be argued that in practice large fixed costs necessarily make exchanges a natural monopoly. Nevertheless, Cybo Ottone et al. (2002) argue that, while exchange scale economies are very difficult to measure, they do not seem to be very important.[5] We choose to be agnostic

[5] See also Malkamaki (1999).

on the issue and analyze the implications of increasing, constant and decreasing returns to scale in exchange services.

Trading procedures and rules To buy or sell a security, investors must place an order with an active broker. In turn, brokers must execute all orders through an exchange, even when the same broker stands at both sides of the transaction.[6] Thus, to complete a transaction each side must use one unit of brokerage services and one unit of exchange services, a fixed–proportions technology. For simplicity we assume that only executed trades generate costs to brokers and exchanges. That is, posting an order is costless, but executing a transaction is not.

To post orders in exchange e a broker must be member of it. We denote broker b's membership in exchange e by $b \in \mathcal{B}_e$, where $\mathcal{B}_e$ is the set of brokers who are members of exchange e. Similarly, we let $\mathcal{E}_b$ be the set of exchanges where broker b is a member. Since exchange services are necessary to make a trade, broker b can trade with broker b' if and only if they meet at some exchange, that is, b, $b' \in \mathcal{B}_e$ for at least one exchange e, or equivalently if $\mathcal{E}_b \cap \mathcal{E}_{b'} \neq \phi$. Henceforth we will call broker b "single" if she is member of only one exchange (i.e. $\mathcal{E}_b$ is a singleton) and "dual" if she is member of two exchanges. Note that $\cup_{e \in \mathcal{E}} \mathcal{B}_e = \mathcal{B}$.

2.2. Liquidity and market integration

In our model liquidity affects the willingness to trade of buyers and sellers. In this subsection we discuss liquidity and define "integrated" and "fragmented" markets in connection to it. Moreover, we obtain sufficient conditions for a market to be integrated. We begin by making explicit what we mean by an 'allocation' in this context.

Definition 1. *An* allocation *is a correspondence*

$$\begin{aligned} \mathsf{E} \quad &: \quad [0, B] \rightarrow 2^{\{1,\ldots,E\}} \\ &: \quad b \rightarrow \mathcal{E}_b \end{aligned}$$

indicating brokers' affiliations, and a function

$$\begin{aligned} \mathsf{X} \quad &: \quad [-1, 1] \times [0, B] \times \mathcal{E} \rightarrow \{0, 1\} \\ &: \quad \theta, b, e \rightarrow x(\theta, b, e) \end{aligned}$$

that assigns where $x(\theta, b, e) = 1$ if investor θ (buyer if $\theta > 0$, seller otherwise) trades with broker b, and the trade is executed in exchange $e \in \mathcal{E}_b$, 0 otherwise.

[6] In many countries there are strict regulations that force brokers to follow this procedure, presumably to protect small investors from fraud.

Remark 1. *Note that we allow brokers to be members of more than one exchange. This is the case, for example, in Chile.*

The fact that trades must be settled in an exchange implies that the volume of buyers and sellers must coincide in each exchange. Therefore, we can restrict attention to allocations such that for all exchanges e

$$\int_{\mathcal{B}} \int_{-1}^{0} x(\theta, b, e) d\theta db = \int_{\mathcal{B}} \int_{0}^{1} x(\theta, b, e) d\theta db. \tag{2.3}$$

Definition 2 (Feasible allocations). *An allocation (E,X) is* feasible *if it satisfies condition (2.3).*

Now given that broker b belongs to all exchanges in $\mathcal{E}_b$, he can reach all brokers in those exchanges: $\cup_{e \in \mathcal{E}_b} \mathcal{B}_e \equiv \mathcal{B}_b$. Hence, for a given allocation (E,X) that satisfies (2.3), he can reach

$$\ell_b^{se} \equiv \sum_{\mathcal{E}} \int_{\mathcal{B}_b} \int_{-1}^{0} x(\theta, b', e) d\theta db'$$

sell orders and

$$\ell_b^{bu} \equiv \sum_{\mathcal{E}} \int_{\mathcal{B}_b} \int_{0}^{1} x(\theta, b', e) d\theta db'$$

buy orders. Let us refer to these numbers $(\ell_b^{se}, \ell_b^{bu})$ as the *liquidities* that broker b offers. They mean the following: broker b can offer liquidity ℓ_b^{se} to a buyer of the security and liquidity ℓ_b^{bu} to a seller of the security. Note that the sum is over all exchanges ($\mathcal{E}$), not only those of which b is a member. What matters for liquidity is the total number of orders executed by brokers that b can reach (i.e. brokers that are in $\mathcal{B}_b$). It does not matter whether or not those orders are executed in exchanges where b is a member. Note, moreover, that broker b does not benefit from the liquidity created by brokers who are only members of exchanges where b is not a member (i.e., trades executed by brokers b' such that $\mathcal{E}_b \cap \mathcal{E}_{b'} = \emptyset$).

The maximum liquidity that any broker may possibly offer is

$$\overline{\ell} \equiv \sum_{\mathcal{E}} \int_{\mathcal{B}} \int_{-1}^{0} x(\theta, b, e) d\theta db = \sum_{\mathcal{E}} \int_{\mathcal{B}} \int_{0}^{1} x(\theta, b, e) d\theta db,$$

where the equality follows from condition (2.3). This occurs when the broker meets almost every broker in at least one exchange. We are now ready to define an integrated market:

Definition 3. *A market is* integrated *if $\ell_b^{su} = \ell_b^{bu} = \overline{\ell}$ a.e. in $\mathcal{B}$. A market is* fragmented *otherwise.*

In a sense, this is to require that (almost) all brokers offer an homogeneous service. Disconnected networks of brokers would produce market segmentation, where each net can only offer a fraction of the total liquidity available in the economy. By contrast, if all networks are connected, everybody enjoys the maximum liquidity available. An integrated market is such that every potential investor is reachable from (almost) any broker, that is, almost any broker meets every other would-be counterpart's broker at some exchange.

Proposition 1 (Sufficient conditions for integration). *A market is integrated if any of the following is satisfied:*

(i) $\exists e^* \in \{1, ..., E\}$ *such that* $\mathcal{B}_{e^*} = \mathcal{B}$.

(ii) $\mathcal{E}_b \cap \mathcal{E}_{b'} \neq \emptyset$ *pairwise a.e. in* $\mathcal{B}$.

Proof. If $\exists e^* \in \{1, ..., E\}$ such that $\mathcal{B}_{e^*} = \mathcal{B}$ then $\cup_{e \in \mathcal{E}_b} \mathcal{B}_e \equiv \mathcal{B}_b = \mathcal{B}$ for almost all b. Then

$$\ell_b^{se} \equiv \sum_{\mathcal{E}} \int_{\mathcal{B}_b} \int_{-1}^{0} x(\theta, b', e) d\theta db' = \sum_{\mathcal{E}} \int_{\mathcal{B}} \int_{-1}^{0} x(\theta, b', e) d\theta db' = \bar{\ell} = \ell_b^{bu}$$

for almost all b, where the last equality follows from condition (2.3). This proves the first part. Part (ii) follows similarly after noting that (ii) implies that $\cup_{e \in \mathcal{E}_b} \mathcal{B}_e \equiv \mathcal{B}_b = \mathcal{B}$. ■

Remark 2. *(i) and (ii) are actually equivalent with two exchanges* ($E = 2$). *With* $E > 3$, *(i) implies (ii), but (ii) does not imply (i). The reason is that with two exchanges, there is only one other exchange to meet if they don't meet in one of them.*

Remark 3. *These conditions are not necessary. On the one hand, it is clear that (i) is stronger than (ii) in general. On the other hand, there might be some allocations where reaching a certain subset of brokers is not necessary because they don't give access to any counterparts for a trade. However, for generic allocations, (ii) is also a necessary condition.*

The previous discussion highlights the fact that it is brokers who offer liquidity to investors, not exchanges. Brokers gain access to counterparts for a trade by meeting other brokers, either in one exchange or in many. Nevertheless, brokers who are members of the same exchange may differ in the amount of liquidity they are able to offer to their clients when the market is not integrated, as the following example shows.

Example 1. *Let broker* b^* *be member of exchange* e^* *only; let* b *be member of all* $e \in \mathcal{E}$; *finally, assume* $\mathcal{B}_{e^*} \neq \mathcal{B}$ *and* $\int_{\mathcal{B}} \int_{-1}^{0} x(\theta, b', e^*) d\theta db' < \sum_{\mathcal{E}} \int_{\mathcal{B}} \int_{-1}^{0} x(\theta, b', e) d\theta db'$. *Then* $\mathcal{B}_{b^*} = \mathcal{B}_{e^*}$ *and*

$\mathcal{B}_b = \mathcal{B}$. *Broker* b^* *offers liquidities*

$$\ell_{b^*}^{se} \equiv \sum_{\mathcal{E}} \int_{\mathcal{B}_{e^*}} \int_{-1}^{0} x(\theta, b', e) d\theta db'$$

and

$$\ell_{b^*}^{bu} \equiv \sum_{\mathcal{E}} \int_{\mathcal{B}_{e^*}} \int_{0}^{1} x(\theta, b', e) d\theta db';$$

broker b*, in turn, offers liquidities* $\ell_b^{se} = \ell_b^{bu} = \overline{\ell}$*, since* $\mathcal{B}_b = \mathcal{B}$*. Moreover,* $\ell_b^{se} = \ell_b^{bu} = \overline{\ell} > \max\{\ell_{b^*}^{se}, \ell_{b^*}^{bu}\}$ *since* $\mathcal{B}_{e^*} \neq \mathcal{B}$*. But both brokers are members of exchange* e^**.*

It is apparent from the example that liquidity is *not* an attribute of the exchange. For instance, from the point of view of a seller, the liquidity of a trade in exchange e is less when executed by b^* than when it is executed by b.

We can define an indicator of average or market liquidity and market integration:

Definition 4 (Market liquidity and integration). *Let* B *be the mass of active brokers. We define 'market liquidity,'* $\ell \in [0, 1]$*, by*

$$\ell^{se} \equiv \frac{1}{B} \int_0^B \ell_b^{se} db$$

and 'market integration,' $\mathcal{I}^{se} \in [0, 1]$*, as*

$$\mathcal{I}^{se} \equiv \frac{\ell^{se}}{\overline{\ell}},$$

and analogously for ℓ^{bu} *and* $\mathcal{I}^{bu}$*. Note that when the market is integrated, market liquidity equals* $\overline{\ell}$ *and integration* 1*. Hence, integration refers to connectedness, whereas liquidity to market size.*

For future reference, it is useful to derive a demand for transactions in an integrated market. A potential buyer would trade only if $\theta + \tau\ell^{se} - p^{bu} \geq 0$, that is, $\theta \geq p^{bu} - \tau\ell^{se}$. Hence, a fraction $\int_{p^{bu}-\tau\ell^{se}}^{1} d\theta = 1 - p^{bu} + \tau\ell^{se}$ of potential buyers will trade at price p^{bu} and liquidity ℓ^{se}. Similarly, a seller will trade if $p^{se} + \tau\ell^{bu} + \theta \geq 0$; thus, at price p^{se} and liquidity ℓ^{bu}, a fraction $\int_{-(p^{se}+\tau\ell^{bu})}^{0} d\theta = p^{se} + \tau\ell^{bu}$ will sell. But since the market is integrated, $\ell^{se} = \ell^{bu} = \overline{\ell}$ and $\overline{\ell} = 1 - p^{bu} + \tau\overline{\ell}$, which implies that $p^{bu} = 1 - (1 - \tau)\,\overline{\ell}$. Moreover, since $\ell^{se} \equiv p^{se} + \tau\ell^{bu}$ and $\overline{\ell} = p^{se} + \tau\overline{\ell}$, it follows that $p^{se} = \overline{\ell}\,(1 - \tau)$; so that for any given liquidity level

$$\begin{aligned} p^{bu} - p^{se} &= 1 - \overline{\ell}\,(1 - \tau) - \overline{\ell}\,(1 - \tau) \\ &= 1 - 2\overline{\ell}\,(1 - \tau)\,, \end{aligned}$$

which gives the demand for transacting. In contrast, if the market is not integrated, the willingness to pay for transactions will not depend on the total volume of trade, but on the fractions being offered on each segment.

3. The social optimum

We begin by studying the problem of a welfare-maximizing planner. We assume that the planner can choose the number of exchanges, E and the exchange fee per transaction, $p_{\mathcal{E}}$. The brokerage market is perfectly competitive and there is free entry, so that the social planner cannot directly choose the mass of active brokers, the allocation of brokers across exchanges and the allocation of trades across brokers and exchanges; it can only choose a price, $p_{\mathcal{E}}$, which is taken as given by brokers and investors.

It is useful to start by showing that the planner always integrates the market. After that we characterize the competitive equilibrium of brokerage when the market is integrated, taking the choices of the planner as parameters. Last, we formulate and solve the planner's problem.

3.1. Some preliminary results

We begin by showing a fundamental result.

Proposition 2 (Integration Pareto-dominates fragmentation). *For any given number of exchanges E and measure of brokers B, any allocation (E,X) such that $\mathcal{I} < 1$ is Pareto dominated by another allocation (E′,X) such that $\mathcal{I} = 1$.*

Proof. If $E = 1$ then the market is trivially integrated. Suppose now that $E > 1$ and that allocation (E,X) is optimal. Since $\mathcal{I} < 1$, it follows that $\ell_b^{se} < \bar{\ell}$ for a positive measure of brokers $b \in \mathcal{B}' \subset \mathcal{B}$. Now suppose that brokers reallocate and become members of all exchanges, so that the market becomes integrated and $\mathcal{I} = 1$. Then all transactions that were formerly made can still be made in the same exchanges. Moreover, becoming a member of an exchange does not use up resources. Thus, even if exactly the same trades are done, a positive measure of the buyers will have now higher utility $\theta + \tau\bar{\ell} > \theta + \tau\ell_b^{se}$, and social welfare will be higher. A similar argument shows that any allocation of buyers such that $\mathcal{I} < 1$ cannot be optimal. Hence, fragmentation is Pareto-inferior to integration. ■

The proof of Proposition 2 only uses the fact that liquidity is valuable. It depends neither on the cost structure of brokers (the proof works without changing the pattern of trades that were done before integration), nor the number of exchanges, nor the cost structure of exchanges. This implies that in the (unconstrained) social optimum $\ell_b = \bar{\ell}$ for all brokers b no matter how many exchanges there are. Hence, all brokers offer exactly the same liquidity and the optimal number of exchanges is determined solely by the technology to execute trades, as the following proposition shows.

Proposition 3. *Fix the number of brokers B. (i) If $\delta = 0$ and $\mu > 0$ (i.e. the technology to execute trades exhibits increasing returns to scale) then having one exchange is socially optimal. (ii) If $\delta > 0$ and $\mu = 0$ (i.e. the technology to execute trades exhibits decreasing returns to scale) then it is optimal to have $\overline{E}$ exchanges and spread volume equally across exchanges.*

Proof. (i) Let $\delta = 0$ and $\mu > 0$ and suppose that $E > 1$. Then Proposition 2 implies that any allocation (E,X) such that $\mathcal{I} < 1$ is Pareto dominated by another allocation (E',X) such that $\mathcal{I} = 1$. Now, choose an allocation $(1,\mathsf{X}')$ (where 1 denotes that only exchange 1 is active) and all brokers do exactly the same trades but all in exchange 1. Then $\mathcal{I} = 1$, total brokerage costs are exactly the same, and total fixed exchange costs are reduced.

(ii) Let $\delta > 0$ and $\mu = 0$, suppose that $E = 1$ and consider an allocation $(1,\mathsf{X})$. Clearly $\mathcal{I} = 1$. Now choose allocation (E,X') which differs from $(1,\mathsf{X})$ in that all brokers are members of all exchanges, all brokers execute the same trades in such a way that volume in each exchange equals $\frac{\bar{\ell}}{\overline{E}}$. Then total brokerage costs are exactly the same, still $\mathcal{I} = 1$, and total exchange costs are reduced since trading costs are convex in volume. ■

Corollary 1. *For any given number of exchanges E, a sufficient condition to minimize execution costs is that volume ℓ is equally spread across exchanges.*

Proof. It follows directly from exchanges' costs being convex in ℓ. ■

Integration is not sufficient to allow the planner to spread trades equally across all exchanges, however, as shown by the trivial case when all brokers are members of only one and the same exchange. Hence, cost minimization requires that brokers are members of a sufficient number of exchanges; trivially, this is satisfied when each broker is member of all exchanges, but weaker membership requirements may also do. For simplicity, we will assume in what follows that each broker is member of each exchange. Last, recall that in an integrated market, we can use the inverse demand for transactions, $p^{bu} - p^{se} = 1 - (1 - \tau)2\ell$, which we derived in the previous section.

3.2. The competitive equilibrium in brokerage

We now characterize the competitive equilibrium in the brokerage market. Suppose that the planner chooses $E \geq 1$ exchanges and an exchange fee $p_{\mathcal{E}}$. Then, we define a competitive equilibrium in the brokerage market as follows:

Definition 5. *A* competitive equilibrium of brokerage services with free entry in an integrated market *is a price of brokerage services* $p_{\mathcal{B}}$*, a vector of executed trades* $[(x_b^*)_0^B, \overline{\ell}]$*, a pair of asset prices* (p^{bu}, p^{se}) *and a feasible allocation* (E,X) *such that, given* E *and* $p_{\mathcal{E}}$*:*

(i) $x_b^* = \arg\max\{p_{\mathcal{B}}x_b - \alpha_b x_b - \frac{\beta_b}{2}x_b^2\}$*;*

(ii) $(p_{\mathcal{B}} - \alpha_b)x_b^* - \frac{\beta_b}{2}(x_b^*)^2 \geq \eta$ *for all* $b \in [0, B]$*, with equality for* B*;*

(iii) sellers sell iff $p^{se} \geq -\theta - \tau\overline{\ell}$ *and buyers buy iff* $p^{bu} \leq \theta + \tau\overline{\ell}$*;*

(iv) $\int_{p^{bu}-\tau\overline{\ell}}^{1} d\theta = \int_{p^{bu}+\tau\overline{\ell}}^{0} d\theta = \overline{\ell} = \frac{1}{2}\int_0^B x_b^* db$*;*

(v) $p^{bu} = p^{se} + 2(p_{\mathcal{B}} + p_{\mathcal{E}})$*.*

(i) and (ii) state that brokers maximize profits in equilibrium; (iii) states the same for investors. (iv) says that in equilibrium the volume of buy orders equals the volume of sell orders; moreover, that number must coincide with the volume intermediated by brokers. The requirement that volumes sold and bought in each exchange must be the same is implicit in the fact that the allocation (E,X) is feasible.

The following lemma establishes sufficient conditions that characterize an equilibrium.

Lemma 1. *Let a competitive equilibrium of brokerage services exist. Then the following conditions are sufficient for* $p_{\mathcal{B}}^*$ *and* $[(x_b^*)_0^B, \overline{\ell}]$ *to be part of an equilibrium:*

$$p_{\mathcal{B}} = \alpha_b + \beta_b x_b^*; \tag{3.1}$$

$$(p_{\mathcal{B}} - \alpha_B)x_B^* - \frac{\beta_B}{2}(x_B^*)^2 = \eta \tag{3.2}$$

$$1 - (1 - \tau)2\overline{\ell} = 2(p_{\mathcal{B}} + p_{\mathcal{E}}) \tag{3.3}$$

$$2\overline{\ell} = \int_0^B x_b^* db \tag{3.4}$$

Proof. Condition (3.1) follows directly from the fact that brokers maximize benefits given the equilibrium price. (3.2) is a standard zero-profit condition that must hold for the marginal entrant B. Since $x_B^* \leq x_b^*$ for all b in $[0, B)$ (since α and β are nondecreasing in b), it immediately implies that $(p_{\mathcal{B}} - \alpha_b x_b^*) - \frac{\beta_b}{2}(x_b^*)^2 \geq \eta$ for all $b \in [0, B]$. Next, we know from the previous section that in an integrated market $p^{bu} - p^{se} = 1 - 2\overline{\ell}\,(1 - \tau)$. From the definition of equilibrium, moreover, $p^{bu} = p^{se} + 2(p_{\mathcal{B}} + p_{\mathcal{E}})$, from which (3.3) follows. This also implies $\int_{p^{bu}-\tau\overline{\ell}}^{1} d\theta = \int_{p^{bu}+\tau\overline{\ell}}^{0} d\theta = \overline{\ell}$, since the demand for transactions $1 - (1 - \tau)2\overline{\ell}$ is derived from investor's profit maximization. Last, condition (3.4) trivially follows from (iv) in the definition of equilibrium. ■

Note that $p_{\mathcal{B}}$, $[(x_b^*)_0^B, \overline{\ell}]$ and B are differentiable functions of $p_{\mathcal{E}}$, because the conditions of the implicit function theorem are met by (3.1)–(3.4). This will considerably simplify the exposition and analysis of the planner's and monopoly problems. Last, the assumptions imply that there exists a well-defined and upward sloping supply curve of brokerage services. Hence, if $p_{\mathcal{E}}$. is not too high a competitive equilibrium in brokerage exists and is unique.

3.3. The planner's problem

We are now ready to study the planner's problem. In practice, exchanges set their fees and brokers compete taking them as given. For this reason, we will study the problem of a planner that can choose the exchange fee $p_{\mathcal{E}}$ and takes the competitive brokerage market as given. Nevertheless, we will also show that choosing $p_{\mathcal{E}}$ is sufficient to implement the allocation that the planner would choose if it could dictate allocations directly.

As said before, we only consider two cases. In the first, $\delta = 0$ and $\mu > 0$: marginal costs of trade execution are constant, but there is a fixed cost of establishing an exchange. Then it is socially optimal to have only one exchange, which is a natural monopoly. The second case assumes that $\delta \geq 0$ and $\mu = 0$: there are non-decreasing marginal costs, but no fixed costs of entry. In this case the optimal number of exchanges is $\overline{E} > 1$.[7] Moreover, we know that the planner will choose to integrate markets and, when $\overline{E} > 1$, choose an allocation such that trades can be spread evenly across exchanges.

Given that the planner will choose to integrate markets and volume will be equally spread across all exchanges when $E > 1$, it follows that he will choose $p_{\mathcal{E}}$ to maximize

$$\mathcal{L} = \int_0^{\overline{\ell}} [1 - (1-\tau)2s]ds - \left\{ \int_0^B [\alpha_b x_b^* + \tfrac{\beta_b}{2}(x_b^*)^2 + \eta]db + E\left[\gamma \tfrac{\overline{\ell}}{E} + \tfrac{\delta}{2}\left(\tfrac{\overline{\ell}}{E}\right)^2 + \mu\right]\right\}, \qquad (3.5)$$

where E is either 1 or $\overline{E}$. In other words, the planner wants to maximize the difference between investor surplus on the one hand, and the sum of brokerage and exchange costs on the other.

The planner's maximization is subject to two different constraints. First, $[(x_b^*)_0^B, \overline{\ell})]$ and B come from a competitive equilibrium in brokerage, and as such are functions of $p_{\mathcal{E}}$. Second, we will impose a self-financing constraint

$$2p_{\mathcal{E}}\overline{\ell} - \left(\gamma + \frac{\delta}{2}\frac{\overline{\ell}}{E}\right)\overline{\ell} - E\mu \geq 0 \qquad (3.6)$$

that is, exchange fees collected from brokers in equilibrium must pay for all costs incurred by

[7] By ruling out cases with $\delta, \mu > 0$ we avoid finding the optimal integer number of exchanges, which is somewhat cumbersome and adds little understanding about the issues we are interested in.

exchanges. This self-financing constraint is standard in the economics of regulation and we introduce it to compare the planner's solution with the optimal solution in standard monopoly regulation.

The next proposition shows that it is optimal to price exchange services at marginal cost when the technology to execute trades exhibits non-increasing returns to scale ($\delta \geq 0$ and $\mu = 0$) and set the exchange fee equal to average costs when exchanges exhibit increasing returns ($\delta = 0$ and $\mu > 0$).

Proposition 4. *(i) If $\delta \geq 0$ and $\mu = 0$, then marginal cost pricing of exchange services is optimal. (ii) If $\delta = 0$ and $\mu > 0$, then average cost pricing of the exchange is optimal.*

Proof. (i) If the self-financing constraint holds with slack, then the first order condition is

$$\frac{d\mathcal{L}}{dp_{\mathcal{E}}} = \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\bar{\ell}}{E}\right)\frac{d\bar{\ell}}{dp_{\mathcal{E}}} + \left[2p_{\mathcal{B}}\frac{d\bar{\ell}}{dp_{\mathcal{E}}} - p_{\mathcal{B}}x_B^*\frac{dB}{dp_{\mathcal{E}}} - p_{\mathcal{B}}K\frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}}\right] = 0.$$

Using the comparative statics derivatives found in Appendix A, we note that

$$2p_{\mathcal{B}}\frac{d\bar{\ell}}{dp_{\mathcal{E}}} - p_{\mathcal{B}}x_B^*\frac{dB}{dp_{\mathcal{E}}} - p_{\mathcal{B}}K\frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}} = 0,$$

from which it follows that $2p_{\mathcal{E}} = \gamma$ satisfies the FOC. Moreover, we know that marginal cost pricing with decreasing returns to scale covers all costs and results in a quasirent. Hence, marginal cost pricing satisfies the self-financing constraint with slack.

(ii) When $\delta = 0$ having just one exchange is optimal and marginal cost pricing would imply $2p_{\mathcal{E}} = \gamma$. Hence, it is optimal to charge the lowest exchange fee that covers the total cost, which is

$$2p_{\mathcal{E}} = \gamma + \frac{\mu}{\bar{\ell}},$$

the average cost. ■

At first sight Proposition 4 looks quite standard. Yet, it is somewhat surprising that either marginal or average cost pricing of exchange services is optimal, despite the fact that there are network externalities in trading—investors gain when more investors trade. Shouldn't network externalities make of exchanges, as is often claimed, natural monopolies? Moreover, isn't entry of brokers suboptimally low in the competitive brokerage equilibrium? In fact, it is shown in the appendix that the optimal allocation we have obtained subject to the constraint that brokerage is competitive, is the same that would be chosen by the planner if he could *directly* choose B, x_b, and $\bar{\ell}$. In other words, a competitive equilibrium in brokerage with free entry plus the appropriate exchange fee is sufficient to decentralize the optimal allocation. Therefore:

Result 3.1. *Optimal pricing of exchange services depends only on the characteristics of the exchange's technology. In an integrated market, liquidity is irrelevant to determine its industrial organization: it does not make of exchanges natural monopolies.*

3.4. Discussion: why does liquidity become irrelevant?

Why is liquidity irrelevant and where did the network externality go? The intuition behind Result 3.1 can be understood in two steps. First, when the market is integrated neither brokers nor exchanges produce a network externality. To see this, suppose that $\overline{\ell}$ is being traded and add another broker or exchange. Liquidity will still be exactly $\overline{\ell}$, because no matter in which exchange or through which broker an investor decides to trade, it will access exactly the same liquidity. Thus, the contribution of an additional broker or exchange is just to reduce the marginal cost of executing a trade, as the same quantity is spread over more units. This is exactly the effect of an additional firm in any competitive market. Second, an additional investor, say a seller, in fact directly increases buyers' utility thus producing an externality. Nevertheless, for a given equilibrium brokerage fee, $p_{\mathcal{B}}$, the added liquidity increases the willingness to pay for the security. If the brokerage market is competitive, this increase willingness to pay will be fully translated into the net price received by the seller. Hence, at the margin the seller exactly internalizes the value of this external effect and the decision to sell is efficient—the externality is in fact *pecuniary*.

It is now clear why we call the industrial organization of stock exchanges "simple": when the market is integrated, optimal pricing of exchanges and brokerage services looks exactly as in any other market. Competition leads to optimal brokerage commissions and optimal exchange fees set by a planner follow exactly the same principles as in any standard market. Therefore, whether stock exchanges are natural monopolies depends only on the shape of the technology to execute trades, not on liquidity. As we will see in section 5, however, structure matters, because under some circumstances liquidity can be used to establish a barrier to entry. Before that, we will study pricing by a monopolistic exchange and compare it with the social optimum.

4. A monopolistic exchange

Until recently it was the norm that only one broker-owned exchange would exist. Indeed, many countries granted a legal monopoly to one exchange. It also used to be the case that exchanges had the right to fix commissions and in many countries it was illegal for a broker to undercut this brokerage fee. Increasingly, brokerage fees are market–determined. But, as we show in this section, exchanges can still exploit monopoly power through fixing $p_{\mathcal{E}}$ and may want to restrict entry into brokerage. In what follows we show that the ownership structure of the exchange affects its pricing and entry policies.

We begin by studying the optimum of a profit maximizing monopoly which is integrated with brokerage. Such structure is not found in practice, but is still provides a useful benchmark to compare the two structures that we observe more frequently, namely, the exchange owned by an outsider, and the brokers' cooperative. Further, it will become apparent that the outside-owned monopolistic exchange is a special case of the broker-owned exchange.

4.1. A vertically-integrated monopoly

We begin by showing that the monopolist is better off integrating the market.

Lemma 2 (Monopolists integrate markets). *For any given number of exchanges E and measure of brokers B, any allocation (E,X) such that $\mathcal{I} < 1$ there exists another allocation (E',X) such that $\mathcal{I} = 1$ and $\pi_M(\mathsf{E},\mathsf{X}) \leq \pi_M(\mathsf{E}',\mathsf{X})$, where π_M is the monopolist profit.*

Proof. If $E = 1$, the market is trivially integrated. Suppose now that $E > 1$ and that allocation (E,X) is optimal. Since $\mathcal{I} < 1$, it follows that $\ell_b^{se} < \overline{\ell}$ for a positive measure of brokers $\mathcal{B}' \subset \mathcal{B}$.

If brokers reallocate and become members of all exchanges, the market becomes integrated and $\mathcal{I} = 1$. Then, all transactions that were formerly made can still be made in the same exchanges. This, however, cannot be a competitive equilibrium in the brokerage market, because in equilibrium investors are indifferent with which broker to trade. This implies that at the old set of prices, brokers in $\mathcal{B}'$ are more attractive than the rest and all buyers would like to trade with them. Hence, fixing $\overline{\ell}$, it follows that total brokerage costs can be reduced if brokers in $\mathcal{B}'$ gain some volume, and brokers in $(\mathcal{B}')^{\mathrm{C}}$ loose some volume. Then (almost) *all* investors are better off. But then the monopolist can increase $p_{\mathcal{E}}$ and still $\overline{\ell}$ will be traded. A similar argument shows that any allocation of buyers such that $\mathcal{I} < 1$ cannot be optimal. Hence, profits for the monopolist are higher in an integrated market. ■

It follows that a vertically-integrated monopolist chooses $p_{\mathcal{E}}$ to maximize the difference between, on the one hand, exchange and membership fees, and, on the other the cost of providing exchange *and* brokerage services, viz.

$$\max_{p_{\mathcal{E}}} \left\{ \Pi_{\mathcal{E}}^{\mathrm{VI}} = 2p_{\mathcal{E}}\overline{\ell} - \left(\gamma\overline{\ell} + \frac{\delta}{2}\frac{\overline{\ell}^2}{E} + \mu E, \right) - \int_0^B [\alpha_b x_b + \frac{\beta_b}{2}(x_b^2) + \eta] db \right\}$$

subject to

$$2\overline{\ell} = \int_0^B x_b db, \tag{4.1}$$

$$2p_{\mathcal{E}} + 2p_{\mathcal{B}} = 1 - 2\bar{\ell}(1-\tau)$$

$$p_{\mathcal{E}}\bar{\ell} + \frac{\delta\bar{\ell}}{2E} - E\mu \geq 0 \tag{4.2}$$

(superscript "VI" denotes that this is a vertically-integrated monopoly). Assuming that the self-financing constraint (4.2) holds with slack at the optimum, the first order conditions of this problem are

$$\frac{\partial \Pi_{\mathcal{E}}^{\mathrm{VI}}}{\partial \bar{\ell}} = [1 - 2\bar{\ell}(1-\tau)] - 2\bar{\ell}(1-\tau) - \left(\gamma + \frac{\delta\bar{\ell}}{E}\right) + 2\lambda = 0,$$

$$\frac{\partial \Pi_{\mathcal{E}}^{\mathrm{VI}}}{\partial x_b} = \alpha_b + \beta_b x_b - \lambda = 0,$$

$$\frac{\partial \Pi_{\mathcal{E}}^{\mathrm{VI}}}{\partial B} = \alpha_b x_b + \frac{\beta_b}{2}(x_b)^2 + \eta - \lambda x_b = 0, \tag{4.3}$$

where λ is the multiplier of constraint (4.1). We note in passing that, as with the social planner, the optimal number of exchanges E depends only on the shape of the technology to execute trades. The monopolist will operate $\overline{E}$ exchanges if returns are decreasing ($\delta \geq 0$ and $\mu = 0$), and only one exchange if there are scale economies ($\delta = 0$ and $\mu > 0$).

Working out the first order conditions of this maximization, and setting $p_{\mathcal{B}} = \lambda$, we obtain

$$2p_{\mathcal{E}}^{\mathrm{VI}} = \left(\gamma + \frac{\delta\bar{\ell}}{E}\right) + 2\bar{\ell}(1-\tau). \tag{4.4}$$

$\gamma + \frac{\delta\bar{\ell}}{E}$ is the marginal cost of brokerage and $2\bar{\ell}(1-\tau)$ is the monopolist's markup. Thus, condition (4.4) is quite familiar: the vertically integrated monopolist adds the standard markup over the exchange's marginal cost, which depends on the elasticity of the demand for trading.[8] Note also that, as condition (4.3) shows, the vertically integrated monopoly does not distort entry, even though the number of brokers is less than with a social planner because the equilibrium volume $\bar{\ell}$ is smaller.

[8] In an integrated market the elasticity of the demand for trading with respect to the exchange fee is $-\frac{p_{\mathcal{E}}}{\bar{\ell}(1-\tau)}$ and the mark-up at the optimum is $\frac{\bar{\ell}(1-\tau)}{p_{\mathcal{E}}}$.

4.2. An independent monopoly

It is currently debated whether stock exchanges should be "demutualized", that is, owned by shareholders who are not brokers at the same time. We now study the choices of an independent monopolist who does not own brokerage. Since the proof of Lemma 2 also applies for an independent monopolist, the optimization problem is now

$$\max_{p_{\mathcal{E}}} \left\{ \Pi^{\mathrm{I}}_{\mathcal{E}} = 2p_{\mathcal{E}}\bar{\ell} - \left(\gamma\bar{\ell} + \frac{\delta}{2}\frac{\bar{\ell}^2}{E} + \mu E \right) \right\}, \tag{4.5}$$

subject to the same self-financing constraint (4.2), where $\bar{\ell}$ is determined in a competitive equilibrium of the brokerage market (superscript "I" denotes that this is an independent monopoly). If the self-financing constraint holds with slack, the FOC of problem 4.5 is

$$\frac{d\Pi^{\mathrm{I}}_{\mathcal{E}}}{dp_{\mathcal{E}}} = 2\bar{\ell} + \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\bar{\ell}}{E} \right) \frac{d\bar{\ell}}{dp_{\mathcal{E}}} = 0,$$

which looks quite standard. At the margin, the monopolist trades off the increase in revenue due to a higher exchange fee with the fall due to reduced volume. Since $\frac{d\bar{\ell}}{dp_{\mathcal{E}}} < 0$, it is apparent that the monopolist wants to set the exchange fee above marginal cost.

Nevertheless, it is convenient to examine the FOC a bit more closely. To do so, note that from Appendix A it follows that in a competitive equilibrium of brokerage $\frac{d\bar{\ell}}{dp_{\mathcal{E}}} = \frac{1}{\Delta}[(x^*_B)^2 + KC]$ with $\Delta \equiv -\{(1-\tau)[(x^*_B)^2 + KC] + 2C\}$, $C \equiv \alpha'_B x^*_B + \frac{1}{2}\beta'_B (x^*_B)^2$ and $K \equiv \int_0^B \frac{1}{\beta_b} db$. Thus,

$$2p^{\mathrm{I}}_{\mathcal{E}} = \left(\gamma + \frac{\delta\bar{\ell}}{E} \right) + 2\bar{\ell} \left[(1-\tau) + \frac{2C}{(x^*_B)^2 + KC} \right]$$

It is interesting to note that the elasticity of the supply curve for brokerage services with free entry is

$$\frac{(x^*_B)^2 + KC}{C} \frac{p_{\mathcal{B}}}{2\bar{\ell}}$$

(see Appendix C). As a benchmark, consider the case when all brokers are identical. Then $C = 0$ (the supply of brokerage services is perfectly elastic), $\frac{d\bar{\ell}}{dp_{\mathcal{E}}} = -\frac{1}{1-\tau}$ and

$$2p^{\mathrm{I}}_{\mathcal{E}} = \left(\gamma + \frac{\delta\bar{\ell}}{E} \right) + 2\bar{\ell}(1-\tau)$$

which is exactly the same exchange fee that would be chosen by a vertically integrated monopolist.

This is a standard result from the vertical control literature.[9] Nevertheless, if $C' > 0$ the supply curve of brokerage services is upward sloping. Then the monopoly exchange is also a monopsony in the market for exchange services. Thus, it fixes an exchange fee which is higher than the one that a vertically integrated monopoly would choose, and the markup is higher the more inelastic the brokerage supply curve.

Result 4.1. *An independent monopolist exploits both monopoly and monopsony power optimally.*

Therefore, for a given number of brokers, an independent monopoly distorts the market both on the demand and the supply side. Would the monopolist gain by restricting entry to brokerage? The answer is no, as the following proposition shows.

Proposition 5. *An independent monopolist does not restrict entry into brokerage.*

Proof. Assume that the monopolist restricts entry to $\overline{B}$ and then chooses $p_{\mathcal{E}}$ optimally. Also, suppose that entry is efficient. Then, simple differentiation yields

$$\frac{d\Pi_{\mathcal{E}}^{\mathrm{I}}}{dB} = \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\bar{\ell}}{E}\right) \frac{d\bar{\ell}}{dB}\bigg|_{B=\overline{B}} > 0,$$

since (see Appendix B) $\frac{d\bar{\ell}}{dB}\Big|_{B=\overline{B}} = \frac{x_B^*}{2+(1-\tau)K}$. ■

Thus, entry is less than what would be chosen by a social planner ($B^{\mathrm{I}} < B^{\mathrm{P}}$, where B^{P} is the number of brokers that optimally enters when the planner chooses $p_{\mathcal{E}}$), but only because the exchange fee set by the monopolist is too high. The intuition is quite simple: entry by an additional broker spreads the same volume among more brokers. Competition in brokerage reduces brokerage fees and thus allows the monopolist to increase the exchange fee and profits without affecting volume. Hence, it is in the monopolist's interest to allow for unrestricted entry, but charge an exchange fee that exploits both its monopoly and monopsony power. Also:

Corollary 2. *With an independent monopoly, the value of an exchange seat is zero.*

At the margin the last broker makes no quasirents, and thus the value of a seat is zero.

4.3. A broker-owned monopoly

Broker-owned exchanges are usually built on the "one member one share" rule and each receives an equal share of the profits made by the exchange. Nevertheless, entry and exchange fee policies

[9] This result dates back to Spengler (1950). See also Tirole (1988).

will in general differ from those of an independent monopolist because broker-members' profits also depend on their quasi rents as brokers. Moreover, we will see that cost differences among brokers introduce a conflict of interest among them. Thus, the objective of the exchange will depend on its internal decision rule. This section confirms formally the assertions formulated by Pirrong (1999) in this respect. Of particular importance is the result that brokers heterogeneity determines the behavior of the exchange.

To begin, we make the following simplifying assumption.

Assumption 1. *Original membership is the interval* $\left[0, \overline{B}\right]$*; that is, the original owners of the exchange are the lowest cost brokers.*

This efficiency assumption is not necessarily realistic, but it is generous in the sense that it endows the exchange with no initial inefficiency; any inefficiency will be due to the distorted incentives that drive a monopoly. Also, note that in a broker-owned exchange a necessary condition to be active as a broker is to have one share. Nevertheless, shareholders may choose not to become active brokers. In what follows we will use the terms "shareholders" and "broker-member" interchangeably. Brokers who choose to execute trades and pay the entry cost η we will call "active".

Consider broker-member b. Given $\overline{B}$ and $p_{\mathcal{E}}$, its profits are

$$\pi_b = \frac{1}{\overline{B}}\Pi_{\mathcal{E}} + \max\left\{p_{\mathcal{B}} x_b^* - \alpha_b x_b^* - \frac{\beta_b}{2}(x_b^*)^2 - \eta, 0\right\}, \tag{4.6}$$

The first term is very similar to the objective function of the independent monopolist, except for the fact that it is decreasing in the measure of members. The second term is the broker's profit, which, as we have seen, is decreasing in $p_{\mathcal{E}}$; it is the maximum between $p_{\mathcal{B}} x_b^* - \alpha_b x_b^* - \frac{\beta_b}{2}(x_b^*)^2 - \eta$ and 0 because the broker can always choose to remain inactive and save the fixed cost η. We start by characterizing pricing when the number of broker-members is large. Then we review the entry policy of a broker-owned exchange. Last, we study pricing in an exchange with a small number of broker-members.

Note that the exchange fee charged by a broker-owned monopoly and its entry policy will be the result of an election. Majority voting is a possible procedure[10], but there are other voting rules that may be included in a charter. For example, the exchange's charter may delegate the decision to a board of directors, restrict the vote to shareholders that become active brokers in equilibrium, and so on. Instead of sticking to one of these alternatives, we will study the optimal fee and entry policy that would be chosen by each broker member. This will reveal the central conflict of interest between broker-members, and enable us to show that a particular ownership

[10] See Hart and Moore (1996).

structure and distribution of voting rights over the exchange fee can lead to average cost pricing of exchange services.

4.3.1. Exchange fees with a large number of broker-members

So suppose that broker b could decide the exchange's policy. Call $p_{\mathcal{E}}(b)$ the exchange fee that maximizes b's profit (4.6) subject to the exchange's self-financing constraint (4.2). We now state two lemmas that will be useful to characterize the optimum that would be chosen by broker b. The first characterizes $p_{\mathcal{E}}(b)$ when the self-financing constraint is slack.

Lemma 3. *There exists a unique* $\widehat{b} > B^{\mathrm{I}}$ *such that (i) for all* $b > \widehat{b}$, $p_{\mathcal{E}}(b) = p_{\mathcal{E}}^{\mathrm{I}}$*; (ii) for all* $b \leq \widehat{b}$, $p_{\mathcal{E}}(b) < p_{\mathcal{E}}^{\mathrm{I}}$*;(iii)* $p_{\mathcal{E}}(b)$ *is increasing in* b *for all* $b \in [0, \widehat{b}]$.

Proof. In the Appendix. ■

Clearly, as Hart and Moore (1996) have pointed out, the exchange fee set by a broker-owned monopoly depends on which shareholder is pivotal in the vote to set the fee —hence on the internal voting procedure. But, in addition, the optimal fee from the point of view of broker-member b also depends on $\overline{B}$. In effect, if $\overline{B}$ is large enough, the highest costs broker-members will choose to remain inactive in equilibrium, and the market will behave as if there were free entry: while the number of shareholders is fixed in $\overline{B}$, the number of active brokers is endogenous. By contrast, if $\overline{B}$ is small enough, then every broker-member will be active in equilibrium.

The following lemma makes precise what we mean by "large" and "small" $\overline{B}$.

Lemma 4. *(i) For all* b *there exists a number* $B(b)$*, such that (i) if* $\overline{B} > B(b)$*, then brokers in* $(B(b), \overline{B}]$ *remain inactive when* $p_{\mathcal{E}} = p_{\mathcal{E}}(b)$*; (ii) if* $\overline{B} \leq B(b)$ *then all brokers are active when* $p_{\mathcal{E}} = p_{\mathcal{E}}(b)$*; (iii)* $B(b)$ *is non-increasing in* b*; (iv) for* $b \geq \widehat{b}$, $B(b) = \widehat{b}$.

Proof. In the Appendix. ■

We are now ready to study the case when $\overline{B}$ is large. If the self-financing constraint holds with slack, some manipulation of the FOC (D.1) yields

$$2p_{\mathcal{E}}(b) = \left(\gamma + \frac{\delta\overline{\ell}}{E}\right) + 2\overline{\ell}\left[(1-\tau) + \frac{2C}{(x_B^*)^2 + KC}\right] - \overline{B}x_b^* \frac{2C}{(x_{B(b)}^*)^2 + KC} < p_{\mathcal{E}}^{\mathrm{I}}, \tag{4.7}$$

where we have used the derivatives which are found in Appendix A. It follows that when the number of brokers is large, an active broker $b \leq B(b)$ chooses an exchange fee which is lower than with an independent monopolist. The reason is that the active broker internalizes in part the cost of a high exchange fee because it extracts part of his quasirent and redistributes it to other brokers. It is useful to manipulate the (4.7) a bit further, to yield

$$2p_{\mathcal{E}}(b) = \left(\gamma + \frac{\delta \bar{\ell}}{E}\right) + 2\bar{\ell}(1-\tau) + \frac{2C\overline{B}}{(x^*_{B(b)})^2 + KC}(\overline{x} - x^*_b),$$

where $\overline{x} \equiv \frac{2\bar{\ell}}{\overline{B}}$ is the average volume per *member* (not active broker). Note first that a broker who optimally would set $p_{\mathcal{E}}(b)$ so that $\overline{x}(b) = x^*_b(b)$ (henceforth call it "the average broker") would set the same exchange fee that a vertically integrated monopoly, $p_{\mathcal{E}}^{\mathrm{VI}}$; an low-cost broker (i.e. one that would optimally set a fee $p_{\mathcal{E}}(b)$ so that $x^*_b(b) > \overline{x}(b)$ in equilibrium) would like to set it below; and a "small" broker ($x^*_b(b) < \overline{x}(b)$), above.

Result 4.2. *There is a conflict of interest between broker-members. The lower the broker's cost, the lower the exchange fee he would like to set.*

The conflict of interest between members is a central feature of broker-owned exchanges. It occurs because the exchange fee plays two roles, exploitation of market power and, given the one member-one share rule, redistribution of profits from relatively low-cost broker-members to relatively high-cost ones. An average broker does not see this conflict, because her contribution to exchange profits exactly matches her share in the dividend. Therefore, at the margin she receives back as profits the additional dollar of exchange fees she generates. Given that the average broker fully internalizes the cost of brokerage at the margin, she would like to set exactly the same fee as a vertically integrated monopoly. By contrast, one with a lower–than-average-cost broker receives a smaller share of exchange profits than his contribution to them, and would like to set an exchange fee lower than $p_{\mathcal{E}}^{\mathrm{VI}}$. At the other extreme, those with higher costs than the average broker would like to set the exchange fee higher than $p_{\mathcal{E}}^{\mathrm{VI}}$ to redistribute profits. In the extreme, a member with too high costs chooses to remain inactive, and would like to act as an independent monopoly and set $p_{\mathcal{E}}^{\mathrm{I}}$.

Now note that $\overline{x}$ is decreasing in $\overline{B}$ but does not affect $B(b)$. Hence, the more shareholders there are, the more likely it is that an active broker will prefer to set the fee below $p_{\mathcal{E}}^{\mathrm{VI}}$. More importantly, when $\overline{B}$ is sufficiently large, then all active brokers prefer that. Thus:

Result 4.3. *If the measure of broker-members is large, and active brokers set the fee, then welfare with a broker-owned monopoly is unambiguously higher than with a vertically-integrated monopoly.*

It is now apparent that when $\overline{B}$ is very large, then all active brokers may want to set the exchange fee as low as possible and the exchange's self-financing constraint will bind. Therefore:

Result 4.4. *If the number of broker members is sufficiently large, then all active brokers would like to set the exchange fee equal to average cost.*

When the mass of inactive members is sufficiently large, there is no conflict of interest among active brokers and they behave "efficiently". The reason is that by then all active brokers contribute a share to exchange profits which is higher than what they get back as dividends and all want to set the lowest possible exchange fee that is consistent with self-financing—i.e. average cost pricing. By contrast, a member who chooses not to become active would like to act as an independent monopoly. It follows that:

Result 4.5. *If the exchange is a natural monopoly ($\delta = 0$ and $\mu > 0$) and $\overline{B}$ is sufficiently large, all active brokers would like to set the same fee as the social planner.*

Of course, if the fee is chosen by majority voting and $\overline{B}$ is very large, then inactive broker-members would vote to set the fee at $p_{\mathcal{E}}^{\mathrm{I}}$. But Result 4.4 is shows that allocating voting rights over the exchange fee only to active brokers but equal cash flow rights to all shareholders yields the planner's allocation when the exchange is a natural monopoly. The key here is that a very large $\overline{B}$ implies that each active broker gets a negligible share of the exchange's profits which is clearly less than its contribution to total profits. If those brokers have the decision power to set the exchange fee they will want to set it as low as possible—that is, equal to average cost. Thus, we have shown that there is a simple way of regulating exchanges optimally when they are natural monopolies: force them to issue a lot of shares, but give the right to vote on the exchange fee only to active brokers.

4.3.2. Entry policy in a broker-owned exchange

We can now evaluate the entry policy of a broker-owned exchange. It is clear that shareholders would like to block further entry when $\overline{B}$ is large: additional shareholders only dilute profits and reduce the dividends received by each current shareholder. But when the number of brokers is small there are two additional effects. On the one hand, additional brokers compete and reduce brokerage profits, which reinforces the profit-dilution effect. But on the other, marginal brokerage costs fall, volume increases and so do exchange profits. The main result of this subsection is that the profit-dilution effect prevails and all brokers would like to block entry.

To show this result, we begin by computing the derivative of b's profit with respect to $\overline{B}$, which yields

$$\frac{d\pi_b}{d\overline{B}} = \frac{1}{\overline{B}}\left[-\frac{\Pi_{\mathcal{E}}}{\overline{B}} + \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\overline{\ell}}{E}\right)\frac{d\overline{\ell}}{d\overline{B}} + \overline{B}x_b^*\frac{dp_{\mathcal{B}}}{d\overline{B}}\right], \tag{4.8}$$

where $\Pi_{\mathcal{E}}$ is the exchange's profit. The first two terms are common to all broker members: on the one hand, adding one member forces to divide the exchange's profit among more members; but, on the other hand, an additional member increases trading volume and the exchange's rents (the second term). The third term, which is clearly negative, is the fall in brokerage income caused by the fall in the commission $p_{\mathcal{B}}$ due to the added competition brought about by additional entry. But we can now prove the following proposition:

Proposition 6. *Assume that the same broker who chooses $p_{\mathcal{E}}$ also decides whether to accept new members. Then, if $\Pi_{\mathcal{E}}(0) > 0$ then all brokers in $[0, \overline{B}]$ would like to block entry.*

Proof. The assumption that $\Pi_{\mathcal{E}}(0) > 0$ implies that the interval $[0, \overline{B}]$ is large enough so that $p_{\mathcal{E}}$ chosen by $b = 0$ meets the exchange's self-financing constraint in equilibrium.[11] To sign (4.8), it is helpful to manipulate it a bit to yield

$$\frac{d\pi_b}{d\overline{B}} = \frac{1}{\overline{B}}\left[-\frac{\Pi_{\mathcal{E}}}{\overline{B}} + \frac{\overline{B}x^*_{\overline{B}}}{K}(\overline{x} - x^*_b)\right],$$

where we have used the derivatives obtained in Appendix B, which assume a fixed measure of brokers. Hence, the average broker (denote it $\overline{b}$) would chose to block entry because

$$\frac{d\pi_{\overline{b}}}{d\overline{B}} = -\frac{\Pi_{\mathcal{E}}}{\overline{B}^2} < 0.$$

To prove the proposition we show that the derivative of $\frac{d\pi_b}{d\overline{B}}$ with respect to b is negative for all $b \in [\overline{b}, \overline{B}]$. This derivative equals

$$\frac{d\pi_b}{d\overline{B}} = \left\{-\frac{1}{\overline{B}}\frac{d\Pi_{\mathcal{E}}}{dp_{\mathcal{E}}} + \frac{\overline{B}}{K}\left[(\overline{x} - x^*_b)\frac{dx^*_{\overline{B}}}{dp_{\mathcal{E}}} + x^*_{\overline{B}}\left(\frac{d\overline{x}}{dp_{\mathcal{E}}} - \frac{dx^*_b}{dp_{\mathcal{E}}}\right)\right]\right\}\frac{dp_{\mathcal{E}}}{db}\frac{db}{d\overline{B}}.$$

Some algebra shows that $\frac{d\Pi_{\mathcal{E}}}{dp_{\mathcal{E}}} = -\frac{2\overline{B}x^*_b}{\Delta}$, with $\Delta = -[2 + (1-\tau)K]$. Moreover, using the derivatives deduced in Appendix B, we write

$$\frac{d\pi_b}{d\overline{B}} = \frac{2}{\Delta}\left[x^*_b + \frac{\overline{B}}{K}(\overline{x} - x^*_b)\frac{1}{\beta_{\overline{B}}} + \frac{\overline{B}x^*_{\overline{B}}}{K}\left(\frac{K}{\overline{B}} - \frac{1}{\beta_b}\right)\right]\frac{dp_{\mathcal{E}}}{db}\frac{db}{d\overline{B}}.$$

Since $\Delta < 0$ and $\frac{dp_{\mathcal{E}}}{db} > 0$, the derivative will be negative if the sign of the expression within brackets is positive. Recalling that equilibrium in the brokerage market implies that $p_{\mathcal{B}} = \alpha_b + \beta_b x^*_b = \alpha_{\overline{B}} + \beta_{\overline{B}}x^*_{\overline{B}}$, so that $\beta_b x^*_b = (\alpha_{\overline{B}} - \alpha_b) + \beta_{\overline{B}}x^*_{\overline{B}}$, this expression can be rewritten as

$$\frac{\overline{B}}{K}(\overline{x} - x^*_b)\frac{1}{\beta_{\overline{B}}} + x^*_{\overline{B}} + \frac{1}{\beta_b}(\alpha_{\overline{B}} - \alpha_b) + \frac{1}{\beta_b}\left(\beta_{\overline{B}} - \frac{\overline{B}}{K}\right)x^*_{\overline{B}}.$$

[11] It can be shown that when there are scale economies and $\overline{B}$ is sufficiently small, then the exchange makes losses, no matter how high $p_{\mathcal{E}}$ is set. Hence, if $\overline{B}$ is too small, brokers must allow some entry until the self-financing constraint is met.

Now the first three terms are clearly positive for $b \in [\bar{b}, \overline{B}]$. To sign the fourth term, recall that $K \equiv \int_0^{\overline{B}} \frac{1}{\beta_{b'}} db'$. Hence, $\beta_{\overline{B}} K - \overline{B} = \int_0^{\overline{B}} \frac{\beta_{\overline{B}}}{\beta_{b'}} db' - \overline{B} > 0$, since $\frac{\beta_{\overline{B}}}{\beta_{b'}} > 1$ for all $b \in [0, \overline{B})$. Hence $\frac{d\pi_b}{d\overline{B}} < 0$ for all $b \in [0, \overline{B}]$. ■

The incentive to block entry suggests a trade off: on the one hand, for a given $\overline{B}$, a broker-owned exchange sets a lower exchange fee because, contrary to an independent monopoly, all brokers confront the marginal cost of brokerage. Nevertheless, an independent monopoly never blocks entry, so that one may think that when $\overline{B}$ is sufficiently small a broker-owned monopoly may lead to higher exchange fees.

5. Competition between exchanges

A natural presumption that often emerges in policy discussions is that competition between exchanges should lead to efficient pricing. Even with scale economies, the argument goes, exploitation of market power would prompt entry by another exchange and discipline the monopoly.

A general analysis of competition between exchanges would require us to distinguish between alternative ownership structures and combinations of broker memberships; that is beyond the scope of this paper. In this section we do two things. First, we analyze the particular but important case of competition with integrated markets, showing that only with constant returns to scale competition leads to efficient pricing of exchange services. Next, by means of examples, we show that an exchange may use liquidity coupled with exclusive membership requirements as a barrier to entry; and that even if entry by a rival is economically viable, fragmentation may be an equilibrium outcome of competition.

It is useful to begin by defining some conditions that must hold in equilibrium when exchanges compete. As before, we assume that brokers are price-takers and that each exchange chooses the exchange fee. Thus, exchanges compete in fees, which seems a natural assumption.

Recall that if brokers b and b' are involved in a transaction in exchange e as, respectively, buyer and seller, then

$$p_b^{bu} = p_{\mathcal{B}} + p'_{\mathcal{B}} + 2p_{\mathcal{E}}^{e} + p_{b'}^{se}. \tag{5.1}$$

Next note that if two brokers b and b' are able to offer the final asset prices p_b^{bu} and $p_{b'}^{bu}$, buyers would be indifferent between trading with one or the other if and only if they obtain the same utility from both: $\theta + \tau \ell_b^{se} - p_b^{bu} = \theta + \tau \ell_{b'}^{se} - p_{b'}^{bu}$, or

$$p_{b'}^{bu} - p_b^{bu} = \tau \left(\ell_{b'}^{se} - \ell_b^{se} \right) \tag{5.2}$$

As a consequence, any price differential must be compensated by a proportional difference in the liquidity offered. If this condition is not met, all buyers would prefer to trade with the cheapest of them. Hence, it must hold for all active brokers. The analogous condition for sellers is:

$$p^{se}_{b'} - p^{se}_{b} = -\tau \left(\ell^{bu}_{b'} - \ell^{bu}_{b} \right) \tag{5.3}$$

Alternatively, one can say that buyers are indifferent between pairs $\left(p^{bu}, \ell^{se}\right)$ if they lie in the same line $p^{bu} = p^{bu}_0 + \tau \ell^{se}$, where p^{bu}_0 is the competitive price associated to a 0-liquidity broker. Similarly, sellers are indifferent between points in the line $p^{se} = p^{se}_0 - \tau \ell^{bu}$. (Of course, p^{bu}_0 and p^{se}_0 are determined in equilibrium.)

A case we will study below is when there are no dual brokers. Then the set of brokers do not intersect across exchanges, the liquidities for buy and sell orders coincide and are the same for all brokers within each exchange. Hence, liquidity becomes inseparable from the exchange, and the notation ℓ_e has a clear meaning: $\ell_e \equiv \ell^{se}_b = \ell^{bu}_b$ for all $b \in \mathcal{B}_e$. Moreover, since within each exchange brokers must be indifferent between processing buy or sell orders, and in view of the fact that the way the exchange's bill is split among buyer and seller is a matter of convention, (5.1) can be written as $p^{bu} = 2p_{\mathcal{B}} + 2p_{\mathcal{E}} + p^{se}$

The indifference conditions (5.2) and (5.3) must be satisfied simultaneously by any exchange that wishes to operate—in equilibrium, there can be no buy order without an associated sell order. Therefore, it is possible to focus on spreads rather than bid and ask prices separately, in the understanding that when spreads $s \equiv p^{bu} - p^{se}$ are competitive, bid and ask prices also are. To do so, substract (5.3) from (5.2), which yields

$$s_{b'} - s_b = \tau(\ell^{se}_{b'} + \ell^{bu}_{b'}) - \tau(\ell^{se}_{b} + \ell^{bu}_{b}).$$

If there are no dual brokers and b is member of exchange 1, then $\ell^{se}_b = \ell^{bu}_b = \ell_1$, the same holds for broker b', and this expression is zero. Hence, one can speak generically about "exchange's 1 spread". It follows immediately that $s_2 - s_1 = 2\tau(\ell_2 - \ell_1)$. Hence, both exchanges are competitive if and only if their spreads ℓ_e, $e \in \{1, 2\}$ lie on the line

$$s_e = s_0 + 2\tau \ell_e, \tag{5.4}$$

where s_0 is an spread that is competitive at a null level of liquidity $\ell = 0$. Henceforth we shall refer to equation (5.4) as the "Liquidity Adjusted Bid-Ask Spread" line (LABAS, for short), for it is the collection of spreads that are indifferent for buyers and sellers to any other offer in the line. As such, it allows comparisons of different liquidity-spread combinations: from an investor's point of view, those above the line are worse, those below are preferred. In the example at hand, we see that (s_1, ℓ_1) and (s_2, ℓ_2) are competitive offers if they both satisfy (5.4):

$$\left.\begin{array}{l} s_1 = s_0 + 2\tau\ell_1 \\ s_2 = s_0 + 2\tau\ell_2 \end{array}\right\} \Rightarrow s_2 - s_1 = 2\tau(\ell_2 - \ell_1)$$

The LABAS line simplifies the analysis of competition, for it reduces the dimension in which competition occurs from two (liquidity and spread) to one (the 0-liquidity spread, i.e. the intercept of the LABAS line). In particular, we can study competition of the Bertrand type by treating this 0-liquidity spread s_0 as the analog of the price variable in conventional oligopoly models (see Vives [2000]).

On the other hand, if liquidity has no value (that is, $\tau = 0$), competition occurs only in prices: lower spreads without any adjustment for liquidity are preferred by investors, regardless of the volume traded in each particular exchange. In this case, the LABAS line is horizontal. This is also what happens in an integrated market, since by definition, every broker offers the same liquidity, leaving spreads as the only variable for competing.

5.1. Competition in an integrated market

The following proposition establishes that brokers and exchanges must charge the same fees in equilibrium in an integrated market.

Proposition 7. *In an integrated market (i) all active brokers charge the same brokerage commission $p_{\mathcal{B}}$ in equilibrium; (ii) all exchange where transactions take place charge the same fee $p_{\mathcal{E}}$.*

Proof. In an integrated market $\ell_b^{se} = \ell_b^{bu} = \overline{\ell}$ for all brokers b. It follows from the indifference conditions (5.2) and (5.3) that all brokers offer the same buy and sell prices, which implies that $p_b^{bu} - p_{b'}^{se} = s$ for all transactions. Therefore, in an integrated market all trades executed in an exchange must have the same spread s in equilibrium. Now assume by way of contradiction that exchange e charges an exchange fee that is higher than that charged by any other exchange. Then brokers who trade in e must charge smaller commissions than brokers in other exchanges, otherwise $s_e > s$. But since the market is integrated, there must be a positive measure of brokers who are members of other exchanges, and they would prefer to settle trades in an exchange that charges a lower exchange fee. ■

Note that this result holds regardless of the ownership structure of exchanges, as long as the market is integrated. Next we show that when there are no restrictions to multiple memberships, one should expect the market to become integrated.

Proposition 8. *Every broker weakly prefers full membership.*

Proof. Consider broker b with membership $\mathcal{E}_b \in 2^{\overline{E}}$ such that $\ell_b^{se} \leq \overline{\ell}$ or $\ell_b^{bu} \leq \overline{\ell}$, and assume that he executes some buy and sell trades. By joining all exchanges she can increase her liquidity to $\overline{\ell}$ and the commission charged by $\tau(\overline{\ell} - \ell_b^{se})$ (to buyers) or by $\tau(\overline{\ell} - \ell_b^{bu})$ and still retain their clients. ■

Again, Proposition 8 does not depend on the ownership structures of exchanges; the proof exploits the fact that all brokers gain individually by increasing their liquidity. This result is important, because it implies that one should expect the market to become integrated if restrictions to cross memberships are banned by the regulator.

We can now explore how competition depends on the exchanges' cost functions. Let's begin with decreasing returns to scale in the exchange technology. Then we saw in section 3 that it is optimal to integrate the market and set the exchange fee equal to marginal cost. The following proposition shows that if exchanges act as price takers, the resulting Walrasian allocation would implement the planner's optimum.

Proposition 9. *In an integrated market, the Walrasian equilibrium allocation is efficient.*

Proof. The Walrasian equilibrium allocation is a feasible allocation that satisfies (i) $p_{\mathcal{B}} = \alpha_b + \beta_b x_b$, since x_b maximizes π_b a.e. in $\mathcal{B}$; (ii) $2p_{\mathcal{E}} = \gamma + \delta y_e$, since y_e maximizes π_e for all $e \in E$; (iii) $\overline{E} y_e = 2\overline{\ell}$; (iv) $2(p_{\mathcal{B}} + p_{\mathcal{E}}) = 1 - 2\overline{\ell}\,(1-\tau)$ (market clearing). These conditions correspond to equations (C.5)-(C.9) that define the Pareto-optimal allocation in the case of several exchanges. ■

Proposition 9 is important, because it shows again that the IO of stock exchanges is "simple": when the market is integrated, liquidity is not a cause of market failure. Instead, as in standard markets, competition may fail to yield the efficient allocation when exchanges are natural monopolies or they are few in numbers. It it thus natural to analyze competition of the Bertrand kind, where exchanges compete by setting $p_{\mathcal{E}}$.

To proceed, assume, as is customary in this model, that the lowest-fee exchange is forced to serve all demand for transactions it faces.[12] Also, as usual, assume that, in the case of a tie, the market is split according to the "capacity sharing rule," this is, the market shares that would obtain if both competitors acted as price takers. Finally, for ease of exposition assume that there are two identical and independent exchanges and that there is free entry into brokerage.

Figure 1 illustrates the nature of equilibria with increasing costs[13]. The solid line corresponds to the profits the exchange would obtain as a monopolist, when charging $p_{\mathcal{E}}$; the dotted line shows the profits the same exchange would obtain as a duopolist, when both exchanges charge $p_{\mathcal{E}}$ and the market is split in half. Note that both curves intersect to the left of $p_{\mathcal{E}}^{\mathrm{M}}$: as $p_{\mathcal{E}}$ falls away from $p_{\mathcal{E}}^{\mathrm{M}}$,

[12] See Vives (2000, ch. 5).
[13] See Vives (2000, ch.5) for an exposition and Dastidar (1995, 1997) for the formal analysis.

duopoly profits fall because two exchanges spread volume thus achieving lower total costs. If, as in this case, both exchanges have the same convex cost structure, Proposition 5.1 in Vives (2001, ch. 5) implies that the set of equilibrium fees is the set $[p_{\mathcal{E}}^{-}, p_{\mathcal{E}}^{+}]$ that satisfies $\pi_d(p_{\mathcal{E}}^{+}) \geq \pi_m(p_{\mathcal{E}}^{+})$ and $\pi_d(p_{\mathcal{E}}^{-}) = 0$—the thick segment in figure 1. Fees higher than $p_{\mathcal{E}}^{+}$ cannot be part of an equilibrium, because an exchange would increase profits by slightly undercutting, becoming a monopoly and serving the whole market. Now it is known that the interval $[p_{\mathcal{E}}^{-}, p_{\mathcal{E}}^{+}]$ contains the Walrasian allocation, but also prices that are higher can be sustained as an equilibrium. The reason is that capacity constraints restrain undercutting. Hence, there is no guarantee that prices will be efficient. More generally, the next proposition characterizes Bertrand competition in an integrated market under increasing, constant and decreasing returns to scale in the technology to execute trades.

Proposition 10. *If the market is integrated and the technology to execute trades exhibits: (i) increasing returns to scale ($\delta = 0$ and $\mu > 0$), then only one exchange can survive in equilibrium; (ii) constant returns to scale ($\delta = \mu = 0$), then $2p_{\mathcal{E}} = \gamma$ in equilibrium; (iii) decreasing returns to scale ($\delta > 0$ and $\mu = 0$), then there are multiple equilibria, one of which is the Walrasian equilibrium.*

Proof. By proposition (7), $p_{\mathcal{B}}$ is the same for all brokers. Each active broker processes

$$\max\left\{\frac{p_B - \alpha_b}{\beta_b}, 0\right\} = \arg\max \pi_b$$

many orders, so that a price of $p_{\mathcal{B}}$ induces brokers to process a volume of $2\ell = \int_{\mathcal{B}} \max\left\{\frac{p_B - \alpha_b}{\beta_b}, 0\right\} db$. This implicitly defines the equilibrium brokerage commission $p_{\mathcal{B}}(\ell)$, an increasing and continuous function of ℓ. As a duopolist, each exchange obtains a profit of $2p_{\mathcal{E}}y_e - (\gamma y_e + \frac{\delta}{2}y_e^2)$. If both duopolists charge the same price, they split the market according to the capacity-sharing rule: $y_e = \frac{y}{2} = \frac{\ell}{2}$. Moreover, since demand is given by $s = 1 - 2\ell(1-\tau)$, an exchange fee of $p_{\mathcal{E}}$ will produce a total volume defined implicitly by $2[p_{\mathcal{B}}(2\ell) + p_{\mathcal{E}}] = 1 - 2\ell(1-\tau)$, which holds iff $\ell^* = \ell(p_{\mathcal{E}})$. In order to determine whether a given $p_{\mathcal{E}}$ is an equilibrium fee, we need to check whether there are incentives for any duopolist to deviate. Higher prices will drive the duopolist out of the market, yielding a profit of 0; lower prices will attract all demand. Hence, $p_{\mathcal{E}}$ is an equilibrium fee iff

$$\begin{gathered}
2p_{\mathcal{E}}\tfrac{\ell}{2} - \left(\gamma\tfrac{\ell}{2} + \tfrac{\delta}{2}\left(\tfrac{\ell}{2}\right)^2 + \mu\right) \geq 0 \\
2p_{\mathcal{E}}\ell - \left(\gamma\ell + \tfrac{\delta}{2}\ell^2 + \mu\right) \leq 2p_{\mathcal{E}}\tfrac{\ell}{2} - \left(\gamma\tfrac{\ell}{2} + \tfrac{\delta}{2}\left(\tfrac{\ell}{2}\right)^2 + \mu\right) \\
\ell = \ell\left(p_{\mathcal{E}}\right)
\end{gathered}$$

Rearranging, we have:

$$\begin{aligned} 2p_{\mathcal{E}} &\geq \gamma + \tfrac{1}{4}\delta\ell + \tfrac{2\mu}{\ell} \\ 2p_{\mathcal{E}} &\leq \gamma + \tfrac{3}{4}\delta\ell \\ \ell &= \ell\left(p_{\mathcal{E}}\right) \end{aligned} \tag{5.5}$$

(ii) Follows immediately from (5.5) taking $\delta = \mu = 0$: its unique solution is $2p_{\mathcal{E}} = \gamma$. In turn, (iii) follows from taking $\delta > 0$ and $\mu = 0$, case in which it is clear that $\gamma + \frac{1}{4}\delta\ell < \gamma + \frac{3}{4}\delta\ell$ for any $\ell \neq 0$. Hence, the set of equilibrium fees is nonempty, non-degenerate, and contains the Walrasian price $2p_{\mathcal{E}} = \gamma + \delta\ell$ (from lemma 9). Finally, in the case (i) ($\delta = 0$ and $\mu > 0$) there is no solution to the system (5.5), for it requires simultaneously the price to cover the average cost ($2p_{\mathcal{E}} \geq \gamma + \frac{2\mu}{\ell}$) and the profit to a monopolist to be lower than the profit to a duopolist ($2p_{\mathcal{E}} \leq \gamma$). The intuition is the usual one: there is no way that a duopolist would earn a profit while at the same time restrain from cutting marginally its fee in order to drive its competitor out. The only plausible situation is one in which there is only one exchange, for a monopolist always has higher profits than a duopolist: $2p_{\mathcal{E}} \geq \gamma + \frac{2\mu}{\ell} \geq \gamma$. ■

Proposition 10 tells us that competition leads necessarily to efficiency only if the technology to execute trades exhibits constant returns to scale. Not surprisingly, integration and competition are not feasible when exchanges are natural monopolies[14]. When returns are increasing or constant, larger volumes do not increase the exchange's unit cost. This yields the standard Bertrand equilibrium when returns are constant, but implies that competition cannot occur with integration and scale economies. On the other hand, decreasing returns to scale do not guarantee an efficient allocation either, as there exist equilibria where exchanges price above marginal costs.

5.2. Liquidity and fragmentation as a barrier to entry

Proposition 8 established that every broker weakly prefers full membership. However, the example that we are about to present shows that when liquidity matters, an exchange might have an incentive to require exclusive membership from its brokers, banning them to trade in a rival exchange. By doing so, it can appropriate the liquidity generated by the trade that takes place in it. This liquidity, in turn, may suffice to preclude entry. In this way, liquidity and fragmentation can act together as a barrier to entry, even when an entrant faces no cost disadvantage whatsoever.

In particular, assume that there are two exchanges: a monopolist (incumbent) and a potential

[14] If one changes the sharing rule in such a way that in case of a tie an exchange is selected randomly to serve the whole market, then the unique Bertrand equilibrium is to set $p_{\mathcal{E}}$ equal the least break even monopoly price. Then the market is contestable (see Vives [2000, p. 119]). Nevertheless, such sharing rule looks very implausible to model competition between exchanges.

entrant. To keep it simple, assume further that the exchange technologies exhibit constant returns to scale (average cost equals γ), and that both exchanges are profit maximizers (independent exchanges). The incumbent $e = 1$ confronts a perfectly elastic supply of identical brokers. As we saw in section 4.2, in this case the monopolist charges

$$2p_{\mathcal{E}} = \gamma + 2\overline{\ell}(1-\tau).$$

Moreover, free entry with identical brokers implies that the marginal cost of brokerage is constant. With a slight abuse of notation, call this constant marginal cost α. It follows that in equilibrium $p_{\mathcal{B}} = \alpha$ and

$$s_1 \equiv 2p_{\mathcal{B}} + 2p_{\mathcal{E}} = 2\alpha + \gamma + 2\overline{\ell}(1-\tau).$$

Some tedious algebra shows that an unrestricted monopolist sets $p_{\mathcal{E}} = \frac{1}{4}(1+\gamma-2\alpha)$, which implies that $\overline{\ell} = \frac{1-\gamma-2\alpha}{4(1-\tau)}$ after noting that in equilibrium, $s_1 = 2p_{\mathcal{B}} + 2p_{\mathcal{E}} = \frac{1}{2}(1+\gamma+2\alpha)$ (see figure 1).

Now consider an identical potential entrant (call it exchange $e = 2$), but assume that the incumbent is successful in enforcing *exclusive* membership to all brokers who have already joined him. Hence, if a broker is already a member of exchange 1, it cannot become member of exchange 2, and brokers who become members of exchange 2 are not accepted into exchange 1. Thus the market is fragmented and whether investors trade with brokers who are members of exchange 2 depends not only on the differences in spreads $s_2 - s_1$, but also on their relative liquidities. Given that upon entry liquidity in exchange 1 equals $\overline{\ell}$, the LABAS curve is now $s_2 = s_1 - 2\tau(\overline{\ell}.-\ell_2)$. $s_1 = \frac{1}{2}(1+\gamma+2\alpha)$

The LABAS curve is depicted in figure 2. In this particular case, it shows the combinations of liquidity (ℓ_2) and spreads (s_2) that leave investors indifferent between trading in exchange 1 or 2, *given* that exchange 1 offers liquidity $\overline{\ell} = \frac{1-\gamma-2\alpha}{4(1-\tau)}$ and spread $\frac{1}{2}(1+\gamma+2\alpha)$. It intercepts the demand curve at the spread set by the incumbent monopolist: if the entrant could somehow offer liquidity $\overline{\ell}$, it could charge spread s_1 an still leave investors indifferent. Yet when investors expect to obtain less than liquidity $\overline{\ell}$ in exchange 2, then s_2 must be smaller than s_1. In the extreme, if investors expect $\ell_2 = 0$, then some tedious algebra shows that exchange 2 can charge at most $1 - \frac{1-\gamma-2\alpha}{2(1-\tau)}$.

Can a potential entrant expect to make profits? An interesting feature of fragmentation is that it makes the profitability of the entrant depend on coordination between *investors*. One Nash equilibrium of the simultaneous coordination game played by investors is that all investors move to trade in exchange 2, which would make an offer of slightly less than s_1 profitable. Nevertheless, such fast and costless coordination among investors (as opposed to brokers) does not seem to square with the common idea that liquidity is a property of each exchange, not easily transferred to others.[15] If, as seems plausible, investors take as given the liquidity $\overline{\ell}$ offered by the incumbent exchange 1, then any investor unilaterally deviating to trade through a broker who is member of exchange 2 expects her to provide liquidity $\ell_2 = 0$. In that case it is profitable for each investor to unilaterally move to exchange 2 only if the spread is at most $1 - \frac{1-\gamma-2\alpha}{2(1-\tau)}$. But if $\tau > \frac{1}{2}$ (i.e. the preference of liquidity by investors is "strong"),

$$1 - \frac{1-\gamma-2\alpha}{2(1-\tau)} < \gamma + 2\alpha,$$

and the exchange makes losses if it chooses to attract business. Therefore, in this case, liquidity together with fragmentation enables the incumbent to erect a de facto barrier to entry, despite of having no cost advantage. Note that in this example the entrant can draw on the pool of brokers who did not become members of exchange 1. Since we have assumed that there are plenty of these, exchange 2 faces no cost disadvantage.

Exclusivity and fragmentation are central in erecting the barrier to entry. To see this, consider now an integrated market. Hence, upon entry all brokers who are members of exchange 1 automatically become members of exchange 2. Then regardless of the actual number of orders executed in exchange 2, integration implies that every broker offers liquidity $\overline{\ell}$ (in other words, the LABAS line is now horizontal at s_1). Moreover, and this would probably be key in practice, integration implies that coordination between investors is no longer needed: because brokers care only about the exchange fee, investors get liquidity $\overline{\ell}$ regardless with whom they choose to trade. Thus, an entrant who offers a spread of slightly less than s_1 would attract all transactions. As we saw in the previous subsection, both exchanges must then charge the same fee in equilibrium, and in the only Nash equilibrium $p_{\mathcal{E}} = \gamma/2$.

Similarly, note that when the preference for liquidity is not too strong ($\tau \leq \frac{1}{2}$), then the second exchange can make a profit even if it charges a spread of $1 - \frac{1-\gamma-2\alpha}{2(1-\tau)}$, and thus it would enter if the incumbent exploits its monopoly power.

5.3. Fragmentation as an equilibrium outcome

The present example illustrates an extreme situation in which, by integrating, exchanges lose all market power, driving profits to zero. It is in their interest, then, to maintain fragmentation; this is done by requiring exclusivity from their brokers.

As before, assume that there are two independent exchanges ($e = 1, 2$), each with total costs

[15] For example, commenting on competition between exchanges in Europe, The Economist recently stated that "Liquidity will thus be key in determining the winners and loosers in this battle of the bourses. [...] On this measure, the LSE is still some way in front of its rivals."

of γy_e. Also, assume that there is a mass μ of identical brokers in each exchange, where each broker has a cost function of $\beta x^2/2$ (it will become evident that it is in the exchanges interest to limit entry into brokerage). If exchange e wants to process a mass of y_e trades, it needs to assure its brokers a brokerage commission of $p_{\mathcal{B}} = \frac{2y_e}{K}$. Note that a mass y_e of trades only generates a liquidity of $\ell_e = y_e$, since the market is fragmented.

>From section 5.1 we know that competition will occur in 0-liquidity spreads, fact that can be represented by the LABAS line $s_e = s_0 + 2\tau\ell_e$. Within this line, we are implicitly assuming that there is a common liquidity-adjusted bid price $p^{bu} = p_0^{bu} + \tau\ell$, that will attract a total volume

$$y^{bu} = \int_{p_0^{bu}}^{1} d\theta = 1 - p_0^{bu}$$

of buy orders, because only those traders whose valuations are such that $\theta + \tau\ell - p^{bu} \geq 0$ are willing to buy, that is, $\theta + \tau\ell \geq p_0^{bu} + \tau\ell$. A similar argument for sell orders leads to $y^{se} = \int_{-p_0^{se}}^{0} d\theta = p_0^{se}$. Adding both expressions, and using the fact that $y^{bu} = y^{se} = y$, we obtain $y^{bu} + y^{se} = 1 - p_0^{bu} + p_0^{se}$, so that $y = \frac{1-s_0}{2}$.

Given that both exchanges have the same cost structure, in the case of a tie they would split the market in halves: $y_1 = y_2 = \frac{y}{2}$. Hence $y = 2y_e = 2\ell_e$, so that

$$\ell_e = \frac{1 - s_0}{4}$$

On the other hand, if one of them charges a smaller liquidity-adjusted spread, it would get the whole market and, as a consequence, offer a total liquidity of $\ell_e = \frac{1-s_0}{2}$.

Hence, the profits to exchange e as a duopolist and as a monopolist are respectively:

$$\pi_{\text{duop}}(s_0) = (2p_{\mathcal{E}} - \gamma)\,\ell_e = \left[s_0 + 2\tau\left(\frac{1-s_0}{4}\right) - 2\frac{2}{K}\left(\frac{1-s_0}{4}\right) - \gamma\right]\frac{1-s_0}{4}$$
$$\pi_{\text{monop}}(s_0) = (2p_{\mathcal{E}} - \gamma)\,\ell_e = \left[s_0 + 2\tau\left(\frac{1-s_0}{2}\right) - 2\frac{2}{K}\left(\frac{1-s_0}{2}\right) - \gamma\right]\frac{1-s_0}{2}$$

Solving for the set of Bertrand equilibrium 0-liquidity spreads (that is, the solution to the system of inequalities $\pi_{duop}(s_0) \geq \pi_{monop}(s_0)$ and $\pi_{duop}(s_0) \geq 0$), one obtains $s_0 \in \left[\frac{(2\gamma-\tau)K+2}{(2-\tau)K+2}, \frac{(2\gamma-3\tau)K+6}{(2-3\tau)K+6}\right]$, with an associated range of $2p_{\mathcal{E}} \in \left[\gamma, \frac{\left(4\tau+\tau\gamma-2\gamma-3\tau^2\right)K^2+(12\tau-8-2\gamma)K-12}{((2-3\tau)K+6)K}\right]$ for the exchange fees. All but the lowest equilibrium fee yield positive profits for each exchange. Even though the exchanges' own average costs are constant, they compete as if they had increasing costs because, by the exclusive membership clause, the supply of brokerage services in each exchange is upward sloping. This fact implies the existence of appropriable quasirents, softening price competition.

In the absence of exclusive membership requirements, by Proposition 8 all brokers would like to become members of both exchanges, for the effect such membership would have on the liquidity they can offer to their clients. As we already saw, Bertrand competition between exchanges would then yield a unique equilibrium where $2p_{\mathcal{E}} = \gamma$, characterized by zero profits for both exchanges.

Essentially, exclusive membership gave exchanges the ability to differentiate products, on the one hand, and on the other to build up quasirents by exploiting their broker-members' decreasing returns. Market integration destroys both. Thus, exclusive membership can be a device to soften price competition which increases exchanges' profits. This certainly occurs at the expense of investors, brokers and efficiency, since Proposition 2 established that integration is a precondition for efficiency. Investors are hurt by higher spreads and lower liquidity. Brokers are hurt because of a lower volume and a smaller brokerage commission and quasirents.

6. Conclusion: liquidity and market structure

A central question driving this paper was: is liquidity (understood as a network externality) such a key determinant of the industrial organization of the markets for transacting assets, that the standard tools do not apply to them? We found that the answer is "it depends": it depends on whether the market is integrated or not. In an integrated market, the network of brokers is fully connected, thereby giving complete reachability to any pair of potential traders. In this case, the network externality becomes pecuniary, is internalized, and the organization of the industry is "simple," by which we mean that the standard tools of analysis carry over this industry as well. In this way, an allocation is efficient if volume is generated at the least cost, and the optimal level of volume is supplied.

In an integrated market the optimal number of exchanges depends only on the characteristics of the technology to execute trades—whether it exhibits increasing, constant or decreasing returns to scale. Such an allocation can be decentralized in a Walrasian equilibrium if technologies are convex, that is, the first welfare theorem holds. In particular, liquidity is not a source of increasing returns when the market is integrated. Arguably, a Walrasian equilibrium may not be the best possible description of actual competition, particularly because exchanges are hardly price takers. For instance, in a Bertrand equilibrium they may earn profits even under decreasing returns to scale. Yet, this is not a consequence of liquidity, as it is neither the fact that under increasing returns to scale exchanges are natural monopolies.

In contrast, in a fragmented market competition is less intense as liquidity acts as a form of product differentiation. In addition, liquidity is underprovided, and the transactions are not carried at the minimum possible cost in general. Thus, a fragmented market is necessarily inefficient.

Fragmentation may be endogenous as exchanges can build it by requiring exclusivity from their broker-members. It helps exchanges to appropriate part of the value of liquidity implied in the volume of orders it processes, and part of any quasirents generated by brokers. As such, a monopolist may use it as a barrier to entry. Or, a market may fail to become integrated in spite of the incentives brokers have to act in their client's interests. Thus, in the presence of liquidity, it is possible that the market outcome fails to be efficient. This points in the direction of making of integration a regulatory requirement.

Another question of interest in the context of stock exchanges is the optimality of their internal structure. An independent, profit-maximizing exchange behaves differently than a members' cooperative. Under decreasing returns to scale, the applicability of the first welfare theorem suggests that on the grounds of efficiency there is no reason to prefer the form of a cooperative over the independent exchange; on the contrary, internal conflicts of interest may have the undesirable consequence of breaking down profit maximization, leading perhaps to inefficient production.

Notably, when the technology to execute trades exhibits increasing returns to scale, and thus profit-maximization does not lead to efficiency, the cooperative may be preferred. The conflict of interest that arises from broker heterogeneity can be exploited in such a way as to make the exchange choose voluntarily to price exchange services efficiently —which in this case corresponds to average cost—. This is achieved by forcing membership to be as large as possible, while at the same time letting only active brokers vote on pricing policies. Any price above average cost implies a redistribution from active to inactive brokers, and will not obtain approval from such a non-representative board.

Our results suggest in the end that regulators should trust the market mechanism even in the presence of liquidity, but at the same time watch carefully membership policies and any other policies that might challenge integration.

References

[1] Banner, S., "The Origin of the New York Stock Exchange," *Journal of Legal Studies* **37**, 113-140, 1998.

[2] Cybo–Ottone, A., C. Di Noia and M. Murgia, "The Consolidation of Securities Exchanges: Is There Any Lesson From the European Experience?, mimeo, 2000.

[3] Dastidar, K., "On the Existence of Pure-Strategy Bertrand Equilibrium," *Economic Theory* **5**, 19-32, 1995.

[4] Dastidar, K., "Comparing Cournot and Bertrand in Homogeneous Product Market," *Journal of Economic Theory* **75**, 205-212, 1997.

[5] Di Noia, C., "Competition and Integration Among Stock Exchanges in Europe: Network Effects, Implicit Mergers and Remote Access," *European Financial Management* **7**, 39-72, 2001.

[6] Economides, N. and A. Siow, "The Division of Markets is Limited by the Extent of Liquidity," *American Economic Review* **78**, 108-121.

[7] Foucault, T. and C. Parlour, "Competition for Listings," CEPR Discussion Paper N^o 2222, 1999.

[8] Grossman, S., and M. Miller, "Liquidity and Market Structure," *Journal of Finance* **43**, 617–637, 1988.

[9] Hart, O. and J. Moore, "The Governance of Exchanges: Members' Cooperatives versus Outside Ownership," *Oxford Review of Economic Policy* **12**, 53–69, 1996.

[10] Khan, A. *The Economics of Regulation*. Volume II: Institutional Issues. New York: John Willey & Sons, 1971.

[11] Lippman, S. and J. McCall, "An Operational Measure of Liquidity," *American Economic Review* **76**, 43-55, 1986.

[12] Lee, R., *What is an Exchange? The Automation, Management, and Regulation of Financial Markets.* Oxford University Press, 1998.

[13] Lo, A. (editor), *The Industrial Organization and Regulation of the Securities Industry* (a NBER conference report.) The University of Chicago Press, 1996.

[14] Malkamaki, M. "Are there Economis of Scale in Stock Exchange Activities?," Bank of Finland Discussion Papers N°4, 1999.

[15] O'Hara, M., *Market Microstructure Theory.* Blackwell, 1995.

[16] Pirrong, C., "The Organization of Financial Exchange Markets: Theory and Evidence," *Journal of Financial Markets* **2**, 329–357, 1999.

[17] Pagano, M., "Trading Volume and Asset Liquidity," *Quarterly Journal of Economics* **104**, 255-274.

[18] Pirrong, C., "A Theory of Financial Exchange Organization," *Journal of Law and Economics* **43**, 437–471, 2000.

[19] Santos, T. and J. Scheinkman, "Competition Among Exchanges", *Quarterly Journal of Economics* **116**, 1027–1071, 2001.

[20] Shy, O., *The Economics of Network Industries.* Cambridge University Press, 2001.

[21] Spengler, J., "Vertical Integration and Anti-trust Policy," *Journal of Political Economy* **58**, 347-352, 1950.

[22] Tirole, J., *The Theory of Industrial Organization.* Cambridge: MIT Press, 1988.

[23] Vives, X., *Oligopoly Theory.* Cambridge: MIT Press, 2000.

Appendix

A. Comparative statics of a brokerage equilibrium with free entry

Totally differentiate (3.1)–(3.4) to obtain

$$dp_{\mathcal{B}} = \beta_b dx_b^*; \tag{A.1}$$

$$-\left[\alpha'_B x_B^* + \frac{1}{2}\beta'_B (x_B^*)^2\right] dB + x_B^* dp_{\mathcal{B}} = 0; \tag{A.2}$$

$$-(1-\tau)d\bar{\ell} - dp_{\mathcal{B}} = dp_{\mathcal{E}}; \tag{A.3}$$

$$2d\bar{\ell} - \left(\int_0^B \frac{1}{\beta_b} db\right) dp_{\mathcal{B}} - x_B^* dB = 0. \tag{A.4}$$

Rewritten in matrix form, this is

$$\begin{bmatrix} 1 & -\beta_b & 0 & 0 \\ x_B^* & 0 & 0 & -C \\ -1 & 0 & -(1-\tau) & 0 \\ -K & 0 & 2 & -x_B^* \end{bmatrix} \begin{bmatrix} dp_{\mathcal{B}} \\ dx_b^* \\ d\bar{\ell} \\ dB \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ dp_{\mathcal{E}} \\ 0 \end{bmatrix}$$

with $C \equiv \alpha'_B x_B^* + \frac{1}{2}\beta'_B (x_B^*)^2$ and $K \equiv \int_0^B \frac{1}{\beta_b} db$. Solving for the vector of exogenous variables, this yields

$$\begin{bmatrix} dp_{\mathcal{B}} \\ dx_B^* \\ d\bar{\ell} \\ dB \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 2C dp_{\mathcal{E}} \\ \frac{1}{\beta_b} 2C dp_{\mathcal{E}} \\ \left[(x_B^*)^2 + KC\right] dp_{\mathcal{E}} \\ 2x_B^* dp_{\mathcal{E}} \end{bmatrix}$$

with $\Delta \equiv -\left\{2C + (1-\tau)\left[(x_B^*)^2 + CK\right]\right\} < 0$. It follows that

$$\begin{bmatrix} \frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}} \\ \frac{dx_b^*}{dp_{\mathcal{E}}} \\ \frac{d\bar{\ell}}{dp_{\mathcal{E}}} \\ \frac{dB}{dp_{\mathcal{E}}} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 2C \\ \frac{2}{\beta_b} C \\ [(x_B^*)^2 + KC] \\ 2x_B^* \end{bmatrix}$$

B. Comparative statics of a brokerage equilibrium with a fixed number of brokers

Equilibrium conditions are now

$$p_{\mathcal{B}} = \alpha_b + \beta_b x_b^*; \tag{B.1}$$

$$1 - (1-\tau)2\bar{\ell} = 2(p_{\mathcal{B}} + p_{\mathcal{E}}) \tag{B.2}$$

$$2\overline{\ell} = \int_0^B x_b^* db \tag{B.3}$$

Totally differentiate (3.1)–(3.4) to obtain

$$dp_{\mathcal{B}} = \beta_b dx_b^*; \tag{B.4}$$

$$-(1-\tau)d\overline{\ell} - dp_{\mathcal{B}} = dp_{\mathcal{E}}; \tag{B.5}$$

$$2d\overline{\ell} - \left(\int_0^B \frac{1}{\beta_b} db\right) dp_{\mathcal{B}} = x_B^* dB. \tag{B.6}$$

Rewritten in matrix form, this is $\begin{bmatrix} 1 & -\beta_b & 0 \\ -1 & 0 & -(1-\tau) \\ -\frac{K}{x_B^*} & 0 & \frac{2}{x_B^*} \end{bmatrix} \begin{bmatrix} dp_{\mathcal{B}} \\ dx_b^* \\ d\overline{\ell} \end{bmatrix} = \begin{bmatrix} 0 \\ dp_{\mathcal{E}} \\ dB \end{bmatrix}$ with $K \equiv \int_0^B \frac{1}{\beta_b} db$.
Solving for the vector of exogenous variables, this yields

$$\begin{bmatrix} dp_{\mathcal{B}} \\ dx_b^* \\ d\overline{\ell} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 2dp_{\mathcal{E}} + x_B^*(1-\tau)dB \\ \frac{2}{\beta_b} dp_{\mathcal{E}} + \frac{x_B^*(1-\tau)}{\beta_b} dB \\ K dp_{\mathcal{E}} - x_B^* dB \end{bmatrix}$$

with $\Delta \equiv -[2 + (1-\tau)K] < 0$. It follows that

$$\begin{bmatrix} \frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}} \\ \frac{dx_b^*}{dp_{\mathcal{E}}} \\ \frac{d\overline{\ell}}{dp_{\mathcal{E}}} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 2 \\ \frac{2}{\beta_b} \\ K \end{bmatrix};$$

$$\begin{bmatrix} \frac{dp_{\mathcal{B}}}{dB} \\ \frac{dx_b^*}{dB} \\ \frac{d\overline{\ell}}{dB} \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} x_B^*(1-\tau) \\ \frac{x_B^*(1-\tau)}{\beta_b} \\ -x_B^* \end{bmatrix}.$$

C. The social optimum

C.1. The planner chooses $p_{\mathcal{E}}$

The planner chooses $p_{\mathcal{E}}$ to maximize

$$\mathcal{L} = \int_0^{\overline{\ell}} [1 - (1-\tau)2s] ds - \left\{ \int_0^B [\alpha_b x_b^* + \tfrac{\beta_b}{2}(x_b^*)^2 + \eta] db + E\left[\gamma \tfrac{\overline{\ell}}{E} + \tfrac{\delta}{2}\left(\tfrac{\overline{\ell}}{E}\right)^2 + \mu\right] \right\},$$

subject to

$$2p_{\mathcal{E}}\overline{\ell} - \left(\gamma + \frac{\delta}{2}\frac{\overline{\ell}}{E}\right)\overline{\ell} - E\mu \geq 0$$

Call ϕ the multiplier associated with the self-financing contstraint. Then the FOC is

$$\frac{d\mathcal{L}}{dp_{\mathcal{E}}} = \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\bar{\ell}}{E}\right)\frac{d\bar{\ell}}{dp_{\mathcal{E}}} + \left[2p_{\mathcal{B}}\frac{d\bar{\ell}}{dp_{\mathcal{E}}} - p_{\mathcal{B}}x^*_B\frac{dB}{dp_{\mathcal{E}}} - p_{\mathcal{B}}K\frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}}\right] - \phi\left[2\bar{\ell} + \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\bar{\ell}}{E}\right)\frac{d\bar{\ell}}{dp_{\mathcal{E}}}\right] = 0.$$

and the complementary slackness condition is

$$-\phi\left[2p_{\mathcal{E}}\bar{\ell} - \left(\gamma + \frac{\delta}{2}\frac{\bar{\ell}}{E}\right)\bar{\ell} - E\mu\right] = 0$$

Using the comparative statics derivatives found in Appendix A, we note that

$$2p_{\mathcal{B}}\frac{d\bar{\ell}}{dp_{\mathcal{E}}} - p_{\mathcal{B}}x^*_B\frac{dB}{dp_{\mathcal{E}}} - p_{\mathcal{B}}K\frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}} = 0,$$

so that the FOC reads

$$\frac{d\mathcal{L}}{dp_{\mathcal{E}}} = \left(2p_{\mathcal{E}} - \gamma - \frac{\delta\bar{\ell}}{E}\right)\frac{d\bar{\ell}}{dp_{\mathcal{E}}} - \phi\left[2\bar{\ell} + (2p_{\mathcal{E}} - \gamma)\frac{d\bar{\ell}}{dp_{\mathcal{E}}} + \frac{\delta\bar{\ell}}{E}\frac{d\bar{\ell}}{dp_{\mathcal{E}}}\right] = 0.$$

Now suppose $\delta \geq 0$ and $\mu = 0$. If $2p_{\mathcal{E}} = \gamma + \frac{\delta\bar{\ell}}{E}$, then the constraint holds with slack, $\phi = 0$, and marginal cost pricing of exchange services is optimal. Next assume $\delta = 0$ and $\mu > 0$. Then the constraint binds and average cost pricing of exchange services is optimal.

C.2. The planner chooses B, x_b, and $\bar{\ell}$

The planner now maximizes

$$\mathcal{L} = \int_0^{\bar{\ell}}[1 - (1-\tau)2s]ds - f\bar{\ell} - \left\{\int_0^B[\alpha_b x_b + \tfrac{\beta_b}{2}(x_b)^2 + \eta]db + E\left[\gamma\tfrac{\bar{\ell}}{E} + \tfrac{\delta}{2}\left(\tfrac{\bar{\ell}}{E}\right)^2 + \mu\right]\right\}, \tag{C.1}$$

subject to

$$2\bar{\ell} = \int_0^B x^*_b db, \tag{C.2}$$

$$f\bar{\ell} + \frac{\delta\bar{\ell}}{2E} - E\mu \geq 0, \tag{C.3}$$

$$f \geq 0, \tag{C.4}$$

where f is a fee charged to investors for using an exchange and constraint (C.3) is the equivalent of the self-financing constraint in the previous problem—it says that .

Proposition 11. *(i) If $\delta \geq 0$ and $\mu = 0$, then $f = 0$ and the optimal allocation is the same as when the planner chooses only $p_{\mathcal{E}}$. (ii) If $\delta = 0$ and $\mu > 0$, then $f = \frac{\mu}{\bar{\ell}}$ and the optimal allocation is the same as when the planner can only choose $p_{\mathcal{E}}$.*

Proof. Let λ be the multiplier associated with constraint (C.2), ϕ with (C.3) and σ with (C.4). The first order conditions of the problem are

$$\frac{\partial \mathcal{L}}{\partial \bar{\ell}} = [1 - (1-\tau)2\bar{\ell}] - f - \left(\gamma + \frac{\delta \bar{\ell}}{E}\right) - 2\lambda + \phi \frac{\delta}{2E} = 0, \tag{C.5}$$

$$\frac{\partial \mathcal{L}}{\partial x_b} = -(\alpha_b + \beta_b x_b) + \lambda = 0, \tag{C.6}$$

$$\frac{\partial \mathcal{L}}{\partial B} = \left(\alpha_B + \frac{\beta_B}{2} x_B + \eta\right) - \lambda x_B = 0, \tag{C.7}$$

$$\frac{\partial \mathcal{L}}{\partial f} = -1 + \phi \bar{\ell} + \sigma = 0. \tag{C.8}$$

with complementary slackness

$$\phi \left(f\bar{\ell} + \frac{\delta \bar{\ell}}{2E} - E\mu\right) = 0, \tag{C.9}$$

$$\sigma f = 0. \tag{C.10}$$

Now it is apparent that when $\lambda = p_{\mathcal{B}}$, conditions (C.6) and (C.7) are equivalent to conditions (3.1) and (3.2) which characterize the competitive equilibrium in the brokerage market (see Lemma 1). Hence, allocations are equivalent if: (i) $f = \phi = 0$ when $\delta \geq 0$ and $\mu = 0$; or (ii) $f = \frac{\mu}{\bar{\ell}}$ when $\delta = 0$ and $\mu > 0$; because then (C.5) becomes equivalent to (3.3) as $p_{\mathcal{E}} = \gamma + \frac{\delta \bar{\ell}}{E}$ in case (i) and $p_{\mathcal{E}} = \gamma + \frac{\mu}{\bar{\ell}}$ in case (ii), and (3.4) is satisfied as a constraint.

Now if $\delta \geq 0$ and $\mu = 0$ then $f\bar{\ell} + \frac{\delta \bar{\ell}}{2E} - E\mu = f\bar{\ell} + \frac{\delta \bar{\ell}}{2E} \geq 0$ with $f = 0$. Hence, $\phi = 0$, $\sigma = 1$ and equation (C.5) reads

$$[1 - (1-\tau)2\bar{\ell}] - \left(\gamma + \frac{\delta \bar{\ell}}{E}\right) - 2\lambda = 0.$$

Next, if $\delta = 0$ and $\mu > 0$ then $f > 0$. Hence $\sigma = 0$, $\phi = \frac{1}{\bar{\ell}}$, constraint (C.3) holds with equality and $f = \frac{\mu}{\bar{\ell}}$. Then, equation (C.5) reads

$$[1 - (1-\tau)2\bar{\ell}] - \frac{\mu}{\bar{\ell}} - \left(\gamma + \frac{\delta \bar{\ell}}{E}\right) - 2\lambda = 0,$$

which is equivalent to (3.3), setting $p_{\mathcal{E}} = \gamma + \frac{\delta \bar{\ell}}{E}$ and $\lambda = p_{\mathcal{B}}$. ■

D. Proof of Lemma 3

The optimal exchange fee chosen by b, $p_{\mathcal{E}}(b)$, must satisfy

$$\frac{d\pi_b}{dp_{\mathcal{E}}} = \frac{1}{\overline{B}} \left[2\bar{\ell} + \left(2p_{\mathcal{E}} - \gamma - \frac{\delta \bar{\ell}}{E}\right) \frac{d\bar{\ell}}{dp_{\mathcal{E}}}\right] + x_b^* \frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}} = 0 \tag{D.1}$$

(if member b would choose to remain inactive, then $x_b^* = 0$ and $\frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}}$ does not affect his decision). Now for all $p_{\mathcal{E}}$, $x_b^* \geq x_{b'}^*$ if $b < b'$, with strict inequality if $x_{b'}^* > 0$. Therefore, if $x_{b'}^* > 0$, $\frac{d\pi_{b'}}{dp_{\mathcal{E}}} > 0$ when evaluated at $p_{\mathcal{E}}(b)$ and $p_{\mathcal{E}}(b') > p_{\mathcal{E}}(b)$. Moreover, since $x_{b'}^* > 0$ by assumption, $p_{\mathcal{E}}^{\mathrm{I}} > p_{\mathcal{E}}(b')$. hence x_b^* is decreasing in $[0, b']$ and hence $p_{\mathcal{E}}(b)$ is increasing in $[0, b']$. On the other hand, if $x_b^* = 0$ given $p_{\mathcal{E}}(b)$, then $p_{\mathcal{E}}(b) = p_{\mathcal{E}}^{\mathrm{I}}$.

Next we show that $\widehat{b}$ exists and is greater than B^{I}. To see this, observe that if $p_{\mathcal{E}} = p_{\mathcal{E}}^{\mathrm{I}}$, then $\frac{d\pi_{B^{\mathrm{I}}}}{dp_{\mathcal{E}}} = x_{B^{\mathrm{I}}}^* \frac{dp_{\mathcal{B}}}{dp_{\mathcal{E}}} < 0$. It follows that broker B^{I} would optimally chose $p_{\mathcal{E}}(B^{\mathrm{I}}) < p_{\mathcal{E}}^{\mathrm{I}}$. Moreover, since the equilibrium brokerage fee $p_{\mathcal{B}}$ increases with a lower exchange fee, broker B^{I} would make brokerage profits

$$[p_{\mathcal{B}}(B^{\mathrm{I}}) - \alpha_{B^{\mathrm{I}}}] x_{B^{\mathrm{I}}}^*(B^{\mathrm{I}}) - \frac{\beta_{B^{\mathrm{I}}}}{2} [x_{B^{\mathrm{I}}}^*(B^{\mathrm{I}})]^2 - \eta > 0,$$

where, with a slight abuse of notation, $p_{\mathcal{B}}(B^{\mathrm{I}})$ denotes the equilibrium brokerage fee and $x_{B^{\mathrm{I}}}^*(B^{\mathrm{I}})$ the optimal quantity brokered given $p_{\mathcal{B}}(B^{\mathrm{I}})$ when $p_{\mathcal{E}} = p_{\mathcal{E}}(B^{\mathrm{I}})$. Continuity implies that there exists $b > B^{\mathrm{I}}$ such that

$$[p_{\mathcal{B}}(B^{\mathrm{I}}) - \alpha_{B^{\mathrm{I}}}] x_{B^{\mathrm{I}}}^*(B^{\mathrm{I}}) - \frac{\beta_{B^{\mathrm{I}}}}{2} [x_{B^{\mathrm{I}}}^*(B^{\mathrm{I}})]^2 > [p_{\mathcal{B}}(b) - \alpha_b] x_b^*(b) - \frac{\beta_b}{2} [x_b^*(b)]^2$$

because

$$[p_{\mathcal{B}}(b) - \alpha_{B^{\mathrm{I}}}] x_{B^{\mathrm{I}}}^*(b) - \frac{\beta_{B^{\mathrm{I}}}}{2} [x_{B^{\mathrm{I}}}^*(b)]^2 > [p_{\mathcal{B}}(b) - \alpha_b] x_b^*(b) - \frac{\beta_b}{2} [x_b^*(b)]^2 > \eta.$$

Moreover, the same argument implies that $[p_{\mathcal{B}}(b) - \alpha_b] x_b^*(b) - \frac{\beta_b}{2} [x_b^*(b)]^2$ is decreasing in b Hence, there exists $\widehat{b}$ such that $[p_{\mathcal{B}}(\widehat{b}) - \alpha_{\widehat{b}}] x_{\widehat{b}}^*(\widehat{b}) - \frac{\beta_{\widehat{b}}}{2} [x_{\widehat{b}}^*(\widehat{b})]^2 = \eta$ with $p_{\mathcal{E}}(\widehat{b}) < p_{\mathcal{E}}^{\mathrm{I}}$.

Last, by a similar argument, $[p_{\mathcal{B}}(b) - \alpha_b] x_b^*(b) - \frac{\beta_b}{2} [x_b^*(b)]^2 < \eta$ for all $b > \widehat{b}$. Hence for all $b > \widehat{b}$, $x_b^*(b) = 0$ and, $p_{\mathcal{E}}(b) = p_{\mathcal{E}}^{\mathrm{I}}$.

E. Proof of Lemma 4

To prove part (i), consider a free–entry equilibrium where broker b is ensured a fraction $\zeta \equiv \frac{1}{B}$ of the exchange's profit and sets

$$p_{\mathcal{E}} = \arg\max \left\{ \zeta \Pi_{\mathcal{E}} + \max \left[p_{\mathcal{B}} x_b^* - \alpha_b x_b^* - \frac{\beta_b}{2} (x_b^*)^2 - \eta, 0 \right] \right\}.$$

Then entry will occur until

$$[p_{\mathcal{B}}(b) - \alpha_{B(b)}] x_{B(b)}^*(b) - \frac{\beta_{B(b)}}{2} [x_{B(b)}^*(b)]^2 = \eta.$$

Hence, if $\overline{B} > B(b)$, then brokers in $(B(b), \overline{B}]$ remain inactive and the equilibrium in brokerage is the same as in a free-entry equilibrium.

To prove part (ii), note first that for $\overline{B} = B(b)$ the same argument as before applies. So consider $\overline{B} < B(b)$

Part (iii) follows just from a higher exchange fee discouraging entry into brokerage. Since $p_{\mathcal{B}}(b)$ is increasing in b, $B(b)$ is decreasing in b for $b \leq \widehat{b}$ and equal to $p_{\mathcal{E}}^{\mathrm{I}}$ for $b > \widehat{b}$. Similarly, part (iv) follows from $p_{\mathcal{E}}(b) = p_{\mathcal{E}}^{\mathrm{I}}$ for $b > \widehat{b}$. ■

Figure 1
Duopoly and competition
between exchanges

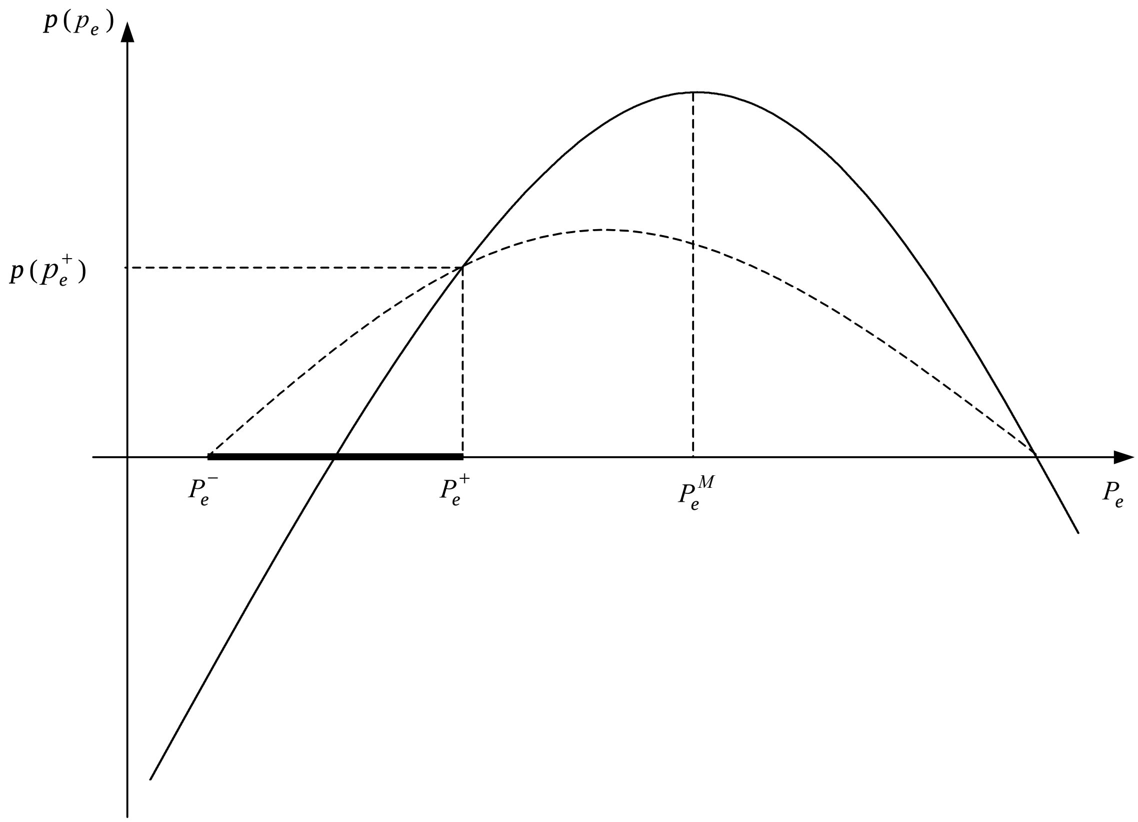

Figure 2
Liquidity and fragmentation as barriers to entry