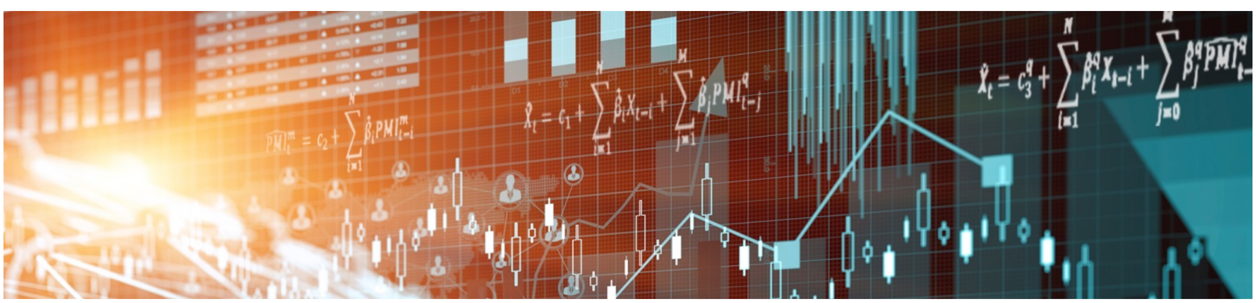
Staff Working Paper/Document de travail du personnel 2016-50

Broker Routing Decisions in Limit Order Markets



by David A. Cimon

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2016-50

November 2016

Broker Routing Decisions in Limit Order Markets

by

David A. Cimon

Financial Markets Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
dcimon@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I'd like to thank Andreas Park, Jordi Mondria, Michael Brolley, Lawrence Glosten, Charles Jones, Robert Battalio and Cheng Zheng for their helpful discussion. I would also like to thank seminar participants at the University of Toronto, Copenhagen Business School, HEC Montreal, Université Laval and the Bank of Canada, as well as conference participants at the 2016 NFA, the 2015 SFS Cavalcade, the 2015 ERIC Doctoral Consortium, the 2015 TADC and the 2015 CEA Annual Conference for their helpful feedback. All errors are my own.

Abstract

The primary focus of this paper is to study conflict of interest in the brokerage market. Brokers face a conflict of interest when the commissions they receive from investors differ from the costs imposed by different trading venues. I construct a model of limit order trading in which brokers serve as agents for investors who wish to access equity markets. I find that brokers preferentially route marketable orders to venues with lower liquidity demand fees, driving up the execution probability at these venues and lowering adverse selection costs. When fees for liquidity supply and demand are sufficiently different, brokers route liquidity supplying orders to separate venues, where investors suffer from lower execution probability and higher costs of adverse selection.

Bank topics: Financial markets; Market structure and pricing
JEL codes: G24; G28

Résumé

Cet article étudie principalement les conflits d’intérêts dans le marché du courtage. Les courtiers se trouvent devant un conflit d’intérêts lorsque les commissions qu’ils reçoivent des investisseurs diffèrent des coûts qu’exigent les différentes plateformes de négociation. L’auteur construit un modèle de négociation d’ordres à cours limité dans lequel les courtiers servent d’intermédiaires aux investisseurs qui souhaitent accéder aux marchés boursiers. Il constate que les courtiers préfèrent acheminer les ordres négociables vers des plateformes dont les frais liés à la demande de liquidité sont moindres, ce qui augmente la probabilité d’exécution sur ces plateformes et réduit les coûts d’antisélection. Lorsque les frais associés à l’offre et à la demande de liquidité sont suffisamment différents, les courtiers transmettent les ordres favorisant l’offre de liquidité à des plateformes distinctes; ainsi, les investisseurs voient diminuer la probabilité d’exécution de leurs ordres et subissent des coûts d’antisélection plus élevés.

Sujets : Marchés financiers; Structure de marché et fixation des prix
Codes JEL : G24; G28

Non-Technical Summary

The primary focus of this paper is to study the routing decisions that maximize brokers' profits. Many investors do not access equity markets directly. Instead, they delegate the decision of which venue to trade in to their brokers. Concerns have been raised in the regulatory and academic communities that brokers may route client orders to a venue that maximizes the brokers' own profitability, rather than the venue that best serves the client. One suggested cause of a conflict of interest between brokers and their clients is the use of rebates by competing venues to attract order flow. Many venues have adopted “maker-taker” pricing frameworks, where traders are given a rebate if they supply liquidity but face a higher fee when demanding liquidity. Other “inverse” or “taker-maker” venues offer rebates for liquidity-demanding orders, with a higher fee for orders supplying liquidity.

To study the effects of exchange fees and rebates on broker routing decisions, I construct a theoretical model of limit order book trading, in which investors leave the routing decision to their brokers and pay a flat commission upon execution. In equilibrium, brokers route marketable orders to exchanges with lower liquidity-demand fees, increasing the volume of uninformed orders at these venues and lowering the risk of adverse selection for limit orders posted there. I find that brokers will route limit orders to the same exchange as market orders only when liquidity-supply fees are very similar. In environments where exchanges have highly different fee structures, I find a trade-off for investors in terms of higher commissions for better market quality. I show that when broker commissions are higher, brokers will switch to routing based primarily on execution probability. In an extension to the model, I find that bid-ask spreads at one exchange can be affected by changes in the fees at other exchanges. In addition, I find that limit order investor welfare increases when liquidity-supply rebates are low, as market makers are no longer subsidized for providing liquidity.

This model has a number of implications for policy regarding trading venues and brokers. First, it implies that the proliferation of trading venues may not necessarily be beneficial for investor welfare, and in fact may be harmful in cases where certain venues face higher adverse selection costs or lower fill rates. Second, it is important to take factors other than price into account when defining concepts such as best execution. It is beneficial to investors if brokers consider factors such as fill rates when selecting venues for their clients. Finally, the change in the fee structure at one exchange can influence the spreads and market conditions at other exchanges, and changes to these fees do not occur in a vacuum.

Many investors do not access equity markets directly. Instead, they delegate the decision of which venue to trade in to their broker. In principle, the broker's and client's interests are aligned, as the broker earns a profit from the commission only if the client's order executes. However, brokers may route to a venue that maximizes their own profitability, rather than the venue that best serves their client. The primary focus of this paper is to study the routing decisions that maximize brokers' profits. Specifically, I investigate how these profit-driven routing decisions affect traders and market quality.

Concerns about order routing by brokers were raised as early as 2000 by the SEC,[1] and have been discussed more recently in 2014 by the United States Senate Committee on Homeland Security & Governmental Affairs,[2] and by the academic community (Battalio, Corwin, & Jennings, 2016). One suggested cause of a conflict of interest between brokers and their clients is the use of rebates by competing venues to attract order flow. To be successful, exchanges want to attract two types of orders: liquidity-supplying orders (limit orders) and liquidity-demanding orders (market orders).[3] Many venues have adopted "maker-taker" pricing frameworks, where traders are given a rebate if they supply liquidity, but face a higher fee when demanding liquidity.[4] Existing models of maker-taker pricing focus on investors who control their own order flow and adjust their prices to account for fees when selecting venues (Colliard & Foucault, 2012). However, when a broker controls the routing

[1]See SEC "Proposed Rule: Disclosure of Order Routing and Execution Practices" `http://www.sec.gov/rules/proposed/34-43084.htm`.

[2]For details regarding the hearings, entitled "Conflicts of Interest, Investor Loss of Confidence, and High Speed Trading in U.S. Stock Markets" see: `http://www.hsgac.senate.gov/subcommittees/investigations/hearings/conflicts-of-interest-investor-loss-of-confidence-and-high-speed-trading-in-us-stock-markets`.

[3]Orders supplying liquidity specify a price and remain available to future traders, while those demanding liquidity remove existing orders at the best price available.

[4]There also exist "inverse" or "taker-maker" exchanges, where traders demanding liquidity are provided a rebate, while those who supply pay a higher fee.

decision for the investor, the effects of exchange fees on investors and markets may be much different.

To study the effects of exchange fees and rebates on broker routing decisions, I construct a two-period model of limit order book trading, in which investors leave the routing decision to their brokers and pay a flat commission upon execution. Brokers have the choice of two possible venues for routing orders. The venues trade a single security at fixed price levels and are modelled following Foucault, Moinas, and Theissen (2007). There are two key differences from their paper. First, there are two venues, differentiated by trading fees for making and taking liquidity. Second, investors do not access the venues directly, but must do so through a broker.

In the first period, an uninformed investor arrives and tries to maximize her expected profit by choosing whether to submit a limit order to her broker. If her broker receives the order, he routes it to the venue that will maximize his expected profit. His profit is defined as the difference between the commission he charges the investor, and the trading fees charged by the exchange. In the second period, there are two potential outcomes. Either an innovation in the security value occurs and an informed trader arrives to remove outstanding limit orders, or a liquidity trader arrives. If a liquidity trader arrives, he submits a market order to his broker, who again routes it to maximize his expected profit.

In equilibrium, brokers route marketable orders in the second period to exchanges with lower taker fees, increasing the volume of uninformed orders at these venues and lowering the risk of adverse selection for limit orders posted there. I find that when the liquidity-making fees levied by the exchanges are very similar, brokers will route limit orders to the same exchange as market orders. Limit orders that are routed to the same exchange as market orders have a higher execution probability and lower adverse selection cost. Intuitively, this routing behaviour follows from the broker's profit maximization problem. When fees are similar, the broker benefits from the increased execution probability at the exchange with lesser-value rebates while incurring a low opportunity cost, as the rebate is only slightly higher at the alternative exchange. Conversely, when the liquidity-making fee structures are very different, the exchange with the higher rebate offers the broker sufficient profit upon filling the order to compensate for the lower execution probability. In this case, limit orders face lower execution probability and higher adverse selection costs.

Decision making by brokers affects their clients and imposes externalities on trading venues. Preferential routing of uninformed market orders to exchanges with lower taker fees lowers the adverse selection costs at these venues. Further, because of a lower concentration of informed trading, I find that the expected value of the trade, conditional on execution, improves for limit orders executed at these exchanges. In this case, the broker's decision on where to route uninformed market orders directly affects the market conditions for their limit order clients. Since informed traders are equally willing to trade at all exchanges, exchanges favoured by brokers for their uninformed market orders have improved fill rates.

In the case where fill rates, rather than rebates, drive brokers' limit order routing decisions, I find that a number of factors improve for investors. Specifically, more investors will choose to submit limit orders, each order will face a lower expected adverse selection cost, and order execution will occur with a higher probability. Intuitively, these factors follow from the improvement of market quality from an increase in the number of uninformed market orders. When brokers route limit orders based on fill rates, they route to the same exchanges as uninformed market orders. When market quality is improved, the expected value of a submitted limit order increases, making these attractive to a larger subset of possible investors.

In environments where exchanges have very different fee structures, I find a trade-off for investors in terms of higher commissions for better market quality. I show that when broker commissions are higher, brokers will switch to routing based primarily on fill rates. When commissions rise, the profit from rebates becomes relatively smaller and the broker's interests become increasingly aligned with those of his clients. In this model, the broker's commissions are set exogenously and his only costs come from the exchange fees. In practice, there is intense commission competition among brokers, especially following the proliferation of low-cost Internet brokerages.[5] Brokers also incur several costs when executing client orders, further reducing profit margins from the commission alone. In a competitive environment, brokers who maximize their profits based on exchange fees should be able to offer the lowest total commissions.

Finally, as an extension to the model, I allow for multiple price levels and endogenous market makers. I find that bid-ask spreads at one exchange can be affected by increases or decreases in spreads at other exchanges. In addition, I find that limit order investor welfare increases when maker rebates are low, as market makers are no longer subsidized for providing liquidity.

A. Regulatory Response

Significant regulatory attention has been paid to trading fees and routing of investor orders by brokers. Since 2001, the SEC has required brokers to make details available regarding their routing practice through Rule 606.[6] Further, the SEC requires brokers to

[5]Existing research on broker competition has argued that commissions from full-service brokerages have dropped dramatically from upwards of $75 USD per 100-share trade in the 1990s (Bakos et al., 2005). Currently, some online brokerages (such as Interactive Brokers) charge as little as $1 per 100 share trade.

[6]Originally SEC Rule 11Ac1-6.

disclose the routing details of their specific orders to investors upon request, as well as statistics related to execution quality.

In 2016, the SEC proposed a pilot on exchange access fees. This pilot would limit the fees for taking liquidity, effectively creating a ceiling for liquidity-making rebates.[7] On the other hand, this pilot would not limit the ability of exchanges to provide rebates for marketable orders.

Robert Battalio raised the concern regarding brokers' routing behaviour before the United States Senate Committee on Homeland Security & Governmental Affairs.[8] He suggested that they may be maximizing their intakes of rebates, paid to them by trading venues in exchange for order flow, rather than obtaining the best execution for their clients.

Regulators have also taken an interest in Canada, where the Ontario Securities Commission (OSC) published proposed regulatory changes that included a pilot study on prohibitions of maker-taker pricing structures, and disclosure of broker routing practices.[9]

B. Existing Literature

Existing work on maker-taker pricing can be divided into three groups. The first focuses on the incentives for brokers who select the trading venue, rather than investors who submit their orders directly. Empirical work by Boehmer, Jennings, and Wei (2007) focuses on changes to broker behaviour in response to the introduction of execution quality reports by SEC Rule 11Ac1-5 (now Rule 605). They find that competition for order flow among

[7]For details, see: `https://www.sec.gov/spotlight/emsac/emsac-regulation-nms-recommendation-61016.pdf`.

[8]For full details and transcripts from the hearings, see: `http://www.hsgac.senate.gov/subcommittees/investigations/hearings/conflicts-of-interest-investor-loss-of-confidence-and-high-speed-trading-in-us-stock-markets`.

[9]For details, see: `http://www.osc.gov.on.ca/en/NewsEvents_nr_20140515_csa-rfc-order-protection-rule.htm`.

broker-dealers drives orders to venues with high fill rates. Work by Bakos et al. (2005) finds that discount brokerages are more likely than traditional brokerages to route orders to third-party dealers as opposed to the primary exchange. Further, they find that low-cost brokers typically offer less price improvement to their clients than traditional brokers.

The work closest to this paper is that of Battalio et al. (2016), who empirically study the problem of the broker-client conflict of interest from trading fees. They find that brokers often make routing decisions based on the presence of liquidity rebates, rather than in the best interests of their clients. Further, they find that clients typically face higher adverse selection costs at exchanges with higher liquidity rebates. This paper provides theoretical confirmation of these two empirical results. First, I find that exchanges with higher liquidity rebates will endogenously have worse fill rates, and that a higher percentage of filled orders at the exchanges will be from informed traders. Second, I find that if trading fees are sufficiently different, brokers will route primarily based on rebates, rather than based on fill rates for their clients.

The second group focuses on the incentives for investors who submit orders directly to exchanges. Colliard and Foucault (2012) and Brolley and Malinova (2013) study maker-taker pricing regimes and their effects on investors from a theoretical perspective. Colliard and Foucault (2012) construct a model of a frictionless market, and study the breakdown of the total exchange fee between maker and taker fees. They find that only the total fee has an effect on investor outcomes, and that the breakdown between maker and taker fees has no effect on the cum fee bid-ask spread or the division of gains from trade. Brolley and Malinova (2013) construct an alternative model, where investors pay only a flat fee to a broker, who then passes the order on to a single exchange. They find that market orders sent to maker-taker exchanges are subsidized by investors who submit limit orders, when these investors pay a flat fee to their brokers. Empirically, several papers have found that exchanges with a maker-taker structure often have a better spread posted (Malinova & Park, 2015; Anand, McCormick, & Serban, 2013). However, exchanges with either a maker-taker structure or a higher taker fee have also been found to have a higher concentration of informed trading (Yim & Brzezinski, 2012; Anand et al., 2013).

The final group is a body of work that focuses on the incentives to exchanges. Theoretical work by Pagnotta and Philippon (2011) focuses the competition among exchanges based on speed. They find that this competition may be beneficial insofar as it increases trading speeds, but that with endogenous exchange entry welfare may be lowered. In relation to this paper, work on endogenous exchange trading fees by Chao, Yao, and Ye (2015) focuses on the profitability of exchanges when they are constrained by fixed tick sizes. They argue that, when tick sizes are fixed, the use of varied fee structures by exchanges improves the welfare of market participants.

In some cases, brokers may have the option of executing client orders against their own inventory rather than exposing them to the market. This allows brokers to bypass exchanges and their associated trading fees, at the cost of holding some inventory. Existing models, such as those by Battalio and Holden (2001) and Chakravarty and Sarkar (2002), describe internalization in varying contexts, including dealer markets. Since internalization requires some inventory to be held, existing work on inventory costs is also relevant.[10]

The remainder of the paper proceeds as follows. In Section 1, I describe the model. In Section 2, I describe the benchmark equilibrium results of this model with a single price level at the bid and the ask, exogenous market making and brokers who pass fees through to their clients. In Section 3, I present a model with a fixed commission. In Section 4, I present an

[10] See Stoll (1978), Amihud and Mendelson (1980) and Ho and Stoll (1981) for some of many examples.

extended model with multiple price levels and endogenous market makers. In Section 5, I conclude.

I. Model

The model borrows from the fixed-tick model in Foucault et al. (2007), and from trading fee models in Colliard and Foucault (2012) as well as Brolley and Malinova (2013).

A. Security

There is a single security that starts $t = 1$ with a value of v and ends $t = 2$ with a value of V. The initial value v is normalized to $v = 0$. With probability δ, an innovation in the security value occurs, raising or lowering the value of the security with equal probability by amount σ, while with probability $1 - \delta$, no change occurs.

The security trades on two exchanges at fixed-price ticks of size Δ. The prices on each exchange are identical, with possible prices at the ask given by $v + x\Delta$, and possible prices at the bid given by $v - x\Delta$. The variable x can only take the form of integers.[11] There is room at each tick on each exchange for a single order of quantity $Q_1 = 1$ (a buy-limit order) or $Q_1 = -1$ (a sell-limit order). Each exchange charges fees M_i and T_i to traders for orders making liquidity (limit orders) and taking liquidity (market orders), respectively. Both exchanges have the same total cost per order: $M_i + T_i = e\ \forall i$. I make three simplifying assumptions on the parameter set.

Assumption 1: The total cost per order, $M_i + T_i$, is set to zero ($e = 0$).

Assumption 2: Exchange 2 charges a higher taker fee ($T_1 < T_2$).

[11] Examples of feasible prices at the ask include $v + \Delta$ and $v + 2\Delta$, where $x = 1$ and $x = 2$, respectively.

Assumption 3: The prices and fees are such that, if an innovation occurs, the value of the security falls outside the grid of prices plus fees at the first tick ($v < v + \Delta + T_i < v + \sigma$ and $v < v + \Delta < v + \sigma$).

Assumption 1 implies that the exchanges have zero marginal cost of processing a trade and do not earn any excess profit from traders. Relaxing this assumption creates a spread between the maker and taker fees at both exchanges. Assumption 2 simplifies the solution for market orders, and implies that the broker will preferentially route market orders to exchange 1, easing interpretation of results. Assumption 3 ensures that that all orders at the closest tick are profitable for the informed trader and removes the case where taker fees are sufficiently high at one market that informed traders will not trade when the security value changes. Further, it ensures that all orders at the closest tick are subject to adverse selection if an innovation occurs.

B. Market Participants and Timing

The first period is the liquidity-supply period, in which agents post limit orders. The second period is the liquidity-demand period, in which agents submit market orders. The timing of the model is illustrated in Figure 1.

i. Limit Order Investors. A utility-maximizing limit order investor arrives at the market at the beginning of period $t = 1$. She is unable to interact directly with the market, and if she wishes to post an order, she must do so through a broker. Investors are risk-neutral, do not discount, and gain utility only from trading profits.

Analogously to Parlour (1998), a limit order investor arrives with a desire to buy or sell with equal probabilities in the form of a quantity signal $Q_1 \in \{-1, 1\}$. As in Parlour

(1998) and Foucault (1999), she also receives a private value $y \in Y$, where Y is a uniform distribution centred on zero.

She may submit a limit order for quantity Q_1, at a given price $x\Delta$, which will remain in the book until the end of $t = 2$. Orders to buy ($Q_1 = 1$) are at the prices $-x\Delta$, while orders to sell ($Q_1 = -1$) are at $+x\Delta$. She pays the broker a commission c upon successful execution of an order. If there is no order execution, the investor pays no commission.

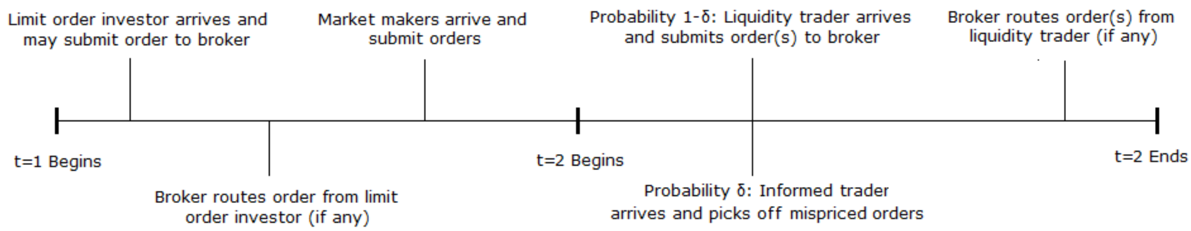
When submitting an order to a broker, she considers three factors: (i) to which exchange the broker will route her limit order; (ii) to which exchange the broker will send market orders; and (iii) at which price levels market makers will post. Combining these three, she determines $\theta_i(x\Delta)$, the probability that her order will execute at a given price $x\Delta$ if routed to exchange i and $E[V|Ex_i]$, the expected value of the asset, conditional on her order being executed at exchange i. A limit order investor who submits a limit buy order at price $-x\Delta$ has expected utility:

$$U_{LO} = \theta_i(-x\Delta)\left(E[V|Ex_i] + y + x\Delta - c\right). \tag{1}$$

ii. Brokers. Risk-neutral, uninformed, profit maximizing brokers exist to provide market access for investors. One broker routes limit orders during $t = 1$, while a second routes market orders during $t = 2$. The brokers receive orders from investors, and route them to one of the trading venues.

When routing orders, brokers are constrained by three rules: (i) brokers must accept and place all orders; (ii) brokers may route limit orders to any venue at the price specified by the order; and (iii) brokers must give market orders the best price available.

Figure 1
Model Timing



This figure illustrates the timing of this model. Investors and market makers provide liquidity in $t = 1$, while liquidity demand takes place in $t = 2$. With probability δ, the liquidity demander is an informed trader, while with probability $1 - \delta$, he is uninformed.

Brokers incur all costs from the exchanges to which they route orders (M_i for limit orders and T_i for market orders at exchange i). In turn, they profit from the difference between these costs and the commission that they charge clients per order executed, c. If no execution occurs, the brokers incur no costs and gain no benefits.

For orders routed to exchange i, the broker's profits for market order and limit orders are given by:

$$\pi_{MO} = [c - T_i] \tag{2}$$

$$\pi_{LO} = \theta_i(x\Delta)[c - M_i]. \tag{3}$$

Under Assumption 2, exchange 1 has a lower taker fee ($T_1 < T_2$) and brokers will preferentially route market orders to exchange 1 if a limit order is available. However, when prices are not equal, they must route market orders to the exchange with the best price.

iii. Market Makers. Uninformed market makers provide liquidity on each exchange. The market makers arrive immediately after the investor's order is routed in $t = 1$, and are able to place further orders. Market makers do not go through the broker and instead, pay fees directly to the exchange.

Market makers view the order, if any, placed by the limit order investor, and immediately have the option to place orders. They do so at any tick, which, given the expected behaviour of all other agents, gives a positive expected value. For buy-limit orders, this is any price $-x\Delta$ such that:

$$E[V|Ex_i] \geq -x\Delta + M_i. \tag{4}$$

iv. Informed Traders. If an innovation occurs, an informed trader arrives at the market and views the current innovation. This trader has direct access to the market, uses market orders, and pays the fees associated with taking liquidity (T_i). This trader presents an adverse selection risk to limit order traders who have already placed orders, similar to Glosten and Milgrom (1985), Easley and O'Hara (1987), Glosten (1994) and others.

If the innovation is positive, such that $V = v + \sigma$, the trader immediately submits market orders for the mispriced orders at both markets at price $v + \Delta$. If the innovation is negative, such that $V = v - \sigma$, all orders at price $v - \Delta$ are filled.

v. Liquidity Traders. If no innovation occurs, a liquidity-demanding investor arrives with a quantity signal, distributed evenly over $Q_2 \in \{-2, -1, 1, 2\}$. He immediately submits market orders for the total amount of his desired quantity. This order is routed by the broker, and the trades execute.[12]

[12]The inclusion of these traders removes the no-trade equilibrium detailed in Milgrom and Stokey (1982).

II. Benchmark Equilibrium

The benchmark equilibrium presented in this model contains three simplifying assumptions. First, I assume that brokers charge a commission of $c = 0$, pass all fees on to their limit order clients and route according to their clients' preferences. Second, I allow for only one tick at the ask $(v + \Delta)$, and one at the bid $(v - \Delta)$. Finally, I assume market makers post exogenously at all empty ticks, following the placement of the order from the limit order investor. One assumption that remains unchanged, is that brokers continue to route market orders. Market order traders are assumed to be random and they have no incentive to choose one exchange or another. Brokers continue to route market orders to exchange 1 prior to exchange 2, because of lower liquidity-taking fees. Assumption 4 is relaxed in Sections 3 and 4, while Assumptions 5 and 6 are relaxed in Section 4. Formally, these assumptions are:

Assumption 4: Brokers charge a commission $c = 0$, pass all liquidity-making fees incurred (M_i) to their limit order clients, and obey all routing instructions from their limit order clients.

Assumption 5: There is only one price available at the ask $(v + \Delta)$, and one at the bid $(v - \Delta)$.

Assumption 6: Following the routing of the limit order from the investor, market makers exogenously place orders at all empty ticks.

This equilibrium is analogous to existing models of limit order submission with exchange fees and fixed tick sizes. The brokers are effectively invisible in the limit order submission process, and costlessly carry out their clients' directions. This will serve as a benchmark for equilibria in which brokers are active in routing their clients' order flow.

In this model, an equilibrium is a solution to the limit order investor's utility maximization problem. This solution is a decision as to whether to submit an order and to which exchange to route this order for every quantity Q_1 and private signal y.

Theorem 1 (Existence of a Threshold Equilibrium):

(i) For fixed parameters M_1, M_2, δ, σ, there exists a unique threshold private value $\overline{y}_1$, such that for all $y \geq \overline{y}_1$ limit order investors with $Q_1 = 1$ will choose to submit a limit buy order and route it to exchange 1.

(ii) For fixed parameters M_1, M_2, δ, σ, there exists a unique threshold private value $\overline{y}_2$. If $\overline{y}_1 \leq \overline{y}_2$, all limit order investors prefer being routed to exchange 1. If $\overline{y}_1 > \overline{y}_2$ then $\overline{y}_2$ is such that all limit order investors with $\overline{y}_1 > y \geq \overline{y}_2$ and $Q_1 = 1$ will choose to submit a limit buy order and route it to exchange 2. Otherwise, if $y < \overline{y}_2$, limit order buyers will abstain.

A. Market Order Routing Decisions

Brokers will route market orders to maximize:

$$\pi_{MO} = c - T_i. \tag{5}$$

Under Assumption 2, brokers route to exchange 1 first, since $T_1 < T_2$ and prices are equal at both exchanges. As there is only a single order available at any given tick, brokers split market orders of size $|Q_2| = 2$ across both venues.

Market order routing has direct consequences for limit orders. All else equal, brokers will route market orders to the exchange with the lower liquidity-taking fee. This decision increases execution probabilities for limit orders at this exchange. Further, because informed orders are always large, both exchanges receive the same absolute quantity of informed orders, but different quantities of uninformed orders. As a result, the expected value for the security, conditional on execution, will be different across the two exchanges. Put differently, the price impact of market orders will differ across the different venues.

Proposition 1 (Market Order Routing and Limit Order Execution Probability):

The execution probability for limit orders at exchange 1, the low taker fee exchange, is always higher than at exchange 2 ($\theta_1 > \theta_2$).

Proposition 1 results directly from the preferential routing of market orders to the venue with the lower taker fee. Since only large orders will be sent to both markets, limit orders posted at the low taker fee market necessarily have a higher fill rate. The difference in quantity of market orders sent to each exchange provides the driving force behind the principal-agent problem between brokers and their clients in Section 3 onwards. When routing limit orders, brokers will always have the choice between an exchange with a higher execution probability for their clients (exchange 1) and an exchange with a lower fee for them (exchange 2).

Arguably, marketable orders from retail traders comprise only a small portion of total order flow. However, retail order flow is especially important in the context of adverse selection risk, as it generally contains less information content. Therefore, any exchange that receives a higher volume of retail orders compared with other orders likely has lower information costs for limit orders posted there.

One issue with Proposition 1 is the implication that the total volume of trading will be much higher at exchanges with an inverse fee structure. In practice, the majority of trading volume remains concentrated at maker-taker markets. Within the simplified model, this stems from the fact that prices are equal at both markets, while in reality this is not the case. One way to resolve this issue is through the introduction of endogenous market making and multiple price levels, as in Section 4. Specifically, if market makers are able to post more aggressively at maker-taker exchanges, volume will be higher.

Proposition 2 (Market Order Routing and Expected Security Value):

For limit buy orders, the expected value of the security conditional on execution, is higher if it is routed to exchange 1, than if it is routed to exchange 2 ($E[V|Ex_1] > E[V|Ex_2]$). The result is reversed for limit sell orders.

Proposition 2 results from the probability of informed trading being higher at the exchange with lower maker fees. Since informed traders are willing to remove mispriced orders from all exchanges, those with a lower number of retail orders face relatively higher adverse selection costs. This result supports the conclusions of empirical work from Anand et al. (2013) and Battalio et al. (2016), who find that exchanges with a maker-taker structure have a greater concentration of informed order flow.

B. Limit Order Investor's Problem

A limit order investor views her quantity signal Q_1, her private value y, and anticipates the broker's market order routing strategy. The limit order investor's decision includes the fees passed through by the broker, altering her utility function to the following:

$$U_{LO} = \theta_i \left(E\left[V|Ex_i\right] + y - (-\Delta) - M_i \right). \tag{6}$$

Proposition 3 (Market Preference with Fee Pass-Through):

When exchange fees are passed on by the brokers, if $M_1 - M_2 \leq \frac{1-\delta}{1+\delta}\delta\sigma$, then $\overline{y}_2 \geq \overline{y}_1$ and all limit order investors prefer being routed to exchange 1. Otherwise, there may be some investors who prefer to be routed to exchange 2.

Proposition 3 compares the expected value of the security at each exchange with the fees at each exchange. If the difference in the expected value of the security is larger than the difference in exchange fees, all investors will have the same preference to be routed to the same exchange.

Alternatively, if $M_1 - M_2 < \frac{1-\delta}{1+\delta}\delta\sigma$, the difference in fees dominates the difference in security value. In this case, depending on the distribution of the private value y, there may be a preference for some or all investors to be routed to exchange 2. These investors would prefer to pay a lower exchange fee and accept a lower execution probability and expected security value.

In equilibrium, investors with the lowest private valuations are the ones who would prefer to be routed to exchange 2. While the private value y can have several possible interpretations, the simplest interpretation of Proposition 3 is that if investors optimally split orders across exchanges, those with the lowest external value of the security accept worse fill rates and worse execution quality.

C. Fee Pass-Through and Broker Routing

It has been argued in the literature that passed-through fees can help alleviate the conflict of interest issues in maker-taker pricing regimes (Angel, Harris, & Spatt, 2011). This baseline model supports that finding with an important caveat. If fees are passed through to investors, but they do not control their own order routing, there is a conflict of interest present.

Assumption 4A: Brokers charge a fixed commission $c > 0$, and pass all liquidity-making fees incurred (M_i) to their limit order clients. Brokers route limit orders to maximize their own profitability.

Assumption 4A modifies Assumption 4 by removing control of limit order routing from investors. In this scenario, brokers receive a positive commission, while continuing to pass through exchange fees to their clients.

Corollary 1 (Pass-Through Fees with Broker Routing):

Under Assumption 4A, brokers optimally route all limit orders to exchange 1. Investors with a private value y such that $\overline{y}_2 \leq y < \overline{y}_1$ are worse off than if they had controlled their own order flow.

Corollary 1 follows from the exchange selection result in Proposition 3. In general, investors prefer exchange fees to be passed through, as opposed to the fixed commission models presented later in this paper. However, this preference only fully holds when they also control their order routing. If brokers pass through fees, but charge additional commission and control the routing decision (as often occurs in practice), there is a group of investors who suffer a loss. Under Assumption 4A, brokers route all limit orders to the venue with the highest execution probability to receive their commission. Since they no longer pay exchange fees, they also pass the higher maker fee at this venue through to the investors. Investors who would have preferred to be routed to exchange 2 under Proposition 3 are worse off.

III. Single Commission

In the benchmark model, brokers are assumed to pass fees through to their clients and route according to their clients' wishes. In this section, I model an arbitrary single commission $c > 0$, where brokers are unable to credibly commit to a routing scheme and must route according to an incentive compatibility constraint.[13] To study the incentive compatibility problem faced by brokers with fixed commissions, I assume brokers exogenously charge a commission c to their clients. This involves the relaxation of Assumptions 4 and 4A and the introduction of Assumption 7.

[13]I do not deal with the commission for market orders, as market order traders are random in this model. Commissions for market orders would then be arbitrary as the market order traders are not endowed with a utility function. Further, execution at all exchanges is equal for market order traders, as limit order investors and market makers are both uninformed.

Assumption 7: The broker commission for limit orders is set exogenously such that $c > |T_i|, |M_i|$.

In this model an equilibrium consists of: (i) a solution to the broker's profit maximization problem; and (ii) a solution to the limit order investor's utility maximization problem. The solution to the broker's problem is a decision of where to route limit orders from investors. The solution to the limit order investor's problem is a decision as to whether or not to submit an order, for every quantity Q_1, private signal y, and broker routing strategy.

Theorem 2 (Existence of a Threshold Equilibrium II):

(i) For fixed parameters M_2, δ, σ, c, there exists a unique threshold maker fee $\overline{M}_1$, such that if $M_1 \leq \overline{M}_1$, brokers will optimally route limit orders to exchange 1. Otherwise, brokers will route limit orders to exchange 2.

(ii) For fixed parameters $M_1, M_2, \delta, \sigma, c$, there exist unique threshold private values $\overline{y}_i$, such that for all $y \geq \overline{y}_i$ limit order investors with $Q_1 = 1$ will choose to submit a limit buy order, given their broker routes limit orders to exchange 1. Otherwise, if $y < \overline{y}_i$, limit order buyers will abstain.

A. Broker's Problem

As in the baseline model, brokers route market orders first to exchange 1, where the taker fee is lower, and subsequently to exchange 2. Unlike the initial model, brokers now choose the exchange for limit orders based on their own profit maximization decision. When routing limit orders, brokers choose exchange i in order to maximize:

$$\pi_{LO} = \theta_i[c - M_i]. \tag{7}$$

Exchange 1 receives more market orders, and therefore has a higher execution probability. For limit orders, brokers will weigh the trade-off between a higher execution probability at exchange 1 and a lower maker fee at exchange 2. This trade-off is at the heart of the broker-client conflict of interest. If one exchange has maker fees that are sufficiently low compared with others, brokers may route there even if the execution probability is low.

B. Limit Order Investor's Problem

A limit order investor views her quantity signal Q_1, her private value y, the commission c and anticipates the brokers' routing strategies. Using these, she decides whether or not to submit a limit order. Given that her order will be routed to exchange i, an investor wishing to buy does so if:

$$\theta_i \left(E\left[V|Ex_i\right] + y - (-\Delta) - c \right) > 0. \tag{8}$$

As described in Theorem 2, in equilibrium there exist values $\overline{y}_i$, such that any limit order buyer with a sufficiently large private value y will optimally choose to submit a limit order, while those with a private value below this cut-off will abstain. These equilibrium actions are illustrated in Figure 2.

Proposition 4 (Limit Order Submission Decision):

Investors require less favourable private valuations to submit an order if their order is routed to exchange 1, than if it is routed to exchange 2 ($\overline{y}_1 < \overline{y}_2$).

Proposition 4 describes the difference in limit order investor behaviour based on broker routing. Limit order traders will optimally choose to submit an order if their expected utility, given routing and execution, is positive. As both the probability of execution and expected value given execution are higher if routed to exchange 1, more traders are willing to submit if their order will be routed there. This is contrary to the fee-pass-through model where $\overline{y}_1 \lesseqgtr \overline{y}_2$, depending on exchange fees and adverse selection. With fixed commissions, all investors prefer to be routed to exchange 1 rather than exchange 2. An illustrated version of this equilibrium is shown in Figure 2.

The decrease in trader volume when routing for rebates also suggests a dilemma for brokers. If they were able to commit to routing to exchange 1, despite a lower profit per order, a larger number of clients would choose to submit orders. Depending on the parametrization, this situation can lead to a larger expected profit for brokers and the preferred routing scheme for their clients. The increase in the number of clients when orders are routed to exchanges with higher fill rates suggests an important role for routing disclosure by brokers, which may ultimately serve as a commitment mechanism to their clients.[14]

By combining Propositions 2 and 4, I am able to make a statement on limit order investor utility.

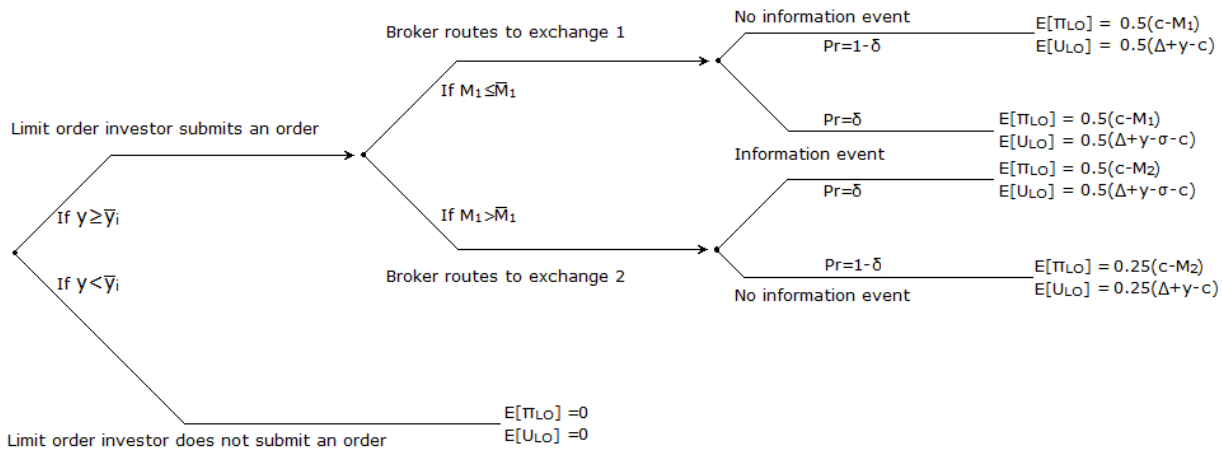
Proposition 5 (Limit Order Investor Utility):

All else equal, if the two exchanges are sufficiently similar ($M_1 \leq \overline{M}_1$), more limit orders would be submitted by investors and each investor would be better off in expectation than if the two exchanges are sufficiently different ($M_1 > \overline{M}_1$). Therefore, the expected utility of limit order investors is higher when exchanges have similar fee structures.

For simplicity, I define welfare such that total investor welfare is simply the sum of utility of each individual investor. However, not only does total welfare increase, but every individual investor also has higher expected utility. The increase in utility corresponds to a first-order dominance relationship, in which the distribution of investor utilities with two similar markets dominates the distribution of utilities under two different markets.

[14]An example is SEC Rule 606, which mandates the partial disclosure of routing information for non-directed orders by brokers. Information related to this is available online at: `http://www.ecfr.gov/cgi-bin/text-idx?SID=6d079725b329f4fc5ad9c80affb45d9f&node=se17.4.242_1606&rgn=div8`.

Figure 2
Model Equilibrium with Fixed Commissions



This figure illustrates the equilibrium actions and payoffs for limit order buyers and their brokers. The equilibrium is determined through backwards induction. Given the expected payoffs, the broker will choose to route limit orders to exchange i. Given the broker's routing decision, the limit order investor will submit an order if her private value is above $\overline{y}_i$. Actions and payoffs for limit order sellers are symmetric.

Lower limit order investor utility has additional implications for market quality, as limit order investors provide additional liquidity to the market. Brogaard, Hendershott, and Riordan (2014) find that in large stocks, non-high-frequency traders (HFTs) may supply as much as 58 percent of liquidity by volume in large-cap stocks and 89 percent of liquidity by volume in small-cap stocks. Fewer liquidity supplying orders from investors could substantially reduce liquidity in markets where HFT is not active.

C. Comparative Statics

An increase in the probability that an information event occurs (δ) reflects an increase in an uninformed trader's risk of being adversely selected. An increased probability of an information event can represent several possible scenarios, including a period of market turmoil, announcement dates, or simply securities that have higher inherent levels of risk.

Proposition 6 (Adverse Selection and Limit Order Investors):

(i) When adverse selection rises, limit order investors receive lower expected values for their trades ($\frac{\partial E[V|Ex_i]}{\partial \delta} < 0$ *for buyers) and fewer investors are willing to submit orders (*$\frac{\partial \overline{y}_i}{\partial \delta} > 0$ *for buyers).*

(ii) When adverse selection rises, exchange fees must be increasingly similar for brokers to route based on fill rates rather than rebates ($\frac{\partial \overline{M}_1}{\partial \delta} < 0$*).*

Proposition 6 demonstrates the two negative effects of adverse selection on limit order investors. First, as the probability of trading against an informed trader increases, the expected value of the trade is worse for limit order investors. The lower expected value lowers the number of limit order investors willing to submit orders to their brokers.

Second, during times when more information is reaching the markets, such as during periods of announcements, brokers are more likely to route based on rebates. The increase in the probability of informed trading causes the total probability of execution to converge across all exchanges and incentivizes brokers to concentrate their orders at exchanges with high maker rebates. Specifically, the equilibrium value $\overline{M}_1$ falls, meaning that the maximum fee for which brokers will route limit orders to exchange 1 is higher. If the actual maker fee at exchange 1 is now above this new value, the broker will alter his routing habits and divert limit orders to exchange 2. This exchange has an even higher adverse selection risk for investors, increasing the impact of the increase in informed trading. In other words, brokers' routing practices increase their clients' risk of being adversely selected during periods when adverse selection is highest.

Proposition 7 (Increase in Broker Commission):

There exists a threshold $\underline{c}$, such that for all $c \geq \underline{c}$, brokers route limit orders to exchange 1. If $c < \underline{c}$, brokers route limit orders to exchange 2.

Proposition 7 reflects one of the basic ideas in principal-agent problems. Given a sufficient incentive, the brokers' interests will be aligned with their clients' interests. In this model, this incentive takes the form of a sufficiently high commission. In conjunction with Proposition 2, this implies that, in some cases a higher commission may increase utility for both brokers and their clients. Though clients generally prefer paying a lower commission, when the commission is sufficiently low, they suffer from a conflict of interest. An increase in commission can cause the broker to begin routing to exchange 1, which lowers the adverse selection costs to investors. If Equation 9 is satisfied, the increase in commission is entirely offset by the decrease in adverse selection costs:

$$\underline{c} - c \leq [V|Ex_1] - E[V|Ex_2]. \tag{9}$$

There is a second interpretation to the broker commission, which stems from the costs incurred by the broker. In this model, the costs to brokers of processing an order are set to zero. If a broker is able to reduce his costs, either through more efficient internal behaviour or through lower external costs, he is effectively increasing the commission he receives per order, without an increase in the commission each client pays. In this case, a decrease in non-trading fee costs to the broker increases his profit from the commission upon execution, and may also cause him to optimally route to the exchange with a higher fill rate.

Incentive-compatible commissions are not simply a transfer from investors to brokers, but represent a loss in gains to trade. The difference between the incentive-compatible price $\underline{c}$ and the fee pass-through value M_1 causes a decrease in the proportion of investors willing to trade. Since the value $\underline{c}$ is characterized by both the difference in exchange fees and adverse selection values, so is the loss of clients. This value represents the minimum loss incurred by investors if commissions must be incentive compatible. Any other commission structure with $c > \underline{c}$ will create a larger deadweight loss to investors.

IV. Extension: Endogenous Market Making

In the benchmark model, the price grid at each side of the market was of size one. In this extension, the prices on each exchange remain identical. However, I allow for two ticks at the ask ($v+\Delta$, $v+2\Delta$), and two at the bid ($v-\Delta$, $v-2\Delta$) with room for a single order of quantity $Q_1 = 1$ (a buy-limit order) or $Q_1 = -1$ (a sell-limit order) at each tick on each exchange. As before, I normalize the initial value of the asset as $v = 0$. Further, I allow market makers to choose whether or not to post at each available price level. This involves the relaxing of Assumptions 5 and 6, and the introduction of Assumptions 8 and 9:

Assumption 8: There are two prices available at the ask (Δ, 2Δ), and two at the bid (Δ, 2Δ).

Assumption 9: The grid of prices is such that $\Delta + T_i < \sigma < 2\Delta + T_i$ and $\sigma < 2\Delta - M_i$.

Assumption 9 ensures that orders placed at the farther tick are not adversely selected against. Informed traders will only pick off orders at the closest tick and market makers will always be willing to post at the farther ticks. Similar to Section 3, I maintain Assumption 7, that the commission c is set exogenously by the broker.

A. Equilibrium

In the extended model an equilibrium consists of: (i) a decision by market makers to post, or not, at every empty tick at each exchange; (ii) a solution to the broker's profit maximization problem; and (iii) a solution to the limit order investor's utility maximization problem.

Theorem 3 (Existence of a Threshold Equilibrium III):

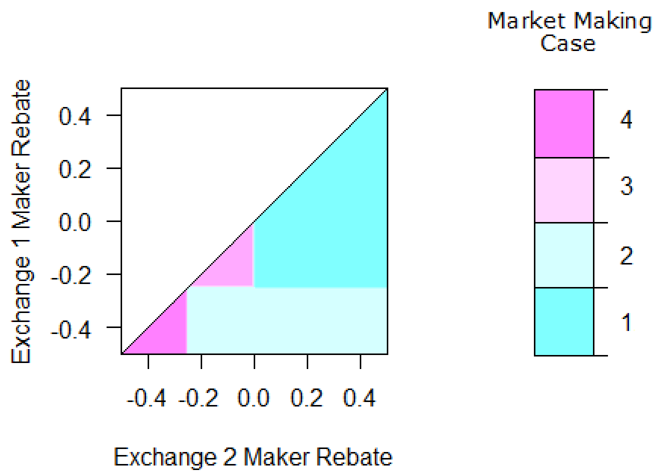
(i) For fixed parameters $M_1, M_2, \delta, \sigma, c$ there exists a unique pair of market making plans $\tilde{M}_1$ and $\tilde{M}_2$, such that market makers will choose to post at prices $\pm\Delta$ at each exchange i if $M_i \leq \tilde{M}_i$. Otherwise if $M_i > \tilde{M}_i$ they will post only at $\pm 2\Delta$ at exchange i.

(ii) For each market making plan and fixed parameters M_2, δ, σ, c, there exist unique threshold maker fees $\overline{M}_1(\Delta)$,$\overline{M}_1(2\Delta)$ such that if $M_1(x\Delta) \leq \overline{M}_1(x\Delta)$ brokers will optimally route limit orders at prices $x\Delta$ to exchange 1. Otherwise brokers will route limit orders at price $x\Delta$ to exchange 2.

(iii) For each market making plan and fixed parameters $M_1, M_2, \delta, \sigma, c$, there exist unique threshold private values $\overline{y}_i(\Delta)$, $\overline{y}_i(2\Delta)$, such that investors with $Q_1 = 1$ will submit an order at price $-\Delta$ if $y \geq \overline{y}_i(\Delta)$ and at price -2Δ if $\overline{y}_i(\Delta) > y \geq \overline{y}_i(2\Delta)$, given that their order will be routed to exchange i. Otherwise, if $y < \overline{y}_i(2\Delta)$, limit order buyers will abstain.

i. Market Makers. Given the expected routing of market orders, market makers choose whether or not to post at each empty tick to solve their profit maximization problem. Since there are many market makers, they individually choose whether each tick they may post at is profitable. Market-maker behaviour depends, in particular, on the difference in fees between the two exchanges. Market makers are aware that brokers will preferentially route market orders to the exchange with lower taker fees, lowering the chance of being adversely selected at these exchanges.

Figure 3
Types of Market Maker Equilibrium



This figure represents the equilibrium for market makers. Case 1: Market makers are willing to post at the narrowest ticks (Δ) at both exchanges. Case 2: Market makers are willing to post at the narrow tick at exchange 2 (Δ), but only at the farther tick (2Δ) at exchange 1. Case 3: Market makers are willing to post at the narrow tick (Δ) at exchange 1, but only at the farther tick (2Δ) at exchange 2. Case 4: Market makers are only willing to post at the farthest ticks (2Δ) at both exchanges.

In equilibrium, there are four possible cases for market making, which will be referred to throughout the remainder of this section. Market makers are always willing to post at the far ticks at both exchanges, and therefore the cases are defined by their willingness to post at the narrow ticks. The cases are: (1) market makers post at all ticks at both exchanges; (2) market makers post only at the narrow tick of the high-maker-rebate exchange; (3) market makers post only at the narrow tick of the high-fill-rate exchange; and (4) market makers post only at the far ticks. These market making cases are defined by fee thresholds $\tilde{M}_1$ and $\tilde{M}_2$.

Proposition 8 (Market-Making Behaviour):

If M_1 increases from $M_1 \leq \tilde{M}_1$ to $M_1 > \tilde{M}_1$:

(i) Market makers will no longer post at prices $\pm\Delta$ at exchange 1;

(ii) $\tilde{M}_2$ rises and market makers will begin to post at prices $\pm\Delta$ at exchange 2, if they were not already doing so.

Proposition 8 describes the results shown in Figure 3. Exchange 1 receives a larger proportion of uninformed market orders, and the market makers' decision to post at this exchange affects the execution probability at exchange 2. The increase in the fee at exchange 1 to $M_1 > \tilde{M}_1$ causes market makers to cease posting orders at the narrow price levels at exchange 1. This increases the number of uninformed orders that reach exchange 2, and decreases $\tilde{M}_2$ such that market makers will optimally post at the narrow price levels. This transition can be seen between Case 3 and Case 2 in Figure 3.

These results are driven by the principle of order protection for market orders and provide policy insight. Specifically, the change in fee structure at one exchange may influence the spreads at the competing exchanges. This can occur even if the changes in exchanges' fees do not change the overall ranking of exchanges by the magnitude of fees. In the transition between Cases 1 and 2 presented in Figure 3, exchange 1 becomes a taker-maker exchange and market makers are no longer willing to post at the best. Order protection moves market orders to exchange 2, because it quotes a narrower spread.

Table I
Equilibrium Execution Probabilities Given Market-Maker Behaviour

MM Case	Case 1	Case 2	Case 3	Case 4
$\theta_1(\Delta)$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\theta_2(\Delta)$	$\frac{1}{4}(1+\delta)$	$\frac{1}{2}$	$\frac{1}{4}(1+\delta)$	$\frac{1}{2}$
$\theta_1(2\Delta)$	0	$\frac{1}{4}(1-\delta)$	$\frac{1}{4}(1-\delta)$	$\frac{1}{2}(1-\delta)$
$\theta_2(2\Delta)$	0	0	0	$\frac{1}{4}(1-\delta)$

This table represents the equilibrium execution probability for each price level given market-maker behaviour at each exchange. Case 1: Market makers are willing to post at the narrowest ticks (Δ) at both exchanges. Case 2: Market makers are willing to post at the narrow tick at exchange 2 (Δ), but only at the farther tick (2Δ) at exchange 1. Case 3: Market makers are willing to post at the narrow tick (Δ) at exchange 1, but only at the farther tick (2Δ) at exchange 2. Case 4: Market makers are only willing to post at the farthest ticks (2Δ) at both exchanges.

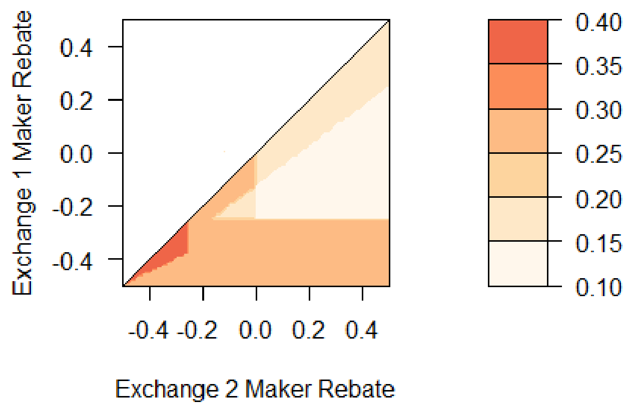
The behaviour of market makers drives the execution probability at both exchanges in this extension. If the trader chooses to submit at the narrowest tick, there is a constant risk that when the security value undergoes an innovation, the order will be picked off. This occurs with probability $\dfrac{1}{2}\delta$. Given the optimal market order routing strategy, the total execution probability $\theta_i(x\Delta)$ is a function of the ticks that market makers are willing to post at. Table I gives the execution probabilities for limit orders, at each ask price, given market-maker behaviour.

Proposition 9 (Volume with Endogenous Market Making):

(i) If market makers post more aggressively at exchange 2, the maker-taker exchange, brokers will always route limit orders at the best to exchange 2.

(ii) If market makers post more aggressively at exchange 2, total volume will be higher at exchange 2.

Figure 4
Limit Order Investor Expected Utility



This figure represents the expected utility of the limit order investor, under varying exchange fees. Investors have higher expected utility values when exchanges are similar and when exchange fees force market makers to post wider spreads.

One result of the baseline model is that volume will be higher at inverse exchanges, a prediction not seen in reality. In this extension, exchange fees may allow for a better spread, and therefore a higher volume, at maker-taker exchanges. Further, if exchange 2 is the only venue where market makers post at the best, brokers will send any limit orders at the best to this exchange. Since their orders will not be covered by market makers posting at exchange 1, this exchange will have both a higher execution probability for the client and a better rebate for the broker.

ii. Limit Order Investors. The presence of endogenous market makers complicates the decision of limit order investors, as their presence affects the execution probability of their limit order, as seen in Table I.

Figure 4 illustrates two separate effects. The first occurs near the 45-degree line when the trading fees are similar at both venues. This region corresponds to where $M_1(\Delta) \leq \overline{M}_1(\Delta)$

and $M_1(2\Delta) \leq \overline{M}_1(2\Delta)$. In this region, brokers optimally route orders to the exchange with the higher fill rate, rather than the higher maker rebate (or lower maker fee), and investor utility is higher. The second effect occurs on the lower border, where maker fees are high at exchange 1. In these regions, investors are better off since market makers no longer post at the narrowest ticks at one or both venues.

Proposition 10 (Limit Order Investor Utility in the Extended Model):

(i) If M_1 increases from $\overline{M}_1 < M_1 \leq \tilde{M}_1$ to $M_1 > \tilde{M}_1$, limit order investor utility increases.

(ii) If M_2 increases from $M_2 \leq \tilde{M}_2$ to $M_2 > \tilde{M}_2$, limit order investor utility increases.

The increase in investor utility described in Proposition 10 comes from two sources. First, when M_1 increases such that market makers no longer post at the narrow ticks at exchange 1, the proportion of uninformed orders increases at exchange 2. There is an increase in both the expected value of limit orders routed to exchange 2 and the proportion of investors willing to submit an order. Second, when either M_1 or M_2 increases, such that market makers no longer post at the narrow ticks of one exchange, liquidity declines at the best. Some investors who were previously unwilling to submit orders will then choose to submit orders at the wider price levels ($\pm 2\Delta$), as these orders now have a meaningful probability of execution, increasing utility.

Rebates allow market makers to post more aggressively, and in turn give them a competitive advantage against limit order investors, who don't receive rebates. It is important to note, that narrower spreads are advantageous to many market participants, notably those trading with market orders. Therefore, while limit order investor utility does increase when market makers choose not to compete at tight spreads, the outcomes for other investors whose utility is not addressed by this model may decline.

V. Conclusion

The principal-agent relationship between brokers and their clients has the potential to affect both the individuals involved and markets as a whole. While existing theoretical literature has addressed the concept of exchange fees, specifically maker-taker pricing, a gap remains in explaining how these fees and rebates drive broker behaviour and affect their clients. To explain these effects, I construct a static model of limit order trading in which brokers route limit orders from their clients to one of two venues.

I show that in an environment with fixed price levels, rebates for making and taking liquidity are able to drive broker routing decisions for both limit and market orders. These routing decisions, in turn, affect both the fill rates and relative probability of informed trading at both exchanges. Fill rates are higher for limit orders placed at exchanges with smaller maker rebates, while the relative probability of facing informed orders is higher at exchanges with larger maker rebates.

I find that when exchanges have similar fee structures, brokers have less incentive to deviate from their clients' interests, and will optimally route to the exchange with a higher fill rate. In this case, their clients will submit more orders, each order will have a higher expected value for the client, and investor welfare will be higher. On the other hand, when exchanges have sufficiently different fee structures, routing will be driven by liquidity rebates, and investor welfare will be lower. The decision to route is also influenced by the broker's commission, given exogenously in this model. I show that when commissions are higher, a broker's interests become aligned with their client's, as they profit from a higher fill rate.

These results are furthered in an extended model with multiple price levels and endogenous market making. In this environment, limit order investors also benefit when both exchanges charge high liquidity-making fees. When maker fees are high, market makers are no longer subsidized when making liquidity, and are less willing to provide liquidity at narrow spreads. Consequently, limit order investors, who pay only flat fees to their brokers, face less competition for their orders and receive a utility improvement. I find that changes in the fee structure at one exchange may influence the spread at a second exchange. This occurs through the shifting of uninformed traders across exchanges.

This model offers a number of empirical predictions that remain available for testing. First, optimal limit order routing depends on the probability of adverse selection. In periods when adverse selection is higher (such as announcement dates), brokers have a greater incentive to send limit orders to maker-taker exchanges. These exchanges should see a greater relative volume of retail limit orders on these dates. Second, the broker's commission size dictates the broker's routing behaviour. Brokers who charge a higher commission should route limit orders to the same exchanges where they route marketable orders, while those with lower commissions will split limit orders and market orders to different exchanges. Finally, any change in the relative fees between exchanges should alter broker routing behaviour in a predictable manner. An exchange that lowers its taker fee should see a higher concentration of retail market orders and a lower concentration of retail limit orders. Consequently, this exchange should see lower levels of adverse selection and higher fill rates. An exchange that lowers its maker fee (or introduces a higher rebate) should see the reverse effect.

This model also has a number of implications for policy regarding both trading venues and brokers. First, it implies that the proliferation of trading venues may not necessarily be beneficial for investor welfare. In some cases, orders may be routed to venues with higher adverse selection costs or lower fill rates. Second, while brokers may be limited in some of their decisions (such as by the Order Protection Rule), it is important to take factors other than price into account when defining concepts such as best execution. It is beneficial to investors if brokers consider factors such as fill rates when selecting venues for their clients. Finally, the change in the fee structure at one exchange can influence the spreads and market conditions at other exchanges, and these changes do not occur in a vacuum.

References

Amihud, Y., & Mendelson, H. (1980). Dealership market: Market-making with inventory. *Journal of Financial Economics*, *8*(1), 31–53.

Anand, A., McCormick, T., & Serban, L. (2013). Incentives for liquidity provision: Is the make-take structure the answer? *SSRN Working Paper*.

Angel, J. J., Harris, L. E., & Spatt, C. S. (2011). Equity trading in the 21st century. *The Quarterly Journal of Finance*, *1*(01), 1–53.

Bakos, Y., Lucas Jr, H. C., Oh, W., Simon, G., Viswanathan, S., & Weber, B. W. (2005). The impact of e-commerce on competition in the retail brokerage industry. *Information Systems Research*, *16*(4), 352–371.

Battalio, R., Corwin, S. A., & Jennings, R. (2016). Can brokers have it all? On the relation between make-take fees and limit order execution quality. *The Journal of Finance*.

Battalio, R., & Holden, C. W. (2001). A simple model of payment for order flow, internalization, and total trading cost. *Journal of Financial Markets*, *4*(1), 33–71.

Boehmer, E., Jennings, R., & Wei, L. (2007). Public disclosure and private decisions: Equity market execution quality and order routing. *Review of Financial Studies*, *20*(2), 315-358.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, *27*(8), 2267–2306.

Brolley, M., & Malinova, K. (2013). Informed trading and maker-taker fees in a low-latency limit order market. *SSRN Working Paper*.

Chakravarty, S., & Sarkar, A. (2002). A model of broker's trading, with applications to order flow internalization. *Review of Financial Economics*, *11*(1), 19–36.

Chao, Y., Yao, C., & Ye, M. (2015). Tick size constraints, two-sided markets, and competition between stock exchanges. *Working Paper*, Retrieved from: http://www.economics.illinois.edu/seminars/documents/Ye.pdf, October 28, 2016.

Colliard, J.-E., & Foucault, T. (2012). Trading Fees and Efficiency in Limit Order Markets. *The Review of Financial Studies*, *25*(11), 3389-3421.

Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, *19*(1), 69–90.

Foucault, T. (1999). Order flow composition and trading costs in a dynamic limit order market. *Journal of Financial Markets*, *2*(2), 99–134.

Foucault, T., Moinas, S., & Theissen, E. (2007). Does anonymity matter in electronic limit order markets? *The Review of Financial Studies*, *20*(5), 1707-1747.

Glosten, L. R. (1994, September). Is the electronic open limit order boon inevitable? *The Journal of Finance*, *49*(4), 1127-1161.

Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*, 71-100.

Ho, T., & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial economics*, *9*(1), 47–73.

Malinova, K., & Park, A. (2015). Subsidizing liquidity: The impact of make/take fees on market quality. *The Journal of Finance*, *70*(2), 509–536.

Milgrom, P., & Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, *26*(1), 17–27.

Pagnotta, E., & Philippon, T. (2011). Competing on speed. *National Bureau of Economic Research*.

Parlour, C. A. (1998). Price dynamics in limit order markets. *Review of Financial Studies*, *11*(4), 789–816.

Stoll, H. R. (1978). The supply of dealer services in securities markets. *The Journal of Finance*, *33*(4), 1133–1151.

Yim, R., & Brzezinski, A. (2012). Limit order placement across multiple exchanges. *2012 IEEE Conference on Computational Intelligence for Financial Engineering and Economics (CIFEr)*.

A Appendix: Proofs and Additional Material

A. Benchmark Equilibrium

Proof of Theorem 1 Part i

Consider a limit order investor wishing to buy the security. This limit order investor will be willing to trade at exchange 1 versus not trading if:

$$E[V|Ex_1] + y + \Delta - M_1 \geq 0. \tag{10}$$

Substituting the expected value term,[15] this will occur when:

$$y \geq \delta \cdot \sigma - \Delta + M_1. \tag{11}$$

Consider the same limit order investor. This investor will be willing to trade at exchange 1, versus trading at exchange 2 if:

$$Pr(Ex_1)(E[V|Ex_1] + y + \Delta - M_1) \geq Pr(Ex_2)(E[V|Ex_2] + y + \Delta - M_2) \tag{12}$$

$$\frac{1}{2}(-\delta \cdot \sigma + y + \Delta - M_1) \geq (\frac{1}{2}\delta + \frac{1}{4}(1-\delta))(\frac{-\delta \cdot \sigma}{\delta + 0.5 \cdot (1-\delta)} + y + \Delta - M_2). \tag{13}$$

This will occur if:

$$y \geq -\Delta + \frac{2M_1 - (1+\delta)M_2}{1-\delta}. \tag{14}$$

Define $\overline{y_1}$ as the threshold over which investors are more willing to trade at exchange 1 than either not trading, or trading at exchange 2. This occurs at:

[15]For proof of expected values and probability, see proofs of Propositions 1 and 2.

$$\overline{y_1} = max\left\{\delta \cdot \sigma - \Delta + M_1, -\Delta + \frac{2M_1 - (1+\delta)M_2}{1-\delta}\right\}. \tag{15}$$

Proof of Theorem 1 Part ii

By the definition in Part i, no trader would be willing to trade at exchange 2 over exchange 1 if $y > \overline{y_1}$. Investors are willing to trade at exchange 2, rather than not trade if:

$$E[V|Ex_2] + y + \Delta - M_2 \geq 0. \tag{16}$$

Traders are willing to trade when this holds with equality and thus $\overline{y_2}$ can be defined as:

$$\overline{y_2} = \frac{\delta \cdot \sigma}{\delta + 0.5 \cdot (1-\delta)} - \Delta + M_2. \tag{17}$$

Since all traders with $y > \overline{y_1}$ are unwilling to trade at exchange 2, if $\overline{y_2} \geq \overline{y_1}$, no traders are willing to trade at exchange 2. If alternatively, $\overline{y_2} < \overline{y_1}$, then there exist some traders with $\overline{y_2} \leq y < \overline{y_1}$ who would prefer to trade at exchange 2 than either trade at exchange 1 or not trade.

Proof of Proposition 1

Given $T_1 < T_2$, brokers will route all small market orders to exchange 1, and all large market orders to both exchanges. Exchange 1 will receive orders any time a market order trader wishes to buy $(\frac{1}{2}(1-\delta))$, while exchange 2 will receive orders any time a market order trader wishes to buy $Q_2 = 2\ (\frac{1}{4}(1-\delta))$. Both exchanges will receive an equal quantity of informed orders $(\frac{1}{2}\delta)$. Thus, the execution probabilities at the two exchanges are:

$$\begin{aligned}\theta_1 &= \frac{1}{2}\delta + \frac{1}{2}(1-\delta) && (18)\\ \theta_2 &= \frac{1}{2}\delta + \frac{1}{4}(1-\delta). && (19)\end{aligned}$$

Since $0 < \delta < 1$, then $\theta_1 > \theta_2$.

Proof of Proposition 2

Consider a limit order buy. By an application of Bayes' Rule, the expected value of the security, conditional on execution at exchange i is:

$$E[V|Ex_i] = \frac{Pr(V=\sigma, Ex_i)\cdot\sigma + Pr(V=0, Ex_i)\cdot 0 + Pr(V=-\sigma, Ex_i)\cdot -\sigma}{Pr(V=\sigma, Ex_i) + Pr(V=0, Ex_i) + Pr(V=-\sigma, Ex_i)}. \qquad (20)$$

With probability $\frac{1}{2}\delta$, $V=\sigma$. An informed trader arrives and removes all orders at the ask. No orders at the bid execute, thus $Pr(V=\sigma, Ex_i)=0$. Alternatively, with probability $\frac{1}{2}(1-\delta)$, $V=0$ and a liquidity trader wishing to sell arrives. He always wishes to sell at least one unit, thus $Pr(V=0, Ex_1)=\frac{1}{2}(1-\delta)$. With probability $\frac{1}{4}(1-\delta)$, he wishes to sell two units, and $Pr(V=0, Ex_2)=\frac{1}{4}(1-\delta)$. Finally, with probability $\frac{1}{2}\delta$, $V=-\sigma$. An informed trader arrives and removes all orders at the bid, thus $Pr(V=-\sigma, Ex_i)=\frac{1}{2}\delta$.

Through substitution of the above probabilities into Equation 20 and algebraic manipulation:

$$\begin{aligned}E[V|Ex_1] &= -\delta\cdot\sigma && (21)\\ E[V|Ex_2] &= \frac{-\delta\cdot\sigma}{\delta + 0.5\cdot(1-\delta)}. && (22)\end{aligned}$$

Proof of Proposition 3

The total execution probability at exchange 1 ($\frac{1}{2}$) is higher than that at exchange 2 ($\delta + \frac{1}{4}(1-\delta)$). If the expected value of the trade, conditional on execution, is higher at exchange 1 for all investors, then the total expected value of the trade is also higher. This is true if the increase in expected value from being routed to exchange 1 is strictly greater than the increase in fees.

$$E[V|Ex_1] - E[V|Ex_2] \geq M_1 - M_2 \tag{23}$$

Substitution of the expected value equation and manipulation of this condition leads to to solution:

$$M_1 - M_2 \leq \frac{1-\delta}{1+\delta}\delta\sigma. \tag{24}$$

Proof of Corollary 1

Consider brokers who pass fees through to their investors, but charge a positive commission c. In equilibrium, brokers will route limit orders to maximize profit, to whichever exchange satisfies:

$$\theta_i c \geq \theta_j c. \tag{25}$$

Since the probability of execution is strictly greater at exchange 1, brokers will optimally route all limit orders to exchange 1. Under Proposition 3, limit order investors with $\overline{y}_2 \leq y < \overline{y}_1$ prefer to be routed to exchange 2 when they pay exchange fees. These investors are now routed to exchange 1, and are worse off when the broker controls their routing decisions.

B. Single Commission

Lemma 1

Lemma 1 (Limit Order Routing Equilibrium):

For limit orders, brokers are able to route, at the price dictated by their client, to any exchange. In equilibrium, brokers will route limit orders to maximize profit, to whichever exchange satisfies:

$$\theta_i[c - M_i] \geq \theta_j[c - M_j]. \tag{26}$$

Proof of Lemma 1

Investors are unable to contract with brokers and brokers choose routing behaviour following receipt of an order.

Suppose a broker were to agree to route to exchange 1 where $\theta_1 > \theta_2$ but $\theta_1[c - M_1] < \theta_2[c - M_2]$. If $\overline{y}_1 < \overline{y}_2$, it is possible that more orders would be submitted and total expected profit over all possible investors would be higher such that $Pr(y \geq \overline{y}_1)\theta_1[c - M_1] > Pr(y \geq \overline{y}_2)\theta_2[c - M_2]$.

However, for any given order, $\theta_1[c - M_1] < \theta_2[c - M_2]$ and the broker has incentive to deviate following the receipt of the order. Since the broker chooses routing behaviour after the receipt of the order, he has no incentive not to deviate. Therefore, the broker's promise to route to exchange 1 is not credible, unless every order sent there is more profitable.

Proof of Theorem 2

The equilibrium is obtained through two steps in backwards induction. First, the routing of market orders determines the broker's expected profit for limit orders at exchanges 1 and 2. This determines the threshold $\overline{M}_1$. Second, given the threshold $\overline{M}_1$, the limit order investor anticipates that her limit order will be routed to exchange i and determines the expected utility of order submission. This determines the threshold $\overline{y}_i$.

Proof of Theorem 2 Part i

By Lemma 1, the broker will optimally route to exchange 1 iff:

$$\theta_1[c - M_1] \geq \theta_2[c - M_2]. \tag{27}$$

By substitution of θ_i and algebraic manipulation this gives the condition that:

$$M_1 \leq c\frac{1-\delta}{2} + M_2\frac{1+\delta}{2}. \tag{28}$$

Denoting $M_1 = \overline{M}_1$ when this condition holds with equality, the broker will optimally route any limit order to exchange 1 if $M_1 \leq \overline{M}_1$.

Proof of Theorem 2 Part ii

Limit order traders take broker routing as given, following the broker's equilibrium market order routing decision and Lemma 1. Following the limit order trader's maximization problem and given routing to exchange i, a limit order trader wishing to buy will submit an order iff:

$$\theta_i \left(E\left[V|Ex_i\right] + y - (-\Delta) - c\right) \geq 0. \tag{29}$$

Through substitution of θ_i from Proposition 1, $E\left[V|Ex_i\right]$ from Proposition 2 and algebraic manipulation, the conditions for exchanges 1 and 2 are:

$$\begin{aligned} y &\geq c + \delta \cdot \sigma - \Delta && (30) \\ y &\geq c + \frac{2\delta \cdot \sigma}{1+\delta} - \Delta. && (31) \end{aligned}$$

Denoting the conditions $\overline{y}_1$ and $\overline{y}_2$ at equality, any limit order trader with $y \geq \overline{y}_i$ will satisfy this condition. These values exist for any parameter set following Assumptions 1 and 2.

Proof of Proposition 4

See proof of Theorem 2, Part ii.

Proof of Proposition 5

Consider two similar exchanges, such that $|T_1 - T_2|$ is small and $M_1 < \overline{M_1}$. Suppose exchange 1 raises its maker fee sufficiently such that $M_1 > \overline{M_1}$, brokers will optimally route limit orders to exchange 2 via Lemma 1 and Theorem 1.

Given that brokers begin routing limit orders to exchange 2, fewer limit order traders will submit orders since $\overline{y}_1 < \overline{y}_2$. Each limit order will have a lower expected profitability since, by Proposition 2, $E[V|Ex_1] > E[V|Ex_1]$. Thus, since fewer orders are submitted and each order earns a lower expected profit, total limit order investor welfare will decline.

Consider instead, two similar exchanges such that $|T_1 - T_2|$ is small and $M_1 > \overline{M_1}$. Suppose exchange 1 lowers its maker fee sufficiently such that $M_1 < \overline{M_1}$, brokers will optimally route limit orders to exchange 1 via Lemma 1 and Theorem 1. The results from above are reversed and total limit order welfare increases.

Lemma 2

Lemma 2 (Increase in Adverse Selection Risk):

An increase in δ:

(1)Decreases the expected value of the security for limit order buyers, given execution at all exchanges.

(2) Lowers threshold $\overline{M}_1$.

Proof of Lemma 2

Given the equilibrium conditions:

$$E[V|Ex_1] = -\delta \cdot \sigma \tag{32}$$

$$E[V|Ex_2] = \frac{-\delta \cdot \sigma}{\delta + 0.5 \cdot (1-\delta)}. \tag{33}$$

Derivation gives:

$$\frac{\partial E[V|Ex_i]}{\partial \delta} < 0. \tag{34}$$

Given the equilibrium condition:

$$\overline{M}_1 = c\frac{1-\delta}{2} + M_2\frac{1+\delta}{2}. \tag{35}$$

The fact that $c \geq M_2$ and derivation shows:

$$\frac{\partial \overline{M}_1}{\partial \delta} < 0. \tag{36}$$

Proof of Proposition 6

Follows from Lemma 2, proof of Theorem 2, Proposition 2 and $0 < \delta < 1$, $c > |M_i|$.

Proof of Proposition 7

As with Theorem 2, by Lemma 1, the broker will optimally route to exchange 1 iff:

$$\theta_1[c - M_1] \geq \theta_2[c - M_2]. \tag{37}$$

By algebraic manipulation this gives the condition that:

$$c \geq \frac{2M_1 - (1+\delta)M_2}{1-\delta}. \tag{38}$$

Denoting $c = \underline{c}$ when this condition holds with equality, the broker will optimally route any limit order to exchange 1 if $c \geq \underline{c}$.

C. Endogenous Market Making

Proof of Theorem 3 Part i

Consider market makers wishing to buy at the bid. Given that brokers will first route market orders to exchange 1, it is optimal for the market makers to post a limit order at exchange 1 iff:

$$-\delta \cdot \sigma - \Delta \geq M_1. \tag{39}$$

$\tilde{M}_1$ is such that the above equation holds with equality.

If $M_1 \leq \tilde{M}_1$, a limit order will always be posted there, either by the market maker or a limit order trader. In this case, it is optimal for the market makers to post at exchange 2 iff:

$$\frac{-\delta \cdot \sigma}{\delta + 0.5 \cdot (1-\delta)} - \Delta \geq M_2. \tag{40}$$

In this case, $\tilde{M}_2$ is such that the above equation holds with equality.

If $M_1 > \tilde{M}_1$, market makers will not post at exchange 1. Further, if no market maker posts at exchange 1, limit order investors will always be routed to exchange 2 since $M_1 > M_2$ and $\theta_1(\Delta) = \theta_2(\Delta)$. Thus, no limit order will ever be posted at the price level $-\Delta$, at exchange 1. In this case, it is optimal for the market makers to post at exchange 2 iff:

$$-\delta\cdot\sigma-\Delta\geq M_2. \tag{41}$$

In this case, $\tilde{M}_2$ is such that the above equation holds with equality.

Proof of Theorem 3 Part ii

Proof is similar to Theorem 2, Part i.

The probabilities $\theta_1(\Delta),\theta_2(\Delta),\theta_1(2\Delta),\theta_2(2\Delta)$ are determined by the market-maker behaviour established in Theorem 3, Part i.

By Lemma 1, the broker will optimally route to exchange 1 iff:

$$\theta_1(\Delta)[c-M_1]\geq\theta_2(\Delta)[c-M_2]. \tag{42}$$

By algebraic manipulation this gives the condition that:

$$M_1\leq\frac{\theta_1(\Delta)-\theta_2(\Delta)}{\theta_1(\Delta)}c+\frac{\theta_2(\Delta)}{\theta_1(\Delta)}M_2. \tag{43}$$

Denoting $M_1=\overline{M}_1(\Delta)$ when this condition holds with equality, the broker will optimally route any limit order to exchange 1 if $M_1\leq\overline{M}_1(\Delta)$.

Proof for the price level 2Δ follows through identical reasoning.

Proof of Theorem 3 Part iii

Proof is similar to Theorem 2, Part ii.

Limit order traders take broker routing as given, following the broker's equilibrium market order routing and Lemma 1. Given that limit orders at price Δ will be routed to exchange i and orders at price 2Δ will be routed to exchange j, limit order traders will submit an order at price Δ if:

$$\theta_i(\Delta)\left(E\left[V|Ex_i\right]+y-(-\Delta)-c\right)\geq\theta_j(2\Delta)\left(y-(-2\Delta)-c\right) \tag{44}$$

$$\theta_i(\Delta)\left(E\left[V|Ex_i\right]+y-(-\Delta)-c\right)\geq 0. \tag{45}$$

In other words, the expected value of submitting at Δ is higher than either submitting at 2Δ or abstaining. Since $\theta_1(\Delta), \theta_2(\Delta), \theta_1(2\Delta), \theta_2(2\Delta), E\left[V|Ex_1\right], E\left[V|Ex_2\right]$ and broker routing decisions are already determined by market-maker behaviour, the condition $\overline{y}_i(\Delta)$ can be determined by evaluating the two above conditions at equality, and selecting whichever is more stringent:

$$\overline{y}_1(\Delta)=max\left\{\frac{\theta_j(2\Delta)(2\Delta-c)-\theta_i(\Delta)(E\left[V|Ex_i\right]-(-\Delta)-c)}{\theta_i(\Delta)-\theta_j(2\Delta)}, -E\left[V|Ex_i\right]+(-\Delta)+c\right\}. \tag{46}$$

Algebraic manipulation can show that:

$$\frac{\theta_j(2\Delta)(2\Delta-c)-\theta_i(\Delta)(E\left[V|Ex_i\right]-(-\Delta)-c)}{\theta_i(\Delta)-\theta_j(2\Delta)}\geq -E\left[V|Ex_i\right]+(-\Delta)+c \tag{47}$$

and thus:

$$\overline{y}_1(\Delta)=\frac{\theta_j(2\Delta)(2\Delta-c)-\theta_i(\Delta)(E\left[V|Ex_i\right]-(-\Delta)-c)}{\theta_i(\Delta)-\theta_j(2\Delta)}. \tag{48}$$

Traders with $y<\overline{y}_1$ will optimally submit at price level 2Δ if:

$$\theta_i(2\Delta)\left(y-(-2\Delta)-c\right)\geq 0. \tag{49}$$

The condition $\overline{y}_i(2\Delta)$ can be determined by evaluating the preceding condition at equality, in order to obtain:

$$\overline{y}_i(2\Delta) = -2\Delta + c. \tag{50}$$

Proof of Proposition 8

Consider an increase from $M_1 \leq \tilde{M}_1$ to $M_1 > \tilde{M}_1$. $\tilde{M}_1$ is defined as the point at which market makers are indifferent from posting, or not, at the first tick on exchange 1. An increase of liquidity-making fees decreases market maker profitability and violates the condition $-\delta \cdot \sigma - \Delta \geq M_1$. Thus, if M_1 increases from $M_1 \leq \tilde{M}_1$ to $M_1 > \tilde{M}_1$, it is no longer optimal for market makers to post at the first tick at exchange 1.

Given that market makers are no longer posting at the first tick at exchange 1, any market order now reaches the first tick at exchange 2 with priority, as a result of order protection. The expected value at exchange 2 changes such that $E[V|Ex_2] = -\delta \cdot \sigma$. Since $M_1 > M_2$, then $-\delta \cdot \sigma - \Delta \geq M_2$. It is now optimal for market makers to post at exchange 2.

Thus, if M_1 increases from $M_1 \leq \tilde{M}_1$ to $M_1 > \tilde{M}_1$. $\tilde{M}_1$, the spread will improve at exchange 2.

Proof of Proposition 9 Part i The case where market makers post more aggressively at exchange 2 represents Case 2 in Table I. Referring to Table I, the probability of execution at the best is equal at both exchanges. Since $M_2 < M_1$, the profitability, given execution, is higher at exchange 2 for the broker than at exchange 1. Therefore, the total expected profit for orders at the best is higher for the broker at exchange 2.

Proof of Proposition 9 Part ii

First, consider the case where a limit order is not posted in $t = 1$. Expected volume at exchange 2 will be $\frac{1}{2}\delta$ for informed orders and $\frac{1}{2}(1-\delta)$ for uninformed orders. Expected volume at exchange 1 will be only $\frac{1}{4}(1-\delta)$ for uninformed orders, as brokers must first send market orders to the best price. Total volume at exchange 2 is strictly greater than that at exchange 1 ($\frac{1}{2} > \frac{1}{4}(1-\delta)$).

Second, consider the case where a limit order is posted in $t = 1$. If it is placed at the best, given the results of Part i, it will be routed to exchange 2. Expected volume is the same as in the case where no limit order is posted. If it is placed at the second tick, it will be routed to exchange 1 since, in market making Case 2:

$$\theta_2(2\Delta) = 0. \tag{51}$$

The market maker will then post at the best, at exchange 2, and volume will be identical to the case where no limit order is posted in $t = 1$.

Therefore, regardless of whether a limit order investor submits an order, volume will always be higher at exchange 2 in market making Case 2.

Proof of Proposition 10 Part i

Consider $\overline{M}_1 < M_1 \leq \tilde{M}_1$. Brokers are routing limit orders to exchange 2. If the fees at exchange 1 increase to $M_1 > \tilde{M}_1$, brokers will continue to route to exchange 2; however, $\theta_2(\Delta)$ increases as there are no longer any limit orders from market makers at exchange 1. The value $E[V|Ex_2]$ also improves for limit order investors, as a higher concentration of uninformed market orders ($\frac{1}{2}(1-\delta)$, as opposed to $\frac{1}{4}(1-\delta)$), reach exchange 2. Therefore, all limit order traders posting at exchange 2 have higher utility in expectation.

In addition, $\theta_1(2\Delta)$ increases from 0 to $\frac{1}{4}(1-\delta)$. Some measure of traders, between $\overline{y}_i(\Delta)$ and $\overline{y}_i(2\Delta)$, will begin submitting orders at the price level 2Δ. Each of these traders will also be better off in expectation.

Proof of Proposition 10 Part ii

Consider $\overline{M}_2 < M_2 \leq \tilde{M}_2$, increasing to $M_2 > \tilde{M}_2$. If $M_1 \leq \tilde{M}_1$, there is still liquidity available at the best. $\theta_1(\Delta)$, $\theta_2(\Delta)$, $E[V|Ex_1]$ and $E[V|Ex_2]$ remain the same; however, $\theta_1(2\Delta)$ increases from 0 to $\frac{1}{4}(1-\delta)$. Some measure of traders, between $\overline{y}_i(\Delta)$ and $\overline{y}_i(2\Delta)$, will begin submitting orders at the price level 2Δ. Each of these traders will also be better off in expectation.

If $M_1 > \tilde{M}_1$, there is now no liquidity available at the best. $\theta_1(2\Delta)$ increases from $\frac{1}{4}(1-\delta)$ to $\frac{1}{2}(1-\delta)$ and $\theta_2(2\Delta)$ increases from 0 to $\frac{1}{4}(1-\delta)$. Each trader submitting an order at price level 2Δ is better off.

B Appendix: Graphical Parameters

All graphs are based on the following parameters:

Table II
Graphical Parameters

Parameter	Notation	Value
Initial Asset Value	v	0
Tick Size	Δ	1
Probability of Informed Trading	δ	0.5
Innovation Size	σ	1.5
Commission	c	0.5
Private Value Maximum Range	y	2