The Political Economy of Stock Exchanges: Exchange Governance and Fee Structures

Zsuzsanna Fluck
Michigan State University
fluck@bus.msu.edu

Alex Stomper
University of Vienna
alexander.stomper@univie.ac.at

March 2003

Preliminar and Incomplete

Please do not quote without the authors' permission

Zsuzsanna Fluck may contacted by phone at 1 517 353 3019 and by email at fluck@bus.msu.edu. Alex Stomper may be contacted by phone at +43 1 4277 38075 or by fax at +43 1 4277 38074. We thank Colin Mayer, Marco Pagano, Duane Seppi and Josef Zechner for helpful discussions. This project started while Alex Stomper was a visiting scholar of the University of Salerno.

1 Introduction

Undoubtedly, stock exchanges provide many economic benefits by allocating capital to the highest value user and by offering ample risk-sharing opportunities. While investors can raise funds and share risks without using an exchange, stock exchanges, by lowering the cost of these services, can substantially improve on the efficiency of capital allocation and on investors' ability to share risks.

The benefits of stock exchanges have been extensively investigated and are reasonably well understood. Yet the question of how much these benefits are worth for exchange participants and how much various parties are willing to pay for the benefits they use is still largely open. On the one hand, exchange participants must be willing to pay as much as they gain from the exchange. On the other hand, implementing a fee structure for the exchange is similar to the problem of the provision of a public good in which participants with conflicting interests are inclined to lobby for prices that benefit them more than the other participants. To avoid the under-provision of the public good, which in our setting is access to the services of an exchange, it is necessary to price discriminate. That is, exchange participants with higher marginal utility will have to be charged proportionally more for the services they use. By charging a separate fee for each service the exchange provides, an optimal fee structure that accomplishes the optimal level of price discrimination can be implemented. The determination of an optimal fee structure of an exchange is in many ways similar to the derivation of the Ramsey prices in the public goods provision problem in social welfare analysis.

To evaluate how much the services of an exchange are worth for the exchange participants we need to assess the value added by the exchange relative to a stand-alone search market where investors or their intermediaries settle trades through direct bargaining with counterparties. In contrast to the search market where liquidity is created in response to an order, the downstairs market of a stock exchange represents a pool of "standby liquidity" created by potential counterparties who are waiting to trade. The downstairs market can therefore replace the "permanent, predictable component" of the search market, but for the "unpredictable, transitory" component that requires more than the standby liquidity the upstairs market can provide a trading venue for. In our model the two markets will naturally coexist and this is how the exchange creates value.[1]

We model the stock exchange as the infrastructure of trading characterized by its market microstructure, its governance structure, and its fee structure. The market microstructure represents the setting in which market participants trade, i.e. the technology, the rules of trading and order execution. Our model of market microstructure for the downstairs market is based on Seppi (1997) and on Keim and Madhavan (1996) for the upstairs market.

We view the governance structure of an exchange as the mechanism for resolving collective action problems and for reconciling conflicts of interests between different interest groups of the exchange.[2] Such conflicts arise between intermediaries on the upstairs and the downstairs market and among active traders and long-term shareholders of listed companies. Some of these participants, like the intermediaries of the search market have direct representation as exchange members, others have indirect representations like shareholders of listed companies who are represented by the exchange via the issuers.

Naturally, the governance structure of the exchange reflects how much the exchange cares about each of its participant and consequently, it determines the objective function

[1] Booth et al, Madhavan et al, Mendelsohn et al. show evidence that the combination of an upstairs and downstairs market represents a Pareto improvement over a downstairs market alone. Our theory is consistent with this evidence while our focus is how the combination of downstairs and upstairs market adds value relative to a stand alone search market.

[2] See Becht, Bolton and Roell (2002) for discussion of this interpretation.

for the exchange. We represent the influence of the interest groups via a social welfare function which is a weighted average of the expected utility that different groups of users derive from the existence of the exchange. In this function there are weights attached to each layer of the nested governance problem: there are weights representing the influence of exchange members relative to issuers, and the influence of the listed companies' shareholders with different trading needs. The first weight can be viewed as a reduced form for the governance structure of the exchange, the second for the governance structure of the issuers.

The exchange sets its fee structure by maximizing its objective function, the welfare function of its participants subject to its budget constraint. Our model allows for the inclusion of the fees that make up for most of the fee revenues in practice: listing fees, trading fees, and market information fees. We will derive how the various fees affect the welfare of each market participant by changing the explicit and implicit costs of trading and the depth of the limit order book. We also show how the volume of trade on the downstairs and the upstairs market depend on the fee structure of the stock exchange.

The model characterizes the optimal fee structure of the exchange conditional on the governance structure. When the fee structure is set only in the interest of the listed companies as representatives of their shareholders, the trading fees and the market information fees are chosen to equalize across all fees the tradeoff between the marginal utility loss imposed on the active traders and the marginal revenue from the fee. This tradeoff will depend on the weight that listed companies attach to the interests of their stockholders with frequent trading needs. However, the more influence upstairs firms exert on the exchange the higher the relative marginal utility loss imposed on the active traders due to fees on downstairs trading and the smaller the relative marginal utility loss due to fees on upstairs trading and market information fees. Depending on whether upstairs facilitation requires wide dissemination of market data, then upstairs firms will lobby for relatively lower or higher market information fees.

Our theory predicts that the fee structure of stock exchange also depends on the number of issuers and on the ownership and the governance structure of listed firms. First, as the number of issuers will grow, the fee structure of the exchange will be less dependent on listing fees and more dependent on trading and market information fees. Second, our model implies that exchanges where the ownership structure of listed firms is more concentrated (dispersed) will be more (less) dependent on trading fees and less (more) dependent on listing fees.

Finally, our theory also highlights the relationship between restructurings of the governance of exchanges and fee structure reforms. Given the recent waves of demutualizations, this is a timely issue to address. Since demutualization typically increases the influence of issuers relative to exchange members, our model predicts that demutualizations of exchanges will induce a fee structure that relies more on trading and market information fees and less on listing fees. This implication is consistent with empirical evidence from The 2001 Cost and Revenue Survey of the World Federation of Exchanges. Furthermore, our theory also predicts that following demutualization the fee structures of stock exchanges will vary according to their ownership structures: exchange listed exchanges will rely somewhat less on trading and market information fees for their revenues than joint stock companies.

2 Stylized Facts

In this section, we will provide a brief survey of how exchange fees are set and how they differ across exchanges. On the first point, consider the following quote from

NYSE's Robert G. Britz, a member of the SEC's Federal Advisory Committee on Market Information:[3]

> "The NYSE does not so much establish prices as allocate costs. Constituent representatives establish NYSE prices by first determining the NYSE's funding needs, and then allocating the resulting costs [...] between listed companies and member broker-dealers [...] by establishing fees denominates as 'listing fees', 'transaction fees' and 'market data fees'."

According to this quote exchange fees are set as the result of a tradeoff between the interests of a number of constituencies, such as the listed companies and the exchange members. Depending on these constituencies' interests and influence, a different balance will be struck between using different fees to raise the fee revenue required to meet the exchange's financing needs.

There are also some stylized facts about reforms of stock exchange fee structures after restructuring of the governance of stock exchanges. Most of these facts are related to the trend towards demutualization of stock exchanges. As an example, consider the Toronto stock exchange. In its notice to participating organizations No. 2000-152, the Toronto Stock Exchange (TSE) states:

> The TSE is now a shareholder-owned, for-profit corporation [...]. As part of this reorganization [...] the formula for charging trading fees is being changed [to] 1/50 of 1% of the value of the trade, to a maximum of $ 80 per trade; [...] only the incoming tradable order will be charged; booked limit orders

[3]This quote is from a letter sent to the Federal Advisory Committee on Market Information by Robert G. Britz, as a Group Vice President of the NYSE. The entire letter is available under www.sec.gov/divisions/marketreg/marketinfo/nyse0201com.htm.

> are filled free of charge; [...] crosses will be treated in the same manner as they are currently."

This quote illustrates also some of the details of stock exchange fee structures, i.e. that trading fees are frequently ad-valorem fees, that exchanges seek to give fee discounts to liquidity providers, and that there may be separate rules for trading fees levied on crosses.

We next consider how stock exchange fee structures differ across exchanges. According to a recent survey of the World Federation of Exchanges, there seem to be quite significant differences between member-owned and demutualized exchanges. While most exchanges charge the same kinds of fees to cover most of their costs, it seems that demutualized exchanges rely more on trading fees and less on listing fees than other exchanges. Among the exchanges responding to the 2001 Cost&Revenue Survey of the World Federation of Exchanges, the member-owned exchanges raised 22.3% of their fee revenue through listing fees while such fees accounted for only about 10% of the fee revenue of demutualized exchanges. Trading fees seemed to contribute slightly more to the fee revenue of demutualized exchanges than to that of member-owned ones, accounting for about 40% and 35% of fee revenue, respectively.[4]

[4] Of course, these differences can also be due to differences across the exchanges in the level of IPO activity, trading volume, etc. Moreover, one cannot identify to which extent these differences are driven by specific outliers.

3 The Stock Exchange

3.1 An Overview

We model the stock exchange as the infrastructure for trading in stocks. Our definition of a stock exchange as infrastructure capital stems from Gramlich (1994) wherein infrastructure capital is described as "capital intensive natural monopolies [...] such as communication systems." In context of a stock exchange the infrastructure capital is the clearing- and settlement system of the exchange.

A stock exchange is characterized by its market microstructure, its governance structure, and its fee structure. The market microstructure represents the setting in which market participants trade, i.e. the technology, the rules of trading and order execution. Our model of market microstructure for the downstairs market is based on Seppi (1997) and on Keim and Madhavan (1996) for the upstairs market. In order to not overwhelm the reader with all the details of the model at once, we will develop the model gradually by adding more structure as the analysis demands it.

The second element of a stock exchange is its governance structure. Following Becht, Bolton and Roell (2002) we define the governance structure of an exchange as the mechanism for resolving collective action problems. Naturally, the governance structure reflects how much the exchange cares about each of its participant and consequently, it determines the objective function for the exchange. We will represent the reduced form model of the governance structure of the exchange by the exchange's objective function. Like a social welfare function, this function represents a weighted average of the expected utility that different groups of users derive from the existence of the exchange.

Finally, the set of fees that the exchange charges for its services will constitute the fee structure of the exchange. Our model will allow for the inclusion of the fees that make up for most of the fee revenues in practice, (i) listing fees, (ii) trading fees, and (iii) market information fees. These fees will be defined below.

In the next sections we will formalize the basics of the market microstructure and the benefits that the exchange provides for its users. Then we will describe the governance structure of the exchange and introduce the fee structure.

3.2 The Basics of the Market Microstructure

There are n stocks listed on the exchange during a given time period Θ. These stocks can be traded in two trading venues, the downstairs and the upstairs market. The downstairs market hosts liquidity providers who stand ready to trade *before* orders get posted to this market. On the upstairs market, there is no such "standby liquidity". Instead, this market relies on financial intermediaries to search for counterparties *after* the orders arrive. We refer to this search as "upstairs facilitation" and to the financial intermediaries as "upstairs firms".

We assume that traders incur a fixed cost when they post orders to the upstairs market. This cost can be interpreted in various ways, such as (1) the cost of delay in the execution of orders due to difficulties in finding counterparties, or (2) the cost of a signal sent by the upstairs traders to make credible that their trading is not informationally motivated.[5] As a consequence of this fixed cost the upstairs market will only attract orders above a critical size.

We denote by Y^d (Y^u) the average volume of downstairs (upstairs) trading in one stock. Furthermore, we denote by $Y^{u \to d}$ the increase in the aggregate volume of downstairs trading if the upstairs market were closed. Notice that $Y^{u \to d}$ can be strictly less

[5] Keim and Madhavan (1996) find evidence that upstairs facilitation may take as long as four weeks of searching for counterparties.

than Y^u if the price impact of a trade on the upstairs market is smaller than the price impact of a trade on the downstairs market (See Fong, Madhavan and Swan (2003) for empirical evidence).

3.3 The Exchange Participants

There are three groups of users of the exchange who may supply or demand liquidity by trading in each of the n listed stocks. These are the active traders, the exchange members, and the value traders. "Active traders" demand liquidity by sending orders to either the downstairs or the upstairs market. Liquidity is supplied by exchange members who stand ready to trade on the downstairs market and by "value traders" who post limit orders. The group of exchange members include the upstairs firms that facilitate the execution of active traders' orders on the upstairs market.

Next we specify how different groups of users benefit from the existence of the exchange. We present a formal model in Appendix A to derive this specification from first principles.

The most important benefit that upstairs firms derive from the exchange is the commission they charge for their services. We model this commission as a fraction ϕ of the expected utility gain realized by active traders who demand upstairs facilitation. The parameter ϕ can be viewed as a measure for the market- and bargaining power of the upstairs firms, with $0 \leq \phi \leq 1$. We denote by μ the aggregate expected utility gain that active traders derive from being able to trade upstairs rather than downstairs, normalized by the number of listed stocks, n. Using this notation, upstairs firms earn an expected payoff of $V = \phi\mu$ per listing.

The risk-averse active traders who demand liquidity, and the risk-averse value traders who supply liquidity by posting limit orders to the downstairs market trade from off the exchange. To post orders to the downstairs market, these traders rely on brokers who offer their services in a perfectly competitive market for brokerage.[6] We assume that there are sufficiently many value traders that the competition between them eliminates any opportunities to gain strictly from posting limit orders. Hence, these traders derive zero expected utility from the existence of the exchange. By contrast, the active traders gain strictly from the existence of the exchange since we assume that the exchange offers them risk-sharing opportunities. During period Θ, each of the active traders may receive random endowments in one or more of the stocks listed on the exchange. To share risk, part or all of these endowments can be sold by posting orders to either the downstairs market or the upstairs market, as modelled in Appendix A. Such risk sharing generates expected utility gains of a magnitude that varies across the active traders corresponding to the size and nature of their stock endowments and the resulting exposure to dividend income risk. We denote by U the aggregate expected utility gains that active traders derive from being able to trade either downstairs or upstairs, normalized by the number of listed stocks, n:

$$U = \nu + (1 - \phi)\mu, \tag{1}$$

where ν and $(1 - \phi)\mu$ denote the aggregate utility gains per listed stock that active traders derive from being able to trade downstairs and upstairs, respectively, net of any explicit trading costs such as trading fees and commissions.

In Appendix A, we derive expressions for U and V in a market microstructure model that is based on the models by Seppi (1997) and on Keim and Madhavan (1996).[7] While we abstain from presenting the full model in the body of the paper, it is consistent with

[6] In a later version of the paper we extend the model so that the downstairs brokers can earn commission for their services.

[7] See result (50) in Section A.3.

all of the above and will repeatedly be referred to below.

3.4 The Governance Structure of the Stock Exchange

In analogy to Becht, Bolton and Roell (2002), we define the governance structure of a stock exchange as the institutional framework for (1) resolving collective action problems among dispersed users of the exchange and (2) reconciling conflicts of interests between different groups of users. The first task arise since the stock exchange serves a large number of different users including many small investors who hold or trade in stocks listed on the exchange. In contrast to the exchange members, these stockholders are not directly involved in exchange governance but they are represented by the management of listed companies.

The problem of reconciling conflicts of interests between different groups of users is the second task of exchange governance. In our model, we will focus on two types of conflicts of interests, (i) those between users of the downstairs and the upstairs market and (ii) those between stockholders with different trading needs. To analyse the latter we assume that the stockholders fall into two groups: active traders (defined above) and long-term (non-trading) stockholders. While the stock exchange benefit the former more than the latter, all of the stockholders share equally in the cost, K of maintaining the listing on the exchange. This cost includes listing fees as well as costs of complying with disclosure requirements, etc.

The governance structure of the stock exchange determines which objective function is being maximized in solving policy problems, such as the determination of optimal the fee structure. We will represent the reduced form model of the governance structure by the objective function of the exchange. This function captures how much the exchange cares about (i) the interests of users of the upstairs and the downstairs market, and (ii) the interests of stockholders with and without trading needs. Formally,

$$- = nW, \text{ for } W = (1-\omega)\,(\tau U + (1-\tau)(-K)) + \omega\, V, \text{ and } \omega, \tau \in (0,1). \tag{2}$$

We refer to the function W as the "social welfare function" of the stock exchange. Like a Bergson-Samuelson social welfare function, W represents a weighted average of the utility that different groups of users derive from the existence of the exchange in each listed stock. The weight ω measures the relative influence of the upstairs firms among the exchange members and the listed companies (as representatives of their stockholders) in the resolution of conflicts of interests between them. Hence, this weight is attached to the expected profit of the upstairs firms, V. The weight τ reflects the relative importance of the active and the nontrading shareholders in the listed companies. The smaller τ, the less the companies care about their stockholders' trading needs and the more concerned they are about the cost of maintaining their listings. As a utility loss incurred by the stockholders, this cost enters the social welfare function with a negative sign.

While ω also depends on the legal form and the ownership structure of the exchange, τ in contrast depends on the ownership structures of the listed companies. For exmaple, τ should be small if these companies are owned by large blockholders who abstain from trading in order to retain control over the companies.

3.5 The Fee Structure of the Stock Exchange

Our model incorporates the fees that make up for most of the fee revenues in practice, (i) listing fees, (ii) trading fees, and (iii) market information fees. We assume that for each listed stock, the exchange can charge a listing fee of L for maintaining the listing during period Θ. This listing fee enters the social welfare function in (2) as part of the cost K.

The revenue raised through trading fees depends also on the average trading volume per listing. We allow for the downstairs and the upstairs market to differ in the extent to which trading fees are levied on transactions since upstairs trading may partly be off-exchange trading.[8] Hence we will use t and f to denote the trading fee on the downstairs and upstairs market, respectively.[9] By allowing for these two fees to be set separately, we can capture effects of rules that govern off-exchange trading of listed stocks, such as the NYSE's rule 390.

Finally, we model the market information fee to be proportional to the number of listed stocks: since trading in each stock generates some market data, the dissemination of such data yields a fee revenue of I per listing.

Formally, the fee revenue is given by,

$$R = nZ, \text{ for } Z = tY^d + fY^u + I + L, \tag{3}$$

The function Z is referred to as the "fee revenue function"; this function determines the expected fee revenue per listing.

4 The Optimal Stock Exchange Fee Structure

We define the optimal fee structure as the one that maximizes the welfare function of the exchange subject to raising sufficient revenue to cover its cost. Formally, an optimal

[8]See Hasbrouck, Sofianos and Sosebee (1993), Section 7.

[9]We introduce two separate trading fees for analytical convenience. It would be equivalent to use only one trading fee t while introducing a variable for that fraction t/f of transactions on the upstairs market that may be executed off the exchange.

fee structure is a vector $(t, f, I, L)^T$ that solves the problem:

$$\begin{aligned} t^*, f^*, I^*, L^* \in \quad & \arg\max \; W \\ & \text{s.t. } Z \geq \underline{R}/n. \end{aligned} \tag{4}$$

At the optimum, one cannot raise social welfare without reducing revenue. Therefore, the fee structure is optimal if no welfare improving reform of the fee structure is "feasible": any such reform must result in a decrease in revenue since the gradients of the social welfare function W and the fee revenue function Z point in opposite directions. Formally, the following condition must be satisfied at any optimum,

$$\nabla \mathbf{W} = -\lambda \nabla \mathbf{Z}, \tag{5}$$

where λ is a constant.

To characterize the optimal fee structure, we need to specify the structure of the gradients ∇Z and ∇W. We start with the gradient ∇Z, given by $\nabla Z = (Z_t \;\; Z_f \;\; Z_I \;\; Z_L)^T$, where the subscripts denote partial derivatives, $Z_t = \partial Z / \partial t$, etc.[10] These partial derivatives measure direct effects of infinitesimal changes in the fee structure, as well as any indirect effects induced by changes in the volume of trading in the downstairs and the upstairs market. Unlike the listing fee L which is sunk cost from the perspective of the exchange users, the trading fees t and f affect the volume of both downstairs and upstairs trading since the latter fees influence the traders choice of markets. The market information fee I affects the liquidity of both trading venues since this fee determines the scale of dissemination of market data. The higher the market information fee, the less liquid the downstairs market becomes. The upstairs market is similarly affected: the smaller the scale of data dissemination, the harder it is for the upstairs firms to find investors with access to sufficient information to be willing to act as counterparties to

[10] Throughout the paper, subscripts refer to derivatives.

transactions in the upstairs market. As a consequence, the cost of upstairs facilitation increases and upstairs trading volume declines.

Differentiating the fee revenue function in (3) yields the following expressions:

$$\begin{aligned} Z_t &= Y^d + tY_t^d + fY_t^u, \\ Z_f &= Y^u + tY_f^d + fY_f^u, \\ Z_I &= 1 \quad + tY_I^d + fY_I^u, \\ Z_L &= 1, \end{aligned} \tag{6}$$

where $Y_t^d = \partial Y^d/\partial t$, $Y_t^u = \partial Y^u/\partial t$, etc.

The effects of changes in the fee structure on the social welfare function are captured by the gradient $\nabla W = (W_t \;\; W_f \;\; W_I \;\; W_L)^T$, where the subscripts again denote partial derivatives, $W_t = \partial W/\partial t$, etc. Differentiating the welfare function in (2) yields the following results:

$$\begin{aligned} W_t &= (1-\omega)\tau U_t + \omega V_t, \\ W_f &= (1-\omega)\tau U_f + \omega V_f, \\ W_I &= (1-\omega)\tau U_I + \omega V_I, \\ W_L &= -(1-\omega)(1-\tau), \end{aligned} \tag{7}$$

where U_t, U_f, U_I and V_t, V_f, V_I denote partial derivatives of the functions U and V with respect to t, f and I, respectively.

Proposition 1: The optimal fee structure

The optimal fee structure is determined by the following system of equations:

$$\begin{aligned} \tfrac{1}{\lambda}W_t &= \tfrac{\tau}{1-\tau}U_t + \tfrac{\omega}{1-\omega}\tfrac{1}{1-\tau}V_t &&= -(Y^d + tY_t^d + fY_t^u) &&= -Z_t, \\ \tfrac{1}{\lambda}W_f &= \tfrac{\tau}{1-\tau}U_f + \tfrac{\omega}{1-\omega}\tfrac{1}{1-\tau}V_f &&= -(Y^u + tY_f^d + fY_f^u) &&= -Z_f, \\ \tfrac{1}{\lambda}W_I &= \tfrac{\tau}{1-\tau}U_I + \tfrac{\omega}{1-\omega}\tfrac{1}{1-\tau}V_I &&= -(1 \quad + tY_I^d + fY_I^u) &&= -Z_I, \\ \tfrac{1}{\lambda}W_L &= \quad U_L &&= -1 &&= -Z_L. \end{aligned} \tag{8}$$

Proof: in Appendix B.

In order to further analyze the optimal fee structure of the stock exchange we need to specify the partial derivatives of the functions U and V with respect to the fees and derive the explicit and implicit costs of trading. This is the subject of the next subsection.

4.1 Stock Exchange Fees and Trading Costs

The fee structure affects the users of the stock exchange both via the explicit costs of trading, and via implicit trading costs, defined as the price impact of trading. In the remainder of this section, we will analyze which fee structure is optimal when such effects are taken into account. Appendix A contains a model that provides a "microfoundation" for this analysis. This model characterizes a specific market microstructure for which the utility functions U and V depend on the trading fees t and f and the market information fee I as described below.

Explicit costs of trading: We start with the explicit trading costs and, more specifically, with the trading fees t and f for downstairs- and upstairs trading, respectively. We will show in Appendix A and the proof of Lemma 1 (stated below) that these trading fees affect the active traders' aggregate expected utility U in proportion to the volume of downstairs trading, Y^d, and that of upstairs trading, Y^u. Due to imperfect competition in the market for upstairs facilitation, these direct effects are partially offset by indirect effects by virtue of changes in the commissions charged by the exchange members which offer upstairs facilitation. To see this, recall that the upstairs firms extract a fraction ϕ of μ, the aggregate expected utility gain per listing that active traders realize since they can trade upstairs rather than downstairs. We show in Appendix A that this utility gain changes in t and f according to the following total derivative,

$$d\mu = (Y^{u\to d}dt - Y^u df). \tag{9}$$

To interpret this total derivative, recall that Y^u denotes the aggregate volume of upstairs trading per listed stock, while $Y^{u\to d}$ is the downstairs trading volume that would result from routing downstairs all of the orders that are executed upstairs. As such, $Y^{u\to d}$ represents the "outside option" of active traders in the upstairs market, that is the option to trade downstairs rather than upstairs. The higher the downstairs trading fee t, the less valuable this outside option, and hence the higher the aggregate expected utility gain per listing μ that active traders obtain from being able to avoid downstairs trading. By contrast, μ is the smaller, the higher the trading fee f charged for upstairs trading. Since μ determines the commissions that upstairs firms charge to active traders, this implies that any changes in t and f have two effects on the explicit trading costs: besides the direct effects (described above), there are indirect effects since the commission payments change in value by $\phi\, d\mu$.

Lemma 1: Explicit trading costs

The trading fees t and f affect the aggregate utility gain per listing, U and V, that active traders and upstairs firms derive from the existence of the exchange, respectively. These effects are given by the total derivatives d^eU and d^eV, respectively:

$$\begin{aligned} d^eU &= d\nu + (1-\phi)d\mu = -(Y^d + Y^{u\to d})dt + (1-\phi)(Y^{u\to d}dt - Y^u df), \\ d^eV &= \phi d\mu = \phi(Y^{u\to d}dt - Y^u df), \end{aligned} \tag{10}$$

Proof: See Appendix A.

Summing up d^eU and d^eV shows the overall effects of changes in the explicit costs of trading on the users of the exchange. The direct effects of changes in the trading fees t and f are proportional to the volume of downstairs trading and upstairs trading, respectively. While the trading fees affect also the commissions charged by the upstairs firms, these effects cancel since the commissions are paid from one group of users of the exchange to another:

$$d^eU + d^eV = -Y^d dt - Y^u df. \tag{11}$$

Implicit costs of trading: The trading fees and the market information fee also affect the depth of the limit order book and thereby the liquidity of the downstairs market. We model this relationship explicitly in our market microstructure model in Appendix A. We show there how the depth of the limit order book, B determines the price impact of orders routed to the downstairs market, and hence the implicit costs of trading. These effects are measured by the product of the derivative $U_B = \partial U/\partial B$ and the derivatives of the depth of the limit order book, $B_t = \partial B/\partial t$, $B_f = \partial B/\partial f$, and $B_I = \partial B/\partial I$.

The prices at which orders get executed in the upstairs market may also be affected by the market information fee I. As it has been discussed above, this fee determines the scale of dissemination of market data, and also the cost of upstairs facilitation as a search for counterparties to transactions in the upstairs market.[11] The higher this cost, the fewer counterparties are found, and hence the worse the prices at which orders get executed. Let $dp^u = p'(I)dI$ denote the average price change by virtue of an infinitesimal change in the market information fee I, where the average is taken across the orders routed upstairs. This price impact affects the aggregate expected utility gain per listing that active traders realize by trading upstairs rather than downstairs, μ. This effect is proportional to the typical size of upstairs (block) transactions, $\bar{b}^u$, measured in terms

[11] In Appendix A, we model the upstairs market in reduced form, inspired by Keim and Madhavan (1996). These authors interpret the search costs as "inversely related to the probability of locating willing counterparties." By analogy to Merton (1987), the scale of dissemination of market data can affect this probability if market participants are only willing to act as counterparties for transactions in securities they know.

of the average number of shares per block traded:

$$d\mu = -\bar{b}^u p'(I)dI. \tag{12}$$

Part of this effect is borne by the exchange members, in the form of a change in the value of the commissions they can charge for providing upstairs facilitation.

Lemma 2: Implicit trading costs

By changing the implicit costs of trading, the trading fees t, f *and the market information fee* I *affect the aggregate utility gain per listing,* U *and* V, *that active traders and upstairs firms derive from the existence of the exchange, respectively. These effects are given by the total derivatives* d^iU *and* d^iV, *respectively:*

$$\begin{aligned} d^iU = \; & d\nu + (1-\phi)d\mu = \; U_B(B_t dt + B_f df + B_I dI) \; -(1-\phi) \; \bar{b}^u p'(I)dI, \\ d^iV = \; & \phi d\mu = \; -\phi \; \bar{b}^u p'(I)dI. \end{aligned} \tag{13}$$

Proof: See Appendix A.

We can now summarize. Proposition 2 states the overall effect of changes in stock exchange fees on the users of the stock exchange.

Proposition 2: The effect of the fee structure

The stock exchange fees affect the aggregate utility gain per listing, U *and* V, *that active traders and upstairs firms derive from the existence of the exchange, respectively. A reform of the fee structure,* $(dt, df, dI, dL)^T$, *has two effects on* U *and* V,

$$\begin{aligned} dU = \; & d^eU + d^iU, \\ dV = \; & d^eV + d^iV, \end{aligned} \tag{14}$$

where d^eU *and* d^eV *capture effects of changes in the explicit costs of trading, given by expression (10), while* d^iU *and* d^iV *capture changes in the implicit costs of trading, given by expression (13).*

Proof: These results follow from those in Lemma 1 and Lemma 2.

4.2 The Optimal Stock Exchange Fee Structure

We will next characterize how the optimal fee structure is set. Like in many similar analyses in the area of public economics, we cannot derive closed-form solutions for the various fees. Instead, we will analyze what marginal expected utility loss is imposed on the active traders under the optimal fee structure. To solve for this marginal expected utility loss, we substitute for the partial derivatives of the function V, $V_t = dV/dt|_{df=dI=0}$, $V_f = dV/df|_{dt=dI=0}$, etc. Upon rearranging, this yields the following system of equations:

$$\begin{aligned} \tfrac{\tau}{1-\tau}U_t &= -Z_t & -\phi\tfrac{\omega}{1-\omega}\tfrac{1}{1-\tau}Y^{u\to d}, \\ \tfrac{\tau}{1-\tau}U_f &= -Z_f & +\phi\tfrac{\omega}{1-\omega}\tfrac{1}{1-\tau}Y^{u}, \\ \tfrac{\tau}{1-\tau}U_I &= -Z_I & +\phi\tfrac{\omega}{1-\omega}\tfrac{1}{1-\tau}\bar{b}^{u}p'(I), \\ U_L &= -Z_L. \end{aligned} \tag{15}$$

The left-hand side of each equation is (a multiple of) the marginal expected utility loss imposed on the active traders. On the right-hand side of each equation, there are two terms. The first term is the marginal fee revenue raised through one of the fees. The second term captures how the optimal fee structure depends on the upstairs firms' interests and influence. To see this, consider first the case in which the fee structure is set only in the interests of the listed companies as representatives of their stockholders, $w = 0$. In this case, each of the first three equations can be written in the form,

$$\left.\frac{U_x}{Z_x}\right|_{w=0} = -\frac{1-\tau}{\tau}, \text{ for } x \in \{t, f, I\}. \tag{16}$$

Hence, the trading fees and the market information fee are set to equalize across all fees the trade-off between the marginal utility loss imposed on the active traders and the marginal revenue from the fee. This trade-off depends on the weight τ that the listed companies give to the interests of stockholders with trading needs (active traders), in representing these stockholders to the exchange. The higher τ, the closer to zero the term on the right-hand side of the above equation, and the higher must be the marginal revenue through each of the fees t, f, and I, for the listed companies to tolerate that a certain marginal expected utility loss is imposed on the active traders.

Now, consider the case in which the upstairs firms exert influence as exchange members, $w > 0$. By inspection of the first three equations of system (15), the upstairs firms lobby for treating *differently* the three fees t, f, and I. While all of these equations can be written in the form,

$$\frac{U_x}{Z_x} = \left.\frac{U_x}{Z_x}\right|_{w=0} \left(1 + \phi \frac{\omega}{1-\omega} \frac{\tau}{(1-\tau)^2} \frac{\mu_x}{Z_x}\right), \tag{17}$$

the term in brackets deviates from one in a way that depends on the sign of $\mu_x = \partial\mu/\partial x$, and hence on the effect of a fee on the aggregate expected utility that active traders derive from being able to trade upstairs rather than downstairs:

$$\mu_x = \begin{cases} Y^{u\to d} & \text{for } x = t, \\ -Y^u & \text{for } x = f, \\ -\bar{b}^u p'(I) & \text{for } x = I. \end{cases}$$

Note that $\mu_x > 0$ for $x = t$ and $\mu_x < 0$ for $x = f$ or I. If upstairs firms exert influence as exchange members, different fees are therefore set with different levels of tolerance towards imposing an expected utility loss on the active traders. While there is more tolerance towards such a utility loss for the trading fee t, there is less tolerance for the fees f and I.

$$\begin{aligned} \tfrac{U_x}{Z_x} &< \left.\tfrac{U_x}{Z_x}\right|_{w=0} \quad \text{for } x = t, \\ \tfrac{U_x}{Z_x} &> \left.\tfrac{U_x}{Z_x}\right|_{w=0} \quad \text{for } x = f \text{ or } I. \end{aligned} \tag{18}$$

To see why this happens, recall that the upstairs market allows upstairs firms to earn commissions with a value of $\phi\mu$. As a measure for the attractiveness of upstairs trading, μ increases in (i) the extent to which such trading allows to avoid trading fees, and (ii) in the extent of price improvement relative to the downstairs market. Our model allows for both of these advantages of upstairs trading to depend on the fee structure. Regarding the first point, more trading fees can be avoided, the higher t is set relative to f. As exchange members, upstairs firms affect that the fee structure is set with more tolerance towards a reduction in active traders' expected utility due to t than due to f. Regarding the second point, the extent of price improvement in the upstairs market depends on the ease with which counterparties can be found. If this requires wide dissemination of market data, upstairs firms are wary of any increase in the market information fee I. Hence, this fee is set such that only a smaller expected utility loss is imposed on the active traders than it would be the case if the upstairs firms had no influence.

Proposition 3: Characterization of the optimal fee structure

Under the optimal fee structure, fees t, f and I are set to impose a relative marginal expected utility loss of U_x/Z_x on the active traders, where Z_x denotes the marginal revenue from each fee $x \in \{t, f, I\}$, and U_x/Z_x is given by equation (18).

2.1 Suppose that exchange members exert no influence in the setting of the fee structure, and hence receive zero weight in the social welfare function (2), $w = 0$. The higher the weight τ given to the active traders among the stockholders of listed companies, the smaller the relative marginal expected utility loss U_x/Z_x imposed on them due to fees $x \in \{t, f, I\}$.

2.2 Suppose that the upstairs firms exert influence as exchange members, measured by the weight w that they receive in the social welfare function (2). The higher w, (i) the higher the relative marginal utility loss U_t/Z_t imposed on the active traders due to fees on downstairs trading, and (ii) the smaller the relative marginal utility loss U_x/Z_x due to fees on upstairs trading and market information fees, $x \in \{f, I\}$.

Proof: The results in claims 2.1. and 2.2. follow from differentiating equation (18).

5 Fee Reforms

We will now analyze how the fee structure should be reformed in response to changes in the governance structure of the exchange. Since the governance structure is represented by the weights of the social welfare function (2), these weights change by virtue of restructuring of exchange governance: we denote the new weights by τ' and ω'.

In order to analyze how a restructuring of exchange governance affects the optimal fee structure, we will assume that the fee structure has been optimally set given the original weights ω and τ. Starting from this formerly optimal fee structure, we will analyze the direction of the marginal reform $\mathbf{d}\rho = (dt, df, dI, dL)$ that maximizes social welfare for the new weights ω' and τ'. The set of feasible reforms depends on the minimum fee revenue that must be raised per listed stock. Since the reform of the fee structure may aim at raising a different fee revenue than before, we denote by dz the change in what fee revenue must be raised per listed stock: dz may be negative due to an increase in the number of listed stocks, n, or positive due to de-listings and an increase in the total revenue $\underline{R}$ that must be raised.[12]

[12] By the constraint of problem (4), $dz = \underline{R}'/n' - \underline{R}/n$, where $\underline{R}' - \underline{R}$ and $n' - n$ capture any change in the required total fee revenue and in the number of listed stocks, respectively, between the time of the reform and the time at which the fee structure was set originally.

The optimal fee reform is defined as follows:[13]

$$\begin{aligned} \mathbf{d}\rho^* \in \arg\max \quad & \mathbf{d}\rho^T \nabla W', \text{ s.t.} \\ & \mathbf{d}\rho^T \nabla Z \geq dz, \\ & \mathbf{d}\rho^T \mathbf{d}\rho = \varepsilon, \end{aligned} \tag{19}$$

where $\nabla W'$ denotes the gradient of the social welfare function W with the weights ω' and τ' instead of ω and τ, respectively. If the fee structure was set optimally before the reform, then $\nabla W'$ is given by,[14]

$$\nabla W' = \begin{pmatrix} W'_t \\ W'_f \\ W'_I \\ W'_L \end{pmatrix} = \begin{pmatrix} -(1-\omega')\tau'\frac{1-\tau}{\tau} Z_t & +\phi\frac{\tau\omega'(1-\omega)-\tau'\omega(1-\omega')}{\tau(1-\omega)} Y^{u\to d} \\ -(1-\omega')\tau'\frac{1-\tau}{\tau} Z_f & -\phi\frac{\tau\omega'(1-\omega)-\tau'\omega(1-\omega')}{\tau(1-\omega)} Y^{u} \\ -(1-\omega')\tau'\frac{1-\tau}{\tau} Z_I & -\phi\frac{\tau\omega'(1-\omega)-\tau'\omega(1-\omega')}{\tau(1-\omega)} \bar{b}^{u} p'(I) \\ -(1-\omega')(1-\tau') & \end{pmatrix}. \tag{20}$$

For the case in which $\omega = \omega'$ and $\tau = \tau'$, the above-stated gradient equals $-\lambda \nabla Z$, as required by condition (5), for $\lambda = (1-\omega)(1-\tau)$.

Within the set of feasible reforms, the optimal reform will be a vector pointing in one of two possible directions, depending on whether or not the revenue constraint restricts the reform. If so, then the optimal reform will point in a direction orthogonal to the gradient ∇Z such that the revenue constraint continues to bind after the reform. If not, then the optimal reform will point in the direction of the gradient $\nabla W'$.

Proposition 4: Fee structure reforms

The optimal reform of the fee structure is given by:

$$\mathbf{d}\rho^* = \frac{1}{2\kappa}(\nabla W' + \Lambda \nabla Z), \textit{ for } \Lambda = \min\left[0, dz - \frac{(\nabla W')^T \nabla Z}{(\nabla Z)^T \nabla Z}\right], \tag{21}$$

[13] In public economics, such a problem is referred to as the problem of "policy reform": for the purpose of formal analysis, reforms are always interpreted as differential changes in a vector of policy instruments; thus, the second constraint. See for example, chapter 6 of Myles (1995).

[14] For each of the fees $x \in \{t, f, I\}$, the expression for W'_x follows from solving system (15) for U_x and substituting the results as well as the derivatives $V_x = dV/dx$ (defined in subsection 3.1) into the gradient (7), while replacing the weights ω and τ by w' and τ'.

where κ satisfies the equation $(\mathbf{d}\rho^*)^T \mathbf{d}\rho^* = \varepsilon$.

Proof: The Lagrangean for problem (19) is $L = (\mathbf{d}\rho^*)^T \nabla W' + \Lambda(\mathbf{d}\rho^*)^T \nabla Z - 2\kappa(\mathbf{d}\rho^*)^T \mathbf{d}\rho^*$. Differentiating yields the first order condition $\nabla W' + \Lambda \nabla Z - 2\kappa \mathbf{d}\rho^* = 0$, as well as the constraints of problem (19). Solving this system of equations yields the result stated above.

In the remainder of this section, we will focus on the case in which the optimal reform of the fee structure depends on the social welfare weights ω' and τ' that characterize the governance structure of the exchange after restructuring. By inspection of the expression for Λ stated above, this happens if dz is sufficiently small; in the other case, the reform of the fee structure is merely determined by the revenue constraint.

Next we will use comparative statics to analyze how the optimal reform of the fee structure depends on the weights ω' and τ'. To obtain these results, we normalize the reform of the trading fees t, f and the market information fee I by the change in the listing fee L. For any fee $x \in \{t, f, I\}$, the ratio dx/dL shows the optimal trade-off between the interests of exchange users and those of non-trading stockholders of listed companies: the latter can benefit from a reform of the fee structure only by virtue of a decrease in the listing fee, $dL < 0$, while they remain unaffected by any other change in the fee structure.

The results are stated as Proposition 4. This proposition has two parts, to be discussed below. First, we consider the effects of increasing the weight ω' that measures the influence that upstairs firms can exert as exchange members after the restructuring of exchange governance. The higher ω', the less the reform reduces the attractiveness of upstairs trading as measured by μ, the aggregate expected utility that active traders derive from being able to trade upstairs rather than downstairs. Hence, the fee on downstairs trading is reduced the less relative to any decrease in the listing fee, while the reform aims the more at reducing fees imposed on upstairs trading and the dissemination of market data.

We next analyze the effects of the weight τ' that is given to the active traders' interests after exchange governance has been restructured. Like the upstairs firms, these traders can benefit from a reform of the fee structure that is aimed at raising more of the fee revenue through the listing fee, and less through fees which affect the volume of trading. However, active traders are also affected by a reform of the fee structure in another way, i.e. via the commissions charged by the upstairs firms. To separate these effects, we first consider the case in which no such commissions are charged: $\phi = 0$. In this case, the higher τ', the higher the optimal reduction of the trading fees t, f and the market information fee I, relative to that in the listing fee, L. These results carry over also to the case in which $\phi > 0$, but only if certain sufficient conditions are met. The conditions capture effects of conflicts of interests that are due to the effects of the fee structure on the attractiveness of trading upstairs rather than downstairs. To see this, consider for example the fee on downstairs trading, t. If $\phi = 0$, we can show that this fee is optimally reduced the more, the higher τ'. If $\phi > 0$, upstairs firms use their influence as exchange members to lobby against such a fee reduction: as discussed above, t will then be reduced the less, the higher the weight ω' that measures upstairs firms' influence. The overall change in t thus depends on the relation of the weights τ' and ω'.

Proposition 5: Comparative statics of fee reforms

Relative to any reduction in the listing fee L, the fee t levied on downstairs trading decreases the less, the higher the weight ω' that measures the influence that upstairs firms can exert as exchange members after the restructuring of exchange governance; the opposite is true for the fees f and I levied on upstairs trading and the dissemination of market data, respectively:

$$\frac{\partial}{\partial \omega'} \frac{dt}{dL} < 0, \; \frac{\partial}{\partial \omega'} \frac{df}{dL} > 0, \; \frac{\partial}{\partial \omega'} \frac{dI}{dL} > 0. \tag{22}$$

If no commissions are paid for upstairs facilitation, $\phi = 0$, the fees t, f, and I decrease the more relative to any reduction in the listing fee L, the higher the weight τ' given to active traders after restructuring of exchange governance. For these results to hold also if commissions are paid for upstairs facilitation, $\phi = 0$, the following conditions must be satisfied:

$$\begin{aligned} \frac{\partial}{\partial \tau'} \frac{dx}{dL} > 0 \quad &\Leftrightarrow \frac{Z_x}{\mu_x} > -\phi \frac{\omega(1-\omega')-\tau\omega'(1-\omega)}{(1-\omega)(1-\omega')(1-\tau)} \quad \Leftarrow \frac{\omega}{\omega'} > \tau \frac{1-\omega}{1-\omega'}, \;\; \textit{for } x = t, \\ \frac{\partial}{\partial \tau'} \frac{dx}{dL} > 0 \quad &\Leftrightarrow \frac{Z_x}{\mu_x} < -\phi \frac{\omega(1-\omega')-\tau\omega'(1-\omega)}{(1-\omega)(1-\omega')(1-\tau)} \quad \Leftarrow \frac{\omega}{\omega'} < \tau \frac{1-\omega}{1-\omega'}, \;\; \textit{for } x \in \{f, I\}. \end{aligned} \tag{23}$$

Proof: in Appendix B.

6 Conclusion

To be added.

References

To be added.

A Appendix A: A Market Microstructure Model

In this appendix, we present a microfoundation for our analysis. We first present a market microstructure model based on the models by Seppi (1997) and Keim and Madhavan (1996). This model will specify how active traders can reduce their exposure to dividend income risk by trading with exchange members and value traders. For simplicity, we will assume that there is only one listed stock, $n = 1$, and only one active trader. Given the players' optimal strategies, we will derive their expected utility gains from being able to trade with each other. Then, we will analyze how these utility gains depend on the trading fees t and f, and the market information fee I.

A.1 The Downstairs Market

The basic framework for this analysis is the microstructure model of liquidity provision by Seppi (1997), i.e. a specialist/limit order market like the NYSE. We extend this model to the case of *elastic* demand for liquidity.

A.1.1 The Model

Players and endowments: There are four types of players, most of which have been introduced in Section 3.3. First, there is an "active trader" A who receives a random stock endowment $\tilde{e}$, distributed according to $F[e]$ with support $[0, \bar{e}]$. Extending Seppi's model, we allow this active trader to demand liquidity *elastically* by selling some of the stock endowment. Liquidity is supplied from off the exchange by "value traders" who can post limit orders, and on the exchange by exchange members, i.e. a single "specialist" S, as well as a "trading crowd" consisting of m other exchange members.

The Stock: The stock which these traders can trade in is characterized by an uncertain dividend per share $\tilde{v}$, a random variable with mean μ and variance σ^2.

Preferences and trading needs: The players differ in their capacity to bear risk. Consider first the exchange members, i.e. the specialist and the trading crowd. As in Seppi's model, all of these players are risk-neutral legal entities and the specialist is the only one of them who can trade without incurring opportunity costs. For the trading crowd, the opportunity cost equals γ, the value of business lost if trading capacity (personnel or capital) is set aside for own-account trading. We assume that each exchange member's trading capacity is large enough to absorb the entire sell-side order flow. We also allow the trading crowd to respond to profitable trading opportunities on short notice. If they do so, they will incur an opportunity cost of $\Gamma > \gamma$ for temporarily reassigning trading capacity.

Next, consider the active trader A and the value traders. Each of these traders seeks to maximize the mean-variance utility function,

$$E[U^i] = E[\tilde{w}^i] - \frac{r}{2} Var[\tilde{w}^i], \tag{24}$$

where $\tilde{w}^i$ denotes the terminal wealth of trader i. This terminal wealth is the sum of i's initial wealth, denoted by $\underline{w}^i$, any payoff from trading, and any dividend income. We follow Seppi (1997) by assuming that all of the value traders incur opportunity costs of o if they post limit orders. The terminal wealth of trader i in (24) is net of this opportunity cost. We assume that the opportunity cost o is small enough that the limit order book is not empty. For notational simplicity, we abstract from any opportunity cost for the active trader A.

We denote by b the block of shares sold by the active trader through a market order. B denotes the depth of the limit order book, defined as the total number of shares that value traders want to buy through limit orders at various prices.

Time line of the trading game: First, each member of the trading crowd decides whether to set aside capacity for own-account trading. This decision is unobservable. Second, the value traders may submit publicly observable limit orders. Third, the active trader's endowment is realized and this trader decides whether to post an order to sell shares and, if so, how many shares. Fourth, the specialist's response to any sale order by the active trader is to announce a price p^*. Fifth, each member of the trading crowd decides whether to post a market order to buy shares at p^*. Finally, the specialist clears the market subject to priority rules discussed below.

Priority rules: In our model like on the NYSE, the specialist S follows three priority rules. First, the rule of "price priority" stipulates that S can only buy from the active trader once all limit orders at prices $p > p^*$ are filled in full. Second, even at the price p^*, the rule of "public priority" means that S can only trade upon filling all executable buy orders from either value traders or the trading crowd. Third, if these buy orders exceed the available sell-signed volume at this price, then the rule of "time priority" dictates execution in the order of arrival.

Price formation: Let $\mathcal{P}$ denote the price grid, i.e. the set of prices from which the specialist can choose the market clearing price p^*. Next we introduce two critical prices in $\mathcal{P}$, $\bar{p}$ and $\underline{p}$. The first $\bar{p} \in \mathcal{P}$ defines the set of prices $\{p \in \mathcal{P} | p \geq \bar{p}\}$ at which the value traders cannot post profitable limit buy orders. The second critical price $\underline{p} \in \mathcal{P}$ defines the set of prices $\{p \in \mathcal{P} | p \geq \underline{p}\}$ at which the trading crowd never buys shares on short notice. At these prices the trading crowd cannot recover the cost Γ of freeing up their trading capacity. To simplify the analysis, without loss of generality, we assume that the price grid $\mathcal{P}$ is sufficiently coarse that there is no price between $\bar{p}$ and $\underline{p}$.

Stock exchange fees: We assume that market orders are subject to the trading fee t. This fee is an ad-valorem fee. Hence any such seller will receive a per-share payoff of $p(1-t)$ and any buyer will pay $p(1+t)$ per share where p is the market clearing price. We assume that there are no trading fees levied on limit orders. This is in line with common practice at many exchange, such as for example the Toronto Stock Exchange.[15]

Trading is also affected by the market information fee I. Since this fee determines how costly it is to obtain data required for trading, we allow for this fee to affect the value traders' opportunity costs of o. We thus write these costs as an increasing function of I, $o[I]$, for $o'[I] > 0$.

A.1.2 The Analysis

We analyze the game recursively starting with the last stage. We will focus on each trader's decision in turn.

The trading crowd's decision to trade: The trading crowd consists of exchange members with and without trading capacity set aside for own-account trading. The former can post market orders at zero cost. This is so because at the time of placing these orders, the cost of setting aside trading capacity is already sunk. Thus, these exchange members buy any number of shares at any price p^* smaller than the expected dividend μ. The remaining trading crowd buys shares only if the specialist sets a price $p^* < \underline{p}$.

[15] The Toronto Stock Exchange charges no fees on booked limit orders ("to reward providers of liquidity") but charges an ad-valorem fee of 55 bp for incoming tradable orders (max. CAD 80 per trade). (See the notices to participating organizations No. 2000-152 and No. 2000-235.)

The specialist's choice of a market clearing price: Suppose that the specialist S receives a market order to sell a block of b shares and that B shares are posted as limit buy orders at price $\underline{p}$. Suppose furthermore that with probability π each member of the trading crowd has set aside capacity in order to trade ahead of S.

In choosing a market clearing price the specialist S seeks to maximize his profit as a liquidity provider. Note that some prices will never be chosen by S. First S has no reason to announce any price $p < \underline{p}$ since the entire trading crowd would trade ahead of him. If S chose such a price, he would fail to buy any shares due to public priority. Second, there is no reason to announce any price $p > \bar{p}$, since S could outbid the value traders even at $\bar{p}$. Hence, we can restrict our analysis to the specialist choosing prices between $\underline{p}$ and $\bar{p}$.

Two cases need to be considered next: (i) $\bar{p} > \underline{p}$ and (ii) $\bar{p} \leq \underline{p}$. Both cases can occur because $\bar{p}$ and $\underline{p}$ depend on the opportunity cost of trading for two different groups of traders, i.e. the value traders and the trading crowd, respectively. We have not yet specified the relation between these opportunity costs. For the limit order book to be non-empty, it must be the case that $\bar{p} > \underline{p}$. Otherwise, S would always announce a price of at least $\bar{p}$ to avoid that the trading crowd trades ahead of him.[16] At these prices, the specialist would always outbid the limit order book and since the value traders would rationally anticipate this, they would not place any limit orders. Consequently, we can ignore the case when $\bar{p} \leq \underline{p}$. Thus, for remainder of the paper we will assume that $\bar{p} > \underline{p}$ holds.

We now derive the specialist's choice of the market clearing price. Suppose first that S sets the market clearing price at $p^* = \bar{p}$. At this price the specialist will receive shares unless the trading crowd trades ahead of him, i.e. the specialist will trade if

[16] For $\bar{p} \leq \underline{p}$, the specialist would set the price $\bar{p}$ if $\bar{p} = \underline{p}$ and $\underline{p}$ if $\bar{p} < \underline{p}$.

no trading capacity has been set aside by any of the m exchange members. This will happen with probability $(1-\pi)^m$. Thus, the expected profit of S at $p^* = \bar{p}$ will be equal to $(1-\pi)^m\, b(\mu - \bar{p})$.

Now suppose that S sets the market clearing price at $p^* = \underline{p}$. In this case priority rules require the specialist to fill all eligible limit orders before he can buy shares. This implies that the specialist will receive $b - B$ shares whenever the block b exhausts the limit order book and no shares otherwise. Thus, the expected profit of S at $p^* = \underline{p}$ will be equal to $(1-\pi)^m \max[0, b - B](\mu - \underline{p})$.

To see which price is chosen by the specialist S, we compare the expected profits at $\bar{p}$ and $\underline{p}$; $(1-\pi)^m\, b(\mu - \bar{p})$ and $(1-\pi)^m \max[0, b-B](\mu - \underline{p})$, respectively. The outcome will depend on b. If b is too small for the block sale to exhaust the limit order book, i.e. $b \leq B$, then S will earn zero profit at $p^* = \underline{p}$. In this case, the market will clear at $\bar{p}$. If the block is large enough to exhaust the limit order book so that $b > B$, then S will clear the market either at $\underline{p}$ or at $\bar{p}$ depending on the size of the block, b. Price $\underline{p}$ will be chosen if and only if

$$(1-\pi)^m (b - B)(\mu - \underline{p}) > (1-\pi)^m\, b(\mu - \bar{p}) \Leftrightarrow b > \underline{b}[B] = \xi B, \tag{25}$$

where $\xi = \frac{\mu - \underline{p}}{\bar{p} - \underline{p}}$. Otherwise, $\bar{p}$ will be chosen.

Now we turn our attention to the active trader's decision.

The active trader's market order: From the analysis above, we know that if the block size b exceeds a critical value, then the market clearing price drops from $\bar{p}$ to $\underline{p}$. Thus, the active trader's block sale *can* have a price impact: A marginal change in b can trigger a price drop. We first analyze the case in which there is no such price impact.

In this case the active trader faces the following maximization problem

$$\max_{b}\ bp^*(1-t)+(e-b)\mu-\frac{r}{2}(e-b)^2\sigma^2 \tag{26}$$

where t denotes the trading fee. Solving the first order condition yields

$$b[e,p^*,t]=e-\frac{\mu-p^*(1-t)}{r\sigma^2}. \tag{27}$$

Equation (27) implies that A sells shares only if her endowment exceeds $\underline{e}$,

$$\underline{e}=\frac{\mu-\bar{p}(1-t)}{r\sigma^2}. \tag{28}$$

Otherwise, A abstains from trading. Thus, no trade will occur with probability $F[\underline{e}]$.

Next we consider the case in which A's order has a price impact. If the block size b exceeds the critical value $\underline{b}[B]$, then the market clearing price drops from $\bar{p}$ to $\underline{p}$. Anticipating this, A will sometimes sell less than $b[e,\underline{p},t]$. Figure A shows how this effect works.[17] It plots the size of the block sold by A as a function of her endowment. For intermediate values of the endowment, $e^s[B]\le e<e^h[B]$, it is optimal for A to sell $\underline{b}[B]<b[e,\underline{p},t]$ shares in order to avoid a drop in the market clearing price.

The boundary values $e^s[B]$ and $e^h[B]$ are derived in Lemma A.1 below. The former is the highest value of A's endowment which induces a block sale of size $b[e,\bar{p},t]$ with no price impact. The latter is the endowment for which A's expected utility from selling $\underline{b}[B]$ shares at price $\bar{p}$ equals her expected utility from selling $b[e,\underline{p},t]$ shares at price $\underline{p}$. Lemma A.1 summarizes this result.

Lemma A.1: *For $e^s[B]\le e<e^h[B]$, the active trader A sells $\underline{b}[B]$ shares of stock,*

[17] Figure 2 includes also upstairs trading, based on results that will be derived in section. We will again refer to Figure 2 as a summary of the trading game.

where $e^s[B]$ *and* $e^h[B]$ *are implicitly defined by*

$$\underline{b}[B] = b[e^s, \bar{p}, t], \quad \text{and} \tag{29}$$

$$(\bar{p} - \underline{p})(1 - t)\underline{b}[B] = \frac{r}{2}\sigma^2(b[e^h, \underline{p}, t] - \underline{b}[B])^2. \tag{30}$$

Proof: Equation (29) formalizes that $e^s[B]$ is the highest value of A's endowment which induces a block sale of size $b[e, \bar{p}, t]$ with no price impact. Equation (30) follows from the fact that selling $\underline{b}[B]$ shares at price $\bar{p}$ raises A's expected utility by $r\sigma^2\underline{b}[B]b[e, \bar{p}, t] - r\sigma^2(\underline{b}[B])^2/2$ while selling $b[e, \underline{p}]$ shares at price $\underline{p}$ raises A's expected utility by $r\sigma^2(b[e, \underline{p}, t])^2/2$. The difference must equal zero for $e = e^h$, i.e.

$$\begin{aligned} 0 &= r\sigma^2\underline{b}[B]b[e^h, \bar{p}, t] - \tfrac{r}{2}\sigma^2(\underline{b}[B])^2 - \tfrac{r}{2}\sigma^2(b[e^h, \underline{p}, t])^2 \\ &= r\sigma^2(b[e^h, \bar{p}, t] - b[e^h, \underline{p}, t])\underline{b}[B] - \tfrac{r}{2}\sigma^2(b[e^h, \underline{p}, t] - \underline{b}[B])^2 \\ &= (\bar{p} - \underline{p})(1 - t)\underline{b}[B] - \tfrac{r}{2}\sigma^2(b[e^h, \underline{p}, t] - \underline{b}[B])^2. \end{aligned} \tag{31}$$

Equation (30) reveals the basic trade-off facing the active trader A. The term on the left-hand side captures the reduction in A's payoff due to the price impact $(\bar{p}-\underline{p})$ of selling more than $\underline{b}[B]$ shares. To avoid this price impact, A must sell less by $b[e, \underline{p}, t] - \underline{b}[B]$ shares. Retaining these shares increases A's exposure to dividend risk. The term on the right-hand side of equation (30) is the reduction in A's expected utility due to the additional exposure. The boundary value $e^h[B]$ is the endowment at which the loss from the price impact of selling a larger block equals the utility loss resulting from bearing additional risk.

The value traders' limit orders: Next we study the decision of each value trader i whether or not to post a limit order to buy l^i shares at price $\underline{p}$. Since posting a limit order is costly, this decision depends on the probability of execution, $\beta[B] = 1 - F[e^h[B]]$. Lemma A.2 below shows that all limit orders will be executed if S sets the market clearing price at $\underline{p}$.

Lemma A.2: *The smallest block sold at price $\underline{p}$, $b[e^h[B], \underline{p}, t]$, exhausts the limit order book for any value of B,*

$$b[e^h[B], \underline{p}, t] \geq B. \tag{32}$$

Proof: By the definition of $e^h[B]$, $b[e^h[B], \underline{p}, t] > \underline{b}[B]$ since the market clearing price equals $\underline{p}$ only if A sells more than $\underline{b}[B]$ shares. This implies that $b[e^h[B], \underline{p}, t] \geq B$ since $\underline{b}[B] = \xi B$ for $\xi > 1$.

Next we will compute the size of the limit order of value trader i. Anticipating the depth of the order book, B, trader i will post a limit buy order of $l^*[B]$ that maximizes his expected utility. Formally,

$$\max_l \beta[B](\mu - \underline{p})l - \frac{r}{2}\left(\beta[B]\ E[((\tilde{v} - \underline{p})l)^2] - (\beta[B])^2\ ((\mu - \underline{p})l)^2\right). \tag{33}$$

The value trader's expected utility is the difference between her expected payoff, $\beta[B](\mu - \underline{p})l$, and the risk-premium due to dividend risk and execution risk. The first order condition yields

$$l^*[B] = \frac{\mu - \underline{p}}{r(\sigma^2 + (1 - \beta[B])(\mu - \underline{p})^2)}. \tag{34}$$

Evaluating the objective function at the maximum shows that trader i's expected utility, net of the opportunity cost, $o[I]$, is

$$E[U^i] - o[I] = \underline{w}^i - o[I] + \beta[B](\mu - \underline{p})\frac{l^*[B]}{2}. \tag{35}$$

We can now analyze the value trader's decision whether to post a limit buy order. Substituting for $l^*[B]$ in (35) shows that such an order is profitable for trader i if the following condition holds

$$0 \leq -o[I] + \beta[B]\frac{(\mu - \underline{p})^2}{2r(\sigma^2 + (1 - \beta[B])(\mu - \underline{p})^2)}. \tag{36}$$

In equilibrium this condition must hold as an equality. Otherwise, value traders could make strictly positive expected profit by posting a limit order. This equation will define implicitly the depth of the limit order book, B^* in equilibrium: for $B = B^*$, limit orders are executed with probability $\beta^* = \beta[B^*]$,

$$\beta^* = \frac{2o[I]r}{1 + 2o[I]r} \left(1 + \frac{\sigma^2}{(\mu - \underline{p})^2} \right). \tag{37}$$

Substituting for $\beta[B]$ in expression (34) yields the equilibrium size of value trader i's order.

The trading crowd's decision to set aside trading capacity: With trading capacity set aside, a member of the trading crowd can absorb any sell-signed order flow. We denote by $\tilde{s}$ the uncleared portion of the active trader's sell order after execution of all eligible limit buy orders. By absorbing this order flow an exchange member will realize the expected profit

$$\Pi = (1 - \beta^*)\ (\mu - \bar{p})E[\tilde{s}|p^* = \bar{p}] + \beta^*\ (\mu - \underline{p})(E[\tilde{s}|p^* = \underline{p}]. \tag{38}$$

Suppose that each member of the trading crowd sets aside capacity for own-account trading with probability π. If more than one member does that, then the order flow $\tilde{s}$ is split equally among them. Let $\varsigma[\pi, m]\tilde{s}$ denote the expected number of shares that each buys. Then their expected profit equals $\varsigma[\pi, m]\Pi$. Hence, the exchange members can break even on the opportunity cost of setting aside trading capacity if each of them does that with probability π^*, given by the condition

$$\varsigma[\pi^*, m]\Pi = \gamma. \tag{39}$$

If the number of exchange members m is high, the law of large numbers implies that $\phi[\pi, m] \approx 1/(\pi m)$. Then, $\pi^* \approx \Pi/(\gamma m)$ such that the order flow $\tilde{s}$ is (almost always) split between $\pi^* m = \Pi/\gamma > 1$ exchange members.

Proposition A.1 summarizes all results of this section.

Proposition A.1: Trading on the downstairs market

(i) Exchange members set aside capacity for own-account trading with probability π^*, given by condition (39).

(ii) Value traders submit limit orders at price $\underline{p}$ for $l^*[B^*]$ shares, where

$$l^*[B^*] = \frac{\mu - \underline{p}}{r(\sigma^2 + (1-\beta^*)(\mu - \underline{p})^2)}. \tag{40}$$

for β^* given by expression (37).

(iii) The active trader A sells $b^*[e]$ shares of stock

$$b^*[e] = \begin{cases} 0 & \text{if } e < \underline{e}, \\ b[e, \bar{p}, t] & \text{if } \underline{e} \leq e < e^s[B^*], \\ \underline{b}[B^*] & \text{if } e^s[B^*] \leq e < e^h[B^*], \\ b[e, \underline{p}, t] & \text{if } e^h[B^*] \leq e < \bar{e}. \end{cases} \tag{41}$$

(iv) The specialist S sets a market clearing price p^* given by,

$$p^* = \begin{cases} \bar{p} & \text{if } b^*[e] \leq \underline{b}[B^*], \\ \underline{p} & \text{if } b^*[e] > \underline{b}[B^*]. \end{cases} \tag{42}$$

A.2 The Upstairs Market

We next extend the model with the upstairs market. On this market the upstairs firms among the exchange members offer to search for risk-averse counterparties to facilitate the active trader's block sale. Our reduced form model of this costly search is inspired by Keim and Madhavan (1996). We assume that selling shares with upstairs facilitation yields a per-share payoff of p^u,[18]

$$p^u = \mu - \delta[I]\sqrt{r\sigma^2}, \tag{43}$$

where $\delta[I]$ denotes the marginal cost of searching for counterparties. Since such counterparties may require access to data in order to be willing to enter into transactions, we allow for this search cost to depend on the market information fee I.

We allow for imperfect competition between the upstairs firms, enabling them to charge commissions that may exceed the cost of searching for counterparties by a mark-up commensurate to the active trader's utility gain from upstairs facilitation. As it has been discussed in Section 3.3, this mark-up amounts to a constant fraction ϕ of this utility gain. Hence, imperfect competition has no effect on either the active trader's demand for upstairs facilitation or the size of her block sale. Instead, the fraction ϕ determines merely the exchange members' ability to extract rents.

The active trader's demand for upstairs facilitation depends on the size of her desired block sale. With upstairs facilitation, the active trader can avoid the price impact of selling a large block in the downstairs market. For the two markets to coexist, it must be the case that $p^u > \underline{p}$.

As it has been discussed in Section 3.1, we assume that the active trader incurs a fixed cost of C in selling shares with upstairs facilitation. We assume that this fixed cost is high enough so that the upstairs market does not crowd out the limit order book as a source of liquidity. In the presence of this fixed cost the active trader sells shares on the upstairs market only if his endowment exceeds the value of e^u, derived below, for $e^u > e^h[B^*]$.

[18]This payoff does not depend on the block size b since the number of counterparties increases in b. In the notation of Keim and Madhavan (1996), this is equivalent to assuming a constant marginal search cost δ. We allow this search cost to depend on the market information fee I.

Suppose that it is optimal for the active trader to sell shares on the upstairs market. In this case the active trader faces the following maximization problem

$$\max_b \; bp^u(1-f) + (e-b)\mu - \frac{r}{2}(e-b)^2\sigma^2, \tag{44}$$

where we have taken into account that an ad-valorem trading fee of f is levied on the active trader's transaction.[19] The optimal solution can be obtained by rearranging the first order condition

$$b[e, p^u, f] = e - \frac{\mu - p^u(1-f)}{r\sigma^2}. \tag{45}$$

In Lemma A.3 we derive when the active trader demands upstairs facilitation for her block sale.

Lemma A.3: *For $e \geq e^u$, the active trader sells $b[e, p^u, f]$ shares of stock with upstairs facilitation, where e^u is implicitly defined by*

$$\frac{r}{2}\sigma^2(b[e^U, p^U, f])^2 - C = \frac{r}{2}\sigma^2(b[e^u, \underline{p}, t])^2 - C. \tag{46}$$

Proof: Selling $b[e, p^u, f]$ shares with upstairs facilitation raises A's expected utility by $(b[e, p^u, f])^2\sigma^2 r/2 - C$. Selling $b[e, \underline{p}, t]$ shares at price $\underline{p}$ in the downstairs market raises A's expected utility by $(b[e, \underline{p}, t])^2\sigma^2 r/s$. Hence, upstairs facilitation increases A's expected utility by $(1-\delta)((b[e, p^u, f])^2\sigma^2 r/2 - C - (b[e, \underline{p}, t])^2\sigma^2 r/s)$. This utility gain equals zero for an endowment of e^u, defined by condition (46). Under the assumption that $p^u > \underline{p}$, the term on the left-hand side of this condition exceeds that on the right-hand side for $e \geq e^u$.

To derive the equilibrium for the upstairs and the downstairs market we need to redefine some of our earlier notation to incorporate the possibility of trading on the

[19] As discussed in Section 3.5, f may be equal to t times that fraction of the active trader's transaction that cannot be executed off the exchange.

upstairs market. Two pieces of notation are affected: $\tilde{s}$ and $\beta[B]$. Earlier, we defined $\tilde{s}$ as the uncleared portion of the active trader's sell order after execution of all eligible limit buy orders. From now on, $\tilde{s}$ will represent only the uncleared portion of a sell order routed to the downstairs market. Similarly, $\beta[B]$ was defined earlier as the probability of execution of limit orders, $\beta[B] = 1 - F[e^h[B]]$. From now on, $\beta[B]$ will stand for the probability of execution of limit orders on the downstairs market when the active trader can also sell shares upstairs. For this reason, $\beta[B]$ will be redefined as $\beta[B] = F[e^u] - F[e^h[B]]$.

Proposition A.2 summarizes the results from above.

Proposition A.2: Trading downstairs and upstairs

(i) Exchange members set aside capacity for own-account trading with probability π^*, *given by condition (39).*

(ii) Value traders submit limit orders at $\underline{p}$ *for* $l^*[B^*]$ *shares defined in expression (40).*

(iii) The active trader A sells $b^d[e]$ *shares of stock in the downstairs market and* $b^u[e]$ *shares with upstairs facilitation,*

$$b^d[e] = \begin{cases} 0 & \text{if } e < \underline{e}, \\ b[e, \bar{p}, t] & \text{if } \underline{e} \le e < e^s[B^*], \\ \underline{b}[B^*] & \text{if } e^s[B^*] \le e < e^h[B^*], \\ b[e, \underline{p}, t] & \text{if } e^h[B^*] \le e < e^u, \\ 0 & \text{if } e^u \le e, \end{cases} \tag{47}$$

$$b^u[e] = \begin{cases} 0 & \text{if } e < e^u, \\ b[e, p^u, f] & \text{if } e^u \le e \le \bar{e}. \end{cases} \tag{48}$$

(iv) The specialist S sets a market clearing price p^* *according to expression (42).*

A.3 The Players' Expected Utility

We can now derive the expressions for the players' expected utility gains from being able to trade. We consider first the active trader. Selling shares increases this trader's expected utility by,

$$U = \nu + (1-\phi)\mu, \tag{49}$$

for

$$\begin{aligned}
\nu \; &= (F[e^h[B^*]] - F[\underline{e}])\, \tfrac{r\sigma^2}{2} E[(b[e,\bar{p},t])^2 | \underline{e} \le e < e^h[B^*]] \\
&- (F[e^h[B^*]] - F[e^s[B^*]])\, \tfrac{r\sigma^2}{2} E[(b[e,\bar{p},t] - \underline{b}[B^*])^2 | e^s[B^*] \le e < e^h[B^*]] \\
&+ (1 - F[e^h[B^*]])\, \tfrac{r\sigma^2}{2} E[(b[e,\underline{p},t])^2 | e^h[B^*] \le e \le \bar{e}], \\
\mu \; &= (1 - F[e^u])(\tfrac{r\sigma^2}{2} E[(b[e,p^u,f])^2 - (b[e,\underline{p},t])^2 | e^u \le e \le \bar{e}] - C).
\end{aligned} \tag{50}$$

In the expression for ν, the second line captures the utility loss from unrealized risk-sharing opportunities if A restricts the size of her block sale to $\underline{b}[B^*]$ shares in order to avoid the price impact of selling more.

We next consider the value traders. Competition between these traders implies that they just break even on the opportunity cost of posting limit orders. Net of this opportunity cost, the expected utility of a value trader i, defined in expression (35), equals zero in equilibrium since condition (36) holds as equation.

Finally, we consider the exchange members.

The upstairs firms among the exchange members can profit from upstairs facilitation of the active trader's block sale. As discussed above, we allow for imperfect competition between the upstairs firms, enabling them to extract a fraction ϕ of the active trader's utility gain μ from being able to sell shares upstairs rather than downstairs. This amounts to a profit of

$$V = \phi\mu, \tag{51}$$

for mu given by the expression stated above.

Now consider exchange members' own-account trading in the downstairs market. We start with the trading crowd. By condition (39), all of these exchange members earn an expected profit equal to the opportunity cost γ. Thus, the downstairs market offers no strictly profitable trading opportunities for them. In approximation, the same is true for the specialist. With a high number of other exchange members m, the specialist can almost never trade on own account: Priority rules allow other exchange members to respond to any trading opportunities before the specialist.

B Appendix B: Proofs

Proof of Lemma 1:

To derive effects of the fees on players' expected utility via explicit trading costs, we hold constant the prices at which trades get executed downstairs and upstairs. For the downstairs market, this means that we need to hold constant the depth of limit order book, B^*, that determines $e^h[B^*]$, the critical value of the active trader's endowment that determines whether this trader's orders have a price impact.

Holding constant B^*, differentiating expressions (49) and (51) yields the following results:

$$
\begin{aligned}
U_t|_{B^*=\text{const.}} &= -(F[e^s[B^*]] - F[\underline{e}])E[\bar{p}b[e,\bar{p},t]|\underline{e} \le e < e^s[B^*]] \\
&\quad -(F[e^h[B^*]] - F[e^s[B^*]])\bar{p}\underline{b}[B^*] \\
&\quad -(1 - F[e^h[B^*]])E[\underline{p}b[e,\underline{p},t]|e^h[B^*] \le e \le \bar{e}] \\
&\quad +(1-\phi)(1 - F[e^u])E[\underline{p}b[e,\underline{p},t]|e^u \le e \le \bar{e}] \\
&= -(Y^d + \phi Y^{u\to d}), \\
U_f|_{B^*=\text{const.}} &= -(1-\phi)(1 - F[e^u])E[p^u b[e,p^u,f]|e^u \le e \le \bar{e}] \\
&= -(1-\phi)Y^u, \\
V_t|_{B^*=\text{const.}} &= \phi(1 - F[e^u])E[\underline{p}b[e,\underline{p},t]|e^u \le e \le \bar{e}] \\
&= \phi Y^{u\to d}, \\
V_f|_{B^*=\text{const.}} &= -\phi(1 - F[e^u])E[p^u b[e,p^u,f]|e^u \le e \le \bar{e}] \\
&= -\phi Y^u,
\end{aligned} \tag{52}
$$

for

$$
\begin{aligned}
Y^d &= (F[e^s[B^*]] - F[\underline{e}])E[\bar{p}b[e,\bar{p},t]|\underline{e} \le e < e^s[B^*]] \\
&\quad +(F[e^h[B^*]] - F[e^s[B^*]])\bar{p}\underline{b}[B^*] \\
&\quad +(F[e^u] - F[e^h[B^*]])E[\underline{p}b[e,\underline{p},t]|e^h[B^*] \le e < e^u], \\
Y^{u\to d} &= (1 - F[e^u])E[\underline{p}b[e,\underline{p},t]|e^u \le e \le \bar{e}], \\
Y^u &= (1 - F[e^u])E[p^u b[e,p^u,f]|e^u \le e \le \bar{e}].
\end{aligned} \tag{53}
$$

The results in (10) follow for

$$
\begin{aligned}
d^eU &= dt\, U_t|_{B^*=\text{const.}} + df\, U_f|_{B^*=\text{const.}}, \text{ and} \\
d^eV &= dt\, V_t|_{B^*=\text{const.}} + df\, V_f|_{B^*=\text{const.}}.
\end{aligned}
$$

□

Proof of Lemma 2:

To derive effects of the fees on players' expected utility via explicit trading costs, we consider only effects on the depth of the limit order book B^* and the price p^u. Consider the expression for d^iU stated in Lemma 2. The first term of this expression is the product of the derivative of expression (49) with respect to B^*, denoted by U_B, times the following total derivative:

$$dB^* = B_t dt + B_f df + B_I dI. \tag{54}$$

The second term is obtained by differentiating expression (49) with respect to p^u,

$$\frac{\partial U}{\partial p^u} = (1-\phi)\bar{b}^u(1-f), \text{ for } \bar{b}^u = (1-F[e^u])E[b[e,p^u,f]|e^u \leq e \leq \bar{e}], \tag{55}$$

and multiplying this expression by the derivative of p^u with respect to I, $\partial p^u/\partial I = -\delta'[I]\sqrt{r\sigma^2}$, in order to obtain the result:

$$\frac{\partial U}{\partial p^u}\frac{\partial p^u}{\partial I} = -(1-\phi)\bar{b}^u p'[I], \tag{56}$$

for $p'[I] = (1-f)\delta'[I]\sqrt{r\sigma^2}$.

The expression for d^iV follows by differentiating expression (51) with respect to p^u and multiplying the result with the above-stated derivative of p^u with respect to I. □

Proof of Proposition 5:

Expressions for dt/dL, df/dL and dI/dL follow from the expression for $\nabla W'$ stated above Proposition 4: $dt/dL = W'_t/W'_L$, $df/dL = W'_f/W'_L$ and $dI/dL = W'_I/W'_L$.. Differentiating these expressions yields the following results:

$$\begin{aligned}
\frac{\partial}{\partial\omega'}\frac{dt}{dL} &= -\frac{\phi Y^{u\to d}}{(1-\tau')(1-\omega')^2} < 0,\\
\frac{\partial}{\partial\omega'}\frac{df}{dL} &= \frac{\phi Y^{u}}{(1-\tau')(1-\omega')^2} > 0,\\
\frac{\partial}{\partial\omega'}\frac{dI}{dL} &= \frac{\phi \bar{b}^u p'[I]}{(1-\tau')(1-\omega')^2} > 0,\\
\frac{\partial}{\partial\tau'}\frac{dt}{dL} &= \frac{Z_t(1-\tau)(1-\omega)(1-\omega')+Y^{u\to d}\phi(\omega(1-\omega')-\tau\omega'(1-\omega)}{\tau(1-\tau')^2(1-\omega)(1-\omega')},\\
\frac{\partial}{\partial\tau'}\frac{df}{dL} &= \frac{Z_f(1-\tau)(1-\omega)(1-\omega')-Y^{u}\phi(\omega(1-\omega')-\tau\omega'(1-\omega)}{\tau(1-\tau')^2(1-\omega)(1-\omega')},\\
\frac{\partial}{\partial\tau'}\frac{dI}{dL} &= \frac{Z_I(1-\tau)(1-\omega)(1-\omega')-\bar{b}^u p'[I]\phi(\omega(1-\omega')-\tau\omega'(1-\omega)}{\tau(1-\tau')^2(1-\omega)(1-\omega')}.
\end{aligned} \tag{57}$$

Substituting for $Y^{u\to d} = \mu_t$, $Y^u = -\mu_f$ and $\bar{b}^u p'[I] = -\mu_I$ and rearranging yields the conditions stated in Proposition 5 for $\partial/\partial\tau' dx/dL > 0$. □