A Theory of High Frequency Market Making in Fragmented Markets*

Soomin "Tomy" Lee†

Department of Economics

University of Toronto

Preliminary

May 26, 2016

Abstract

I examine the impact of high frequency ("fast") market making on liquidity, price discovery and institutional traders' returns in a setting with market fragmentation. Faster market makers more precisely monitor orderflows across trading venues than slower market makers. If every market maker's speed increases, I find that liquidity and price discovery improve at all trading venues while institutional traders' returns decline. When the market makers' speeds increase at a specific trading venue, liquidity and price discovery at that venue improve and spillovers affect the other venues. If the market makers at the other trading venues are sufficiently slow, the spillovers positively affect liquidity and price discovery at those venues; otherwise, the spillovers harm liquidity and price discovery.

JEL-Classification: D4, D62, G1, G20, L1

Keywords: high frequency trading, market making, liquidity, price discovery, institutional traders

*For invaluable advice and comments, I thank Andreas Park and Ekaterina Malinova, as well as the participants of the 4060 PhD Research Seminars and the Financial Economics Seminar at the University of Toronto, including Carolyn Pitchik, Albert Menkveld, Laura Turner, Jordi Mondria, Peter Cziraki, Angelo Melino, Mathieu Gilbert-Gonthier, Daniel Chippin and many others.

†150 St. George St. 229, Toronto, Ontario M5S 3G7, Canada; +1 (647)470-6719; tomy.lee@mail.utoronto.ca

Trading in today's financial markets is fragmented across many trading venues that operate in the mili- and microsecond frequencies.[1] Markets operate at such speeds that the IEX's recent attempt to register as an exchange with the US Securities and Exchange Commission (SEC) received dozens of comments concerning the 350 *micro*second delay it imposes on orders (SEC (2015)).[2] In 2014, the SEC published separate white papers on market fragmentation and high frequency trading in the US (SEC (2013, 2014)). However, the pervasiveness of market fragmentation and high frequency trading they document suggests that, rather than market fragmentation and high frequency trading being separate phenomena, the *interaction* between them may play an important role in the markets.[3]

In this paper, I develop a model in which high frequency traders are market makers and the speed of a market maker determines her ability to monitor orderflows at multiple trading venues. A high frequency market maker can quickly update her quotes to reflect this information, thereby allowing the market maker to avoid adverse selection by informed traders. The informed traders in the model represent buy-side institutional traders, such as hedge funds or asset management firms, who earn returns by exploiting time-sensitive private information. For instance, an institutional trader could trade on his forecast of an imminent announcement, price movements in related securities, or activity data from social networks.[4] The model explains how high frequency market makers can increase the trading costs of the institutional traders, while lowering the trading costs of retail traders. In addition, the model generates testable implications about the *spillovers* that a trading venue imposes on others when the speed of the venue's market makers changes. These results depend crucially on the interaction between market fragmentation and the speed of market making, rather than either in isolation.

Empirical evidence indicates that high frequency traders (HFTs) are net providers, not takers, of liquidity (Hagströmer and Norden (2013); Bershova and Rakhlin (2013); Korajczyk and Murphy (2015)).[5] Many HFTs specialize in market making strategies, fulfilling roles similar to traditional market makers (Hagströmer and Norden (2013); Menkveld (2013)). In addition, the HFTs now supply liquidity for the majority of trades at several trading venues in Europe and North America (Chordia, Goyal, Lehmann, and Saar (2013); Malinova and Park (2015); Riordan and Storkenmaier (2012)). This rise in high frequency trading has coincided with substantial improvements to overall liquidity and price discovery (Chordia, Roll, and Subrahmanyam (2011)), improvements which are likely caused by the HFTs (Hendershott, Jones, and Menkveld (2011); Brogaard, Hendershott, and Riordan (2014); Riordan and Storkenmaier (2012)). Despite the apparent benefits of HFTs to market quality, the HFTs

[1] Market fragmentation refers to the phenomenon in which the same asset is traded at multiple trading venues.

[2] e.g. comments submitted by NYSE, Southeastern Asset Management, Larry Tabb (CEO, TABB Group), and others.

[3] According to O'Hara (2015), "current market structure is thus highly competitive, highly fragmented, and very fast" (O'Hara, 2015, p. 258).

[4] Willmer (2015) provides an example of a hedge fund that uses data from social networks.

[5] For an overview of today's market structure and high frequency traders' strategies, see O'Hara (2015).

may harm the returns to institutional traders (Malinova, Park, and Riordan (2013); Tong (2015)).[6] My model helps to explain this dichotomy.

Two assumptions are crucial for my results: (i) information is fragmented across multiple trading venues; and (ii) faster traders process and act on the fragmented information more effectively than slower traders. The second assumption is grounded in several stylized facts regarding high frequency traders. First, the quotes of HFTs at a given trading venue reflect information at other trading venues (and from related securities) faster than the quotes of slower traders (Hendershott and Riordan (2013)). Price discovery in modern financial markets is consequently driven by the quotes of the HFTs, rather than trades or the quotes of slower traders (Brogaard, Hendershott, and Riordan (2015)). Second, HFTs monitor information more consistently than slower market participants (Chaboud, Chiquoine, Hjalmarsson, and Vega (2014); Malinova and Park (2015)), and process hard (e.g. orderflow) information more efficiently than human traders (Zhang (2012)). Third, HFTs quickly cancel their limit orders in response to new information (Subrahmanyam and Zheng (2015)).

My model is in the tradition of Kyle (1985), and is closely related to Foster and Viswanathan (1996).[7] As in Foster and Viswanathan (1996), there are multiple informed traders, noise traders, and competitive market makers who set prices. The informed traders represent institutional traders, and receive heterogenous (but correlated) private signals about the asset value, which the institutional traders use to extract information rents during trading.[8] On the other hand, the noise traders represent retail traders whose market orders reflect random beliefs and liquidity requirements.[9] Unlike Foster and Viswanathan (1996), there are multiple trading venues to represent market fragmentation. The price at each trading venue is set competitively by a market maker such that the price is equal to the market maker's best estimate of the asset value given the orderflows she observes. Each market maker precisely observes the orderflows at the trading venue where she is present, but she noisily observes the orderflows at the other venues. Faster market makers observe cross-venue orderflows more precisely than slower market makers.

The mechanics of my model may be interpreted as follows. Institutional traders possess an exogenous source of short-lived and time-sensitive private information (e.g. movement in prices of related assets). With the knowledge that private signals relevant to the asset value are forthcoming, the institutional traders select the trading venues where they wish to trade based on the speeds of the venues' market makers. Since the value of private information is short-lived, an institutional trader immediately submits

[6] Brogaard et al. (2014) finds that high frequency traders have no statistically significant effect on the returns to institutional traders.

[7] Specifically, the one-period case of Foster and Viswanathan (1996), which is solved in Pasquariello and Vega (2007).

[8] To emphasize that the informed traders (in Kyle (1985) sense of the term) represent institutions in my model, I will refer to the informed traders as "institutional traders" in the remainder of the paper.

[9] Noise traders (in Kyle (1985) sense) are referred to as "retail traders".

market orders at his preferred venues upon receiving private information. To illustrate, consider a hedge fund with analysts researching an asset. The fund first selects a trading venue at which to obtain direct or sponsored access. Then, once the analysts provide a new prediction on the asset's value, the fund attempts to exploit the prediction at the selected venue before others trade away the opportunity. The market makers observe orderflow from each trading venue and revise their quotes accordingly. Faster market makers process and act on the orderflow information more quickly than slower market makers so that the faster market makers' quotes incorporate a greater proportion of the orderflow information than the quotes of the slower market makers.

The interaction between market fragmentation and speed in my model helps to explain how high frequency traders reduce trading costs for retail traders, even as institutional traders earn smaller returns. Suppose market makers' speeds increase at all trading venues due to, for instance, the spread of colocation services (which are primarily purchased by market makers (Brogaard et al. (2015))) and, insofar as HFTs tend to be liquidity providers, regulations that encourage automated trading (e.g. Reg NMS in the US (Hendershott and Moulton (2011))). Then the precision with which market makers at each trading venue observes other venues' ("cross-venue") orderflows increases, thereby improving price discovery at every venue. Improved price discovery shrinks the information rents available to institutional traders, whose returns decline. In addition, with more precise cross-venue information, the market makers rely less on the orderflows at one's own trading venue to set their prices. Thus, price impact declines at each trading venue and liquidity improves overall.[10]

Suppose instead that market makers' speed increased at *particular* trading venue ("shocked venue") and speeds elsewhere were unaffected. Examples include the entry of a high frequency market maker (Menkveld (2013)) and technical upgrades to the venue's trading system (Hendershott et al. (2011)). In this case, the institutional traders face a tradeoff: continue trading at the shocked venue with faster market makers; or trade at other venues and intensify competition among the institutional traders at those venues. Either choice harms the institutional traders' returns, and the institutional traders substitute away from participating at the shocked venue just enough so that the reduction in their returns is minimized. This causes spillovers on non-shocked venues as follows. First, fewer institutional traders participate at the shocked venue which reduces the information content of the venue's orderflows, decreasing the cross-venue information available at the non-shocked venues. These venues' liquidity and price discovery then decline as the venues' market makers rely more on their own venues' orderflows to set prices — thereby raising price impact and deceasing liquidity — while the total information available at the venues decreases. Second, more institutional traders participate at the non-shocked venues, which

[10]Price impact is the amount by which the price moves against a trader once he submits his orders and is a measure of trading cost (a form of illiquidity).

intensifies competition among the institutional traders at those venues. Thus, the institutional traders at the non-shocked venues trade more aggressively and market makers become less sensitive to the venues' orderflows in response. Overall, the second effect improves liquidity and price discovery. Where trading venues are sufficiently fast, cross-venue information is important and the first effect dominates. Otherwise, the second effect dominates so that the spillovers are negative where trading venues are sufficiently fast, but are positive where the venues are slower.

The remainder of the paper is organized as follows. Section 1 provides an overview of the relevant literature. Section 2 details the model, including the structure of the game in the model. Section 3 solves the model for a linear and symmetric Perfect Bayesian equilibrium and discusses several equilibrium properties. Section 4 presents the main results of the paper. Section 5 discusses the limitations, and the market design and policy implications of the model, then concludes.

1 Related Literature

In this section, I summarize the theoretical literature on high frequency traders, and discuss how my model compliments the existing models.

HFTs in Multiple Venues: Several models feature high frequency traders in settings with multiple trading venues. In Foucault, Kozhan, and Tham (2014), the arrival of information at a trading venue creates opportunity for HFTs to exploit stale quotes at other venues. Baldauf and Mollner (2015) includes an analyst who endogenously acquires private signals then submits orders that are split across multiple venues. Some of the analyst's orders execute earlier than others, which allows the HFTs to trade ahead of the analyst's slower orders and, thereby, discourages information acquisition. Biais, Foucault, and Moinas (2015) has HFTs that always match with liquid venues and have superior information about the asset value. As the share of traders that are HFTs increases, adverse selection on market makers rises and liquidity is reduced. These models treat HFTs as liquidity takers that impose adverse selection costs on others. By contrast, fast traders in my model are market makers, not liquidity takers, whose speed allows them to *avoid* adverse selection.

HFTs in a Single Venue: A large theoretical literature examines high frequency traders in settings with a single trading venue. Many models from this literature assume that HFTs are liquidity takers who: anticipates the trades of other participants (Foucault, Hombert, and Rosu (2015),Rosu (2014)); free-rides on information obtained by others (Yang and Zhu (2015)); or cause volatility as the HFTs react simultaneously to a common signal (Budish, Cramton, and Shim (2013)[11],Jarrow and Protter

[11] Budish et al. (2013) suggests batch auctions as a solution to the excess volatility caused by the HFTs. Trading in my model is a batch auction, yet faster market makers have an informational advantage over slower market makers since the faster market makers have more precise cross-venue information than the slower market makers.

(2012)).

Conversely, Jovanovic and Menkveld (2011), Ait-Sahalia and Saglam (2014), Hoffmann (2014) and Menkveld and Zoican (2014) contain fast traders who provide liquidity in a single venue setting. In Jovanovic and Menkveld (2011) and Ait-Sahalia and Saglam (2014), fast traders are market makers, as in this paper. The fast traders in Jovanovic and Menkveld (2011) facilitate trading between those who arrives at the trading venue at different times while the fast trader in Ait-Sahalia and Saglam (2014) post limit orders based on a private signal. Unlike Jovanovic and Menkveld (2011), whose results pertain to welfare gains from trade, this paper focuses on the implications of fast market making for market quality. Further, unlike Ait-Sahalia and Saglam (2014), the market makers in my model only observe public information. Hoffmann (2014)'s fast traders can avoid adverse selection by cancelling limit orders after asset value shocks and, thus, are more likely to provide liquidity. Menkveld and Zoican (2014) contains fast traders whose type determines whether one is a market maker or a liquidity taker; as the trading venue becomes saturated with fast traders, liquidity declines as market makers become more likely to trade against fast traders. Hoffmann (2014) and Menkveld and Zoican (2014) are not primarily concerned with the price discovery process and the asset values in the models are public knowledge. No agent in my model knows the true asset value and the model's price discovery process is fully endogenous. Moreover, the presence of multiple trading venues in my model allows for some novel analyses of spillovers between the venues.

2 Model

In this section, I specify a model based on Kyle (1985) and its extensions, most relevant being Foster and Viswanathan (1996). Basic details of the model, including the asset, agents and their actions, timing, and the information structure are described. Afterwards, I provide the necessary and sufficient conditions for a linear and symmetric Bayesian Perfect equilibrium.

Overview: The model is a game played in two stages with one asset traded in multiple trading venues. During the first stage ("entry stage"), the institutional traders, who trades on private information, choose the trading venues to enter. At the start of the second stage ("trading stage"), the institutional and retail traders at each venue simultaneously submit market orders to the venue's market maker. The market maker chooses a price based on the orderflows she observes, then clears the net orderflow (the sum of all orders) sent to her trading venue at the chosen price.

The "speed" of a market maker in my model captures the differential abilities of high and low frequency liquidity providers to integrate cross-venue information into prices. In particular, the market maker at a given trading venue more precisely observes the net orderflows submitted at the other venues as her speed increases (i.e. becomes "faster"). The prices at the trading venues with very fast market makers thereby closely reflect the orderflow information available across the trading venues, whereas the prices at the venues with slower market makers reflect noisier cross-venue information. As in Kyle (1985), each trading venue contains one competitive market maker who perfectly observes the net orderflow at her own trading venue.[12]

The institutional traders trade on private information in search of returns, and their trades can significantly affect market prices. Since my paper concerns trading in high frequency environments, the private information received by the institutional traders are in the form of short-lived (one-period) signals, and are therefore time-sensitive. I assume that the institutional traders' signals are sufficiently time-sensitive such that a signal is profitably exploited only when its recipient immediately trades on the signal. Thus, the institutional traders must choose the trading venues in which to trade *prior* to receiving their signals in order to extract information rents. Accordingly, the institutional traders receive their signals at the end of the entry stage, before trading begins.[13]

Asset and Trading Venues: There is a single asset with value v, which is unknown to the agents at the start of the game. Value v is distributed $\mathcal{N}(\mu, \sigma_v^2)$, and v becomes public knowledge after trading ends.[14] The distribution of v and the forms of all other distributions and the structure of the game are public knowledge from the start of the game. Market fragmentation is captured in the model by the presence of two venues, indexed by $h \in \{i, j\}$, wherein the asset is traded.

Traders, Market Makers and Timing: There are three kinds of market participants: $N \geq 2$ profit-maximizing and risk-neutral *institutional traders* (ITs) indexed by $k \in \{1, ..., N\}$; *retail traders* who trade randomly; and competitive and risk-neutral *market makers*. The game is partitioned into an initial stage, called the *entry stage*, and the subsequent stage, called the *trading stage*. The institutional traders are identical and have no private information in the entry stage. Only the ITs act in the entry stage, where each IT either exits or chooses a venue to enter.[15]

The trading stage begins once every institutional trader enters a trading venue and the number of

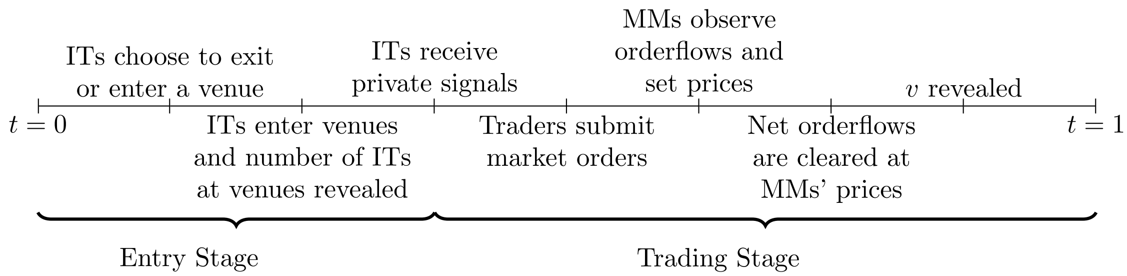
[12] By allowing every market maker to perfectly observe their own venue orderflows ensures that my results are driven by the ability of fast market makers to integrate cross-venue information rather than from being able to observe orderflow information in general.

[13] Timing the arrival of the signals in this way provides two major benefits: first, each subgame in the trading stage (corresponding to each possible realized distribution of the institutional traders among the trading venues) can be solved independently of the institutional traders' entry decisions in the entry stage, greatly simplifying the solution. Second, the solutions to the subgames in the trading stage are now easily comparable to the equilibria of several existing Kyle (1985)-type models. The number of institutional traders at each trading venue become public (as in Kyle (1985)) at the beginning of the trading stage for the same reasons.

[14] One can interpret μ as the realized value of the asset from the previous period in a model when the two-stage game of my model is repeated (Chowdhry and Nanda (1991)).

[15] There are two implicit assumptions being made about the institutional traders: (i) each IT can direct his market orders to particular trading venues; and (ii) the ITs would not trade in every trading venue. Because the ITs represent buy-side institutions, the first assumption is reasonable as many institutions would have direct or sponsored market access to the venues where they trade (rather than trading through a broker as in Cimon (2015)). For the second assumption, recall that many venues trade the same assets in modern markets. Therefore, at least for the most heavily traded assets, a strategy that evaluates every venue trading a given asset for possible order submission is likely infeasible.

Figure 1: Model Timing



ITs at each venue becomes public knowledge. During the trading stage, the kth IT receives a private signal, denoted by s_k. Signals $\{s_k\}_{k=1}^{N}$ are drawn from a symmetric multivariate normal distribution with mean zero, variance σ_s^2, and covariance $\sigma_{ss} > 0$.[16] Following Foster and Viswanathan (1996), the covariance between s_k and v is chosen such that $v = \mu + \sum_{k=1}^{N} s_k$.[17] After the ITs receive the signals, the retail traders and the ITs at each trading venue simultaneously submit market orders to the venue's market maker. Each IT's order reflects public knowledge and his signal while the retail traders receive no private information and trades randomly.[18]

The market orders sent by traders at each venue are aggregated and become the venue's net orderflow, which is cleared by the venue's market maker (MM). Every trading venue contains one MM who has no private information. The MM at a venue sets the venue's price competitively, after observing her own venue's net orderflow perfectly and the cross-venue net orderflow noisily. Competitive price setting implies that the price at each venue is set equal to the expected value of v conditional on the orderflow information observed by the venue's MM. Finally, the net orderflows are cleared and v is revealed. Figure 1 summarizes the timing in the game.

Institutional Traders' Actions: Every institutional trader performs one or two actions in the game: one action in the entry stage and up to one action in the trading stage. In the entry stage, each IT may choose to enter *either* of the two trading venues or exits the game. Then the (pure strategy) entry stage action set for the ITs is $\{\text{venue } i, \text{venue } j, \text{exit}\}$. The ITs may use mixed strategies in their entry decisions, and the probabilities that the kth IT assigns to $\{\text{venue } i, \text{venue } j, \text{exit}\}$ are denoted by the vector $\left\{n_k^i, n_k^j, 1 - n_k^i - n_k^j\right\}$. At the end of the entry stage, any IT that employed a pure strategy in his entry decision (and did not exit) enters his chosen venue. If the IT employed a mixed strategy, he realizes

[16]Negative σ_{ss} may be interpreted as, "each informed trader view[ing] that the other informed traders as collectively pulling the [...] price in the wrong direction" (Foster and Viswanathan, 1996, p.1459). Limiting σ_{ss} to non-negative values reduces the number of cases I need to consider in this paper's analyses. Moreover, it is unlikely that institutional traders would view each other as collectively trading in the wrong direction.

[17]This implies that $\sigma_v^2 = N\sigma_s^2 + N(N-1)\sigma_{ss}$ and $Cov(v, s_k) = \sigma_s^2 + (N-1)\sigma_{ss} = \sigma_v^2/N$, and ensures that the total information available among all the institutional traders is independent of N and the correlation between the signals (Foster and Viswanathan, 1996, Sec. VI).

[18]Chowdhry and Nanda (1991), Bernhardt and Hughson (1997) and Mendelson and Tunca (2003) provide Kyle (1985)-type models with endogenous liquidity traders in place of exogenous noise traders. I opted for the latter for tractability.

one of {venue i, venue j, exit} according to a single draw from the probabilities $\left\{n_k^i, n_k^j, 1-n_k^i-n_k^j\right\}$. The kth IT's location after the entry stage is denoted by $T(k) \in \{\varnothing, i, j\}$, where $T(k)=i$ indicates that the kth IT is at venue i and $T(k)=\varnothing$ indicates that the kth IT has exited the game.

Any kth institutional trader receives the private signal s_k at the beginning of the trading stage, but only those with $T(k) \neq \varnothing$ trades.[19] Then the pair $\{T(k), s_k\}$ is the kth IT's type in the trading stage. After $\{T(k), s_k\}_{k=1}^N$ is realized, the kth IT whose $T(k) \neq \varnothing$ submits a market order, $x_k \in \mathbb{R}$, at trading venue $T(k)$ to maximize his expected profit given s_k and the number of ITs at each trading venue.[20]

Retail Traders' Actions: The retail traders are not assigned to any specific trading venue, receive no information and trade randomly. The total orderflow submited by the retail traders at venue h is denoted by $u_h \overset{iid}{\sim} \mathcal{N}(0, \sigma_u^2)$, and all traders submit their orders simultaneously.

Market Makers' Actions and Speed: Each trading venue contains a competitive risk-neutral market maker who clears the venue's net orderflows at the price that she sets during the trading stage. Prior to setting the price, venue h's MM perfectly observes the net orderflow at venue h, but she observes the net orderflow at the other trading venue with noise, a_h. Define the set of institutional traders at trading venue i (and symmetrically for venue j) as $K_i = \{k | T(k) = i\}$. Then the net orderflow submitted at venue i is given as $y_i := \sum_{k \in K_i} x_k + u_i$, and the net orderflow at venue j observed by venue i's MM is $z_i := y_j + a_i$. Therefore, for $h \in \{i, j\}$, the price at venue h, p_h, is set by the venue's MM after she observes $\{y_h, z_h\}$. The speed of the MM at venue h, $S_h > 0$, reduces the magnitude of $a_h \overset{i.i.d.}{\sim} N(0, S_h^{-1})$ such that the MM at venue h observes the other venue's net ordeflow more precisely as S_h increases. Hence, where $S_h \to \infty$ the MM at venue h exactly observes both venues' net orderflows, while $S_h = 0$ implies that the MM only observes y_h. The speeds $\{S_h\}_{h \in \{i,j\}}$ are public knowledge.

Entry Stage Decisions: In the entry stage, each kth institutional trader chooses the probabilities $\{n_k^i, n_k^j, 1-n_k^i-n_k^j\}$ that he enters a trading venue or exits the game to maximize his expected profit. Denoting the kth IT's prior about $n_{l \neq k}^h$ by $n_{k,l}^h$, the kth IT's beliefs in the entry stage are summarized by the $(N-1)$-by-2 matrix $\mathcal{B}_{e,k} := \{n_{k,l}^h\}_{l \neq k, h \in \{i,j\}}$. Since, at the end of the game, the kth IT at venue h earns $v - p_h$ for each unit of asset he owns, the IT's expected profit to entering venue h is given by:

$$\Pi_{h,k} := \mathbb{E}[x_k(v - p_h) | \mathcal{B}_{e,k}, k \in K_h]. \tag{1}$$

[19]For simplicity, institutional traders with $T(k) = \varnothing$ also receive signals.

[20]As in Kyle (1985), neither the ITs or the MMs face a liquidity constraint. Liquidity constraints, while an important aspect of market participants, would distract from the analysis of market maker speed and market fragmentation.

The kth IT, for some x_k, thereby solves the following in the entry stage:

$$\max_{n_k^i, n_k^j} n_k^i \Pi_{i,k} + n_k^j \Pi_{j,k} \tag{2}$$
$$s.t. \quad n_k^i, n_k^j \in [0,1]\,,\; n_k^i + n_k^j \leq 1.$$

Institutional Traders' Trading Decisions: In the trading stage, each institutional trader chooses his order to maximize expected profits given his private signal and the number of ITs at each trading venue. The number of ITs at venue h is denoted by N_h, and the order that the kth IT believes the lth IT at the same venue (i.e. $T(k) = T(l) = h$) will submit is denoted $x_{k,l}$. Then the trading stage beliefs of the kth IT are summarized by the $(N-1)$-element vector $\mathcal{B}_{\tau,k} := \{x_{k,l}\}_{l \neq k}$, and each IT at venue h (i.e. $\forall k \in K_h$) solves the following in the trading stage:

$$\max_{x_k} \mathbb{E}[x_k(v - p_h)|N_i, N_j, s_k, \mathcal{B}_{\tau,k}, k \in K_h]\,,\; h \in \{i,j\}. \tag{3}$$

Market Makers' Price Setting: The market makers competitively set prices such that the MM at venue h earns zero expected profits given $\{y_h, z_h\}$. Then the MM at venue h sets p_h such that the following holds:

$$p_h = \mathbb{E}\left[v|y_h, z_h\right]\,,\; h \in \{i,j\}. \tag{4}$$

Equilibrium Type: As in Foster and Viswanathan (1996), I search for a linear and symmetric Bayesian Perfect equilibrium. Applied to the model, linearity implies that the price setting rule and the kth institutional trader's trading strategy are affine functions of $\{y_h, z_h\}_{h\in\{i,j\}}$ and s_k, respectively. Symmetry implies that, for any h and $k \neq l$, $n_k^h = n_l^h = n^{h*}$ in the entry stage, and $x_k = x_l$ whenever $T(k) = T(l)$ and $s_k = s_l$ in the trading stage. Hence, the equilibrium strategies at venue h in the trading stage are of the form:

$$p_h = \delta_h + \lambda_{y,h} y_h + \lambda_{z,h} z_h \tag{5}$$
$$x_k = \alpha_h + \beta_h s_k \tag{6}$$
$$\forall k \in K_h \text{ and } \forall h \in \{i,j\}.$$

The Bayesian Perfect equilibrium is solved backwards. The trading stage is solved for each possible pair $\{N_i, N_j\}$. Then the entry stage is solved given the solution to the trading stage. In any equilibrium, the beliefs $\{\mathcal{B}_{e,k}, \mathcal{B}_{\tau,k}\}_{k=1}^N$ must be consistent with public knowledge, and with every player adhering to his or her equilibrium strategy.

Off-Equilibrium Beliefs: Suppose the institutional traders' equilibrium strategies assign zero probability to exiting the game in the entry stage. Then any realization of $\{N_i, N_j\}$ such that $N_j \neq N - N_i$ constitutes an off-equilibrium path. However, every player's beliefs remain the same as in the equilibrium path since the ITs have no private information in the entry stage. If the ITs' equilibrium entry strategies assign nonzero probability to exiting the game, every possible realization of $\{N_i, N_j\}$ is on the equilibrium path. Similarly, every possible $\{y_h, z_h\}_{h\in\{i,j\}}$ is on the equilibrium path, and there is no off-equilibrium path that can arise during trading.

3 Equilibrium

In this section, I define a linear and symmetric Bayesian Perfect equilibrium, and solve the model for such equilibrium. First, a Bayesian Perfect equilibrium is defined, then the notion of equilibrium is restricted to the linear and symmetric case. After an equilibrium is defined, I solve the model backwards, beginning with the trading stage.[21]

Equilibrium Definition: A Bayesian Perfect equilibrium consists of entry strategies $\{n_k^h\}_{h\in\{i,j\}}$, trading strategies $\{x_k\}_{k\in\{1,\ldots,N\}}$, price setting rules $\{p_h\}_{h\in\{i,j\}}$ and beliefs $\{\mathcal{B}_{e,k}, \mathcal{B}_{\tau,k}\}_{k\in\{1,\ldots,N\}}$ such that: the beliefs are consistent with public knowledge and equilibrium strategies; the optimal entry condition (2) is satisfied for all institutional traders; and the price setting condition (4) and the optimal trading condition (3) are satisfied for all possible realizations of $\{N_i, N_j\}$ at all trading venues. Restricting the Bayesian Perfect equilibrium to the linear and symmetric case requires that $n_k^h = n^{h*}$ for all h and k, and that the trading stage strategies of each IT and market maker are in the form (5)-(6).

Definition 1 (Equilibrium)**.** *A linear and symmetric Bayesian Perfect equilibrium is a strategy profile* $\{n^{h*}, \delta_h, \alpha_h, \lambda_{y,h}, \lambda_{z,h}, \beta_h\}_{h\in\{i,j\}}$ *and beliefs* $\{\mathcal{B}_{e,k}\mathcal{B}_{\tau,k}\}_{k\in\{1,\ldots,N\}}$ *consistent with the strategy profile such that:*

1. *(Optimal Entry) For all* $k, l \in \{1, ..., N\}, k \neq l$, $\{n^{i*}, n^{j*}\}$ *solves (2) given* $n_k^i = n_l^i = n^{i*}$ *and* $n_k^j = n_l^j = n^{j*}$;

2. *(Optimal Trading) For all possible realization of* $\{N_i, N_j\}$, *for all* $h \in \{i, j\}$ *and* $k \in K_h$, $x_k = \alpha_h + \beta_h s_k$ *satisfies (3); and*

3. *(Price Setting) For all possible realization of* $\{N_i, N_j\}$ *and* $h \in \{i, j\}$, $p_h = \delta_h + \lambda_{y,h} y_h + \lambda_{z,h} z_h$ *satisfies (4).*

[21] All proofs are in the Appendix.

Institutional Traders' Participation: The institutional traders are initially assumed to exit the game with zero probability in the entry stage. This is subsequently shown to be optimal as the ITs expect to earn strictly positive profits in any equilibrium with full participation.

Trading Stage Solution: The method for solving the trading stage closely follows Kyle (1985) and Pasquariello and Vega (2007). In the trading stage, the kth institutional trader at venue h solves (3) given the beliefs $\mathcal{B}_{e,k}$ consistent with the linearity conditions (5)-(6). Substituting (5)-(6) into (3) yields a quadratic objective function, whose solution implies:[22]

$$\beta_i = \frac{1 + (N-1)\rho - N_j \rho \lambda_{z,i} \beta_j}{[2 + (N_i - 1)\rho]\lambda_{y,i}} \tag{7}$$

$$\alpha_i = \frac{\mu - \delta_i - N_j \lambda_{z,i} \alpha_j}{(N_i + 1)\lambda_{y,i}} \tag{8}$$

$$\lambda_{y,i} > 0. \tag{9}$$

Once the traders at venue h have submitted their orders, the venue's market maker sets the price according to the competitive price setting and linearity conditions, (4) and (5). These conditions imply that the price is determined by the linear regression of the asset value, v, on the MM's orderflow information, $\{y_h, z_h\}$:

$$\mathbb{E}[v|y_h, z_h] = p_h = \delta_h + \lambda_{y,h} y_h + \lambda_{z,h} z_h \; , \; h \in \{i, j\}. \tag{10}$$

Because the unconditional expected values of y_h and z_h are zero, δ_h must equal μ, the unconditional expected value of v. The terms $\lambda_{y,h}$ and $\lambda_{z,h}$, which represent the MM's sensitivity to her orderflow information, are regression coefficients. Denote the vector of the MM's orderflow sensitivities $\{\lambda_{y,h}, \lambda_{z,h}\}$ and $\{y_h, z_h\}$ by λ and $\mathbf{y}$, respectively. Then λ is obtained from the projection theorem:

$$\lambda = \mathbb{E}\left[v\mathbf{y}'\right]\left[\mathbb{E}(\mathbf{y}\mathbf{y}')\right]^{-1}. \tag{11}$$

The relations (7)-(9) and (11) are the necessary and sufficient conditions for a solution to any possible realization of $\{N_i, N_j\}$ in the trading stage.[23]

Entry Stage Solution: Assume that the kth institutional trader, at the beginning of the entry stage, believes that each other IT will assign probability n^h to entering venue h and will not exit the game. Then the conditions for an equilibrium in the trading stage, (7)-(9) and (11), imply that the kth IT's expected profit to entering venue h, $\Pi_{h,k}$, depends on n^h as well as the venue's speed, S_h.

[22] Derivations are symmetric for venue j.

[23] Except the case with $N_h = 0$, in which case β_h is undefined. Where $N_h = 0$, I set β_h equal to an arbitrary constant to ensure the equilibrium is well defined even in cases with $N_h = 0$, $h \in \{i, j\}$.

Denoting the correlation between the ITs' signals by $\rho := \sigma_{ss}/\sigma_s^2$:

$$\begin{aligned}\Pi_{h,k} &= \mathbb{E}\left[x_k(v-p_h)|k \in K_h\right] \\ &\propto \mathbb{E}\left[\left.\frac{1}{[2+(N-1)\rho](2-\rho)+[2+(N_h-1)\rho]\frac{1}{S_i\sigma_u^2}}\right| n^h\right]. \end{aligned} \tag{12}$$

Relation (12) depends on n^h only because the number of institutional traders who enters venue h, N_h, is expected to rise as n^h increases. Since N_h cannot be negative for any $n^h \in [0,1]$ and $S_h > 0$, $\Pi_{h,k}$ is strictly positive and the ITs never exit. Moreover, $\Pi_{i,k} < \Pi_{j,k}$ whenever $n^i = 1$, which implies that the ITs' equilibrium entry strategy is a mixed strategy. The optimal entry condition (2) is then reduced to an equality constraint:

$$\mathbb{E}\left[\Pi_{i,k} - \Pi_{j,k}|n^{i*}\right] = 0\,,\ \forall k \in \{1,...,N\} \tag{13}$$

There always exists n^{i*} in $(0,1)$ such that the equality constraint (13) is satisfied. Hence, (13) is the necessary and sufficient condition for a solution to the entry stage.

Theorem 1 (Equilibrium)**.** *There exists a linear and symmetric Bayesian Perfect equilibrium. For any $h \in \{i,j\}$, and for all possible realizations of $\{N_i, N_j\}$ and $k \in K_i$, the equilibrium strategy in the entry stage is given by $n_k^i = n^{i*} = 1 - n^{j*}$, and the equilibrium strategies in the trading stage at venue h are in the form (5)-(6). Define:*

$$A(S_i) := \frac{(1+(N-1)\rho)\left(2-\rho+\frac{1}{S_i\sigma_u^2}\right)}{(2+(N-1)\rho)(2-\rho)+(2+(N_i-1)\rho)\frac{1}{S_i\sigma_u^2}}. \tag{14}$$

Then, if $N_i, N_j > 0$, the equilibrium is characterized by:

$$n^{i*} = 1 - n^{j*}, \quad n^{i*} \in [0,1] \tag{15}$$

$$\sum_{m=0}^{N-1} \frac{\binom{N-1}{m}(n^{i*})^m(1-n^{i*})^{N-1-m}}{\sqrt{m+1}\left[(2+(N-1)\rho)(2-\rho)+(2+m\rho)\frac{1}{S_i\sigma_u^2}\right]} = \frac{2-\rho+\frac{1}{S_j\sigma_u^2}}{2-\rho+\frac{1}{S_i\sigma_u^2}} \sum_{q=0}^{N-1} \frac{\binom{N-1}{q}(1-n^{i*})^q(n^{i*})^{N-1-q}}{\sqrt{q+1}\left[(2+(N-1)\rho)(2-\rho)+(2+q\rho)\frac{1}{S_j\sigma_u^2}\right]} \tag{16}$$

$$\alpha_i = 0 \tag{17}$$

$$\delta_i = \mu \tag{18}$$

$$\beta_i = \sqrt{\frac{\sigma_u^2}{N_i\sigma_s^2}} \tag{19}$$

$$\lambda_{y,i} = A(S_i)\sqrt{\frac{N_i\sigma_s^2}{\sigma_u^2}} \tag{20}$$

$$\lambda_{z,i} = A(S_i)\frac{2-\rho}{2-\rho+\frac{1}{S_i\sigma_u^2}}\sqrt{\frac{N_j\sigma_s^2}{\sigma_u^2}}. \tag{21}$$

where the equilibrium trading strategies at venue j are symmetric. If $N_h = 0$ for some $h \in \{i,j\}$, the equilibrium is characterized by the same relations except β_h is set equal to an arbitrary constant. Further, n^{h*} that solves (16) always exists and is unique in the domain $(0,1)$.

The equilibrium is in closed form excluding condition (16) whose numerical solution is easily obtained. For any $h \in \{i,j\}$, the term β_h determines the size of an institutional trader's market order given his signal and measures how aggressively the ITs' trade on their signals. $\lambda_{y,h}$ is the slope of the price schedule offered by the market maker at venue h, and measures the sensitivity of p_h to venue h's net orderflow, y_h. Similarly, $\lambda_{z,h}$ measures the sensitivity of p_h to the net orderflow at the other trading venue as observed by MM at venue h, z_h.

To explain the economic meaning of $A(S_h)$, consider the kth institutional trader who is assigned to venue h in the trading stage. His price impact is given by $\lambda_{y,h}x_k = A(S_h)s_k$. Then $A(S_h)$ determines how intensely each IT at venue h impounds his signal into the price. Collectively, the ITs at venue h with average signal $\bar{s} := \frac{1}{N_h}\sum_{k\in K_h} s_k$ has total price impact equal to $N_hA(S_h)\bar{s}_h$. If $h = i$, the relationship between $N_hA(S_h)$ and the kth IT's expected profit in the trading stage is given by:

$$\mathbb{E}[\Pi_{i,k}|N_i] = \mathbb{E}[x_k(v-p_i)|N_i] = \beta_i\mathbb{E}\left[s_k\left(\sum_{l=1}^{N} s_l - N_iA(S_i)\left[\bar{s}_i + \frac{2-\rho}{2-\rho+\frac{1}{S_i\sigma_u^2}}\bar{s}_j\right]\right)\middle| N_i\right]. \tag{22}$$

Intuitively, higher $N_h A(S_h)$ implies that a larger proportion of the information held by venue h's institutional traders is revealed during trading, and the rents available for the ITs correspondingly fall. The term $N_h A(S_h)$ is thereby a measure of competition between the ITs at venue h.

Relationship to Existing Models: The trading stage of the model is closely related to the one-period case of Foster and Viswanathan (1996). The trading stage encompasses Foster and Viswanathan (1996) as follows. Suppose venue h's market maker is extremely slow ($S_h \to 0$). Then the MM cannot observe the orderflow at the other trading venue and the solution to the trading stage at venue h becomes equivalent to the equilibrium of the one-period Foster and Viswanathan (1996).[24] Now, instead consider the case in which $S_h > 0$ and every institutional trader is assigned to venue h ($N_h = N$). In this case, the solution to venue h's trading stage returns the exact one-period Foster and Viswanathan (1996) result.[25] Therefore, the absence of *either* multiple trading venues (i.e. market fragmentation) or speed reverts my model to an existing one, such that both elements of the model are necessary to generate the results in the paper.

Equilibrium Properties: A series of propositions describe the properties of Theorem 1 that are important to the mechanisms underlying my results. The first proposition is vital to generating spillovers from venue i to venue j as S_i changes:

Proposition 1 (Entry Substitution). *n^{i*} is strictly decreasing in S_i and strictly increasing in S_j, $i \neq j$.*

Proposition 1 states that fewer institutional traders enter a trading venue when its speed increases, while more ITs enter the venue as the *other* venue's speed increases. The intuition is as follows: suppose S_i rises but the entry probabilities remain unchanged. Then the expected number of ITs entering venue i is unaffected, and the ITs are expected to compete at venue i as intensely as before. Consequently, the informativeness of y_i is unchanged in expectation, although the increase in S_i improves the informativeness of z_i. Venue i's market maker then receives strictly superior information than she would prior to the increase in S_i, and her prices become more informative. This implies smaller information rents for the ITs at venue i, and they compensate by substituting away from entering venue i.

This intuition can be shown mathematically. Let $I(y_i)$ denote the proportion of the variance in venue i's net orderflow, y_i, attributable to the institutional traders' orders, $I(y_i) := \frac{Var\left(\sum_{k\in K_i} x_k\right)}{Var(y_i)}$. Since the only orders in y_i that are informative to the market makers are those sent by the ITs,

[24] Specifically, $\lambda_{z,h} = 0$ while $\lambda_{y,h}$ becomes equal to λ in Pasquariello and Vega (2007) scaled by a constant that depends on N and N_h.

[25] When $N_h = N$ and the institutional traders' signals are perfectly correlated ($\rho = 1$, which implies every IT knows v), the solution to the trading stage at venue h becomes identical to the equilibrium in Foster and Viswanathan (1993). The solution to venue h's trading stage recovers the original Kyle (1985) equilibrium if, in addition, there is a single IT ($N = N_h = 1$).

$I(y_i)$ measures the *informativeness* of y_i. Analogously, $I(z_i)$ is the informativeness of the orderflow information at venue j observable by venue i's MM, z_i. It is simple to show that:

$$I(y_i) = \frac{1 + (N_i - 1)\rho}{2 + (N_i - 1)\rho} \tag{23}$$

$$I(z_i) = \frac{1 + (N - N_i - 1)\rho}{2 + (N - N_i - 1)\rho + \frac{1}{S_i \sigma_u^2}}. \tag{24}$$

If S_i increases and n^{i*} remains constant, the distributions of N_i and $N_j = N - N_i$ do not change and $\mathbb{E}[I(y_i)]$ is unaffected while $\mathbb{E}[I(z_i)]$ increases. Thus, venue i's institutional traders choose a smaller n^{i*} to reduce the impact of the increase in S_i on the informativeness of their orderflow.

The informativeness of venue i's orderflow, $I(y_i)$, is independent of S_i since the venue's market maker always observes y_i perfectly. On the other hand, $I(z_i)$ is strictly increasing in S_i as the MM observes the other venue's orderflow more precisely. When N_i increases, competition among the institutional traders at venue i intensifies and $I(y_i)$ improves. Because a larger N_i means smaller N_j, $I(z_i)$ declines as N_i increases.

Proposition 2 (Entry and Orderflow Information). *For any $h \in \{i, j\}$ and given $\rho > 0$:*

- *$I(y_h)$ is independent of S_h and strictly increasing in N_h; and*
- *$I(z_h)$ is strictly increasing in S_h and strictly decreasing in N_h.*

The relations (23)-(24) demonstrate the intuition to another important property of the equilibrium: if the trading venues have the same speed (i.e. $S_i = S_j$), the total information at venue i, $I(y_i) + I(z_i)$, is greater than $I(y_j) + I(z_j)$ whenever $N_i > N_j$. Hence, the institutional traders will be indifferent between the venues only if the ITs assign equal probability to entering either venue ($n^{i*} = \frac{1}{2}$) so that the distributions of N_i and $N_j = N - N_i$ are identical.

Proposition 3 (Equal Entry). *Given $S_i = S_j$ and $\rho > 0$:*

- *$I(y_i) + I(z_i) > I(y_j) + I(z_j)$ if and only if $N_i > N_j$; and*
- *$n^{i*} = n^{j*} = \frac{1}{2}$.*

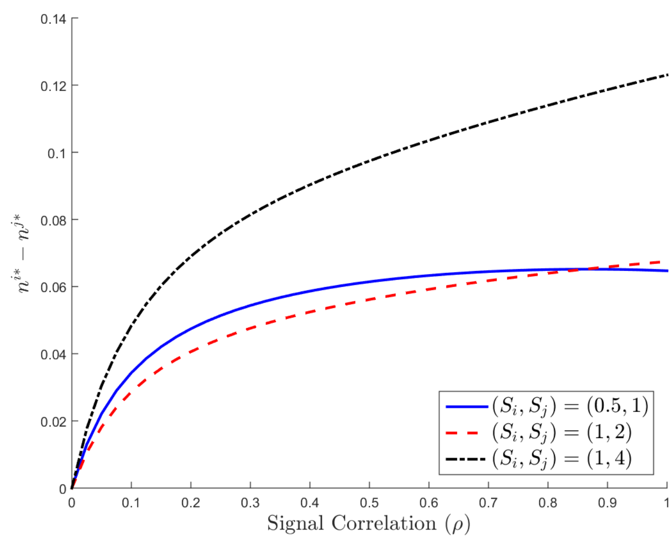
One mechanism causing competition among the institutional traders is from Foster and Viswanathan (1996): positive correlation between the ITs' signals ($\rho > 0$) means that a part of the value in the kth IT's signal is shared with the other ITs. Therefore, trading by the other ITs reduces the profit the kth IT can earn on his signal, s_k, which leads him to trade more intensely as N_i increases. Note that $N_i A(S_i)$ is increasing in $\rho \geq 0$, which implies that $I(y_i)$ and $I(z_i)$ are increasing in ρ as more private information is impounded into the price.

Proposition 4 (Competition and Signal Correlation)**.** *For any $h \in \{i, j\}$, given $\rho \geq 0$ and $N_h < N$, the terms $N_h A(S_h)$, $I(y_h)$ and $I(z_h)$ are strictly increasing in ρ.*

Because the speed of venue i affects $I(z_i)$ and $N_i A(S_i)$, interactions between S_i and ρ exist. When ρ is higher, there is greater overlap between the ITs' signals and $N_i A(S_i)$ is greater. As venue i's speed increases, the effect of ρ on $N_i A(S_i)$ is amplified as faster market makers more precisely observe their cross-venue orderflows. Then competition between the ITs at venue i intensifies further and reduces information rents available to the venue's ITs; thereby reduing the ITs' total profit. Hence, the ITs' substitute away from the venue with increased speed by a larger magnitude where ρ is higher.

Numeric Result 1 (Signal Correlation and Speed)**.** *For any $h \in \{i, j\}$ and given $\rho \geq 0$, n^{h*} decreases more steeply as S_h increases when ρ is larger.*

Figure 2: Signal Correlation and Speed



Numeric solutions to $n^{i*} - n^{j*}$ over ρ as determined by (16) and for cases $\{S_i, S_j\} \in \{\{0.5, 1\}, \{1, 2\}, \{1, 4\}\}$. Venue i is the slower venue in each case and N, σ_s^2 and σ_u^2 are set to 15, 1 and 1, respectively.

Figure 2 demonstrates numeric result 1. The figure shows that the entry probabilities to entering either venue are the same ($n^{i*} = \frac{1}{2}$) when there is zero correlation ($\rho = 0$) between the institutional traders' signals. Without correlation, each IT is a monopolist over his own signal, and the ITs do not compete over information. Consequently, the $N_i A(S_i)$ and the informativeness of orderflows ($\{I(y_i), I(z_i)\}$) no longer depend on the number of ITs (N) or their venue assignments (N_i):

$$A(S_i|\rho = 0) = I(y_i|\rho = 0) = \frac{1}{2}$$
$$I(z_i|\rho = 0) = \frac{1}{2 + \frac{1}{S_i \sigma_u^2}}.$$

Despite $I(z_i)$ remaining dependent on S_i, the institutional traders' entry probabilities are unimpacted by S_i because z_i is orthogonal to y_i as $\rho = 0$. Thus, the speeds at the trading venues cannot influence the ITs' entry probabilities (n^{i*}) when $\rho = 0$. This leaves the retail traders' orderflows (u_i) as the sole determinant of n^{i*}.

As in Foster and Viswanathan (1996), the orderflow sent by the retail traders disguises the institutional traders' orders so that the market makers cannot invert orderflows to obtain the aggregate signal. Where N_h is large, there is less retail trading with which each IT at venue h could disguise his orders and the IT thereby trade less aggressively (smaller β_h). Conversely, $\beta_h = \sqrt{\frac{\sigma_u^2}{N_h \sigma_s^2}}$ is increasing as the expected magnitude of u_h *per* IT, given by $\frac{\mathbb{E}[|u_h|]}{N_h} = \sqrt{\frac{2}{\pi}} \frac{\sigma_u}{N_h}$, increases, which occurs whenever $\frac{\sigma_u^2}{N_h}$ increases. Returning to n^{h*}, when $\rho = 0$, $\mathbb{E}\left[\frac{\sigma_u^2}{N_h}\right]$ solely determines the ITs' expected profits such that an equilibrium is attained only if $n^{i*} = n^{j*} = \frac{1}{2}$:

Proposition 5 (Entry and Uncorrelated Signals). *If $\rho = 0$, then $n^{i*} = n^{j*} = \frac{1}{2}$.*

Entry Stage Tradeoffs: In summary, the institutional traders' face trade offs along three dimensions in the entry stage:

(i) Speeds at the trading venues, S_i and S_j;

(ii) Correlation between the ITs' signals, ρ; and

(iii) Magnitude of retail trading as determined by σ_u^2.

Across dimension (i), the institutional traders are more likely to enter the slower venue (proposition 1). Across dimension (ii), n^{i*} becomes more sensitive to the difference between S_i and S_j as ρ increases (proposition 1). Lastly, across dimension (iii), $\mathbb{E}\left[\frac{\sigma_u^2}{N_h}\right]$ is decreasing in n^{h*} so that n^{h*} is close to $\frac{1}{2}$ whenever ρ is low.

Trading Stage Properties: Liquidity at venue h is determined by the sensitivity of the venue's market maker to the venue's net orderflow, $\lambda_{y,h}$. As S_h increases, the MM's cross-venue orderflow becomes more informative (i.e. $I(z_i)$ increases) and the MM relies less on y_h to determine her price. Thus, $\lambda_{y,h}$ declines while $\lambda_{z,h}$ increases. The liquidity at venue h therefore is increasing in S_h.

Proposition 6 (Price Sensitivity). *For any $h \in \{i, j\}$ and given $\rho > 0$, $\lambda_{z,h}$ is increasing in S_h and $\lambda_{y,h}$ is decreasing in S_h.*

Further, the trading stage solution at a given venue, characterized by (14) and (17)-(21), is unaffected by the other venue's speed conditional on $\{N_i, N_j\}$:

Proposition 7 (Speed Irrelevance). *Given $\{N_i, N_j\}$, the solution to the trading stage at venue i (venue j) is independent of S_j (S_i).*

Proposition 7 implies that the effect of speed is loaded on the entry stage. For the intuition, recall that each market maker observes the net orderflow at her trading venue perfectly while the MM's speed allows her to observe cross-venue orderflows noisily. As a result, any information the MM at venue j obtains due to S_j is already available at venue i with greater precision. Consider $I(y_i)$ and $I(z_j)$: it is straightforward to show that $I(y_i) \geq I(z_j)$. Therefore, z_j is superfluous for venue i's MM and, because venue i's ITs care only about p_i, S_j does not affect the venue i's trading strategies given $\{N_i, N_j\}$.

4 Comparative Statics

In this section, I examine two counterfactuals: (i) an increase in speed at all trading venues ("aggregate shock"); and (ii) an increase to speed at a given venue ("local shock"). The aggregate shock may be interpreted as a regulatory or technological change that improves the ability of all market makers to incorporate cross-venue information into their quotes. The spread of colocation services is an example: majority of colocation services are purchased by market makers and, by shortening the physical distance to trading venues' servers, increase the speeds of the market makers (Brogaard et al. (2015)). A regulatory example is the establishment of the Regulation National Market System (Ref NMS) in the US. Reg NMS encouraged trading venues to automate, thereby increasing high frequecy traders' access to those venues (Hendershott and Moulton (2011)).

The local shock may represent, for example, latency update to the trading venue's server or a reduction in trading message fees.[26] Other relevant examples are speed bumps[27] which trading venues may impose on all or specific types of orders. TSX Alpha and Aequitas NEO are trading venues that impose speed bumps against market orders (TSXA (2015); Schmitt (2015)), thereby lengthening the time market makers have to incorporate cross-venue information into their quotes. IEX is a venue that places speed bumps on all orders with the aim of reducing participation by high frequency traders (Picardo (2014)). The introduction of IEX-type speed bump would then correspond to a decrease in speed as HFTs tend to act as market makers (Hagströmer and Norden (2013); Malinova and Park (2015)).

The remainder of the section analyses liquidity, price discovery and the institutional traders' profits for each counterfactual. The ITs' profits provide insight into the allocative aspects of high frequency market making; specifically, there is an ongoing debate regarding the impact of high frequency trading on buy-side institutions and retail investors.[28] I begin the section by defining market quality measures.

[26] Conrad, Wahal, and Xiang (2015), Hagströmer and Norden (2013) and Malinova, Park, and Riordan (2013) argues that algorithmic market making is disproportionately affected by message-based fees.

[27] Processing delay on orders.

[28] For instance, see the comments submitted to the US Securities and Exchange Commission regarding the application by IEX to become an exchange (SEC (2015)).

Then I examine the aggregate shock, followed by the local shock's effect on the shocked venue. The section ends with the analysis of the local shock's effect on the *non*-shocked venue ("spillovers").

Market Quality Measures: Each measure is an unconditional expectation computed at the beginning of the game (i.e. $t = 0$). Liquidity at venue h, as in Kyle (1985), is measured by the inverse of the sensitivity of the venue's price to the venue's net orderflow, $L_h := \mathbb{E}[\lambda_{y,h}^{-1}]$. When p_h is sensitive to y_i, price impact is greater such that $\mathbb{E}[\lambda_{y,h}^{-1}]$ is the inverse of trading costs at venue h.

The price discovery measure should reflect the information content of the prices. One such measure is the proportion of variation in the true asset value explained by the price. If p_h explains most of the variation in v, price discovery at venue h is high as one can generate an accurate estimate of v using only p_h. Denote the price discovery measure for venue h by PD_h:

$$PD_h := 1 - \mathbb{E}\left[\frac{Var(v|p_h)}{\sigma_v^2}\right].$$

Applying the law of total variance[29] gives $\mathbb{E}[Var(v|p_h)] = \sigma_v^2 - \mathbb{E}[Var(p_h)]$ and:[30]

$$PD_h = \frac{\mathbb{E}[Var(p_h)]}{\sigma_v^2}. \tag{25}$$

To analyse allocative impact, I measure the institutional traders' *total* expected profit at each venue. Calculating the ITs' individual expected profit from the total measure is trivial and the total measure has the advantage that, due to competitive price setting, it is equal to the retail traders' total expected losses. Theorem 1 and (22) together imply:

$$PRFT_h := \mathbb{E}[\Pi_{i,k}] = \mathbb{E}[A(S_h)\sqrt{N_h}]\sigma_u\sigma_s. \tag{26}$$

Table 1 summarizes the market quality measures.

Table 1: Market Quality Measures

Measure	**Definition**
Liquidity	$L_h := \mathbb{E}[\lambda_{y,h}^{-1}]$
Price Discovery	$PD_h := \frac{\mathbb{E}[Var(p_h)]}{\sigma_v^2}$
IT Profit	$PRFT_h := \mathbb{E}[A(S_h)\sqrt{N_h}]\sigma_u\sigma_s$

[29] Law of total variance and the relation (5) implies $Var(v) = \mathbb{E}[Var(v|p)] + Var(\mathbb{E}[v|p]) = \mathbb{E}[Var(v|p)] + Var(p)$.

[30] $PD_h = \frac{[1+(N-1)\rho]\left[D(2-\rho)^2 + C(2-\rho)\frac{1}{S_h\sigma_u^2} + \frac{N_h}{N}B\left(\frac{1}{S_h\sigma_u^2}\right)^2\right]}{\left[D(2-\rho)+B\frac{1}{S_h\sigma_u^2}\right]^2}$, where $D := 2 + (N-1)\rho$, $B := 2 + (N_h - 1)\rho$ and $C := \frac{N_h}{N}D + \frac{1}{N}[2 + (N_h(N - N_h) - 1)\rho]$.

4.1 Aggregate Shock

This subsection presents the market quality effects of increasing speeds at all trading venues (i.e. the aggregate shock). The analytical result relies on imposing an equality constraint on the venues' speeds which, by proposition 3, fixes the institutional traders' entry probabilities to $n^{i*} = n^{j*} = \frac{1}{2}$. Then the distributions of $\{N_i, N_j\}$ are independent of the venues' speeds and the aggregate shock's impact on each market quality measure has the same sign whether the measure are unconditioned or conditioned on $\{N_i, N_j\}$. I exploit this to prove proposition 8.

Proposition 8 (Aggregate Shock and Market Quality). *Suppose $S_i = S_j = S > 0$ and $N_i, N_j > 0$. Then:*

- *L_i and L_j are increasing in S;*
- *PD_i and PD_j are increasing in S; and*
- *$PRFT_i$ and $PRFT_j$ are decreasing in S.*

The aggregate shock improves liquidity and price discovery but decreases the institutional traders' total expected profit at every trading venue. The intuition is as follows: by proposition 2, the increase in S_h, given N_h, does not impact the informativeness of venue h's net orderflow ($I(y_h)$) whereas the informativeness of its cross-venue information ($I(z_h)$) increases, implying that orderflows observed by the market makers are strictly more informative. Moreover, by proposition 6, greater $I(z_h)$ causes the MM at venue h to rely more on z_h in price setting, which reduces her sensitivity to y_h (i.e. $\lambda_{y,i}$ decreases). The former effect leads to increased price discovery while the latter effect improves liquidity as the venues' speeds increase, and indicate reduced information rents and trading costs, respectively. The ITs thereby earn smaller profits even as retail traders benefit from lower trading costs.

Empirical Evidence: Findings from large, multi-venue data are consistent with proposition 8. Conrad, Wahal, and Xiang (2015) and Boehmer, Fong, and Wu (2014) find that high frequency trading improves liquidity and price discovery in equities markets. Malinova, Park, and Riordan (2013) and Tong (2015) observe that HFT activity and institutional traders' returns are negatively related.

Numerical Results: Proposition 8 imposes an equality constraint the speeds of the trading venues. For the numerical results, the equality constraint is relaxed to test whether the findings in proposition 8 hold in cases where the venues' have unequal speeds. The numerical methodology is specified in procedure 1.[31]

[31] In procedure 1, the liquidity measure for venue h is undefined whenever realized $N_h = 0$. Accordingly, I set $N_{irq} = 1$ (and $N_{jrq} = N - 1$) whenever $N_{irq} = 0$ is drawn.

Procedure 1: Aggregate Shock Numerical Methodology

1. Choose vectors of parameters $\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\}$, initial speeds $\{S_{i,initial}, S_{j,initial}\}$ and aggregate shocks $\{\omega_r\}_{r=1}^R$;
2. For each $r \in \{1, ..., R\}$, set $\{S_i, S_j\} = \{S_{i,initial} + \omega_r, S_{j,initial} + \omega_r\}$ and use (16) to compute n^{i*}, then stored the output as n_r^{i*};
3. For each $r \in \{1, ..., R\}$ and each $q \in \{1, ..., Q\}$, take N draws from the binomial distribution with probability n_r^{i*}, then store the sum of the output as N_{irq} and set $N_{jrq} = N - N_{irq}$;
4. For each $\{N_{irq}, N_{jrq}\}_{\substack{r\in\{1,\dots,R\}\\ h\in\{1,\dots,H\}}}$, compute $\{PD_{hrq}, L_{hrq}, PRFT_{hrq}\}_{h\in\{i,j\}}$; and
5. For each $r \in \{1, ..., R\}$ and $h \in \{i, j\}$, compute $\overline{PD}_{hr} := \frac{1}{q}\sum_{q=1}^{Q} PD_{hrq}$, $\overline{L}_{hr} := \frac{1}{q}\sum_{q=1}^{Q} L_{hrq}$ and $\overline{PRFT}_{hr} := \frac{1}{q}\sum_{q=1}^{Q} PRFT_{hrq}$.

Parameter values are $\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\} = \{15, 0, 1, 0.5, 1\}$.[32] Initial speeds are $\{S_{i,initial}, S_{j,initial}\} \in \{\{0.5, 1\}, \{1, 2\}, \{1, 4\}\}$ as in figure 2 where venue i is always the slower venue. Aggregate shocks, $\{\omega_r\}_{r\in\{1,\dots,R\}}$, range from 0 to 5 in increments of 0.1 so that $R = 51$, and $Q = 20,000$. For notational convenience, $\overline{PD}_{ir}$, $\overline{L}_{ir}$ and $\overline{PRFT}_{ir}$ are written as PD_i, L_i and $PRFT_i$, respectively.

Figure 3 shows that liquidity and price discovery are monotonically increasing, and the institutional traders' total expected profit is monotonically decreasing at the slower venue (venue i) as the speeds of all trading venues increase. The same holds for the faster venue (venue j) except in the case where the initial difference in speed is very large. The figure suggests that proposition 8 is robust to relaxing the equality constraint on the venues' speeds.

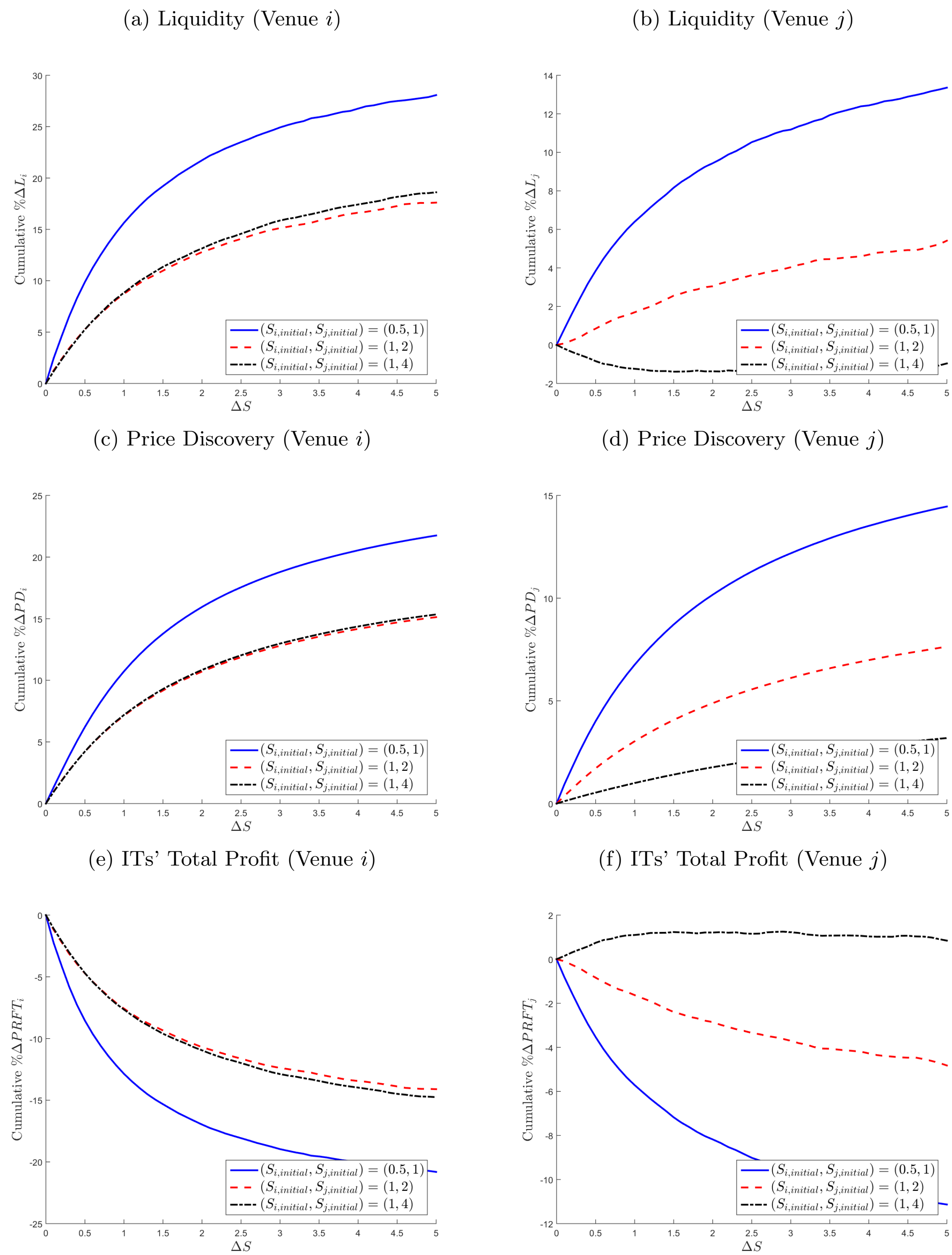
4.2 Direct Effect of the Local Shock

This subsection presents the effect of increasing the speed at a particular venue (i.e. the local shock) on that venue's market quality ("direct effects" of the local shock). The intuition underlying the direct effects of the local shock is close to the intuition behind the effects of the aggregate shock, except that the institutional traders substitute away from the shocked venue (proposition 1). To analyse the direct effect, a numerical methodology similar to procedure 1 is used. Corresponding results, with venue i as the shocked venue, are provided in figure 4.

Figure 4 shows that the local shock's direct effects are qualitatively identical to the aggregate shock's effect as per proposition 8; liquidity and price discovery improve while the institutional traders' total expected profit decreases. The results suggest that impact of the change in $\{N_i, N_j\}$ due to the local

[32] Parameters imply that $\sigma_v^2 = 120$. For all numerical analyses, parameterizations using every combination of $N = \{5, 15, 30\}$, $\sigma_s^2 = \{.5, 1, 5\}$, $\rho = \{.01, .5, .99\}$ and $\sigma_u^2 = \{.5, 1, 5\}$ were also used and found to generate similar results. Figures using alternative parameterizations are available upon request.

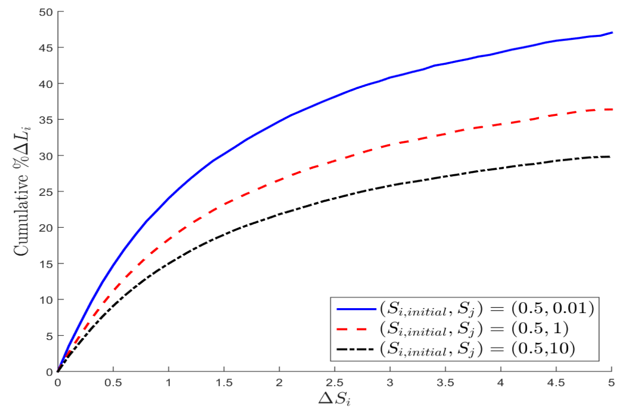
Figure 3: Effects of Aggregate Shock with Unequal Speeds



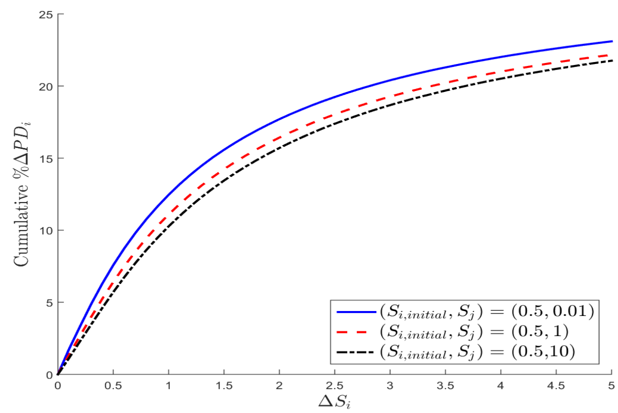
Market quality measures as cumulative percent change computed where the value at $\Delta S_i = 0$ as the base. Parameter values are $\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\} = \{15, 0, 1, 0.5, 1\}$ and the initial speeds are $\{S_{i,initial}, S_{j,initial}\} \in \{\{0.5, 1\}, \{1, 2\}, \{1, 4\}\}$. The vertical axes are in percentages, venue i is the slower market across all specifications and the results are averaged across $Q = 20,000$ draws.

Figure 4: Direct Effects of the Local Shock

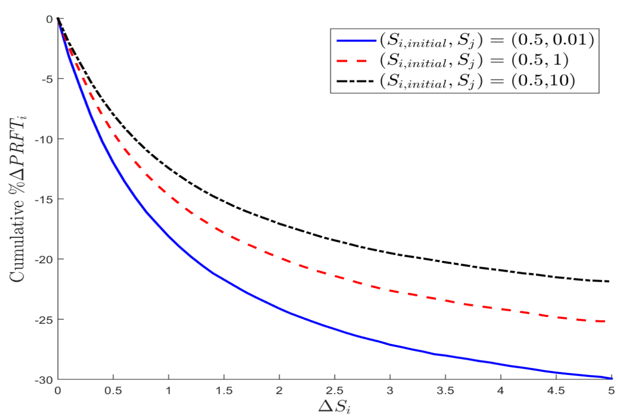
(a) Liquidity



(b) Price Discovery

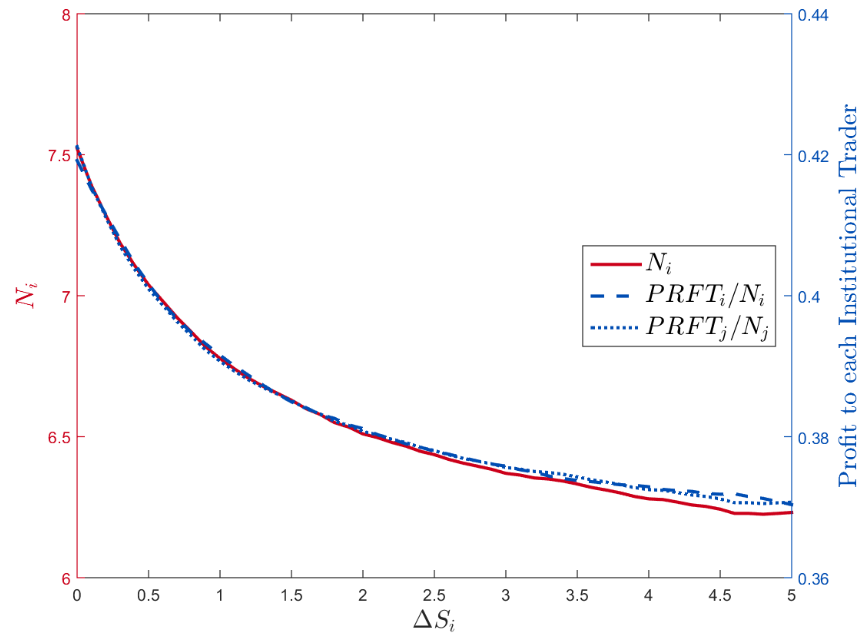


(c) ITs' Total Profit



Market quality measures as cumulative percent change computed where the value at $\Delta S_i = 0$ as the base. Parameter values are $\left\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\right\} = \{15, 0, 1, 0.5, 1\}$, the initial speed at the shocked venue (venue i) is $S_{i,initial} = 0.5$ while venue j's speeds are set equal to $S_j \in \{0.01, 1, 10\}$. The vertical axes are in percentages and results are averaged across $Q = 20,000$ draws.

Figure 5: Individual Institutional Trader Profit



Expected profits for each institutional trader at each venue and N_1.[33] Individual IT profit is computed by $PRFT_i/N_i$. Local shocks only affect the speed of venue 1, and the initial speed at venue 1 equals 0.25. The speed at venue 2 is set to 1. Parameter values are $\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\} = \{15, 0, 1, 0.5, 1\}$. $H = 20,000$ draws are made for each numerically computed n^{i*}. Values in the graphs are averages across the H draws.

shock are second order in the direct effects. Intuitively, this is due to the trade off faced by the ITs as S_i increases: given n^{i*}, the ITs' expected profits shrink when S_i increases; yet, if the ITs respond by reducing n^{i*} and raising n^{j*}, the competition between ITs at venue j is expected to intensify and the ITs' expected profits *still* decrease. The ITs therefore do not substitute away from venue i sufficiently to reduce the venue's liquidity or price discovery on net as S_i increases. Figure 5 illustrates this dynamic.

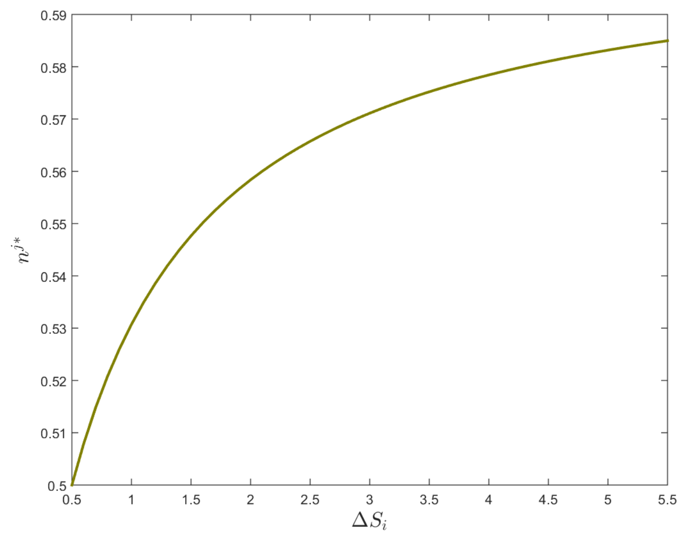
Empirical Evidence: My results are consistent with the large literature on high frequency traders which suggests that HFTs improve liquidity and price discovery. Papers in this literature exploit technical upgrades at trading venues (Hendershott et al. (2011); Riordan and Storkenmaier (2012); Hendershott and Riordan (2013)), introduction of colocation (Brogaard et al. (2015); Boehmer et al. (2014)), datasets that identify HFTs' transactions (Carrion (2013); Brogaard et al. (2014); Hasbrouck and Saar (2013)), and other strategies[34] to establish that HFT activity improves liquidity and price discovery. The allocative aspect of the results are consistent with Malinova et al. (2013) and Tong (2015) which finds that HFTs decrease institutional traders' returns.[35]

[33] Notice that the Expected profits for individual institutional traders are equal across the trading venues.

[34] E.g. Menkveld (2013) examines the entry of a high frequency market maker into an European venue.

[35] In contrast, Brogaard et al. (2014) detects no statistically insignificant effect from HFTs on the institutional traders' returns.

Figure 6: Equilibrium Entry Probability into Venue 2



Probability that an institutional trader enters venue j (i.e. n^{j*}) as the speed at venue i changes. Initial speeds are $S_{i,initial} = S_{j,initial} = 1$ and parameter values are $\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\} = \{15, 0, 1, 0.5, 1\}$.

4.3 Spillovers from the Local Shock

This subsection presents the effects of increasing the speed at a given venue on the venue whose speed has *not* changed ("spillovers" from the local shock). By proposition 7, the local shock on venue i does not affect venue j if $\{N_i, N_j\}$ are held constant. The spillovers from the local shock must therefore be entirely driven by the change in the distributions of $\{N_i, N_j\}$ as institutional traders substitute away from venue i. Figure 6 shows the rise in n^{j*} as S_i increases.

As more institutional traders enter venue j, two opposing effects occur: (i) competition between the ITs intensifies at venue j and the venue's ITs trade more aggressively ("entry effect"); and (ii) ITs' competition at venue i declines and venue j's cross-venue information (i.e. z_j) becomes less informative. The first effect reduces the market maker's sensitivity to y_j at venue j (i.e. $\lambda_{y,j}$ decreases) since y_j becomes more sensitive to the signals held by the venue's ITs. Thus, both liquidity and price discovery improve, and the ITs' total profit decreases, at venue j. The second effect reduces the information available to venue j's MM and worsens the venue's price discovery and increases its ITs' total profit. Furthermore, the MM increases her reliance on y_j as the informativeness of z_j deteriorates, raising $\lambda_{y,j}$ and decreasing venue j's liquidity. Because cross-venue information is more useful to the venues with faster MMs, the second effect is dominates at sufficiently fast venues whereas the first effect dominates at slower venues. Proposition 9 captures this intuition:

Proposition 9 (Spillovers from the Local Shock). *Define:*

$$S_j^* := \frac{(N_j+1)\rho - 2}{[2+(N-1)\rho](2-\rho)\sigma_u^2}. \tag{27}$$

Then $\mathbb{E}[L_j|N_j]$ *is strictly decreasing in* S_i *and* $\mathbb{E}[PRFT_j|N_j]$ *is strictly increasing in* S_i *whenever* $S_j > S_j^*$*, and* $\mathbb{E}[L_j|N_j]$ *is strictly increasing in* S_i *and* $\mathbb{E}[PRFT_j|N_j]$ *is strictly decreasing in* S_i *whenever* $S_j < S_j^*$*.*

Proposition 9 states that, conditional on N_j, the spillovers from the local shock reduces liquidity and increases ITs' total profit at venue j whenever the venue is sufficiently fast, but otherwise improves the venue's liquidity and decreases ITs' total profit. However, the distribution of N_j shifts right as S_i increases. Numerical results in figure 7 show that the intuition behind proposition 9 holds nonetheless.

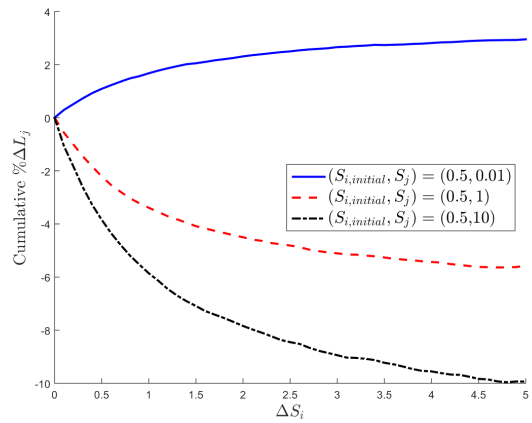
Empirical Implications and Evidence: The spillovers from changes to the speed at a trading venue on *other* venues provide some novel empirical implications. Specifically, when a given trading venue's speed increases, other fast (slow) venues will experience:

- More (more) institutional (i.e. informed) trading;
- Less (more) price discovery;
- Less (more) liquidity;
- Higher (lower) institutional traders' total returns; and
- Lower (higher) retail traders' total returns.

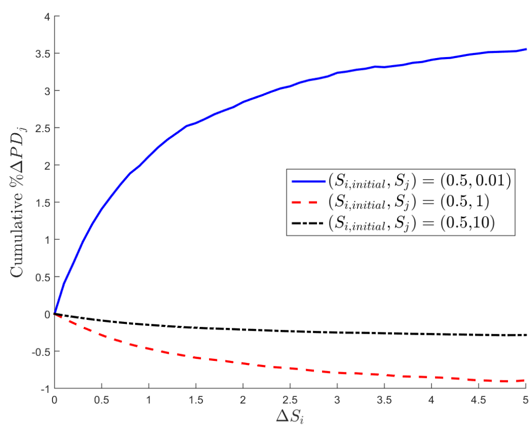
Empirical Evidence: Chen, Sean, Michael, and Thomas (2016) examines spillovers between trading venues and finds results consistent with the model's empirical implications. Chen et al. (2016) analyses the 2015 introduction of a speed bump on *marketable* orders at TSX Alpha, a Canadian exchange, and concludes that: (i) informed orderflows decreased at TSX Alpha but increased at other Canadian venues; and (ii) liquidity at the other venues declined. TSX Alpha's speed bump would improve the ability of its market makers to revise stale quotes in response to cross-venue information since only liquidity taking orders are affected; therefore, in the language of my model, the speed bump is equivalent to an increase in the speed at TSX Alpha. Then conclusion (i) in Chen et al. (2016) is consistent with my results. Moreover, Malinova and Park (2015) shows that high frequency traders provide the majority of liquidity at Canadian exchanges, which suggests that market makers at Canadian trading venues are very fast. Conclusion (ii) in Chen et al. (2016) may thereby be consistent with my results.

Figure 7: Spillovers from the Local Shock

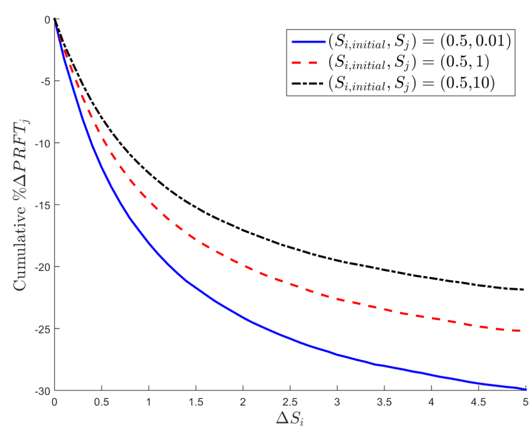
(a) Liquidity



(b) Price Discovery



(c) ITs' Total Profit



Market quality measures at venue j as cumulative percent change computed where the value at $\Delta S_i = 0$ as the base. Parameter values are $\{N, \mu, \sigma_s^2, \rho, \sigma_u^2\} = \{15, 0, 1, 0.5, 1\}$, the initial speed at the shocked venue (venue i) is $S_{i,initial} = 0.5$ while venue j's speeds are set equal to $S_j \in \{0.01, 1, 10\}$. The vertical axes are in percentages and results are averaged across $Q = 20,000$ draws.

5 Conclusions

I conclude with a discussion of my model's limitations as well as a its potential implications for market design and regulatory policy.

Limitations: The model captures a few aspects of the interaction between market fragmentation and high frequency trading that arise due to high frequency market makers. Thus, this paper complements existing research on high frequency traders that employ liquidity taking strategies, including latency arbitrage (e.g. Foucault et al. (2014) and Budish et al. (2013)). The model only considers short-lived and time-sensitive private information and does not explicitly feature arbitrage. This paper should then be viewed as developing a theory of how prices evolve at high frequencies wherein arbitrage opportunities are common (Chaboud et al. (2014)). In addition, retail traders and information acquisition are exogenous in the model. Endogenous liquidity trading as in Chowdhry and Nanda (1991), Bernhardt and Hughson (1997) or Mendelson and Tunca (2003), and endogenous information acquisition as in Baldauf and Mollner (2014) would allow richer analyses of volume and the welfare implications. For instance, if information acquisition is costly for institutional traders, the reductions in their expected profits would also reduce information gathering and may harm price discovery.

Market Design Implications: The model explains how high frequency traders may integrate information in otherwise fragmented markets. As HFTs do so, they avoid adverse selection by institutional traders, reduce trading costs and generate spillovers between trading venues. Some resulting comments on market design:

Speed Bumps: TSX Alpha and Aequitas Neo exchanges apply a processing delay against liquidity taking orders (TSXA (2015); Alpha (2014)). A delay on liquidity takers improves the ability of market makers to incorporate cross-venue information and is therefore expected to improve market quality while decreasing returns to institutional traders. Speed bumps on *all* orders, such as those by IEX (Picardo (2014)), may harm market quality and increase returns to institutional traders if HFTs are chiefly engaged in market making and order delay affects HFTs more severely than slower traders.

Fees: Fees on cancellations or messages would reduce market makers' ability update quotes on cross-venue information (Malinova et al. (2013)), and thereby decrease market quality and increase the returns to institutional traders.

Colocation: Brogaard et al. (2015) finds that most colocation services are allocated to market makers. Then, colocation services would improve market quality and reduce the returns to institutional traders.

Regulatory Implications: The model predicts that increasing the speed of market makers at a given trading venue negatively affects other venues' market quality if those venues' market makers are already very fast. Since high frequency traders are pervasive (O'Hara (2015)) and HFTs tend to provide liquidity (Brogaard et al. (2015, 2014); Malinova and Park (2015); Hagströmer and Norden (2013), the model implies that technical upgrades or other changes that increase the speed of market makers at a given venue may reduce liquidity and price discovery at other venues.

Appendix

Proof of Theorem 1: The first part of this proof closely follows Kyle (1985) and Pasquariello and Vega (2007). Steps of the proof are as follows: (i) solve the trading stage conditional on $\{N_i, N_j\}$; then (ii) solve the entry stage given the solution to the trading stage.

Solution to the Trading Stage: In the trading stage, an equilibrium solution is found if and only if the institutional traders are maximizing expected profits given their information set and the price in each trading venue is equal to the expected value of v given the information set of the venue's market maker. The kth IT ($k \in \{1, ..., N\}$) assigned to venue h ($h \in \{i, j\}$) has an information set that includes $\{N_i, N_j\}$ and s_k such that he solves:

$$\max_{x_k} \mathbb{E}[x_k(v - p_h)|N_i, N_j, s_k, k \in K_h].$$

In a linear equilibrium, the kth IT believes that the other ITs strategies and the price setting rules will be in the form (5)-(6) so that where $k \in K_i$ ($k \in K_j$ case is symmetric):

$$\begin{aligned}
\max_{x_k} \mathbb{E}[x_k(v - p_i)] = \max_{x_k} x_k \mathbb{E}\Bigg[\mu + \sum_{l=1}^{N} s_l - \delta_i - \lambda_{y,i}\left((N_i - 1)\alpha_i + \beta_i \sum_{l \in K_i \setminus k} s_l + x_k + u_i\right) \\
- \lambda_{z,i}\left(N_j\alpha_j + \beta_j \sum_{l \in K_j} s_l + a_i + u_j\right)\Bigg] \\
= \max_{x_k} \left[\mu + (1 + (N-1)\rho)\, s_k - \delta_i + \lambda_{y,i}(N_i - 1)\alpha_i + \lambda_{z,i} N_j \alpha_j\right] x_k \\
- \lambda_{y,i}\left[\beta_i(N_i - 1)\rho s_k x_k + x_k^2\right] - \lambda_{z,i}\beta_j N_j \rho s_k x_k
\end{aligned}$$

whose first order condition is satisfied if and only if:

$$2\lambda_{y,i}x_k = \mu - \delta_i + \left[1 + (N-1)\rho - \lambda_{y,i}\beta_i(N_i - 1)\rho - \lambda_{z,i}\beta_j N_j \rho\right] s_k$$

and the second order condition for a maximum is satisfied if and only if $\lambda_{y,i} > 0$. Then, the kth IT follows a linear strategy in the form (6) (i.e. $x_k = \alpha_i + \beta_i s_k$) if and only if:

$$\alpha_i = \frac{\mu - \delta_i - N_j \lambda_{z,i}\alpha_j}{(N_i + 1)\lambda_{y,i}} \tag{28}$$

$$\beta_i = \frac{1 + (N-1)\rho - \lambda_{z,i}\beta_j N_j \rho}{[2 + (N_i - 1)\rho]\lambda_{y,i}} \tag{29}$$

which are identical to (7)-(8).

For equilibrium price setting, note that the market maker at venue i knows $\{N_i, N_j\}$ and $\{y_i, z_i\}$ before she chooses the price. Then:

$$
\begin{aligned}
p_i &= \mathbb{E}[v|N_i, N_j, y_i, z_i] \\
&= \delta_i + \lambda_{y,i} y_i + \lambda_{z,i} z_i \\
&= \delta_i + \lambda_{y_i} \left[N_i \alpha_i + \beta_i \sum_{k \in K_i} s_k + u_i \right] + \lambda_{z,i} \left[N_j \alpha_j + \beta_j \sum_{k \in K_j} s_k + a_i + u_j \right]
\end{aligned}
$$

where δ_i equals $\mathbb{E}[v|N_i, N_j, y_i = 0, z_i = 0] = \mu$. Then, the projection theorem implies that:

$$
\lambda = \mathbb{E}\left[v\mathbf{y}'\right] \left[\mathbb{E}(\mathbf{y}\mathbf{y}')\right]^{-1}
$$

where λ is the vector $\{\lambda_{y,i}, \lambda_{z,i}\}$, and $\mathbf{y}$ is the vector $\{y_i, z_i\}$. First, expand the term $\mathbb{E}\left[v\mathbf{y}'\right]$ (recall that $\mathbb{E}(s_k) = 0$, $\forall k \in \{1, ..., N\}$):

$$
\begin{aligned}
\mathbb{E}\left[v\mathbf{y}'\right] &= \mathbb{E}[vy_i\,,\, vz_i] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^{N} s_k \right) \left(N_i \alpha_i + \beta_i \sum_{k \in K_i} s_k + u_i \right) ,\, \left(\sum_{k=1}^{N} s_k \right) \left(N_j \alpha_j + \beta_j \sum_{k \in K_j} s_k + a_i + u_j \right) \right] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^{N} s_k \right) \left(\sum_{k \in K_i} s_k \right) \beta_i \,,\, \left(\sum_{k=1}^{N} s_k \right) \left(\sum_{k \in K_j} s_k \right) \beta_j \right] \\
&= [1 + (N-1)\rho]\sigma_s^2 [N_i \beta_i \,,\, N_j \beta_j].
\end{aligned}
$$

Now, compute the term $[\mathbb{E}(\mathbf{y}\mathbf{y}')]^{-1}$ (recall that $Var(a_i) = S_i^{-1}$):

$$
\begin{aligned}
[\mathbb{E}(\mathbf{y}\mathbf{y}')]^{-1} &= \begin{bmatrix} Var(y_i) & Cov(y_i, z_i) \\ Cov(z_i, y_i) & Var(z_i) \end{bmatrix}^{-1} \\
&= \begin{bmatrix} \beta_i^2[N_i\sigma_s^2 + N_i(N_i-1)\sigma_{ss}] + \sigma_u^2 & \beta_i\beta_j N_i N_j \sigma_{ss} \\ \beta_i\beta_j N_i N_j \sigma_{ss} & \beta_j^2[N_j\sigma_s^2 + N_j(N_j-1)\sigma_{ss}] + S_i^{-1} + \sigma_u^2 \end{bmatrix}^{-1} \\
&= \left(\sigma_s^2\right)^{-1} \begin{bmatrix} N_i\beta_i^2[1+(N_i-1)\rho] + \frac{\sigma_u^2}{\sigma_s^2} & \beta_i\beta_j N_i N_j \rho \\ \beta_i\beta_j N_i N_j \rho & N_j\beta_j^2[1+(N_j-1)\rho] + \frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2} \end{bmatrix}^{-1} \\
\iff \sigma_s^2\,[\mathbb{E}(\mathbf{y}\mathbf{y}')]^{-1} &= \frac{1}{(\beta_i\beta_j)^2(1-\rho)[1+(N-1)\rho] + \left[\beta_i^2\frac{1+(N_i-1)\rho}{N_j} + \frac{1}{N_iN_j}\frac{\sigma_u^2}{\sigma_s^2}\right]\left(\frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2}\right) + \beta_j^2\frac{1+(N_j-1)\rho}{N_i}\frac{\sigma_u^2}{\sigma_s^2}} \\
&\quad \times \begin{bmatrix} \beta_j^2\frac{1+(N_j-1)\rho}{N_i} + \frac{1}{N_iN_j}\left(\frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2}\right) & -\beta_i\beta_j\rho \\ -\beta_i\beta_j\rho & \beta_i^2\frac{1+(N_i-1)\rho}{N_j} + \frac{1}{N_iN_j}\frac{\sigma_u^2}{\sigma_s^2} \end{bmatrix}.
\end{aligned}
$$

Thus (after some simplification):

$$
\begin{aligned}
\lambda &= \mathbb{E}[v\mathbf{y}']\,[\mathbb{E}(\mathbf{y}\mathbf{y}')]^{-1} \\
&= \frac{1+(N-1)\rho}{(\beta_i\beta_j)^2(1-\rho)[1+(N-1)\rho] + \left[\beta_i^2\frac{1+(N_i-1)\rho}{N_j} + \frac{1}{N_iN_j}\frac{\sigma_u^2}{\sigma_s^2}\right]\left(\frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2}\right) + \beta_j^2\frac{1+(N_j-1)\rho}{N_i}\frac{\sigma_u^2}{\sigma_s^2}} \\
&\quad \times \left[\beta_i\left(\beta_j^2(1-\rho) + \frac{1}{N_j}\left(\frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2}\right)\right) \quad , \beta_j\left(\beta_i^2(1-\rho) + \frac{1}{N_i}\frac{\sigma_u^2}{\sigma_s^2}\right)\right]'. \qquad (30)
\end{aligned}
$$

Substituting (30) into (29) and simplifying gives:

$$
(\beta_i\beta_j)^2(1-\rho) + \beta^2\frac{1}{N_j}\left(\frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2}\right) = \frac{\frac{\sigma_u^2}{\sigma_s^2}}{N_iN_j}\left(\frac{1}{S_i\sigma_s^2} + \frac{\sigma_u^2}{\sigma_s^2}\right) + \beta_j^2(1-\rho)\frac{\frac{\sigma_u^2}{\sigma_s^2}}{N_i}
$$

which, together with (30) and (28), implies the solutions for the trading stage conditional on $\{N_i, N_j\}$, (17)-(21). Note that (20) ensures that the second order condition ($\lambda_{y,i} > 0$) is met. Finally, where $N_i = 0$, β_i as determined by (19) does not exist. Since there are no institutional traders in any subgame at venue i with $N_i = 0$, I set β_i equal to an arbitrary constant whenever $N_i = 0$.

Given the solution to the trading stage, I solve the entry stage in three steps given that institutional traders never exit the game during the entry stage: (i) solve for the kth IT's expected profit to entering venue i given each other IT assigns probability n^i to entering the venue; (ii) show that $n^i \notin \{0, 1\}$ in any linear and symmetric equilibrium; and (iii) show that there exists a unique $n^i \in (0, 1)$ such that the kth IT's expected profit to entering either venue are equal. First, the solution to the trading stage implies that the kth IT's expected profit to entering venue i is (after some simplification):

$$\begin{aligned}
\Pi_{i,k} &:= \mathbb{E}[x_k(v-p_i)|k \in K_i] \\
&= \mathbb{E}\left[\beta_i s_k(\mu + \sum_{l=1}^{N} s_l - \delta_i - \lambda_{y,i} y_i - \lambda_{z,i} z_i)\big| k \in K_i\right] \\
&= \mathbb{E}\left[A(S_i)|k \in K_i\right]\sigma_u\sigma_s \\
&= (1+(N-1)\rho)\left(2-\rho+\frac{1}{S_i\sigma_u^2}\right)\sigma_u\sigma_s \sum_{l=0}^{N-1} \frac{\binom{N-1}{l}(n^i)^l(1-n^i)^{N-1-l}}{\sqrt{l+1}\left[(2+(N-1)\rho)(2-\rho)+(2+l\rho)\frac{1}{S_i\sigma_u^2}\right]}. \quad (31)
\end{aligned}$$

Note that the denominators of the terms in the series (31) are strictly increasing in l. In addition, $\frac{\partial}{\partial n^i}(n^i)^l(1-n^i)^{N-1-l}$ is strictly increasing in l, with $\frac{\partial}{\partial n^i}(n^i)^l(1-n^i)^{N-1-l} > 0$ for all $l > (N-1)/2$ but $\frac{\partial}{\partial n^i}(n^i)^l(1-n^i)^{N-1-l} < 0$ for all $l < (N-1)/2$, which imply that (31) is strictly decreasing in n^i. Then, showing that $\Pi_{i,k}(n^i=0) > \Pi_{j,k}(n^i=0)$ and $\Pi_{i,k}(n^i=1) < \Pi_{j,k}(n^i=1)$ is sufficient to prove that $n^{i*} \notin \{0,1\}$:

$$\begin{aligned}
& \Pi_{i,k}(n^i=0) > \Pi_{j,k}(n^i=0) \\
\iff \quad & \frac{2-\rho+\frac{1}{S_i\sigma_u^2}}{(2+(N-1)\rho)(2-\rho)+2\frac{1}{S_i\sigma_u^2}} > \frac{1}{\sqrt{N}(2+(N-1)\rho)} \\
\iff \quad & \frac{\sqrt{N}(2-\rho+\frac{1}{S_i\sigma_u^2})}{2-\rho+\frac{2}{2+(N-1)\rho}\frac{1}{S_i\sigma_u^2}} > 1
\end{aligned}$$

which is always true ($\Pi_{i,k}(n^i=1) < \Pi_{j,k}(n^i=1)$ is proved symmetrically). This as well as that $\Pi_{i,k}$ is continuous and strictly decreasing in $n^i \in (0,1)$ (and, by symmetry, $\Pi_{j,k}$ is continuous and strictly increasing in $n^i = 1-n^j$) together imply that the solution to $\Pi_{i,k}(n^i) = \Pi_{j,k}(n^i = 1-n^j)$ always exists and is unique in $(0,1)$ (i.e. n^{i*} exists and is unique in $(0,1)$). Finally, it is obvious that $\Pi_{i,k}(n^{i*}) > 0$. Therefore, no institutional trader leaves the game in the entry stage (i.e. $n^{i*} = 1-n^{j*}$). □

Proof of Proposition 1: By Theorem 1:

$$\sum_{m=0}^{N-1} \frac{\binom{N-1}{m}\left(n^{i*}\right)^m(1-n^{i*})^{N-1-m}}{\sqrt{m+1}\left[E\cdot(2-\rho)+F(m)\cdot G_i\right]} = \frac{2-\rho+G_j}{2-\rho+G_i}\sum_{q=0}^{N-1} \frac{\binom{N-1}{q}\left(1-n^{i*}\right)^q(n^{i*})^{N-1-q}}{\sqrt{q+1}\left[E\cdot(2-\rho)+F(q)\cdot G_j\right]} \quad (32)$$

where $G_h = \frac{1}{S_h\sigma_u^2}$, $E := 2+(N-1)\rho$ and $F(m) := 2+m\rho$. To prove that n^{i*} is strictly decreasing in S_i, I want to show that the $(2-\rho+G_j)/(2-\rho+G_i)$ term decreases by a larger proportion than the left-hand-side of (32), when S_i decreases such that G_i increases (holding n^{i*} constant). This implies, because (from the proof to Theorem 1) the left-hand-side of (32) is continuous and strictly decreasing while the summation in the right-hand-side side is continuous and strictly increasing, that n^{i*} must decrease to reattain (32). Therefore, it is sufficient to show that $\frac{2-\rho+G_i'}{2-\rho+G_i} \geq \frac{E\cdot(2-\rho)+F(m)\cdot G_i'}{E\cdot(2-\rho)+F(m)\cdot G_i}$ holds for all $m \in \{0, ..., N-1\}$ and $G_i' := G_i + \epsilon$, $\epsilon > 0$, and holds with strict inequality for some m. Since $E > F(m)$, $\forall m < N-1$, and $E = F(N-1)$:

$$\frac{2-\rho+G_i+\epsilon}{E\cdot(2-\rho)+F(m)\cdot(G_i+\epsilon)} > \frac{2-\rho+G_i}{E\cdot(2-\rho)+F(m)\cdot G_i}\,, \forall m < N-1$$

$$\text{and} \quad \frac{2-\rho+G_i+\epsilon}{E\cdot(2-\rho)+F(N-1)\cdot(G_i+\epsilon)} = \frac{1}{E} = \frac{2-\rho+G_i}{E\cdot(2-\rho)+F(N-1)\cdot G_i}.$$

Then, because $N \geq 2$, n^{i*} is strictly decreasing in S_i, and a symmetric argument proves that n^{i*} is strictly increasing in S_j. □

Proof of Proposition 3: If $S_i = S_j = S > 0$:

$$I(y_i) + I(z_i) = \frac{1+(N_i-1)\rho}{2+(N_i-1)\rho} + \frac{1+(N_j-1)\rho}{2+(N_j-1)\rho+\frac{1}{S\sigma_u^2}}.$$

Trivially, $N_i = N_j$ implies $I(y_i)+I(z_i) = I(y_j)+I(z_j)$. If $N_i \neq N_j$, then $I(y_i)+I(z_i) = I(y_j)+I(z_j)$ holds if and only if:

$$\frac{1+(N_i-1)\rho}{2+(N_i-1)\rho} - \frac{1+(N_j-1)\rho}{2+(N_j-1)\rho} = \frac{1+(N_i-1)\rho}{2+(N_i-1)\rho+\frac{1}{S\sigma_u^2}} - \frac{1+(N_j-1)\rho}{2+(N_j-1)\rho+\frac{1}{S\sigma_u^2}}$$

which is untrue for any finite S. Now, given $S_i = S_j = S > 0$, (32) becomes:

$$\sum_{m=0}^{N-1} \frac{\binom{N-1}{m}\left(n^{i*}\right)^m (1-n^{i*})^{N-1-m}}{\sqrt{m+1}\,[E\cdot(2-\rho)+F(m)\cdot G]} = \sum_{q=0}^{N-1} \frac{\binom{N-1}{q}\left(1-n^{i*}\right)^q (n^{i*})^{N-1-q}}{\sqrt{q+1}\,[E\cdot(2-\rho)+F(q)\cdot G]}$$

which holds when $n^{i*} = 1/2$. Then, by Theorem 1, $n^{i*} = 1/2$ is the unique solution to (32) in $(0,1)$, given $S_i = S_j = S > 0$. □

Proof of Proposition 4: Suppose ρ increases from ρ_0 to $\rho' := \rho_0 + \epsilon$, $\epsilon > 0$, and define (only for this proof) a vector $\{A(\rho), B(\rho), C(\rho), D(\rho), E\} := \left\{1+(N-1)\rho\,,\ 2-\rho\,,\ 2+(N-1)\rho\,,\ 2+(N_i-1)\rho\,,\ \frac{1}{S_i\sigma_u^2}\right\}$. As a short-hand, for some function $X(\rho)$, I denote $X(\rho')$ by X' and denote $X(\rho_0)$ by X. Then, by definition

(14):

$$
\begin{aligned}
& A(S_i, \rho = \rho') - A(S_i, \rho = \rho_0) = \frac{A'(B'+E)}{C'B'+D'E} - \frac{A(B+E)}{CB+DE} > 0 \\
\iff \quad & A'(B'+E)(CB+DE) - A(B+E)(C'B'+D'E) \\
& \qquad = (A+(N-1)\epsilon)((B-\epsilon)+E)(CB+DE) \\
& \qquad\quad - A(B+E)((C+(N-1)\epsilon)(B-\epsilon)+(D+(N_i-1)\rho)E) > 0 \\
\iff \quad & [-A(CB+DE)+(N-1)(B+E)(CB+DE) \\
& \qquad + A(B+E)C - (N-1)AB(B+E) - (N_i-1)A(B+E)E]\epsilon \\
& \qquad > (N-1)[(CB+DE)-A(B+E)]\epsilon^2. \qquad (33)
\end{aligned}
$$

By choosing a small ϵ, it is sufficient to show that the left-hand-side of the inequality (33) is strictly positive:

$$
\begin{aligned}
& -A(CB+DE)+(N-1)(B+E)(CB+DE)+A(B+E)C \\
& \qquad - (N-1)AB(B+E) - (N_i-1)A(B+E)E \\
& \qquad = A(C-D)E + (N-1)(B+E)\left[(CB+DE) - AB - \frac{N_i-1}{N-1}AE\right].
\end{aligned}
$$

Since $A(C-D)E = A(N-N_i)E \geq 0$, it is sufficient to show that the other terms are strictly positive:

$$
\begin{aligned}
(CB+DE) - AB - \frac{N_i-1}{N-1}AE &= (C-A)B + \left(D - \frac{N_i-1}{N-1}A\right)E \\
&= B + \left(2 - \frac{N_i-1}{N-1}\right)E
\end{aligned}
$$

which is strictly positive. The proof from venue j is symmetric, and the $I(y_i)$ and $I(z_i)$ results are trivial. $\square$

Proof of Proposition 5: If $\rho = 0$, equation (32) reduces to:

$$
\sum_{m=0}^{N-1} \frac{\binom{N-1}{m}\left(n^{i*}\right)^m \left(1-n^{i*}\right)^{N-1-m}}{\sqrt{m+1}} = \sum_{q=0}^{N-1} \frac{\binom{N-1}{q}\left(1-n^{i*}\right)^q \left(n^{i*}\right)^{N-1-q}}{\sqrt{q+1}}
$$

which is attained if $n^{i*} = 1/2$. Then, by Theorem 1, $n^{i*} = 1/2$ is the unique solution in $(0,1)$. $\square$

Deriving PD_i (equation (25)): The law of total variance and the identity $\mathbb{E}[v|p_i] = p_i$ imply:

$$Var(v) = \mathbb{E}[Var(v|p_i)] + Var(\mathbb{E}[v|p_i]) = \mathbb{E}[Var(v|p_i)] + Var(p_i)$$
$$\iff \mathbb{E}[Var(v|p_i)] = \sigma_v^2 - Var(p_i).$$

Since $PD_i := 1 - \mathbb{E}[Var(v|p_i)]/\sigma_v^2$:

$$\begin{aligned} PD_i &= Var(p_i) \\ &= Var\left(\mu + \lambda_{y,i} y_i + \lambda_{z,i} z_i\right) \\ &= \lambda_{y,i}^2 Var(y_i) + \lambda_{z,i}^2 Var(z_i) + 2\lambda_{y,i}\lambda_{z,i} Cov(y_i, z_i) \end{aligned} \tag{34}$$

Substituting $Var(y_i)$, $Var(z_i)$ and $Cov(y_i, z_i)$ from the proof to Theorem 1, as well as $\sigma_v^2 = N(1+(N-1)\rho)\sigma_s^2$ into (34) gives the result. □

Proof of Proposition 8: Define (for this proof) a vector $\{A, B, C, D, E, F\} := \{1 + (N-1)\rho, 2 + (N_i - 1)\rho, \frac{N_i}{N} D + \frac{1}{N}[2 + (N_i N_j - 1)\rho], 2 + (N-1)\rho, \frac{1}{S_i \sigma_u^2}, 2 - \rho\}$. Then:

$$PD_i = \frac{A\left[DF^2 + CFE + \frac{N_i}{N} BE^2\right]}{(DF + BE)^2}.$$

To show that PD_i is strictly increasing in S_i (given N_i), it is sufficient to show that the denominator of PD_i increases by a larger proportion than the numerator of PD_i when S_i decreases so that E increases to $E' := E + \epsilon$, $\epsilon > 0$. Therefore, it is sufficient to show that:

$$\begin{aligned} \left(\frac{DF + BE'}{DF + BE}\right)^2 &> \frac{DF^2 + CFE' + \frac{N_i}{N} B(E')^2}{DF^2 + CFE + \frac{N_i}{N} BE^2} \\ \iff \frac{F^2 + d_1 E' + d_2 (E')^2}{F^2 + d_1 E + d_2 E^2} &> \frac{F^2 + m_1 E' + m_2 (E')^2}{F^2 + m_1 E + m_2 E^2} \end{aligned} \tag{35}$$

where $d_1 := 2\frac{FB}{D}$, $d_2 := \frac{B^2}{D^2}$, $m_1 := \frac{CF}{D}$, and $m_2 := \frac{N_i}{N}\frac{B}{D}$. Then, the conditions $d_1 > m_1$ and $d_2 > m_2$ are sufficient to show that (35) holds. Specifically, when the conditions:

1. $\frac{CF}{D} < \frac{2FB}{D}$; and

2. $\frac{N_i}{N}\frac{B}{D} < \frac{B^2}{D^2}$

are satisfied, then (35) is satisfied, which implies that PD_i is strictly increasing in S_i (but not the other way around). The first condition implies:

$$\begin{aligned} &C < 2DB \\ \iff \left(\frac{N_i}{N} - B\right) &D < DB - \frac{1}{N}[2 + (N_i N_j - 1)\rho] \\ &= N(N - N_i)\rho + N_i^2\rho + (N-1)(2-\rho). \end{aligned}$$

Since the left-hand-side of the inequality is negative and the right-hand-side is positive, the first condition always holds. The second condition implies:

$$\begin{aligned} &\frac{N_i}{N} D < B \\ \iff &\frac{N_i}{N}(2 + (N-1)\rho) < 2 + (N_i - 1)\rho \end{aligned}$$

which implies $\rho < 2$. Therefore, PD_i is strictly increasing in S_i (given any possible N_i).

For measures L_i and $PRFT_i$, it is sufficient to show that $A(S_i)$ is strictly increasing in E (thereby strictly decreasing in S_i), given N_i. Recall that $A(S_i) := \frac{A(F+E)}{DF+BE}$. Then, it is sufficient to show that, for $E' := E + \epsilon$, $\epsilon > 0$:

$$\begin{aligned} &\frac{A(F+E')}{DF+BE'} > \frac{A(F+E)}{DF+BE} \\ \iff &DF + BE > B(F+E) \end{aligned}$$

which is satisfied whenever $D > B \iff 2 + (N-1)\rho > 2 + (N_i - 1)\rho$, which is true whenever $N_i < N$. □

Proof of Proposition 9: For this proof, it must be shown that the first derivative of $\sqrt{N_i}A(S_i)$ is strictly positive if and only if $S_i > S_i^*$, and is strictly negative if and only if $S_i < S_i^*$. The first derivative of $\sqrt{N_i}A(S_i)$ is:

$$\frac{1}{\sqrt{N_i}}A(S_i) - \frac{\rho\frac{1}{S_i\sigma_u^2}}{[2 + (N-1)\rho](2-\rho) + [2 + (N_i - 1)\rho]\frac{1}{S_i\sigma_u^2}}\sqrt{N_i}A(S_i)$$

which is strictly positive if and only if $S_i > S_i^*$, and is strictly negative if and only if $S_i < S_i^*$. □

References

Ait-Sahalia, Y. and M. Saglam (2014). High frequency traders: Taking advantage of speed. *Available at SSRN 2331613*.

Alpha, T. (2014, Jun). Alpha exchange inc. notice of proposed rule amendments and request for comments.

Baldauf, M. and J. Mollner (2014). High-frequency trade and market performance. *Working Paper*.

Baldauf, M. and J. Mollner (2015). *Trading in fragmented markets*.

Bernhardt, D. and E. Hughson (1997, 1997). Splitting orders. *The Review of Financial Studies 10*(1), 69–101.

Bershova, N. and D. Rakhlin (2013). High-frequency trading and long-term investors: A view from the buy-side. *Journal of Investment Strategies 2*(2), 25–69.

Biais, B., T. Foucault, and S. Moinas (2015). Equilibrium fast trading. *Journal of Financial Economics 116*(2), 292–313.

Boehmer, E., K. Y. Fong, and J. J. Wu (2014). International evidence on algorithmic trading. In *AFA 2013 San Diego Meetings Paper*.

Brogaard, J., B. Hagströmer, L. L. Norden, and R. Riordan (2015). Trading fast and slow: Colocation and liquidity. *Forthcoming in Review of Financial Studies*.

Brogaard, J., T. Hendershott, S. Hunt, and C. Ysusi (2014). High-frequency trading and the execution costs of institutional investors. *Financial Review 49*(2), 345–369.

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

Brogaard, J., T. Hendershott, and R. Riordan (2015). Price discovery without trading: Evidence from limit orders. *Available at SSRN 2655927*.

Budish, E. B., P. Cramton, and J. J. Shim (2013). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Fama-Miller Working Paper*, 14–03.

Carrion, A. (2013, 11). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets; High-Frequency Trading 16*(4), 680–711.

Chaboud, A. P., B. Chiquoine, E. Hjalmarsson, and C. Vega (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance 69*(5), 2045–2084.

Chen, H., F. Sean, G. Michael, and R. Thomas (2016, Jan). The value of a millisecond: Structural segmentation of uninformed order flow.

Chordia, T., A. Goyal, B. N. Lehmann, and G. Saar (2013, 11). High-frequency trading. *Journal of Financial Markets; High-Frequency Trading 16*(4), 637–645.

Chordia, T., R. Roll, and A. Subrahmanyam (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics 101*(2), 243–263.

Chowdhry, B. and V. Nanda (1991). Multimarket trading and market liquidity. *Review of Financial Studies 4*(3), 483–511.

Cimon, D. A. (2015). Broker routing decisions in limit order markets.

Conrad, J., S. Wahal, and J. Xiang (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics 116*(2), 271–291.

Foster, F. D. and S. Viswanathan (1993). The effect of public information and competition on trading volume and price volatility. *Review of Financial Studies 6*(1), 23–56.

Foster, F. D. and S. Viswanathan (1996). Strategic trading when agents forecast the forecasts of others. *The Journal of Finance 51*(4), 1437–1478.

Foucault, T., J. Hombert, and I. Rosu (2015). News trading and speed. *The Journal of Finance*.

Foucault, T., R. Kozhan, and W. W. Tham (2014). Toxic arbitrage. *Working Paper*.

Hagströmer, B. and L. Norden (2013). The diversity of high-frequency traders. *Journal of Financial Markets 16*(4), 741–770.

Hasbrouck, J. and G. Saar (2013, 11). Low-latency trading. *Journal of Financial Markets; High-Frequency Trading 16*(4), 646–679.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *The Journal of Finance 66*(1), 1–33.

Hendershott, T. and P. C. Moulton (2011, 11). Automation, speed, and stock market quality: The nyse's hybrid. *Journal of Financial Markets 14*(4), 568–604.

Hendershott, T. and R. Riordan (2013). Algorithmic trading and the market for liquidity. *Journal of Financial and Quantitative Analysis 48*(04), 1001–1024.

Hoffmann, P. (2014, JUL). A dynamic limit order market with fast and slow traders. *Journal of Financial Economics 113*(1), 156–169. PT: J; NR: 31; TC: 0; J9: J FINANC ECON; PG: 14; GA: AI1VL; UT: WOS:000336643200008.

Jarrow, R. A. and P. Protter (2012). A dysfunctional role of high frequency trading in electronic markets. *International Journal of Theoretical and Applied Finance 15*(03).

Jovanovic, B. and A. J. Menkveld (2011). Middlemen in limit-order markets. *Western Finance Association (WFA)*.

Korajczyk, R. A. and D. Murphy (2015). High frequency market making to large institutional trades. *Available at SSRN*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Malinova, K. and A. Park (2015). Liquidity provision and market making by hfts.

Malinova, K., A. Park, and R. Riordan (2013). Do retail traders suffer from high frequency traders? *Available at SSRN 2183806*.

Mendelson, H. and T. I. Tunca (2003). Strategic trading, liquidity, and information acquisition. *Rev. Financ. Stud. Review of Financial Studies 17*(2), 295337.

Menkveld, A. J. (2013, 11). High frequency trading and the new market makers. *Journal of Financial Markets; High-Frequency Trading 16*(4), 712–740.

Menkveld, A. J. and M. A. Zoican (2014). Need for speed? exchange latency and liquidity.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics 116*(2), 257–270.

Pasquariello, P. and C. Vega (2007). Informed and strategic order flow in the bond markets. *Review of Financial Studies 20*(6), 1975–2019.

Picardo, E. (2014, 4/23/2014). How iex is combating predatory types of high-frequencytraders.

Riordan, R. and A. Storkenmaier (2012, 11). Latency, liquidity and price discovery. *Journal of Financial Markets 15*(4), 416–437.

Rosu, I. (2014). Fast and slow informed trading. In *AFA 2013 San Diego Meetings*.

Schmitt, J. (2015). The ultimate best execution conflict of interest?

SEC (2013). Equity market structure literature review part i: Market fragmentation. Technical report, Staff of the Division of Trading and Markets, U.S. Securities and Exchange Commission.

SEC (2014). Equity market structure literature review part ii: High frequency trading. Technical report, Staff of the Division of Trading and Markets, U.S. Securities and Exchange Commission.

SEC (2015). Comments on investors' exchange llc; notice of filing of application, as amended, for registration as a national securities exchange under section 6 of the securities exchange act of 1934.

Subrahmanyam, A. and H. Zheng (2015, Nov). Limit order placement by high-frequency traders.

Tong, L. (2015). A blessing or a curse? the impact of high frequency trading on institutional investors. *Working Paper (October 5, 2015)*.

TSXA (2015). Overview of tsx alpha exchange.

Willmer, S. (2015, Jan). Hedge fund mines twitter for stock tips.

Yang, L. and H. Zhu (2015). Seeking and hiding fundamental information in order flows. *Available at SSRN 2583915*.

Zhang, S. S. (2012). Need for speed: An empirical analysis of hard and soft information in a high frequency world. *Available at SSRN 1985951*.