Equilibrium Fast Trading*

Bruno Biais[†], Thierry Foucault[‡], and Sophie Moinas[§]

September 2014

Abstract

High–speed market connections improve investors' ability to search for attractive quotes in fragmented markets, raising gains from trade. They also enable fast traders to observe market information before slow traders, generating adverse selection, and thus negative externalities. When investing in fast trading technologies, institutions do not internalize these externalities. Accordingly, they overinvest in equilibrium. Completely banning fast trading is dominated by offering two types of markets: one accepting fast traders, the other banning them. However, utilitarian welfare is maximized by having i) a single market type on which fast and slow traders coexist and ii) Pigovian taxes on investment in the fast trading technology.

JEL classification: D4, D62, G1, G20, L1.

Keywords: high–frequency trading, externalities, welfare.

---

*Many thanks for helpful comments to the referee, Alex Guembel, Terry Hendershott, Carolina Manzano, Albert Menkveld, Anna Obizhaeva, Emiliano Pagnotta, Rafael Repullo, Ailsa Roëll, Xavier Vives and participants in the High Frequency Conference (Madrid, 2012), the 2012 SFS Cavalcade, the 2013 AFA meeting, and the 2014 SED meeting. Moinas acknowledges the support of the ANR (Grant 09 - JCJC - 0139 - 01.) Biais acknowledges the support of the ERC (Grant 295484 – TAP.)

†Corresponding author. Toulouse School of Economics (CNRS-CRM and FBF IDEI Chair on Investment Banking and Financial markets.) Adress: Manufacture des Tabacs, 21 allées de Brienne, 31000 Toulouse, France. Tel: (33) 5 61 12 85 98. E-mail: bruno.biais@ut-capitole.fr

‡HEC, Paris, GREGHEC, and CEPR. Adress: 1 rue de la Liberation, 78351 Jouy en Josas, France. Tel: (33) 1 39 67 95 69. E-mail: foucault@hec.fr

§Toulouse School of Economics (University of Toulouse 1 Capitole and FBF IDEI Chair on Investment Banking and Financial markets.) Adress: Manufacture des Tabacs, 21 allées de Brienne, 31000 Toulouse, France. Tel: (33) 5 61 12 85 73. E-mail: sophie.moinas@ut-capitole.fr

1. Introduction

Investors must process very large amounts of information, in particular about trades and quotes, which are relevant both for the valuation of securities and the identification of trading opportunities. Timely collection of this information has become increasingly difficult due to the fragmentation of the markets, e.g., for U.S. equities, there are now more than 50 trading venues: 13 registered exchanges and 44 so called Alternative Trading Systems.[1]

In fragmented markets, investors must search for quotes across markets. This can result in delayed or partial execution, which is costly. Chiyachantana and Jain (2009) find that delays in execution account for about 1/3 of total costs borne by institutional investors in their sample.[2] To reduce these costs, traders can invest in fast trading technologies. For instance, they can use smart routers that instantaneously compare quotes across trading venues and allocate their orders accordingly. Furthermore, to better inform their routing decisions, they can buy fast access to exchange data feed, using colocation rights (the placement of their computers next to the exchange's servers), or high-speed connections via fiber optic cables or microwave signals.

By the same token, however, fast trading technologies also accelerate access to value relevant information for an asset, conveyed by recent transaction prices and quote changes for this asset or related ones. Numerous empirical studies document that orders placed by fast traders reflect advance information.[3] This informational advantage generates adverse selection costs for other market participants. For example, Baron, Brogaard, and Kirilenko (2014) observe that aggressive, liquidity–taking, high–frequency traders earn short–term profits at the expense of other market participants, and Brogaard, Hendershott, and Riordan (2014) write: *"Our results are consistent with ... high–frequency traders imposing adverse selection on other investors"*. Thus, firms investing in fast trading technologies generate adverse–selection costs for the other market participants.

[1] See, for instance, O'Hara and Ye (2011), or, http://fragmentation.fidessa.com/ which provides statistics on market fragmentation in the U.S. and Europe.

[2] In practice, delay costs stem from (i) a worsening of price conditions between an order arrival and its completion and (ii) opportunity costs due to partial execution. Margin constraints could also make delayed execution costly (see Zhu (2014)).

[3] For instance, Brogaard, Hagströmer, Norden, and Riordan (2014), Brogaard, Hendershott, and Riordan (2014), Hendershott and Riordan (2013), Zhang (2013), and Kirilenko, Kyle, Samadi, and Tuzun (2011).

Fast trading firms have no incentives to internalize these costs when making their investment decisions, which can generate a wedge between privately and socially optimal investment in fast trading technologies.[4] In this paper, we analyze equilibrium investment decisions in fast trading technologies, their consequences for welfare, and possible policy interventions (taxation and slow markets) to achieve the socially optimal level of investment in fast trading technologies.

To examine these issues, we consider a simple model suitable for welfare and policy analysis. Financial institutions have i) heterogeneous private valuations, e.g., due to differences in tax or regulatory status, and ii) private information about common values. The latter is a source of adverse selection, whereas the former creates gains from trade.[5] Before trading, institutions decide to invest or not in a fast trading technology. Then, institutions seek to trade in a fragmented market. At each round of trade, a fraction λ of the trading venues offer attractive quotes, while the others do not. Fast institutions can instantaneously search across all markets, and consequently always find attractive quotes. Slow institutions cannot do so. For simplicity we assume they can visit only one market venue per period. Correspondingly, at each period, they execute their desired trade with probability λ. Otherwise they must continue to search for quotes, and find this delay costly. Moreover, in addition to speeding up execution, fast institutions' ability to scan markets ultra rapidly enables them to obtain advance information (e.g., from observing prices of other correlated assets), generating adverse selection costs for the other market participants.

First, we analyze equilibrium allocations and prices for a given fraction (α) of fast institutions. The larger α, the greater the information content, and hence the price impact, of trades. Now, institutions prefer to abstain from trading when their price impact exceeds their private gain from trade. Hence, an increase in α lowers gains from trade for all market participants. Thus, fast institutions exert a negative externality upon the others, by increasing adverse selection in the marketplace.

[4] As written by Hirshleifer (1971), *"the distributive aspect of access to superior information... provides a motivation for the acquisition of private information that is quite apart from any social usefulness of that information... There is an incentive for individuals to expend resources in a socially wasteful way in the generation of such information."*

[5] The differences in private values in our setting are similar to those in Duffie, Garleanu, and Pedersen (2005). Our assumption is also in line with Bessembinder, Hao, and Zheng (2013), in which private valuation shocks induce gains from trade and hence transactions between rational agents.

Second, we study equilibrium investment in fast trading technologies, i.e., we endogenize α. Financial institutions invest only if the cost of the fast technology is smaller than the relative value of being fast, i.e., the difference between the expected profit of a fast and a slow institution. Now, the relative value of being fast depends on the fraction of institutions who choose to be fast. Hence, the equilibrium level of investment in the fast trading technology is the solution of a fixed point problem: if institutions expect the level of fast trading to be α^*, then exactly this fraction find it optimal to be fast. When the relative value of being fast declines with the level of fast trading (i.e., if institutions' decisions are substitutes), the equilibrium is unique. Otherwise, there can be multiple equilibria. This happens when entry of a new fast institution reduces the profit of slow institutions more than that of fast institutions. In this case, institutions' investment decisions are complements: they reinforce each other, because the technology becomes increasingly attractive as more institutions invest in it. As a result, all institutions can end up investing in the fast technology, even though other equilibria with less or no investment in fast trading exist as well. This outcome has the flavour of an arms' race, as in Glode, Green, and Lowery (2012).

Third, we show that, because of the negative externality induced by fast traders, equilibrium investment in the fast trading technology exceeds its utilitarian–welfare maximizing counterpart.[6] This problem arises whether institutions' investment decisions are substitutes or complements. However, complementarities in investment decisions tend to worsen overinvestment because institutions can be trapped in an investment race, even if the socially optimal level of investment is low.

Fourth, we analyze various possible policy interventions to mitigate this inefficiency. A ban on fast trading precludes reaping the benefits of the technology. This approach is too harsh because the socially optimal level of investment is not necessarily zero. We therefore focus on less heavy-handed approaches.

The first approach is to let "slow markets" (on which fast trading is banned) coexist with fast markets. This approach always dominates a complete ban on fast trading or "laissez-faire".

[6] In practice, trading firms invest significant amounts to obtain fast access to markets. For example, the cost of Project Express, which drew a new and faster fiber optic cable across the Atlantic, to connect Wall Street to the City, was $300 million. For 2013 alone, the Tabb Group estimates the investment in fast trading technologies at $1.5 billion, twice the amount invested in 2012.

However, it can lead to *underinvestment* in the fast trading technology. Slow institutions migrate to slow markets where there is no adverse selection, and this reduces the expected profits of fast institutions. In this context, there are only two possible equilibrium outcomes: either all institutions trade in slow markets, or all of them are fast. The "All-Slow" equilibrium naturally arises when the technological cost is higher than a threshold. However, this threshold is lower than the threshold below which investment in the fast trading technology is socially desirable. When the technological cost is between the two thresholds, the introduction of a slow platform lowers investment in the fast trading technology relative to the utilitarian optimum.

The second approach is to have only fast markets with Pigovian taxation of the fast trading technology. This approach Pareto dominates the former. Indeed, equating the tax to the negative externality generated by fast trading leads to the level of investment that maximizes utilitarian welfare. Redistribution of this tax among all institutions (fast and slow) enables them to share the social gains.

Our theoretical analysis has several empirical implications. Trades become more informative when the level of fast trading increases. Hence, a reduction in the cost of fast trading raises the informational content of trades. This has an ambiguous effect on trading volume, however. Indeed, investment in fast trading technologies increases the chances that institutions are able to carry out desired trades, which tends to increase volume, but it also raises price impact, which tends to reduce trading volume. Consequently, trading volume can be non monotonic in the level of fast trading. The model also implies that an increase in market fragmentation lowers the profitability of fast institutions, because it increases the price impact of trades. Yet, for a high cost of fast trading, increased market fragmentation might stimulate investment in fast trading because market fragmentation hurts slow institutions even more than fast institutions, so that the relative value of being fast increases. For a low cost of fast trading, this prediction is reversed.

The next section discusses the relation between our analysis and the theoretical literature. Section 3 presents the model and Section 4 derives equilibrium prices and trades, for a given level of investment in the fast trading technology. This level is endogenized in Section 5. We then show that the equilibrium level of investment in fast trading technologies is excessive and study policy responses in Section 6. Section 7 describes empirical implications of the model and Section 8 concludes.

2. Related theoretical literature

2.1. Equilibrium information acquisition

Our paper is in line with the seminal analysis of private information acquisition in financial markets by Grossman and Stiglitz (1980). One major difference is that, while in Grossman and Stiglitz (1980) trading occurs because of noise traders, in our framework all agents optimize, so that trading endogenously responds to gains from trade, adverse selection, and information acquisition. This is necessary to perform a welfare analysis of information acquisition in financial markets. We are not aware of any other welfare analysis of information acquisition in financial markets, in the absence of noise traders. Moreover, since all trades are optimally chosen by investors, we can study how uninformed investors' trading reacts to an increase in the fraction of informed traders, which enables us to to characterize equilibrium trading volume. Finally, while in Grossman and Stiglitz (1980) investments in information acquisition are always strategic substitutes, in our model they can also be strategic complements. In Ganguli and Yang (2009) and Breon–Drish (2013), complementarity in information acquisition arises when prices become less informative as the number of informed investors increases. This interesting mechanism is completely different from that at play in our model, whereby financial institutions can decide to be fast because they anticipate many others to also be fast, and thus fear to obtain very low profits if they remain slow.

Llosa and Venkateswaran (2012) and Colombo, Femminis, and Pavan (2014) study the wedge between social and private optimality of information acquisition. These models, however, do not apply to trading in financial markets. For instance, Colombo, Femminis, and Pavan (2014) rely on exogenous technological externalities, negative as in the case of pollution, or positive if agents have a taste for conformity. This differs from our analysis in which the negative externality, induced by information acquisition, arises because of endogenous adverse selection costs for financial market participants.

Hu and Qin (2013) show that investors' ex-ante expected utility declines with the number of informed investors in Grossman (1976)'s model. In their set-up, prices are fully revealing. Hence, investors are equally informed, whether or not they acquire information, and information has therefore no private value. In contrast, in our model, information has both private and social value and investors' average welfare is not necessarily maximized when all investors are uninformed. This is indeed the reason why banning fast trading is in general inefficient in our model.

2.2. Market microstructure

Budish, Cramton, and Shim (2014) develop a model in which market makers invest in speed to be first to react to, and profit from, public information arrival. In their analysis, however, investors are noise traders and gains from trade are not modelled. Fast trading simply generates transfers of resources from investors to market makers, without bringing any social benefit. From a utilitarian point of view, the cost of fast trading is just the cost of investing in the fast technology. In this context, trading slowly (e.g., in periodic batch auctions) is always socially optimal. In contrast, our analysis emphasizes the dual role of fast trading technologies, which facilitate the search for quotes at the same time as they generate adverse selection. Thus, we show that the socially optimal level of investment in fast trading technology is in general not zero, and we analyze the tradeoff giving rise to the socially optimal level of that investment.

Du and Zhu (2013) study the welfare consequences of changing the frequency of uniform price auctions for a risky asset. In their model, trading is faster when auctions are more frequent. Investors are strategic with interdependent and decreasing marginal valuations for owning a risky asset. They privately observe their inventory and signals about their valuations before trading. In this context, as in Vayanos (1999), slowing trading can be beneficial, because it reduces the scope for strategic behavior. Yet, with stochastic news arrivals, fast trading can be socially useful because it reduces the delay until the asset can be reallocated in response to news. This is in line with the finding, by Pagnotta and Phillipon (2013), that in faster markets, investors can realize gains from trade more rapidly.

In our analysis, each investor chooses the speed at which it operates on a given market. This contrasts with Pagnotta and Phillipon (2013) and Du and Zhu (2013), in which the frequency of trades is determined, for all investors, at the market level. Excessive market investment in speed can arise in Pagnotta and Phillipon (2013) because competing *markets* seek to differentiate from one another. In contrast, in our model, excessive investment arises because *investors* do not internalize the adverse selection cost they inflict on others. This problem arises even when there is no competition between markets – a case in which investment in speed is always socially optimal in Pagnotta and Phillipon (2013).

Our analysis of investors' choices between fast and slow markets echoes the analysis by Zhu (2014) of investors' choices between lit exchanges and dark pools. In both models, one market segment (dark pools in Zhu (2014) or slow markets in our case) is relatively more attractive for uninformed traders. However, the two papers focus on different economic mechanisms and different issues. A key driving force in Zhu (2014) is that the price formation mechanism is different in the dark–pool and the lit venue. In contrast, a key driving force in our model is the dual role of fast trading technologies (improving traders' ability to find attractive quotes and obtaining advance information about asset payoffs). In this context, Zhu (2014) focuses on price discovery, while we focus on gains from trade and welfare.

3. Model

Asset: Consider a risky asset trading at dates $\tau = 1, 2, ..., t, ..., \infty$. At the end of each trading round τ, the asset pays off cash-flow θ_τ, equal to $+\sigma > 0$ or $-\sigma$.[7] Across periods, cash-flows are i.i.d. For simplicity, we normalize to zero the unconditionally expected stream of cash-flows by setting $\Pr(\theta_\tau = +\sigma\ ) = \frac{1}{2}$.

Markets: To capture the fragmentation of the market, we assume there is a size–one continuum of trading venues, distributed on a circle and indexed clockwise from 0 to 1. Moreover, to model variations in liquidity conditions across trading venues, we assume that, at each period, only a fraction $\lambda < 1$ of the trading venues are liquid, in the sense that they offer

[7] Alternatively, one can assume that the asset pays off $V_T = \sum_{\tau=1}^{\tau=T} \theta_\tau$ at some random date $\tilde{T}$ and that θ_τ is publicly observed by all participants at date $\tau + 1$. With this interpretation, the $\theta_\tau s$ are innovations in the expected payoff of the asset due to the arrival of information over time. Results are identical with this specification but notations are slightly more complex because the expected payoff of the asset varies over time (it is $\sum_{\tau=1}^{\tau=t} \theta_\tau$ at date t).

attractive quotes. The other venues are illiquid, i.e., they post spreads that are too large to warrant trades, which, for simplicity, we model as lack of quotes. The set of venues offering attractive quotes varies from one period to the other. At date τ, the liquid venues are located on the circle in the interval of size λ, starting at x_τ; x_τ is uniformly distributed on the $[0, 1]$ circle, and i.i.d across periods. If $x_\tau + \lambda \leq 1$ then the set of trading venues offering attractive quotes is $[x_\tau, x_\tau + \lambda]$. If $x_\tau + \lambda > 1$ then the set of venues offering attractive quotes is $[x_\tau, 1] \cup [0, \lambda - (1 - x_\tau)]$.

Investors: In each period, a continuum of institutions enter the market. All are risk neutral and discount the future at rate r. For simplicity, each institution can only i) buy one unit of the asset, or ii) sell one unit of the asset, or iii) refrain from trading. To execute its desired trade, an institution must find a trading venue posting attractive quotes. When they enter the market, institutions are uncertain about which venues are liquid, i.e., they do not know x_τ. There are two types of institutions: fast and slow, with different abilities to search for quotes.

Fast institutions have extremely rapid connections with the markets. Thus, they can inspect all trading venues instantaneously and find a liquid one with certainty upon arrival. Rapid connections with markets also enable a fast institution to simultaneously observe prices of other assets with payoffs correlated with θ_τ [e.g., futures as suggested by Zhang (2013)]. These prices provide a signal on θ_τ, which, for simplicity, we assume to be perfect.

Slow institutions are less efficient at receiving price information from markets and searching for quotes. Hence, unlike fast institutions, they do not observe recent value relevant prices from other market, i.e., they do not observe signals about θ_τ. Moreover, it takes them longer to detect liquid trading venues. To capture this we assume that, within one period, a slow institution can inspect only one of the trading venues in the circle. Since it does not know x_τ, a slow institution randomly sends its orders to one of the trading venues, uniformly drawn from the unit circle. With probability λ, this trading venue is liquid, and the slow institution can trade. With the complementary probability, it is illiquid and the slow institution cannot trade during this period.

If a slow institution does not find a liquid trading venue during the period, then, with probability π, it can wait until the next period to search again for quotes. With the complementary probability, $1-\pi$, the institution exits the market and obtains a zero payoff. Thus, the likelihood, λ^s, that a slow institution eventually finds a liquid market is:

$$\lambda^s(\lambda,\pi) = \sum_{t=\tau}^{t=\infty} ((1-\lambda)\pi)^{t-\tau}\lambda = \lambda(1-(1-\lambda)\pi)^{-1}. \tag{1}$$

λ^s increases in π and λ and is equal to one when π or λ equal one.

Once an institution has found a liquid market, it decides optimally whether to trade or not. Then it leaves the market. Denote by α (resp. $1-\alpha$) the mass of new fast (resp. slow) institutions entering the market at each period and let I_τ be the mass of slow institutions that entered the market before date τ and are still in the market, searching for quotes at date τ. Given our assumptions, the law of motion for I_τ is:

$$I_{\tau+1} = F(I_\tau) \equiv (1-\alpha)(1-\lambda)\pi + I_\tau(1-\lambda)\pi. \tag{2}$$

The stationary level of I_τ, is the fixed point, I^*, of $F(\cdot)$, i.e.,

$$I^* = \frac{(1-\alpha)(1-\lambda)\pi}{(1-(1-\lambda)\pi)}. \tag{3}$$

We hereafter focus on the stationary regime, in which, at each period, a mass

$$((1-\alpha)+I^*)\lambda = (1-\alpha)\lambda^s \tag{4}$$

of slow institutions find quotes (where the equality follows from Eq. (3)).[8]

Valuations: We assume institutions' preferences are linear in common and private values. Formally, an institution with position $y \in \{-1,0,1\}$ and private value δ obtains utility flow

[8]The process by which slow institutions search for quotes in our model has the flavor of Duffie, Garleanu, and Pedersen (2005)'s model of search in over the counter (OTC) markets. The reason is that finding a good quote takes time both in fragmented lit markets or in OTC markets.

$y \times (\theta_\tau + r(1+r)^{-1}\delta)$ in period τ.[9] The common value, θ_τ, reflects the stream of cash flows. In addition, as in Duffie, Garleanu, and Pedersen (2005), each institution is endowed with its own private valuation, δ.[10] Differences in δ, across institutions capture in a simple way that other considerations than expected cash–flows affect investors' willingness to hold assets. For example, regulation can make it costly or attractive for certain investors, such as insurance companies, pension funds, or banks to hold certain asset classes. Differences in tax regimes can also induce differences in private values.

Denote by $v(\delta, i\theta_\tau)$ the total valuation for the asset of an institution at date τ, with $i = 0$ if the institution is slow and $i = 1$ if it is fast. For a slow institution,

$$v(\delta, 0) = \mathrm{E}(\Sigma_{t=\tau}^{t=\infty}(1+r)^{-(t-\tau)}(\theta_t + r(1+r)^{-1}\delta)) = \delta, \tag{5}$$

and for a fast institution,

$$v(\delta, \theta_\tau) = \mathrm{E}(\Sigma_{t=\tau}^{t=\infty}(1+r)^{-(t-\tau)}(\theta_t + r(1+r)^{-1}\delta)\,|\theta_\tau) = \delta + \theta_\tau. \tag{6}$$

We assume private valuations, δ, are i.i.d. across institutions and continuously distributed on $[-\overline{\delta}, \overline{\delta}]$ with cumulative distribution function $G(\cdot)$ and density $g(\cdot)$; $g(.)$ is symmetric around zero so that $G(0) = \frac{1}{2}$, $\mathrm{E}(\delta) = 0$, and $G(\delta) = \Pr(\delta_\tau \leq \delta) = \Pr(\delta_\tau \geq -\delta)$. Furthermore, we assume

$$\overline{\delta} \geq 2\sigma. \tag{7}$$

As shown below, Condition (7) implies that institutions with large valuations are always willing to trade at the equilibrium bid or ask price. This feature simplifies the exposition without qualitatively affecting results. In several examples, we shall consider the limit case in which $\overline{\delta} \to \infty$ and private valuations are normally distributed with standard deviation σ_δ.

[9] The scaling factor $r(1+r)^{-1}$ is just a convenient way to simplify computations. Removing it would not change any qualitative result.

[10] For simplicity, unlike in Duffie, Garleanu, and Pedersen (2005), the private valuation of an institution (δ) does not evolve stochastically through time.

Trading: We focus on stationary symmetric equilibria. In these equilibria, quotes are the same across all periods and all liquid trading venues and are symmetrically positioned around the unconditional expected payoff of the asset (zero). This is natural since all probability distributions are symmetric around zero and constant over time in our model. Hence, in each liquid venue, the institutions buy the asset at the ask price S and sell it at the bid price, $-S$, i.e., S is the half–spread or effective spread.

Since equilibrium quotes are the same in all liquid venues, fast institutions are indifferent between all of them. Correspondingly, we assume fast institutions are uniformly distributed across all liquid venues. Similarly, we assume that slow institutions who find liquid venues are distributed uniformly across these venues. Thus, each liquid venue is contacted by a mass α of fast institutions and a mass $(1-\alpha)\lambda^s$ of slow institutions.

Denote by $\omega^a(S,\theta_\tau)$ (resp. $\omega^b(S,\theta_\tau)$) the mass of institutions buying (resp. selling) the asset at date τ in each liquid venue. An institution that finds a liquid market optimally chooses to buy the asset if its valuation is greater than the ask price ($v(\delta, i\theta_\tau) \geq S$), to sell if its valuation is smaller than the bid price ($v(\delta, i\theta_\tau) \leq -S$), and to refrain from trading otherwise. Hence, if $\theta_\tau = \sigma$, we have:

$$\begin{aligned}
\omega^a(S,\sigma) &= \alpha \Pr(v(\delta,\sigma) \geq S) + (1-\alpha)\lambda^s \Pr(v(\delta,0) \geq S) \\
&= \alpha \overline{G}(S-\sigma) + (1-\alpha)\lambda^s \overline{G}(S), \qquad (8) \\
\omega^b(S,\sigma) &= \alpha \Pr(v(\delta,\sigma) \leq -S) + (1-\alpha)\lambda^s \Pr(v(\delta,0) \leq -S) \\
&= \alpha \overline{G}(S+\sigma) + (1-\alpha)\lambda^s \overline{G}(S), \qquad (9)
\end{aligned}$$

where $\overline{G}(x) = 1 - G(x)$. By symmetry, we have $\omega^b(S,-\sigma) = \omega^a(S,\sigma)$ and $\omega^a(S,-\sigma) = \omega^b(S,\sigma)$.

Our modeling of the trading process in each liquid market is similar to Zhu (2014). At the beginning of each period, in each liquid venue, quotes are posted by risk neutral competitive market makers, with zero private valuation for the asset. Then institutions' market orders arrive simultaneously and, in each venue, all buy orders are executed at the ask price, S, while all sell orders are executed at the bid price, $-S$. The equilibrium condition is that the competitive market makers break even in expectation.[11] Thus, the equilibrium (break–even) half–spread,

[11] This is slightly different from Glosten and Milgrom (1985) in which orders arrive one at a time and dealers

S^*, solves:

$$\mathrm{E}\left(\omega^a(S^*,\theta_\tau)(S^* - \Sigma_{t=\tau}^{t=\infty}(1+r)^{t-\tau}\theta_t) + \omega^b(S^*,\theta_\tau)(\Sigma_{t=\tau}^{t=\infty}(1+r)^{t-\tau}\theta_t + S^*)\right) = 0, \tag{10}$$

where the expectation is taken over $\theta_t, t \geq \tau$.

That fast institutions send informed market orders is in line with stylized facts. For example, Brogaard, Hendershott, and Riordan (2014) find that high–frequency traders trade in the direction of permanent price changes with market orders, and Baron, Brogaard, and Kirilenko (2014) that most of high–frequency traders' profits are generated by aggressive, liquidity–taking, trades.

Investment in the fast trading technology: All institutions entering the market at a given date simultaneously decide whether to be fast or slow, before observing their valuation for the asset. This choice determines the *level of fast trading* in the market, i.e., α (see Section 5). To be fast, an institution must invest in a fast trading technology, at cost C. This is the cost of the investment in infrastructures (computers, colocation, etc.) and intellectual capital (skilled traders, codes, etc.) required for quickly receiving information from markets, processing this information, and acting upon it.

Fig. 1 summarizes the description of the model by showing the sequence of play within one period.

[**Insert Fig. 1 About Here**]

4. Trading with fast and slow investors

In this section, we analyze equilibrium prices and trading decisions in a given trading round, for a given α. This sets the stage for studying the equilibrium level of fast trading, which is the focus of Section 5.

expect zero–profit on each order. As in Zhu (2014), while simplifying, this modeling choice is innocuous. In a previous version of this paper, trading was modeled as in Glosten and Milgrom (1985) and results were identical to those presented here.

4.1. Equilibrium Bid-Ask Spread: Existence, and Uniqueness

As (i) $\omega^b(S, -\sigma) = \omega^a(S, \sigma)$, (ii) $\omega^a(S, -\sigma) = \omega^b(S, \sigma)$, and (iii) cash-flows are i.i.d with mean zero, Eq. (10) is equivalent to:

$$\left[\omega^a(S^*, \sigma) + \omega^b(S^*, \sigma)\right] S^* = \left[\omega^a(S^*, \sigma) - \omega^b(S^*, \sigma)\right] \sigma. \tag{11}$$

The left–hand–side is the gross profit of a market maker, which is strictly positive whenever the spread is. The right–hand–side is the adverse selection cost borne by the market maker, which is strictly positive, as soon as $\alpha > 0$ and $\sigma > 0$. This adverse selection cost reflects that the net order flow in period τ $(\omega^a(S^*, \theta_\tau) - \omega^b(S^*, \theta_\tau))$ is positively correlated with the innovation in the cash–flow process (θ_τ). That is, there are more buyers than sellers $(\omega^a(S^*, \sigma) > \omega^b(S^*, \sigma))$ when the period cash–flow is high (and, symmetrically, fewer buyers than sellers when it is low: $\omega^a(S^*, -\sigma) < \omega^b(S^*, -\sigma)$).

In our trading environment, an equilibrium is a spread S^*, such that Eq. (11) holds. In equilibrium, market makers must charge a bid-ask spread $(S^* > 0)$ to cover the adverse selection cost. Thus, institutions with valuations in $[-S^*, S^*]$ choose not to trade because their expected gain from trade is smaller than their trading cost, S^*. This generates a welfare loss because, for all $\delta \neq 0$, gains from trade exist between market makers and institutions.

As shown in the appendix, Eq. (11) always has at least one solution, $0 \leq S^* \leq \sigma$. Hence, we can state the following lemma.

Lemma 1. *Equilibrium exists. When $\alpha = 0$ or $\sigma = 0$, the unique equilibrium is $S^* = 0$. Otherwise, equilibrium is not necessarily unique but in all equilibria $0 < S^* < \sigma$.*

The equilibrium bid-ask spread is not necessarily unique because an increase in the spread can generate an increase in both the revenue and the adverse selection cost for the market maker. For instance, suppose that fast institutions receive a good signal. An increase in the spread, S, decreases the fraction of fast institutions who buy and sell the asset but the effect can be stronger for those who decide to sell. In this case, the adverse selection cost $((\omega^a(S, \sigma) - \omega^b(S, \sigma))$ increases with S. When this happens, market makers' net expected profit (the difference between the left and right hand sides of Eq. (11)) is not necessarily monotonic in the half spread and, for this reason, there might be multiple spreads for which market makers break even, i.e., for which Eq. (11) holds.[12] The next example illustrates this point.

Example 1. Suppose that institutions' private valuations are normally distributed with standard deviation σ_δ. Fig. 2 plots market makers' net expected profit when $\alpha = 0.1$, $\sigma = 3$, $\lambda = 0.8$, and $\pi = 1$, for $\sigma_\delta = 0.9$ or $\sigma_\delta = 2$. Equilibrium bid-ask spreads are those for which market makers' net expected profit is zero. When $\sigma_\delta = 0.9$, the market makers' net expected profit is non monotonic in S. For this reason, there are several values of S^* such that Eq. (11) holds: $S^* = 0.47$, $S^* = 1.37$, and $S^* = 2.96$. In contrast, when $\sigma_\delta = 2$, market makers' net expected profit decreases in S everywhere and, as a result, there is a unique equilibrium spread, $S^* = 0.29$.■

[**Insert Fig. 2 about here**]

When there exist multiple solutions to Eq. (11), economic reasoning suggests to select spreads that cannot be profitably undercut, as other spreads would attract competition. Consider Fig. 2 again. Bid-ask spreads $S^* = 1.37$ and $S^* = 2.96$ satisfy the zero net profit condition (i.e., they solve Eq. (11)) but they can be profitably undercut because any spread sufficiently close to and above $S^* = 0.47$ yields a strictly positive expected profit. This is a more general principle: When several bid-ask spreads solve Eq. (11), only the smallest one cannot be profitably undercut.

Lemma 2. *Let $S^*_{\min}(\alpha)$ be the smallest solution to Eq.* (11)*. This equilibrium bid-ask spread is the only one that cannot be profitably undercut.*

Hence, if one adds the natural economic requirement that an equilibrium spread should not be profitably undercut, then equilibrium is always unique. Therefore we hereafter focus on

[12] Glosten and Milgrom (1985) and Dow (2005) underscore the possibility of multiple equilibria in financial markets because of virtuous circles (traders anticipate the market will be liquid, hence they submit lots of orders, hence the market is liquid) or vicious circles (where illiquidity is a self–fulfilling prophecy). The same phenomenon is at play here.

$S^*_{\min}(\alpha)$, to which we refer when we write "the equilibrium spread". For our purpose, focusing on $S^*_{\min}(\alpha)$ is conservative because we analyze welfare losses generated by excessive investment in the fast trading technology. Larger spreads would increase these losses.

4.2. Bid-ask spreads, Trading Volume, and Fast Trading

The next proposition spells out how the level of fast trading, α, affects the spread and trading volume.

Proposition 1. *The equilibrium bid-ask spread increases in the level of fast trading, α, and the volatility of the asset fundamental value, σ. It (weakly) decreases with the likelihood that a slow institution finds a trading opportunity, λ.*

When α increases or λ decreases, orders are more likely to come from fast institutions. Hence, the adverse selection cost is higher for market makers. Thus, the bid-ask spread increases. Fig. 3 illustrates this testable implication when the distribution of traders' private valuation is normal.

[Insert Fig. 3 about here]

Denote by $\Delta\text{Vol}(S^*(\alpha), \alpha)$ the difference between the likelihood of a trade by a fast and a slow institution. As institutions buy the asset if their valuation is higher than $S^*(\alpha)$ and sell it if their valuation is lower than $-S^*(\alpha)$, we have:

$$\begin{aligned}\Delta\text{Vol}(S^*(\alpha), \alpha) &= [\Pr(v(\delta, \theta_\tau) \geq S^*(\alpha)) - \lambda^s \Pr(v(\delta, 0) \geq S^*(\alpha))] \\ &+ [\Pr(v(\delta, \theta_\tau) \leq -S^*(\alpha)) - \lambda^s \Pr(v(\delta, 0) \leq -S^*(\alpha))]. \end{aligned} \tag{12}$$

Because of the symmetry of institutions' private valuations around zero, we have

$$\Delta\text{Vol}(S^*(\alpha), \alpha) = 2\left[(\Pr(v(\delta, \theta_\tau) \geq S^*(\alpha)) - \lambda^s \Pr(v(\delta, 0) \geq S^*(\alpha))\right]. \tag{13}$$

$\Delta\text{Vol}(S^*(\alpha),\alpha)$ reflects the difference between the respective contributions of fast and slow institutions to trading volume. As such it offers a measure of the toxicity of the order flow. Moreover, since the gains from trade of each of the two categories of institutions are related to their trading volume, $\Delta\text{Vol}(S^*(\alpha),\alpha)$ is also related to the difference between the gains from trade of fast and slow institutions. Because of the important role played by $\Delta\text{Vol}(S^*(\alpha),\alpha)$ in our analysis (see, e.g., Eq. (24)), it is useful to analyze its economic determinants.

Straightforward manipulations of Eq. (13) yield

$$\Delta\text{Vol}(S^*(\alpha),\alpha) = 2(1-\lambda^s)(1-G(S^*)) + (G(S^*) - G(S^*-\sigma)) + (G(S^*) - G(S^*+\sigma)). \quad (14)$$

The first two terms in Eq. (14) are positive while the latter is negative. Thus, the sign of ΔVol is ambiguous. To explain why, we now discuss the economic interpretation of Eq. (14).

Consider an institution with private valuation δ and without advance information on the period cash flow. This institution is willing to trade if $|\delta| > S^*$, which happens with probability $2(1-G(S^*))$. The institution is able to trade if it finds a liquid venue before leaving the market. The probability of this event is one if the institution is fast and λ^s if it is slow. Thus, the first term in Eq. (14), $2(1-\lambda^s)(1-G(S^*))$, reflects the increase in the likelihood of a trade for an institution due to more efficient search for quotes.

Now, the fast trading technology also provides advanced information on cash-flows, which affects an institution's valuation for the asset. This effect cuts both ways in term of incentives to trade for a fast institution. First, cash–flow news can trigger trading by an institution that would not have traded without information. This effect plays out when $|\delta| \in [S^*-\sigma, S^*]$ and corresponds to the second term in Eq. (14). On the other hand, cash–flow news can prevent trading by an institution that would have traded without information. This happens when $|\delta| \in [S^*, S^*+\sigma\ ]$ and corresponds to the third, negative, term in Eq. (14).[13]

There exist specifications of G, the distribution of institutions' private valuations (see Ex-

[13] For instance, consider an institution with $S^* < \delta < S^*+\sigma$. If it is slow, this institution buys the asset. However, if it is fast and learns that the cash-flow of the asset is low, this institution does not trade because its valuation, $\delta-\sigma$, is then within the bid-ask spread. The same effect happens when the asset cash flow is high and $-(S^*+\sigma) < \delta < -S^*$. Thus this effect reduces the likelihood of trading for a fast institution by the probability that $|\delta| \in [S^*, S^*+\sigma\ ]$, which is $G(S^*+\sigma) - G(S^*)$.

ample 2 in Section 5.2) such that, for some values of α, the third (negative) term in Eq. (14) dominates the two other (positive) ones. In these cases, $\Delta\text{Vol}(S^*(\alpha), \alpha) < 0$. That is, in spite of their efficacy at searching for quotes, fast institutions trade less than slow ones in equilibrium. This is because they are often conflicted between a positive (resp. negative) private valuation and a negative (resp. positive) cash–flow signal. As a result, their valuation is neither large, nor small enough to overcome the cost of trading, S^*.

The next lemma provides a sufficient condition on the distribution of institutions' private valuations such that fast institutions are more likely to trade than slow institutions (i.e., $\Delta\text{Vol}(S^*(\alpha), \alpha) > 0$) for all values of α. Let $h_g(.)$ be the hazard rate of the distribution of institutions' private valuations, that is, $h_g(\delta) = g(\delta)/(1 - G(\delta))$.

Lemma 3. *In equilibrium* $\Delta\,Vol(S^*(\alpha), \alpha) > 0$*,* $\forall\alpha$*, i.e., fast institutions are more likely to trade than slow institutions, if one of the following conditions is satisfied:*

1. $h_g(\delta)$ *decreases in* δ*.*
2. $h_g(\delta)$ *increases in* δ *, and either (i)* $\lambda^s \leq \frac{1}{2}$*, or (ii)* $\lambda^s > \frac{1}{2}$ *and*

$$2G(\sigma) + \left(\frac{1 - G(2\sigma)}{1 - G(\sigma)}\right) \geq 2\lambda^s. \tag{15}$$

The second case in Lemma 3 is maybe more relevant because, for a large class of probability distributions (e.g., all log concave distributions such as the normal distribution or the uniform distribution), the hazard rate is increasing [see Bagnoli and Bergstrom (2005)]. Condition (15) is satisfied if $\sigma \geq Max\{G^{-1}(\lambda^s), 0\}$, i.e., when the asset volatility is sufficiently high.

Equilibrium trading volume is

$$\begin{aligned} \text{Vol}(S^*(\alpha), \alpha) &= \omega^a(S^*(\alpha), \sigma) + \omega^b(S^*(\alpha), \sigma) \qquad (16) \\ &= \alpha(2 - (G(S^*(\alpha) - \sigma) + G(S^*(\alpha) + \sigma)) + \lambda^s(1 - \alpha)(1 - G(S^*(\alpha)), \end{aligned}$$

where the second equality comes from Eq. (8) and (9). We deduce that:

$$\frac{d\mathrm{Vol}(S^*,\alpha)}{d\alpha}=\frac{\partial\mathrm{Vol}(S^*,\alpha)}{\partial\alpha}+\frac{\partial\mathrm{Vol}(S^*,\alpha)}{\partial S^*}\frac{\partial S^*}{\partial\alpha}=\Delta\mathrm{Vol}(S^*(\alpha),\alpha)+\underbrace{\frac{\partial\mathrm{Vol}(S^*,\alpha)}{\partial S^*}\frac{\partial S^*}{\partial\alpha}}_{<0}. \tag{17}$$

An increase in the level of fast trading, α, has two effects. First, it shifts some institutions from the pool of slow to the pool of fast investors. This increases trading volume iff fast institutions are more likely to trade than slow institutions (i.e., $\Delta\mathrm{Vol}(S^*(\alpha),\alpha)>0$). Second, the increase in the level of fast trading raises the bid-ask spread ($\frac{\partial S^*}{\partial\alpha}>0$). This effect leads a larger fraction of institutions to abstain from trading, which reduces trading volume. Thus, the effect of an increase in the level of fast trading on volume is ambiguous, and, for this reason, trading volume is in general non monotonic in this level.

[**Insert Fig. 4 about here**]

Fig. 4 illustrates this point when investors' private valuations are normally distributed. It depicts equilibrium trading volume ($\mathrm{Vol}(S^*(\alpha),\alpha)$) as a function of α for various values of λ. The possibility of a negative effect of fast trading on the volume of trade is in line with Jovanovic and Menkveld (2011), who find that, for Dutch stocks, the entry of a fast trader on Chi-X led to a drop in volume.[14]

5. Equilibrium investment in fast trading technologies

We now turn to the equilibrium determination of α. To do so, we first analyze the gains of fast and slow institutions, which are then compared to determine investment decisions.

[14]Anecdotal evidence also suggests that, as high–speed trading expands, trading volume can increase or decrease. For example, an article entitled "*Electronic trading slowdown alert*" published in the Financial Times on September 24, 2010 (page 14) describes a sharp drop in trading volume in 2010 from a high of about $7,000 billions in April 2010 to a low of $4,000 billions in August 2010. The article explicitly points to changes in market structures as a cause for this reversal in trading volume.

5.1. *Comparing the gains of fast and slow institutions*

Denote the ex-ante expected gains of slow and fast institutions by $\psi(\alpha)$ and $\phi(\alpha)$, respectively.[15] Let $\omega(\delta, i\theta_\tau)$ be the market order of an institution with private valuation δ and type $i \in \{1, 0\}$ ($i = 1$ if the institution is fast and $i = 0$ if slow), in equilibrium. Thus, for fast traders,

$$\phi(\alpha) = \mathrm{E}((v(\delta, \theta_\tau) - S^*(\alpha))\omega(\delta, \theta_\tau)). \tag{18}$$

Recall that $\omega(\delta, i\theta_\tau) = 1$ if $v(\delta, i\theta_\tau) \geq S^*$, $\omega(\delta, i\theta_\tau) = -1$ if $v(\delta, i\theta_\tau) \leq -S^*$, and $\omega(\delta, i\theta_\tau) = 0$, otherwise. Hence, using the symmetry of the distribution of institutions' private valuations around zero,

$$\phi(\alpha) = \int_{S^*(\alpha)-\sigma}^{\bar{\delta}} (\delta + \sigma - S^*(\alpha))g(\delta)d\delta + \int_{S^*(\alpha)+\sigma}^{\bar{\delta}} (\delta - \sigma - S^*(\alpha))g(\delta)d\delta. \tag{19}$$

The first term in Eq. (19) is the gain of fast institutions when they trade in the direction of the asset cash-flow (e.g., buy it when its cash-flow is high). The second term is their gain when they trade against the asset cash flow. For instance, they might buy the asset even when its cash-flow is low when $v(\delta, -\sigma) \geq S^*(\alpha)$. The mass of such institutions is never zero when Condition (7) holds, that is, when $\bar{\delta} \geq 2\sigma$, because $S^*(\alpha) + \sigma \leq 2\sigma$.

In a given period, a slow institution finds a liquid venue with probability λ. With probability $(1 - \lambda)$, it does not. In that case, with probability π, it keeps searching for quotes at the next period. The discounted expected gains from trade of a slow institution are therefore:

$$\psi(\alpha) = \sum_{t=\tau}^{t=\infty} \left(\frac{(1-\lambda)\pi}{1+r}\right)^{t-\tau} \lambda\mathrm{E}((v(\delta, 0) - S^*(\alpha))\omega(\delta, 0)). \tag{20}$$

Using the symmetry of the distribution of institutions' private valuations around zero and

[15] Ex-ante means just before institutions learn their private valuations and enter the market. Alternatively, all institutions could choose to be fast or slow at date $\tau = 0$. In this case, one must discount their expected payoff at arrival date, $\psi(\alpha)$ and $\phi(\alpha)$, appropriately. As the discount factor is identical for all institutions, results in this case are identical to those obtained when institutions decide to be slow or fast just before entering the market.

simplifying, we obtain:

$$\psi(\alpha) = 2\mu(\lambda, \pi, r) \int_{S^*(\alpha)}^{\overline{\delta}} (\delta - S^*(\alpha))g(\delta)d\delta, \tag{21}$$

where

$$\mu(\lambda, \pi, r) = \frac{\lambda^s(\pi, \lambda)(1 - (1 - \lambda)\pi)}{(1 - (1 - \lambda)\pi(1 + r)^{-1})}. \tag{22}$$

The integral in Eq. (21) accounts for the gains from trade of the institution when it finds a liquid venue. The scaling term, $\mu(\lambda, \pi, r)$, reflects the cost of delayed execution induced by the search for quotes in fragmented markets. This cost is large (μ is small) when the discount rate, r, is high or when the risk of exiting the market without trading is high (i.e., π is small). In contrast, an increase in λ reduces the cost of delayed execution because it increases the speed (λ^s) at which an institution finds a counterparty.

It is clear from Eq. (21) that the risk free rate, r, affects institutions' payoffs only through its effect on μ. This is the only effect of r on the variables of interest in the model. Thus, the economic effect of a decrease in r is similar to that of an increase in π: it makes delays in execution less costly for institutions. For simplicity, from now on, we assume $r = 0$. In this case, the expression for $\mu(\lambda, \pi, r)$ simplifies to $\mu(\lambda, \pi, 0) = \lambda^s(\pi, \lambda)$. This lightens the exposition without affecting any findings (see the on-line appendix).

Fast institutions obtain higher expected gains than slow institutions because (i) they have zero delay costs and (ii) they obtain speculative profits by trading on advance information. These speculative profits, however, generate adverse selection costs for other market participants. Thus, fast traders generate a negative externality for the other market participants. The larger the level of fast trading, α, the larger this negative externality. The next proposition summarizes the above discussion.

Proposition 2. *In equilibrium, the expected profit of fast institutions, gross of the technological cost, is always higher than the expected profit of slow institutions: $\phi(\alpha) > \psi(\alpha)$. Moreover, an increase in the level of fast trading, α, reduces the expected gains of slow and fast institutions.*

5.2. *Strategic substitutability or complementarity*

For a given level of α, the net expected profits of fast and slow institutions are $\phi(\alpha) - C$ and $\psi(\alpha)$, respectively. Thus, an institution is better off investing if and only if:

$$\phi(\alpha) - \psi(\alpha) \geq C. \tag{23}$$

As ϕ and ψ vary with α, the profitability of investment for one institution depends on other institutions' decisions. Thus, investment choices are interdependent. If $\phi - \psi$ decreases in α, then fast institutions loose more than slow ones when α goes up. In this case, institutions' investment decisions are strategic substitutes: the greater the level of fast trading, the lower the relative value of fast trading. In contrast, if $\phi - \psi$ increases in α, slow institutions are hurt more than fast ones by an increase in α. Institutions' investment decisions are then strategic complements and mutually reinforcing: the greater the level of investment in the fast trading technology, the more profitable it is to invest in it.

Let $\Delta(\alpha) = \phi(\alpha) - \psi(\alpha)$ denote the relative value of being fast. Institutions' decision to be fast are substitutes if $\frac{\partial \Delta(\alpha)}{\partial \alpha} < 0$ and complements if $\frac{\partial \Delta(\alpha)}{\partial \alpha} > 0$. Using Eq. (19) and (21), we obtain after some algebra (see the on-line appendix) that:

$$\frac{\partial \Delta(\alpha)}{\partial \alpha} = -\frac{\partial S^*}{\partial \alpha} \times \Delta \mathrm{Vol}(S^*(\alpha), \alpha), \tag{24}$$

where $\Delta \mathrm{Vol}(S^*(\alpha), \alpha)$ (given in Eq. (14)) is the difference between the likelihood of a trade for a fast and a slow institution in equilibrium.

The equilibrium bid-ask spread increases with α (Proposition 1). Hence, by Eq. (24), institutions' decisions are locally substitutes if $\Delta \mathrm{Vol}(S^*(\alpha), \alpha) > 0$ and locally complements if $\Delta \mathrm{Vol}(S^*(\alpha), \alpha) < 0$. This is intuitive: the increase in the cost of trading ($S^*(\alpha)$) due to an increase in the level of fast trading hurts more those institutions that trade more. If the distribution of institutions' private valuations satisfies one of the conditions in Lemma 3 then $\Delta \mathrm{Vol}(S^*(\alpha), \alpha) > 0$ for all values of α. Thus, we obtain the following result.

Corollary 1. *Under the conditions of Lemma 3, $\Delta Vol(S^*(\alpha),\alpha) > 0$, $\forall\alpha$. In this case, the relative value of being fast ($\Delta(\alpha)$) decreases in α for all values of α: $\frac{\partial\Delta(\alpha)}{\partial\alpha} < 0$, $\forall\alpha$.*

Hence, under fairly general conditions (given in Lemma 3), institutions' decisions to invest in the fast trading technology are *globally* (i.e., for all values of α) substitutes. In contrast, institutions' investment decisions are never globally complements because they are always substitutes for α sufficiently close to zero.[16] Yet, when the conditions of Lemma 3 are not satisfied, institutions' decisions can be complements for some range of α, as illustrated by the next example.

Example 2. Define $\gamma = \left(\overline{\delta} - \varphi\left(\overline{\delta} - \sigma\right)\right)/\sigma$ with $\varphi \in [1, \frac{\overline{\delta}}{\overline{\delta}-\sigma}]$. Assume $g(\delta) = \varphi(2\overline{\delta})^{-1}$ if $-\overline{\delta} \leq \delta \leq -\sigma$, $g(\delta) = \gamma(2\overline{\delta})^{-1}$ if $-\sigma \leq \delta \leq \sigma$, and $g(\delta) = \varphi(2\overline{\delta})^{-1}$ if $\sigma \leq \delta \leq \overline{\delta}$. The conditions on γ and φ guarantee that the cumulative probability distribution of δ is symmetric around zero and that it is equal to one when $\delta = \overline{\delta}$. If $\varphi = 1$ then $\gamma = 1$ and private valuations are uniformly distributed. If $\varphi > 1$ then $\gamma < 1$. In this case, the mass of institutions with extreme valuations (between $[-\overline{\delta}, -\sigma]$ or $[\sigma, \overline{\delta}]$) is greater than the mass of traders with intermediate private valuations (in $[-\sigma, \sigma]$). In this context we obtain the following corollary.

Corollary 2. *Suppose the distribution of institutions' private valuations is as defined in example 2. If $\lambda^s(\lambda,\pi) > Min\{1, \frac{2\overline{\delta}-(\gamma+\varphi)S^*(1)}{2(\overline{\delta}-\gamma S^*(1))}\}$ then there exists a threshold α_0, such that $\Delta Vol(S^*(\alpha),\alpha) < 0$ iff $\alpha > \alpha_0$, i.e., institutions' investment decisions are substitutes for $\alpha \leq \alpha_0$ and complements for $\alpha > \alpha_0$. If $\lambda^s(\lambda,\pi) \leq Min\{1, \frac{2\overline{\delta}-(\gamma+\varphi)S^*(1)}{2(\overline{\delta}-\gamma S^*(1))}\}$ then $\Delta Vol(S^*(\alpha),\alpha) > 0$ for all α and, therefore, institutions investment decisions are substitutes for any level of fast trading.*

Fig. 5 illustrates the corollary when $\varphi = 1.5$, $\sigma = 3$, $\overline{\delta} = 7$, $\lambda = 0.5$, and $\pi = 0.99$ (so that $\lambda^s \approx 0.99$). In this case, $\alpha_0 \approx 0.25$. Thus, institutions' decisions are complements for $\alpha > 0.25$ and substitutes when $\alpha \leq 0.25$.■

[Insert Fig. 5 about here]

[16]Eq. (14) yields $\Delta\text{Vol}(S^*(0), 0) = (1-\lambda^s) > 0$ because $S^*(0) = 0$. Thus, at least at $\alpha = 0$ (and by continuity for values of α close to zero), a small increase in fast trading always reduces the relative value of being fast.

5.3. Equilibrium fast trading

If

$$\phi(1) - \psi(1) > C, \tag{25}$$

then institutions prefer to invest when they expect all the others to do so. Hence, $\alpha^* = 1$ is an equilibrium if Condition (25) holds. Symmetrically, if

$$\phi(0) - \psi(0) < C, \tag{26}$$

then institutions prefer not to invest when they expect the others also will not. Hence, $\alpha^* = 0$ is an equilibrium if Condition (26) holds. Finally, α^* is an interior equilibrium if, when institutions expect that a fraction α^* of institutions will invest, they are indifferent between investing and not investing:

$$\phi(\alpha^*) - \psi(\alpha^*) = C. \tag{27}$$

As $\phi(\alpha) - \psi(\alpha)$ is continuous in α, at least one of these three equilibrium conditions must hold. Thus, an equilibrium level of fast trading always exists. Furthermore, if $\Delta(0) = \phi(0) - \psi(0) > C$, then each institution is better off being fast if it expects others to be slow. Thus, in this case, it cannot be an equilibrium that all institutions prefer to be slow. The next proposition summarizes the above discussion.

Proposition 3. *An equilibrium level of fast trading exists. Moreover, there is some investment in the fast trading technology ($\alpha^* > 0$) if $\Delta(0) > C$.*

To gain insights into the economics of the decision to be fast, it is useful to write $\Delta(0)$ explicity:

$$\Delta(0) = (1 - \mu(\lambda, \pi, 0))E(|\delta|) + 2(2G(\sigma) - 1)(\sigma - E(|\delta|\,|\delta| \leq \sigma)). \tag{28}$$

Hence, the increase in expected profit for an institution that becomes fast when all others are slow, $\Delta(0)$, is the sum of two terms, which can be interpreted as the "search value" and the "speculative value" of the fast trading technology, respectively (the formal derivation of Eq. (28) is given in the on-line appendix for brevity).

First consider the "search value". When $\alpha = 0$, the bid-ask spread is zero ($S^*(0) = 0$). Thus, expected gains from trade are $E(|\delta|)$ for all institutions finding a trading venue because the spread is zero in this case. Slow institutions, however, can only appropriate a fraction μ of this gain because of delayed execution. In contrast, fast institutions obtain 100% of the expected gains from trade because they bear no delay costs. Thus, adoption of the fast trading–technology generates an increase in expected profit due to more efficient search equal to $(1-\mu)E(|\delta|)$ for the first adopter. Hence, the first term in (28) measures the "*search value*" of the trading technology (when $\alpha = 0$).

Now turn to the "speculative value" of the fast trading technology, which arises when $\sigma > 0$ and is given by the second term in Eq. (28). To grasp the economic intuition of that term, first observe that the technology has speculative value *only* when it leads an institution to trade differently than if it were slow. Suppose that a fast institution learns that the asset cash flow is high ($\theta_\tau = \sigma$). As $\alpha = 0$, it buys the asset iff $\delta + \sigma - S^*(0) > 0$, i.e., iff $\delta \geq -\sigma$. However, the institution would have purchased the asset anyway if slow when $\delta \geq 0$. Thus, the technology has speculative value only when $-\sigma \leq \delta < 0$. In this case, if fast, the institution buys the asset and earns $\delta + \sigma$ whereas if slow it *sells* the asset and earns $-(\delta + \sigma)$. Thus, the net speculative gain of the technology is $\delta + \sigma - (-(\delta + \sigma)) = 2(\delta + \sigma)$, conditional on $-\sigma \leq \delta < 0$ and $\theta_\tau = \sigma$. This generates an average speculative gain of $2(G(0) - G(-\sigma))\,(\sigma - E(|\delta|\,||\delta| \leq \sigma))$ when $\theta_\tau = \sigma$.[17] By symmetry, this is also the average speculative gain when $\theta_\tau = -\sigma$. Thus, the total average speculative value of the fast trading technology is $2(G(0) - G(-\sigma))\,(\sigma - E(|\delta|\,||\delta| \leq \sigma)) = (2G(\sigma) - 1)(\sigma - E(|\delta|\,||\delta| \leq \sigma))$ because $G(.)$ is symmetric around 0.

The previous calculations hold for $\alpha = 0$ (i.e., for very first adopters of the fast trading technology). More generally, for any value of α, the gain of being fast, $\Delta(\alpha) = \phi(\alpha) - \psi(\alpha)$, has a search value and a speculative value component. Closed-form expressions for these components, however, cannot be obtained for $\alpha > 0$ because they depend on the equilibrium spread, $S^*(\alpha)$, which in general cannot be computed in closed-form for $\alpha > 0$. However, if one of the conditions of Lemma 3 holds, then $\partial\Delta(\alpha)/\partial\alpha < 0, \forall\alpha$ and we have the following proposition.

Proposition 4. *When $\Delta(\alpha)$ decreases for all values of α, then there exists a unique equilibrium*

[17] Indeed, $E(\theta_\tau + \delta\,|\theta_\tau = \sigma, -\sigma \leq \delta < 0) = \sigma - E(|\delta|\,||\delta| \leq \sigma)$ because of the symmetry of institutions' private valuations.

level of fast trading. This level is such that: if (a) $C \geq \Delta(0)$, $\alpha^* = 0$, *if (b)* $\Delta(1) < C < \Delta(0)$, $0 < \alpha^* < 1$, *and if (c)* $C \leq \Delta(1)$, $\alpha^* = 1$. *Furthermore, as* C *increases, the level of fast trading declines in equilibrium.*

Fig. 6 illustrates the determination of α^* when institutions' private valuations are normally distributed. This level is obtained at the intersection of i) the horizontal line that gives the value of C and ii) the downward sloping curve reprsenting $\Delta(\alpha)$. In this example, $\Delta(0) = 3.64$ and $\Delta(1) = 1.42$. Thus, for $C = 3$, there is an interior equilibrium, $\alpha^* \approx 0.26$. As the cost of fast trading increases (the horizontal line shifts up in Fig. 6), the level of fast trading declines.

[Insert Fig. 6 about here]

Now consider the case in which, at least for some ranges of α, institutions' decisions are complements. In that case, $\Delta(\alpha)$ does not decrease everywhere and, for this reason, there might be multiple equilibrium levels of fast trading. This is particularly striking when $\Delta(0) \leq C < \Delta(1)$. In this case, there are at least two equilibria: one in which no institution finds it optimal to invest because each expects others not to invest and one in which all institutions find it optimal to invest because each expects others to invest. In each of the two equilibria, institutions' beliefs about other institutions' decisions are self-fulfilling. Yet, all institutions would prefer to coordinate on not being fast because their expected profit ($\psi(0)$) in the "All-Slow" equilibrium is larger than their expected profit ($\phi(1) - C$) in the "All-Fast" equilibrium when $\Delta(0) \leq C$. Indeed, this condition implies that: $\psi(0) \geq \phi(0) - C$, which is strictly larger than $\phi(1) - C$ since $\phi(.)$ is decreasing. Thus, the All-Slow equilibrium Pareto dominates the All-Fast equilibrium. The All-Fast equilibrium can be interpreted as the outcome of an arms race, similar to Glode, Green, and Lowery (2012): every institution chooses to invest in the fast trading technology in the fear that others do so. This belief is indeed self-fulfilling.

Fig. 5 illustrates these points. In this case, $\Delta(\alpha)$ is a U-shape function of α with a minimum in $\alpha_{\min} = 25\%$ and a maximum in $\alpha_{\max} = 1$. Institutions' decisions are complements for $\alpha \in [\alpha_{min}, 1]$. Furthermore, $\Delta(0) = 0.25$ and $\Delta(1) = 0.27$. Thus, for any C, in $(0.25, 0.27)$, there are three equilibria: the two corner equilibria and one interior equilibrium. For instance, as Fig. 5 shows, when $C = 0.264$, there are three possible equilibrium levels of fast trading: (i)

All Slow ($\alpha^* = 0$), (ii) All Fast ($\alpha^* = 1$), and (iii) $\alpha_3^* = 83.5\%$. Following Manzano and Vives (2012), we say that an equilibrium α^* is stable if when one slightly perturbates α around α^* and, at this point, (i) reduces α if $\Delta(\alpha) < C$ or (ii) increases α if $\Delta(\alpha) > C$ then one is brought back to α^*. Inspecting Fig. 5, one can immediately see that the interior equilibrium, α_3^*, is not stable: a small increase in the fraction of fast institutions at this point triggers a domino effect that leads all institutions to be fast. This would appear as an investment wave in fast trading technologies, as if fast trading were contagious. In contrast, the corner equilibria are stable.[18]

6. Social Optimum and Policy Intervention

As explained in the previous section, the decision to become fast by one institution exerts a negative externality on other institutions (see Proposition 2). As institutions do not internalize this externality in making their investment decision, one expects the equilibrium level of fast trading to be too high relative to the level that maximizes social welfare. We show that this is indeed the case in Section 6.1. We then analyze possible policy responses to this problem in Section 6.2.[19]

6.1. Excessive Fast Trading

Utilitarian welfare is equal to[20]

$$W(\alpha) = \alpha\left(\phi(\alpha) - C\right) + (1-\alpha)\psi(\alpha). \tag{29}$$

[18]When an interior equilibrium, α^*, is stable, $\Delta(\alpha)$ must necessarily be decreasing at $\alpha = \alpha^*$. This is consistent with Manzano and Vives (2012), who find that only equilibria in which agents' actions are strategic substitutes are stable. This principle, however, does not apply for corner equilibria. For instance, in Fig. 6, $\Delta(\alpha)$ is increasing at $\alpha = 1$. Yet, $\alpha^* = 1$ is a stable equilibrium.

[19]All institutions optimally decide to trade or not (see Section 3). Hence, there are no noise traders in our setting and all investors optimally adjust their trading strategies when the market structure (e.g., the fraction of fast institutions) changes. We can therefore conduct welfare and policy analyses because all investors' welfare and responses to changes in market structure are well defined.

[20]Market makers are risk neutral and obtain zero–expected profits. Hence, their contribution to utilitarian welfare is equal to zero.

Thus, a marginal increase in the level of fast trading has the following effect on welfare:

$$\frac{\partial W(\alpha)}{\partial \alpha} = \Delta(\alpha) - C - \left[-\alpha \frac{\partial \phi(\alpha)}{\partial \alpha} - (1-\alpha) \frac{\partial \psi(\alpha)}{\partial \alpha} \right]. \tag{30}$$

The term within brackets is positive because an increase in α reduces fast and slow institutions' expected gains ($\frac{\partial \phi(\alpha)}{\partial \alpha} \leq 0$ and $\frac{\partial \psi(\alpha)}{\partial \alpha} \leq 0$; see Proposition 2). It measures the externality cost incurred by all institutions when α increases. Denoting this cost by $C_{ext}(\alpha)$, we have

$$\frac{\partial W(\alpha)}{\partial \alpha} = \underbrace{\Delta(\alpha)}_{\text{Social Value of Fast Trading}} - \underbrace{(C + C_{ext}(\alpha))}_{\text{Social Cost of Fast Trading}}. \tag{31}$$

Thus, a marginal increase in α has two opposite effects on social welfare. On the one hand, institutions who become fast are better off. This benefit is captured by the first term in Eq. (31). On the other hand, institutions who become fast pay a cost C and exert a negative externality on all institutions.

The socially optimal level of fast trading, α^{SO}, (i.e., the value of α maximizing $W(\alpha)$) trades off the social benefit of fast trading ($\Delta(\alpha)$) and its social cost ($C + C_{ext}(\alpha)$). This level is not necessarily zero because, at $\alpha = 0$, the social value of the fast trading technology can exceed its social cost (see Proposition 6 below). Yet, when $\sigma > 0$, the socially optimal level of fast trading is always smaller (and in most cases strictly smaller) than the equilibrium level of fast trading, as stated in the next proposition.

Proposition 5. *When $\sigma > 0$, the socially optimal level of fast trading is smaller than the equilibrium one ($\alpha^{SO} \leq \alpha^*$), with a strict inequality when equilibrium is interior ($0 < \alpha^* < 1$). When $C \geq \Delta(0)$, the socially optimal level of fast trading is zero and this level is also an equilibrium (but not necessarily the unique equilibrium). When $C \leq \Delta(1)$, $\alpha^* = 1$ is an equilibrium and the socially optimal level of fast trading is either lower than or equal to the equilibrium level.*

Thus, in general, there is overinvestment in the fast trading technology in equilibrium.

Fig. 7 illustrates this result. It depicts social welfare when $\sigma = 5$ and institutions' private valuations are normally distributed with $\sigma_\delta = 10$ for $C = 4.77$ or 5 ($\pi = 0.9$ and $\lambda = 0.0356$ so that $\lambda^s = 0.27$). For $C = 4.77$, the social optimum is strictly positive and equal to $\alpha^{SO} \approx 25\%$ whereas for $C = 5$, the social optimum is zero. In either case, however, the unique equilibrium is such that all institutions inefficiently choose to be fast ($\alpha^* = 1$).

[Insert Fig. 7 about here]

Overinvestment in the fast trading technology arises as soon as $\alpha^* \in (0, 1)$, whether institutions' decisions are substitutes or complements. Complementarity in institutions' decisions, however, aggravates the overinvestment problem because it tends to disconnect investment decisions from the technological cost. For instance, suppose that $\Delta(0) \leq C < \Delta(1)$, which can occur when there is complementarity in institutions' decisions. In this case, the socially optimal level of fast trading is $\alpha^{SO} = 0$ because $C \geq \Delta(0)$ (Proposition 5). Yet there are two possible stable equilibria in this case: $\alpha^* = 0$ and $\alpha^* = 1$. There is no overinvestment in the former but maximal overinvestment in the latter. This happens because each institution anticipates that if it remains slow when others are fast then it will obtain a very low profit. This makes the value of being fast relatively high, despite the fact that the technological cost is so large ($\Delta(0) \leq C$) that any investment in the fast trading technology is inefficient.

Overinvestment in the fast trading technology does not mean that one should necessarily bar institutions from using it. In fact a necessary and sufficient condition for $\alpha^{SO} > 0$ is:

$$\frac{\partial W(0)}{\partial \alpha} = \Delta(0) - C_{soc}(0) > 0. \tag{32}$$

Using the expression for $\Delta(0)$ in Eq. (28), we deduce the following result.

Proposition 6. *The socially optimal level of investment in fast trading technologies is strictly larger than zero if and only if $\lambda < \widehat{\lambda}(\sigma, C, \pi)$ where $\widehat{\lambda}(\sigma, C, \pi)$ is a threshold strictly smaller than one (the expression of this threshold is given in the proof of the proposition). This threshold decreases with C, σ, and π. It is zero when σ or C are large enough or when $\pi = 1$.*

The social value of the fast trading technology comes from its search value, i.e., the reduction in the cost of delayed execution for those using the technology. As explained previously, at $\alpha = 0$, this efficiency gain is equal to $(1 - \mu(\lambda, \pi, 0))E(|\delta|)$. Investment in the fast trading technology is socially optimal only if this gain is large enough relative to the social cost of the technology. As slow institutions' delay cost decreases in λ, the socially optimal level of investment is strictly positive only if λ is smaller than a threshold $\widehat{\lambda}$. This threshold decreases in π because an increase in π reduces delay costs, other things equal.[21] The threshold $\widehat{\lambda}$ also decreases with σ because a higher σ enlarges the range of private valuations for which institutions make socially inefficient trading decisions (i.e., do not trade or sell when they should buy and vice versa). Last, it decreases in C because the reduction in delay costs due to the fast trading technology must at least exceed its cost for investment in this technology to be socially optimal.

Interestingly, $\widehat{\lambda}(\sigma, 0, \pi) < 1$. Hence, overinvestment can arise even when the fast trading technology costs nothing ($C = 0$). Indeed, when $C = 0$, the only equilibrium is $\alpha^* = 1$. Yet, if $\lambda > \widehat{\lambda}(\sigma, 0, \pi)$ then $\alpha^{SO} = 0$ and *even when* $\lambda < \widehat{\lambda}(\sigma, 0, \pi)$, α^{SO} will be positive but strictly less than one. The reason is that the social cost of fast trading includes the negative externality generated by fast trading, not just real resources invested in the technology.

Proposition 5 focuses on $\sigma > 0$. For completeness, the next corollary considers the particular case in which the fast trading technology has no speculative value because there is no uncertainty about the cash flow of the asset ($\sigma = 0$).

Corollary 3. *(Benchmark: no adverse selection): When $\sigma = 0$, the socially optimal level of fast trading is $\alpha^{SO} = 1$ if $\lambda < \widehat{\lambda}(0, C, \pi)$ and $\alpha^{SO} = 0$ if $\lambda \geq \widehat{\lambda}(0, C, \pi)$. Furthermore, in this case, the equilibrium level of fast trading is unique and it coincides with the socially optimal level of fast trading.*

In the absence of adverse selection, the cost of fast trading is just the technological cost, C. As this cost is independent of the level of fast trading, the socially optimal level of fast

[21] If $\pi = 1$ then $\widehat{\lambda} = 0$ when $r = 0$ because there is then no cost of delaying execution ($\mu(\lambda, 1, 0) = 1$). However, if $r > 0$, the cost of delayed execution is strictly positive even if $\pi = 1$ ($\mu(\lambda, 1, r) < 1$) for $r > 0$). Hence, in general, $\widehat{\lambda}$ can be strictly positive even when $\pi = 1$ (see the on-line appendix).

trading is either zero or one, depending on whether the social value of fast trading (i.e., $(1 - \mu(\lambda, \pi, 0))E(|\delta|)$) is less than or higher than the technological cost. This comparison is exactly that made by institutions in choosing to invest or not and as a result there is no divorce between institutions' investment decisions and social optimality.

6.2. *Policy Responses*

The previous section shows that investment in fast trading is in general too high relative to the efficient level. In this section, we analyze two possible responses: (i) Pigovian taxes, and (ii) "slow markets".[22]

6.2.1. *Pigovian taxation*

Suppose that the social planner can levy a lump sum tax T on fast institutions. This tax raises the total cost of being fast to $C + T$. Thus, for a tax T, let $\alpha^{**}(T)$ be the equilibrium level of fast trading, determined as in Section 5 with $C + T$ replacing C. The central planner wants to set T so that $\alpha^{**}(T) = \alpha^{SO}$.

When $0 < \alpha^{SO} < 1$, the socially optimal level of fast trading solves:

$$\frac{\partial W(\alpha^{SO})}{\partial \alpha^{SO}} = \phi(\alpha^{SO}) - \psi(\alpha^{SO}) - C + \underbrace{(\alpha^{SO}\frac{\partial \phi(\alpha^{SO})}{\partial \alpha^{SO}} + (1 - \alpha^{SO})\frac{\partial \psi(\alpha^{SO})}{\partial \alpha^{SO}})}_{-C_{ext}(\alpha^{SO})} = 0. \tag{33}$$

Set $T^* = C_{ext}(\alpha^{SO})$. Using Eq. (33), we have:

$$\phi(\alpha^{SO}) - \psi(\alpha^{SO}) = C + T^*. \tag{34}$$

Thus, with the tax T^*, there is an equilibrium in which the fraction of institutions choosing to be fast is $\alpha^{**}(T^*) = \alpha^{SO}$.[23] This tax is such that fast institutions bear the cost they impose on

[22] Another policy response would be to ban fast trading altogether. This approach can be implemented by setting sufficiently high taxes for using the fast trading technology. In Section 6.2.1, we show that this is optimal only when $\lambda \geq \widehat{\lambda}(\sigma, C, \pi)$. See Proposition 7.

[23] If institutions' decisions are substitutes everywhere, this is the unique equilibrium whereas if institutions'

other institutions, $C_{ext}(\alpha^{SO})$, when the level of fast trading is α^{SO}. Thus, institutions internalize this cost, which aligns private incentives with social optimality.

When $\alpha^{SO} = 1$, $\frac{\partial W(1)}{\partial \alpha^{SO}} > 1$. Hence, using Eq. (33), $\phi(1) - \psi(1) > C + C_{ext}(1)$. For this reason, if $\alpha^{SO} = 1$ then $\alpha^{**}(T^*) = 1$ is the unique equilibrium when the tax is $T^* = C_{ext}(1)$.

In sum, the tax $T^* = C_{ext}(\alpha^{SO})$ implements the socially optimal level of fast trading when $\alpha^{SO} > 0$, i.e., when $\lambda < \widehat{\lambda}(\sigma, C, \pi)$ (see Proposition 6). When $\lambda \geq \widehat{\lambda}(\sigma, C, \pi)$, $\alpha^{SO} = 0$. Thus, the social planner wants to prevent any investment in the fast trading technology. This can be achieved either with a ban on fast trading or equivalently with a tax that exceeds the largest possible value of $\Delta(\alpha)$. The next proposition summarizes these results.

Proposition 7. *When $\lambda < \widehat{\lambda}(\sigma, C, \pi)$, a tax equal to $T^* = C_{ext}(\alpha^{SO})$, implements the socially optimal level of fast trading, which, in this case, is strictly positive. When $\lambda \geq \widehat{\lambda}(\sigma, C, \pi)$, a tax that exceeds the largest possible value of $\Delta(\alpha)$ implements the socially optimal level of fast trading, which in this case is zero.*

Proposition 7 provides an economic rationale for recent proposals to tax fast traders.[24] Calibrating the optimal tax is difficult, however, as it requires estimating the negative externality generated by fast institutions at the socially optimal level of fast trading. Yet, our analysis provides some insights on what optimal taxes should look like: the tax should be higher for assets in which the negative externality of fast trading is higher, that is, more volatile assets (σ higher) or assets in which gains from trade are smaller, i.e., assets for which the dispersion of traders' private valuations is smaller. Interestingly, this suggests that taxes on fast traders should be asset specific.

decisions are complements for some range of α, there might be other equilibria. However, in these other equilibria, the level of fast trading must be greater than α^{SO} because, in any equilibrium, the level of fast trading either coincides with or exceeds the socially optimal level (Proposition 5). This implies that the equilibrium in which $\alpha^{**} = \alpha^{SO}$ is Pareto dominant. Indeed, in any interior equilibrium α^{**}, the expected profit of fast institutions *net* of the cost of being fast (including a tax if any) is $\psi(\alpha^{**})$, which is also the expected profit of slow institutions. When $\alpha^{**} > \alpha^{SO}$, we have $\psi(\alpha^{**}) < \psi(\alpha^{SO})$. Hence, all institutions prefer the equilibrium in which the level of fast trading is $\alpha^{**} = \alpha^{SO}$. If $\alpha^{**} = 1$, all institutions obtain $\phi(1) - C - T^* = W(1) - T^*$. In contrast, in the equilibrium in which $\alpha^{**} = \alpha^{SO}$, institutions obtain: $W(\alpha^{SO}) - T^*$. This is stricly larger than $W(1) - T^*$ by definition of α^{SO} when $\alpha^{SO} < 1$.

[24]See "*Robin Hood Tax: A Long Shot,*" Financial Times, May 2013 and "*Italy introduces tax on high speed traders in equity derivatives*," Financial Times, September 1, 2013.

Furthermore, a per trade tax is unlikely to be optimal, even if it affects fast traders only. Indeed, a per trade tax is similar to an increase in the bid-ask spread. Thus, it widens the range of private valuations for which fast institutions decide not to trade. This effect results in a welfare loss (unrealized gains from trade) that does not arise with a lump sum tax. Thus, taxing investment in fast trading technology is more efficient than taxing fast institutions' trades.

Finally, observe that tax proceeds can be redistributed among all institutions so that they all eventually share the welfare gain associated with fast trading (even if they remain slow). Indeed, suppose that $0 < \alpha^{SO} < 1$ and that the tax proceeds are redistributed equally among all institutions, so that slow institutions receive in aggregate $(1-\alpha^{SO})T^*$. Slow institutions' aggregate welfare is therefore:

$$(1-\alpha^{SO})\psi(\alpha^{SO}) + (1-\alpha^{SO})\alpha^{SO}T^* = (1-\alpha^{SO})W(\alpha^{SO}), \tag{35}$$

where the second equality follows from the definition of $W(.)$ and Eq. (34). Using the same reasoning, fast institutions obtain $\alpha^{SO}W(\alpha^{SO})$. Thus, per capita, both fast and slow institution obtain the same expected profit, $W(\alpha^{SO})$, after redistribution. They therefore equally benefit from the improvement in welfare relative to the case in which fast trading is forbidden.

6.2.2. Slow and Fast Markets

Another way to alleviate the negative externality of fast trading is to create "slow–friendly" markets. There are several ways in which trading venues can be "slow-friendly," i.e., limit the ability of fast traders to adversely select slower ones. For instance, they could batch incoming orders (see "*High-frequency traders face speed limits*," Financial Times, April 28, 2013), delay the execution of market orders, provide no-colocation services, or simply deny entry to fast traders.[25] In this section, we show that introduction of slow-friendly markets can lead to underinvestment in fast trading.

To this end, consider the following extension of our baseline model. A continuum of slow

[25]For instance, the Investors Exchange (IEX) (see http://www.iextrading.com/about/) is an electronic limit order book market that delays execution of incoming market orders by 350 microseconds and offers no-colocation services. According to IEX's CEO, this is sufficient to deter "predatory" behavior by high frequency trading firms. See "How IEX Is Combating Predatory Types Of High-Frequency Traders," Forbes, April 23, 2014.

markets (the "slow segment") coexist with fast markets ("the fast segment"). In either segment, a slow institution has a probability λ to execute its trade in a given period. Fast institutions cannot access slow markets. In the context of our model this means that, in slow markets, participants do not have private information. Before entering the market, institutions decide: (i) whether to be fast or slow and (ii) if they are slow, whether to trade in the fast segment or the slow segment.

Let β denote the fraction of slow institutions trading on slow markets and let $S^*_{Fast}(\alpha,\beta)$ be the equilibrium spread in fast markets. As in the baseline model, $S^*_{Fast}(\alpha,\beta)$ is the smallest spread for which market makers in the fast segment break even. The only difference with the baseline model is that market makers receive fewer buy and sell orders from slow institutions because a fraction β of these institutions trade in the slow market. Thus, market makers in fast markets bear a higher adverse selection cost than without slow markets and must therefore charge a larger bid-ask spread to break-even, as the next lemma shows.

Lemma 4. *When a fraction β of slow institutions trade in slow markets, the equilibrium bid-ask spread in fast markets is:*

$$S^*_{Fast}(\alpha,\beta) = S^*(\alpha^{Fast}(\alpha,\beta)), \tag{36}$$

where $S^*(\cdot)$ is the equilibrium bid-ask spread in the baseline model and

$$\alpha^{Fast}(\alpha,\beta) = \frac{\alpha}{\alpha + (1-\beta)(1-\alpha)}. \tag{37}$$

The equilibrium bid-ask spread on fast markets increases with α and β.

The bid-ask spread on slow markets is zero because they only attract uninformed institutions.

Institutions' expected gain on the fast market are obtained by replacing $S^*(\alpha)$ by $S^*(\alpha^{Fast})$ in Eq. (19) and (21). Thus:

$$\phi^{Fast}(\alpha,\beta) = \phi(\alpha^{Fast}) \text{ and } \psi^{Fast}(\alpha,\beta) = \psi(\alpha^{Fast}), \tag{38}$$

where (i) α^{Fast} is defined in Lemma 4 and (ii) $\phi^{Fast}(\alpha, \beta)$ and $\psi^{Fast}(\alpha, \beta)$ are respectively fast and slow institutions' expected gains when they trade on fast markets. Institutions' expected gain on the slow market, ψ^{Slow}, is identical to that obtained on the fast market when all investors are slow in the baseline model. That is:

$$\psi^{Slow}(\alpha, \beta) = \psi(0). \tag{39}$$

Institutions' expected gains on the fast market, $\phi(\alpha^{Fast})$ or $\psi(\alpha^{Fast})$, decrease with α^{Fast} and therefore β. Hence, *slow* institutions migrating to the slow market exert a *negative externality* on those who remain on the fast market. Indeed, an increase in β results in a larger bid-ask spread on the fast market because it increases the likelihood (α^{Fast}) that trades on this market come from fast informed traders. This is stated in the next corollary.

Corollary 4. *In equilibrium, an increase in β or α reduces the expected gain of fast and slow institutions on fast markets and has no effect on the expected gain of slow institutions on slow markets.*

Now, consider institutions' decisions to trade in fast or slow markets. For simplicity, and because this is not key for our conclusions, we assume that there is no cost of joining a market.[26] Slow institutions trade at the same speed in either type of markets but there is no adverse selection on slow markets. Thus, trading exclusively on the slow market is a dominant strategy for slow institutions. This implies that $\beta = 1$ in equilibrium.[27] Hence, in any equilibrium with $\alpha^* > 0$, fast institutions cannot make speculative profits at the expense of slow institutions, which considerably reduces their expected gain from trade. Indeed, for any α, they obtain:

$$\phi^{Fast}(\alpha, 1) - C = \phi(1) - C, \tag{40}$$

[26] In reality, markets compete in trading fees and differentiation in speed is a way to sustain non competitive fees [see Pagnotta and Phillipon (2013)]. Analyzing this competition is beyond the scope of our paper. Furthermore, by assuming zero fee on the slow market, we bias the model against finding that slow markets are inefficient, which is the main finding of this section.

[27] When $\alpha = 0$, institutions are indifferent between slow and fast markets because they obtain an expected gain of $\psi(0)$ in either case. In this case, any β is an equilibrium.

which is the lowest possible expected gain for fast institutions because ϕ is minimal in $\alpha = 1$.

Thus, the choice between being fast and slow boils down to a comparison between $\phi^{Fast}(\alpha, 1) - C$ and $\psi^{Slow}(\alpha, 1)$, that is, $\phi(1) - C$ and $\psi(0)$. If $\phi(1) - C > \psi(0)$, each institution is better off investing in the fast technology and exploiting it on the fast market, independently of other institutions' choices (because, in this case, $\phi^{Fast}(\alpha, \beta) > \phi^{Fast}(\alpha, 1) > \psi(0)$). Thus, all institutions choose to be fast in equilibrium. If instead $\phi(1) - C \leq \psi(0)$, all institutions are better off being slow and trading on slow markets only, for all values of α.

Proposition 8. *If $\phi(1) - \psi(0) \leq C$ no institution becomes fast and all trade on slow markets ($\alpha^* = 0$ and $\beta^* = 1$), while if $C < \phi(1) - \psi(0)$ all institutions are fast and only trade on fast markets ($\alpha^* = 1$ and $\beta^* = 0$).*

Thus, the introduction of slow markets significantly affects equilibrium investment decisions. In particular, equilibria in which an interior fraction of institutions invest in the fast technology unravel and one ends up with only two corner equilibria: (i) the "All Fast" equilibrium with no activity in slow markets or (ii) the "All Slow" equilibrium with no activity in fast markets.

For given values of α and β, utilitarian welfare is:

$$\begin{aligned} W(\alpha, \beta) &= \beta(1-\alpha)\psi^{Slow}(\alpha, \beta) + (1-\beta)(1-\alpha)\psi^{Fast}(\alpha, \beta) + \alpha(\phi^{Fast}(\alpha, \beta) - C) \\ &= \beta(1-\alpha)\psi(0) + (1-\beta)(1-\alpha)\psi(\alpha^{Fast}) + \alpha(\phi(\alpha^{Fast}) - C). \end{aligned} \tag{41}$$

Using Proposition 8, we deduce that with slow and fast markets, utilitarian welfare in equilibrium is either:

$$W(0, 1) = \psi(0), \text{ when } \phi(1) - \psi(0) \leq C, \tag{42}$$

or

$$W(1, 0) = \phi(1) - C, \text{ when } \phi(1) - \psi(0) > C. \tag{43}$$

When there is no slow market, $\beta = 0$ and the equilibrium level of fast trading is α^*. Social welfare in equilibrium is therefore $W(\alpha^*, 0)$. We have $W(\alpha^*, 0) = \psi(0)$ when $\alpha^* = 0$, $W(\alpha^*, 0) = \phi(1) - C$ when $\alpha^* = 1$, and $W(\alpha^*, 0) = \psi(\alpha^*)$ when $0 < \alpha^* < 1$ because in this case $\phi(\alpha^*) - \psi(\alpha^*) = C$. When $\phi(1) - \psi(0) < C$, equilibrium social welfare without slow markets is always less than or equal to equilibrium social welfare with slow markets ($\psi(0)$) because $\psi(\alpha)$ is maximal at $\alpha = 0$. When $\phi(1) - \psi(0) \geq C$, we have $\phi(\alpha) - \psi(\alpha) > C$ for all α because ϕ and ψ decrease with α. Thus, with or without slow markets, all institutions choose to be fast ($\alpha^* = 1$) and social welfare is $\phi(1) - C$. This yields the following result.

Corollary 5. *In equilibrium, social welfare with slow and fast markets is greater than social welfare with fast markets only.*

Thus, if one cannot tax fast institutions, opening slow markets *improves* welfare. However, this market structure does not necessarily *maximize* social welfare. In fact, in general, it does not, because it induces too many (all) institutions to remain slow, relative to the social optimum.

To analyze this point, suppose that $\lambda \leq \widehat{\lambda}(\sigma, C, \pi)$ and $\phi(1) - \psi(0) < C$. As $\widehat{\lambda}(\sigma, C, \pi)$ decreases with C, these two conditions are satisfied when C is small enough for investment in the fast trading technology to be socially optimal but large enough for each institution to prefer trading on slow markets (see Proposition 8). Thus, in equilibrium, social welfare is $W(0,1) = W(0,0)$. It is not maximal. Indeed, without slow markets, the socially optimal level of fast trading, α^{SO}, is strictly between zero and one (Proposition 6). Hence:

$$W(0,0) < W(\alpha^{SO}, 0). \tag{44}$$

This means that, if regulators could pick α and β, they could improve social welfare by imposing $\beta = 0$ and setting the level of fast trading at α^{SO}. Intuitively, in equilibrium, there is "too much" trading on slow markets ($\beta = 1$) because institutions joining the slow market exert a negative externality on those on the fast market.[28]

Thus, regulators are between a rock and a hard place: with only fast markets, there is overinvestment in the fast trading technology in equilibrium, whereas with slow markets, there can be underinvestment in the fast trading technology. To solve this conundrum, one should tax both investment in the fast technology and access to the slow market. The next proposition states that optimal taxation should preclude trading on the slow market.

[28] In other cases ($\lambda > \widehat{\lambda}(\sigma, C, \pi)$ or $C \leq \phi(1) - \psi(0)$), the equilibrium outcome with slow markets coincide with the outcome maximizing social welfare in the absence of slow markets.

Proposition 9. *If the regulator can use Pigovian taxes, it should choose (i) a tax larger than $\psi(0)$ for institutions trading on the slow market to preclude trading on this market and (ii) a tax chosen as explained in Proposition 7 for fast institutions.*

The next example illustrates the results obtained in this section.

Example 3: Consider the same parameters as in Fig. 7 with $C = 4.77$, so that $\alpha^{SO} = 25\%$. Without slow markets, all institutions choose to be fast in equilibrium. Utilitarian welfare is $W(1, 0) = \phi(1) - C = 2.142$. With slow markets, this equilibrium unravels and all institutions choose to be slow. Investors' welfare improves and becomes $W(0, 1) = \psi(0) = 2.15$. Yet, social welfare is not maximal. As implied by Proposition 9, it can be improved by charging a tax larger than 2.15 for trading in slow markets (so that no trader chooses to do so) and a tax equal to $T^* = 2.03$ for investing in the fast trading technology. With this tax, $\alpha^* = \alpha^{SO} = 25\%$ and social welfare is $W(\alpha^{SO}, 0) = 2.16$. If the tax is equally redistributed among all institutions they all obtain an expected profit of $2.16 > \psi(0)$ after redistribution (see Section 6.2.1).■

7. Empirical implications

Our model implies that the informational content of trades should be inversely related to the cost of fast trading (e.g., colocation fees). Indeed, at any stable equilibrium, an increase in C triggers a drop in the level of fast trading (α) and hence in the informational content of trades. In contrast, an increase in the cost of fast trading can have ambiguous effects on trading volume (see the analysis in Section 4.2).

Anecdotal evidence suggests that the profitability of high frequency traders decreased in recent years. For instance, the profits of GETCO, one of the early adopter of fast trading technologies have constantly declined since 2007 (see "*GETCO profit drops 82% on weak US market*" Financial Times, February 13, 2013). One simple explanation for this evolution is that as the number of fast institutions increases, the profitability of fast trading declines. Another possibility is that the cost of fast trading has increased. Anecdotal evidence suggests that this is indeed the case (see "*High-Speed Trading no Longer Hurtling Forward,*" New-York Times, October 14, 2012 and "*High-speed stock traders turn to laser beams*", WSJ, March 10, 2014). Both explanations are consistent with our model in which the net expected profit of fast institutions, $\phi(\alpha) - C$ decreases in the level of fast trading, α, and the cost of fast trading, C. In addition, the model suggests two other, less obvious, explanations, highligted in Implications 1 and 2 below.

Implication 1: *Holding the level of fast trading (α) constant, the expected gains of fast institutions decrease in market fragmentation and in the fraction of trading that takes place on slow markets.*

As market fragmentation increases (i.e., λ goes down), it is more difficult for slow institutions to quickly find attractive quotes. This hurts fast traders in our model because it increases the spread. Accordingly, fast and slow institutions' expected gains decline (see Eq. (19) and (21)). Similarly, an increase in the fraction of institutions trading on slow markets (β) raises the spread on the fast market, which lowers the profitability of fast trading.

Over the counter and dark markets are, by design, slower than centralized electronic limit order book markets. The trading volume on these markets has grown in recent years and, in line with the logic behind Proposition 8, this growth is in part driven by slow investors' desire to insulate themselves from high frequency traders. For instance, a 2013 New-York times article ("*As markets heat up, trading slips into shadows*") notes that: "*Investors also have said that they have moved more of their trading into the dark because they have grown more distrustful of the big exchanges like the NYSE and the Nasdaq. Those exchanges have been hit by technological mishaps and become dominated by so-called high-frequency traders.*" Consistent with Implication 1, this evolution might also be responsible for the drop in fast trading profitability in recent years.

Implication 2: *The expected profit of fast institutions increases in the volatility of the asset when the level of fast trading is low. However, it can decrease with volatility when the level of fast trading is large.*

In our model, the volatility of the asset payoff is σ.[29] Its effect on the profitability of fast

[29] In reality high frequency traders seem to obtain information on short term price movements. Hence, for empirical tests, σ should be proxied by the volatility of short term changes in asset fair values [see Hasbrouck (2005) for various methods to estimate this volatility].

trading are ambiguous. On the one hand, holding the bid and ask prices constant, an increase in volatility raises the speculative value of fast trading. On the other hand, an increase in volatility raises the spread, which lowers fast institutions' expected gains. The former effect dominates the latter when α is small but not necessarily when α is large (see the on-line appendix for an example).

Now consider the effects of variations in market fragmentation (λ) and volatility (σ) on the equilibrium level of fast trading, α^*. The model suggests that analyzing the effects of these factors on the (net) expected profit of fast institutions is not sufficient to predict entry or exit of fast institutions. Indeed, the decision to become fast is determined by the difference between the profit of being fast and the profit of being slow rather than just the profit of being fast. As a result, the effect of a parameter that negatively affects the profitability of fast traders can, counterintuitively, increase the equilibrium level of fast trading if its negative impact on slow institutions is stronger. Consider, for example, an increase in market fragmentation, i.e., a decrease in λ. For low levels of fast trading, an increase in market fragmentation reduces the expected profit of fast *and* slow institutions, but the negative impact is more severe for slow institutions. This follows directly from Eq. (28) and the continuity of $\Delta(\alpha)$ with respect to λ. Accordingly, for α^* close enough to zero (i.e., C high), an increase in market fragmentation should trigger entry of new fast institutions, even though it decreases fast institutions' expected profit. Similarly, for high values of C, an increase in volatility should raise the level of fast trading in this asset.

Implication 3: *For high values of the cost of trading fast, C, equilibrium investment in fast trading should increase when (a) trading becomes more fragmented or (b) volatility increases.*

This implication fits well with the idea that market fragmentation and volatility fostered the development of fast trading technologies. However, as the level of fast trading grows, further increases in market fragmentation or volatility can lower the profitability of fast trading and force some fast trading firms to exit. This highlights the importance of controlling for the cost and the level of fast trading in empirical studies considering the effects of market fragmentation or volatility on fast trading. To illustrate this point, consider the following example: Private valuations are normally distributed with $\sigma_\delta = 4$; other fixed parameters are $\sigma = 7$; and $\pi = 0.9$.

For a large cost of fast trading, $C = 4$, an increase in λ (from $\lambda = 0.6$ to $\lambda = 0.3$), generates an increase in α^*, from 15.5% to 17.5%. In contrast, for a lower cost of fast trading, $C = 3$, the same increase in fragmentation generates a decrease in α^*, from 32% to 28.2%.

8. Conclusion

Investment in fast trading technology helps financial institutions cope with market fragmentation. To the extent that this technology enhances their ability to reap mutual gains from trade, it improves social welfare. However, fast trading technology also provide advance access to value relevant information, which creates adverse selection, lowering welfare. Thus, fast trading generates a negative externality. Because financial institutions do not internalize this negative externality, equilibrium investment in fast trading technologies is in general excessive.

We show that, for some parameter values, institutions' investment decisions can be strategic complements. In this case, the overinvestment problem is particularly acute because the value of being fast relative to remaining slow becomes increasingly large as the amount invested in fast trading escalates. This leads to an arms' race in which all institutions end up investing in the fast technology, in the same spirit as in Glode, Green, and Lowery (2012).

One way to mitigate the adverse consequences of fast traders is to create "slow–only" markets. Yet, we show that this market structure can lower equilibrium investment in fast trading technologies below its socially optimal level. On the other hand, if the regulator can impose Pigovian taxes on investments in fast trading technology (equal to the externalities they generate), the socially optimal level of investment in fast trading technology can be implemented. To the extent that institutions can game taxes on fast trading, the efficacy of Pigovian taxes will be reduced. Some investments in fast trading (such as fiber–optic cables, microwave signals, or colocation) are easily observable, however. To reduce the scope for gaming, the tax base could be contingent on such easily observable investments.[30]

[30] Fast institutions could be tempted to collude with tax–exempt slow institutions by sharing their signals with them. The attractiveness of this strategy, however, would be greatly reduced by the time it would take to disseminate the signals to slow institutions.

Appendix A. Proofs

Propositions 3, 7, 8, and Corollaries 1, 4, 5 follow directly from the arguments in the text.

Proof of Lemma 1. Let $\Pi(S;\alpha,\lambda,\sigma)$ denote a market maker's net expected profit, i.e., the difference be the Right Hand Side (R.H.S) and the Left Hand Side (L.H.S) of Eq. (11). Using Eq. (8) and (9), we obtain the expected profit of a market maker:

$$\begin{aligned} \Pi(S;\alpha,\lambda,\sigma) &= \left[\alpha(\overline{G}(S-\sigma)+\overline{G}(S+\sigma))+2\lambda^s(1-\alpha)\overline{G}(S))\right] S \\ &- \left[\alpha(\overline{G}(S-\sigma)-\overline{G}(S+\sigma))\right] \sigma. \end{aligned} \tag{45}$$

Equilibrium spreads solve Eq. (11), i.e., they are such that $\Pi(S^*;\alpha,\lambda,\sigma)=0$. When $\alpha>0$ and $\sigma>0$, we have $\Pi(0;\alpha,\lambda,\sigma)<0$ and $\Pi(\sigma;\alpha,\lambda,\sigma)=[2\alpha(1-G(2\sigma))+2\lambda^s(1-\alpha)(1-G(\sigma))]\,\sigma>0$. Therefore, as $\Pi(\cdot)$ is continuous in S, there always exists at least one $S^*\in(0,\sigma)$, such that $\Pi(S^*;\alpha,\lambda,\sigma)=0$. When $\alpha=0$ or $\sigma=0$, $\Pi(0;\alpha,\lambda,\sigma)=0$ and, for $S>0$, $\Pi(S;\alpha,\lambda,\sigma)>0$. Hence, in these cases, $S^*=0$ is the unique equilibrium.

Proof of Lemma 2. By definition, $S^*_{\min}(\alpha)$ is the smallest positive spread, S^*, such that $\Pi(S^*;\alpha,\lambda,\sigma)=0$ where $\Pi(.)$ is defined in (45). Therefore, as $\Pi(0;\alpha,\lambda,\sigma)\le 0$, we have $\Pi(S;\alpha,\lambda,\sigma)<0$ for all $S<S^*_{\min}(\alpha)$, which in turn implies (i) that $S^*_{\min}(\alpha)$ cannot be profitably undercut and (ii) that:

$$\frac{\partial \Pi}{\partial S}\Big|_{S=S^*_{\min}(\alpha)} > 0. \tag{46}$$

Hence, by continuity of $\Pi(\cdot)$, there is always a bid-ask spread S_0 arbitrarily close to but larger than $S^*_{\min}(\alpha)$ such that market makers' net expected profit when they charge S_0 is strictly positive. Thus, any equilibrium spread above $S^*_{\min}(\alpha)$ can be profitably undercut.

Proof of Proposition 1. Remember that the equilibrium spread is the smallest S^* such that market makers break even, i.e., such that $\Pi(S^*;\alpha,\lambda,\sigma)=0$. Using the definition of $\Pi(\cdot)$

in Eq. (45) and the implicit function theorem, we obtain

$$S^{*\prime}(\alpha) = -\frac{\frac{\partial \Pi}{\partial \alpha}\Big|_{S=S^*(\alpha)}}{\frac{\partial \Pi}{\partial S}\Big|_{S=S^*(\alpha)}}, \tag{47}$$

where $S^{*\prime}(\alpha)$ denotes the first derivative of $S^*(\alpha)$ with respect to α. We know from Eq. (46) that $\frac{\partial \Pi}{\partial S}\big|_{S=S^*(\alpha)} > 0$. From Eq. (45), we have that $\Pi(.)$ is linear in α with an intercept equal to $2\lambda^s\overline{G}(S)S$. Thus:

$$\Pi(S;\alpha,\lambda,\sigma) = \frac{\partial \Pi}{\partial \alpha} \times \alpha + 2\lambda^s\overline{G}(S)S. \tag{48}$$

At equilibrium, $S = S^*$, and $\Pi(S^*;\alpha,\lambda,\sigma) = 0$. Thus, Eq. (48) implies that $\frac{\partial \Pi}{\partial \alpha}\big|_{S=S^*(\alpha)} \leq 0$ for $\alpha > 0$ since $S^* \geq 0$ and therefore $2\lambda^s\overline{G}(S^*)S^* \geq 0$. When $\alpha = 0$, $S^* = 0$ and partial differentiation of Eq. (45) with respect to α yields

$$\frac{\partial \Pi}{\partial \alpha}\Big|_{S=S^*(0)} = -(\overline{G}(-\sigma) - \overline{G}(\sigma))\sigma = (2G(\sigma) - 1)\sigma \leq 0. \tag{49}$$

Thus, $\frac{\partial \Pi}{\partial \alpha}\big|_{S=S^*(\alpha)} \leq 0$ for all α. Hence, $\frac{\partial S^*}{\partial \alpha} > 0$. Proceeding in a similar way, we also obtain (i) $\frac{\partial S^*}{\partial \lambda} < 0$ if $\pi < 1$ because then $\frac{\partial \Pi}{\partial \lambda}\big|_{S=S^*(\alpha)} > 0$ and (ii) $\frac{\partial S^*}{\partial \sigma} > 0$ because $\frac{\partial \Pi}{\partial \sigma}\big|_{S=S^*(\alpha)} < 0$. When $\pi = 1$, $\lambda^s = 1$ and therefore the bid-ask spread does not depend on λ. Thus, the bid-ask spread weakly decreases with λ.

Proof of Lemma 3. Define $f(x,y) = \frac{1-G(x+y)}{1-G(x)}$. Using Eq. (13), $\Delta\text{Vol}(S^*(\alpha),\alpha) > 0$ iff: $\frac{1-G(S^*-\sigma)}{1-G(S^*)} + \frac{1-G(S^*+\sigma)}{1-G(S^*)} \geq 2\lambda^s$, that is, iff:

$$f(S^*,-\sigma) + f(S^*,\sigma) \geq 2\lambda^s. \tag{50}$$

We have:

$$\frac{\partial f(x,y)}{\partial x} = \frac{-g(x+y)(1-G(x)) + g(x)(1-G(x+y))}{(1-G(x))^2}. \tag{51}$$

Thus, $f(x,y)$ increases with x iff $\frac{g(x)}{1-G(x)} \geq \frac{g(x+y)}{1-G(x+y)}$, that is, iff $h_g(x) \geq h_g(x+y)$. Suppose first that $h_g(.)$ decreases with x. Thus, $h_g(x+\sigma) < h_g(x) < h_g(x-\sigma)$. Hence, setting $y = -\sigma$,

$f(S,-\sigma)$ decreases with S and is therefore minimal in $S=\sigma$. Symmetrically for $y=+\sigma$, $f(S,\sigma)$ increases with S and is therefore minimal in $S=0$. Thus, Condition (50) is satisfied if $f(\sigma,-\sigma)+f(0,\sigma)\geq 2\lambda$, that is:

$$\frac{1-G(0)}{1-G(\sigma)}+\left(\frac{1-G(0)}{1-G(\sigma)}\right)^{-1}\geq 2\lambda^s. \tag{52}$$

This condition is always satisfied if $\lambda^s\leq\frac{1}{2}$ because the first term in the L.H.S of this equation is larger than 1 and the second term is positive. It is also satisfied when $\lambda^s>\frac{1}{2}$ because the L.H.S of the previous equation reaches its minimum for $\sigma=0$, for which it is equal to 2. This proves the first part of the proposition.

Now suppose that $h_g(.)$ increases with x. Thus, $h_g(x+\sigma)>h_g(x)>h_g(x-\sigma)$. Hence, setting $y=-\sigma$, we deduce that $f(S,-\sigma)$ increases with S and is therefore minimal in $S=0$. Symmetrically, $f(S,\sigma)$ decreases with S and is therefore minimal in $S=\sigma$. Thus, Condition (50) is satisfied if $f(0,-\sigma)+f(\sigma,\sigma)\geq 2\lambda^s$, that is, using the symmetry of $g(.)$:

$$2G(\sigma)+\left(\frac{1-G(2\sigma)}{1-G(\sigma)}\right)\geq 2\lambda^s. \tag{53}$$

As $G(\sigma)\geq\frac{1}{2}$, this condition is always satisfied for $\lambda^s\leq\frac{1}{2}$. This proves the second part of the proposition.

Proof of Proposition 2. As $S^*(\alpha)$ increases with α, it is immediate from Eq. (19) and Eq. (21) that fast and slow institutions' expected profits decrease with α. This is the second part of the proposition. For the first part, observe that:

$$\begin{aligned}\phi-\psi &= \int_{S^*(\alpha)-\sigma}^{\bar{\delta}}(\delta+\sigma-S^*(\alpha))g(\delta)d\delta+\int_{S^*(\alpha)+\sigma}^{\bar{\delta}}(\delta-\sigma-S^*(\alpha))g(\delta)d\delta\\ &- 2\mu\int_{S^*(\alpha)}^{\bar{\delta}}(\delta-S^*(\alpha))g(\delta)d\delta.\end{aligned} \tag{54}$$

Thus:

$$
\begin{aligned}
\phi-\psi &= \int_{S(\alpha)-\sigma}^{\overline{\delta}}(\delta+\sigma-S^*(\alpha))g(\delta)d\delta+\int_{S^*(\alpha)+\sigma}^{\overline{\delta}}(\delta-\sigma-S^*(\alpha))g(\delta)d\delta \\
&\quad -2\mu\left[\frac{1}{2}\int_{S^*(\alpha)}^{\overline{\delta}}(\delta+\sigma-S^*(\alpha))g(\delta)d\delta+\frac{1}{2}\int_{S^*(\alpha)}^{\overline{\delta}}(\delta-\sigma-S^*(\alpha))g(\delta)d\delta\right], \quad (55)
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\phi-\psi &= (1-\mu)\int_{S^*(\alpha)-\sigma}^{\overline{\delta}}(\delta+\sigma-S^*(\alpha))g(\delta)d\delta+(1-\mu)\int_{S^*(\alpha)+\sigma}^{\overline{\delta}}(\delta-\sigma-S^*(\alpha))g(\delta)d\delta \\
&\quad +\mu\int_{S^*(\alpha)-\sigma}^{S^*(\alpha)}(\delta+\sigma-S^*(\alpha))g(\delta)d\delta-\mu\int_{S^*(\alpha)}^{S^*(\alpha)+\sigma}(\delta-\sigma-S^*(\alpha))g(\delta)d\delta. \quad (56)
\end{aligned}
$$

The two first terms on the R.H.S of the previous equation are clearly positive. The last two terms are strictly positive as well because $\int_{S^*(\alpha)}^{S^*(\alpha)+\sigma}(\delta-\sigma-S^*(\alpha))g(\delta)d\delta<0$ and $\int_{S^*(\alpha)-\sigma}^{S^*(\alpha)}(\delta+\sigma-S^*(\alpha))g(\delta)d\delta>0$. Thus, $\phi-\psi>0$.

Proof of Corollary 2. Using the definition of $g(.)$ in Example 2, we have $G(S^*+\sigma)=\left(\overline{\delta}+\gamma\sigma+\varphi S^*\right)/(2\overline{\delta})$, $G(S^*-\sigma)=\left(\gamma S^*+\varphi(\overline{\delta}-\sigma)\right)/(2\overline{\delta})$, and $G(S^*)=\frac{1}{2}+\frac{\gamma}{2\overline{\delta}}S^*$. Hence, using Eq. (14) and $\left(\gamma\sigma+\varphi\left(\overline{\delta}-\sigma\right)\right)/2\overline{\delta}=1/2$, we obtain:

$$
\Delta\text{Vol}(S^*(\alpha),\alpha)=(1-\lambda^s)-\frac{\left(\varphi+\gamma(1-2\lambda^s)\right)S^*(\alpha)}{2\overline{\delta}}. \quad (57)
$$

As $S^*(0)=0$, $\Delta\text{Vol}(S^*(0),0)=1-\lambda^s\geq 0$. Furthermore $\Delta\text{Vol}(S^*(\alpha),\alpha)$ decreases with α because S^* increases with α and $\varphi\geq\gamma$. Thus, there are two cases to consider. If $\lambda^s<$Min$\{1,\frac{2\overline{\delta}-(\gamma+\varphi)S^*(1)}{2(\overline{\delta}-\gamma S^*(1))}\}$ then $\Delta\text{Vol}(S^*(1),1)>0$. In this case, $\Delta\text{Vol}(S^*(\alpha),\alpha)>0$ for all α and institutions' decisions are globally substitutes ($\frac{\partial\Delta(\alpha)}{\partial\alpha}<0$, $\forall\alpha$).[31] If instead, $\lambda^s>$Min$\{1,\frac{2\overline{\delta}-(\gamma+\varphi)S^*(1)}{2(\overline{\delta}-\gamma S^*(1))}\}$ then $\Delta\text{Vol}(S^*(1),1)<0$. Therefore, by continuity of $\Delta\text{Vol}(S^*(\alpha),\alpha)$, there is one value of α, denoted α_0, such that $\Delta\text{Vol}(S^*(\alpha),\alpha)<0$ iff $\alpha>\alpha_0$.

[31] This is always the case if $\gamma=\varphi=1$, that is, if the distribution of institutions' private valuation is uniform.

Proof of Proposition 4. When institutions' decisions are substitutes everywhere, $\Delta(\alpha)$ is decreasing for all α. Thus, if $\Delta(0) - C \leq 0$ then $\Delta(\alpha) - C < 0$ for all α and $\alpha^* = 0$ is the unique equilibrium. If $\Delta(1) - C \geq 0$ then $\Delta(\alpha) - C > 0$ for all α. Thus, $\alpha^* = 1$ is the unique equilibrium. If $C \in (\Delta(0), \Delta(1))$, $\Delta(0) - C > 0$ and $\Delta(1) - C < 0$. As $\Delta(\alpha) - C$ is continuous and decreasing, there is a unique $\alpha^* \in (0,1)$ such that $\Delta(\alpha^*) - C = 0$.

Proof of Proposition 5. Consider first an interior equilibrium ($0 < \alpha^* < 1$). If $\sigma > 0$, then such an equilibrium is never a social optimum because the social cost of fast trading necessarily exceeds the benefit of fast trading. Indeed, such an equilibrium is characterized by $\Delta(\alpha^*) = C$ (see Eq. (27)). Thus, $\Delta(\alpha^*) < C + C_{ext}(\alpha^*)$ when $\sigma > 0$, because $C_{ext}(\alpha^*) > 0$ in this case. Hence, at equilibrium, the social cost of fast trading strictly exceeds the social benefit, which implies $\alpha^{SO} \neq \alpha^*$.[32] Furthermore, as $\phi(\alpha^*) - \psi(\alpha^*) = C$, Eq. (29) yields:

$$\begin{aligned} W(\alpha^{SO}) - W(\alpha^*) &= \alpha^{SO}\left(\phi(\alpha^{SO}) - C\right) + (1-\alpha^{SO})\psi(\alpha^{SO}) - \psi(\alpha^*), \\ &= \alpha^{SO}\left(\phi(\alpha^{SO}) - \phi(\alpha^*)\right) + (1-\alpha^{SO})(\psi(\alpha^{SO}) - \psi(\alpha^*)) > 0, \end{aligned} \tag{58}$$

where the inequality is strict because $\alpha^{SO} \neq \alpha^*$ and α^{SO} maximizes $W(.)$ by definition. As $\phi(\cdot)$ and $\psi(\cdot)$ decrease with α, this implies that $\alpha^{SO} < \alpha^*$ when $0 < \alpha^* < 1$.

Now, let us analyze the corner equilibria. Suppose that there is an equilibrium level of fast trading such that $\alpha^* = 0$. This implies that $\Delta(0) < C$ and therefore, given that $\Delta(0) = \phi(0) - \psi(0)$: $0 \leq W(\alpha^{SO}) - W(0) \leq \alpha^{SO}\left(\phi(\alpha^{SO}) - \phi(0)\right) + (1-\alpha^{SO})(\psi(\alpha^{SO}) - \psi(0))$. As $\phi(\cdot)$ and $\psi(\cdot)$ decrease with α, the terms in parentheses on the R.H.S of the second inequality are strictly less than zero if $\alpha^{SO} > 0$, which is impossible since the first inequality imply that the sum of the terms in the second inequality is positive. Thus, the only possibility in this case is that $\alpha^{SO} = 0$. In other words, if there is at least one equilibrium such that $\alpha^* = 0$ then $\alpha^{SO} = 0$, so that $\alpha^{SO} \leq \alpha^*$. Finally, if there is an equilibrium in which $\alpha^* = 1$ (which requires $C \leq \Delta(1)$), we obviously have $\alpha^{SO} \leq \alpha^*$.

[32]When $\Delta(\alpha) < C + C_{ext}(\alpha^*)$, a small decrease in α makes social welfare larger because $\frac{\partial W(\alpha)}{\partial \alpha} < 0$ in this case (see Eq. (31)).

Proof of Proposition 6. We first show that $C_{ext}(0) = (2G(\sigma) - 1)$. By definition $C_{ext}(0) = \psi'(0)$. Hence, using Eq. (21)

$$C_{ext}(0) = -2\mu(\lambda, \pi, 0)S^{*\prime}(0)(1 - G(S^*(0)) = -\mu(\lambda, \pi, 0)S^{*\prime}(0), \tag{59}$$

because $G(S^*(0)) = G(0) = \frac{1}{2}$ ($S^{*\prime}(\alpha)$ denotes the first derivative of $S^*(\alpha)$ with respect to α). Furthermore, using Eq. (45), we have:

$$\frac{\partial \Pi}{\partial S} \Big|_{S=S^*(0)} = \lambda^s(\lambda, \pi). \tag{60}$$

Thus, using Eq. (47) and (49),

$$S^{*\prime}(0) = -\frac{\frac{\partial \Pi}{\partial \alpha} \big|_{S=S^*(0)}}{\frac{\partial \Pi}{\partial S} \big|_{S=S^*(0)}} = \frac{(2G(\sigma) - 1)\sigma}{\lambda^s(\lambda, \pi)}. \tag{61}$$

Finally, using Eq. (59),

$$C_{ext}(0) = (2G(\sigma) - 1)\sigma, \tag{62}$$

because $\mu(\lambda, \pi, 0) = \lambda^s(\lambda, \pi)$. Using Eq. (62) and the expression of $\Delta(0)$ in Eq. (28), Eq. (32) is equivalent to:

$$\frac{\mathrm{E}(|\delta|) - C - (2G(\sigma) - 1)\mathrm{E}(|\delta|\,||\delta| \leq \sigma)}{\mathrm{E}(|\delta|)} > \mu(\lambda, \pi, 0). \tag{63}$$

As $\mu(\lambda, \pi, 0) = \lambda^s(\lambda, \pi)$, we deduce from Eq. (1) that Eq. (63) is equivalent to: $\lambda < \widehat{\lambda}(\sigma, C, \pi)$, where:

$$\widehat{\lambda}(\sigma, C, \pi) = Max\{\frac{(1 - \pi)(\mathrm{E}(|\delta|) - C - (2G(\sigma) - 1)\mathrm{E}(|\delta|\,||\delta| \leq \sigma))}{\mathrm{E}(|\delta|) - \pi\,(\mathrm{E}(|\delta|) - C - (2G(\sigma) - 1)\mathrm{E}(|\delta|\,||\delta| \leq \sigma))}, 0\}. \tag{64}$$

Thus, $\alpha^{SO} > 0$ iff $\lambda < \widehat{\lambda}(\sigma, C, \pi)$. Clearly, $\widehat{\lambda}(\sigma, C, \pi)$ decreases with C and π. Furthermore, it decreases with σ because $(2G(\sigma) - 1)E(|\delta|\,||\delta| \leq \sigma)$ increases with σ.

Proof of Corollary 3. Using Eq. (64) for $\sigma = 0$,

$$\widehat{\lambda}(0, C, \pi) = Max\{\frac{(1-\pi)(\mathrm{E}(|\delta|) - C)}{\mathrm{E}(|\delta|)(1-\pi) + \pi C}, 0\}. \tag{65}$$

When $\sigma = 0$, the bid-ask spread is zero for all values of α. Hence, $S^{*\prime}(\alpha) = 0$ and therefore $C_{ext}(\alpha) = 0$ for all α. Moreover, Eq. (24) and (28) yield $\Delta(\alpha) = \Delta(0) = (1 - \mu(\lambda, \pi, 0))E(|\delta|)$, for all α. Hence, $\Delta(\alpha) - (C + C_{ext}(\alpha)) = \Delta(0) - C = (1 - \mu(\lambda, \pi, 0))E(|\delta|) - C, \forall\alpha$. Thus, using Eq. (31), the fact that $\mu(\lambda, \pi, 0) = \lambda^s(\lambda, \pi)$, and the expression of $\lambda^s(\lambda, \pi)$ in Eq. (1), we deduce that

$$\frac{\partial W}{\partial \alpha} = \Delta(0) - C = (1 - \lambda^s(\lambda, \pi))E(|\delta|) - C > 0, \forall\alpha, \tag{66}$$

iff $\lambda < \widehat{\lambda}(0, C, \pi)$. This implies that $\alpha^{SO} = 1$ if $\lambda < \widehat{\lambda}(0, C, \pi)$ and $\alpha^{SO} = 0$ if $\lambda \geq \widehat{\lambda}(0, C, \pi)$.

Now consider an institution's investment decision. For any level of α, an institution invests in the fast trading technology if $\Delta(\alpha) - C = \Delta(0) - C > 0$ and does not invest otherwise. Thus, in equilibrium, $\alpha^* = 1$ when $\lambda < \widehat{\lambda}(0, C, \pi)$ and $\alpha^* = 0$ if $\lambda \geq \widehat{\lambda}(0, C, \pi)$. This proves the second part of the corollary.

Proof of Lemma 4. Proceeding as in the baseline model, in each period, there is a mass $(1-\alpha)(1-\beta)\lambda^s$ of slow institutions trading in liquid fast markets. Hence, the masses of institutions buying and selling the asset in fast markets at date τ when $\theta_\tau = \sigma$ are:

$$\omega^a(S, \sigma) = \alpha\overline{G}(S - \sigma) + (1-\alpha)(1-\beta)\lambda^s\overline{G}(S), \tag{67}$$
$$\omega^b(S, \sigma) = \alpha\overline{G}(S + \sigma) + (1-\alpha)(1-\beta)\lambda^s\overline{G}(S). \tag{68}$$

Also as in the baseline model, $\omega^a(S, -\sigma) = \omega^b(S, \sigma)$ and $\omega^b(S, -\sigma) = \omega^a(S, \sigma)$ and the equilibrium spread in fast markets solves:

$$\Pi(S^*; \alpha, \lambda) = \left[\omega^a(S^*, \sigma) + \omega^b(S^*, \sigma)\right] S^* - \left[\omega^a(S^*, \sigma) - \omega^b(S^*, \sigma)\right] \sigma = 0. \tag{69}$$

Therefore, replacing $\omega^a(S^*,\sigma)$ and $\omega^b(S^*,\sigma)$ by their expressions given by Eq.(67) and (68) in Eq.(69), we deduce that $S^*(\alpha,\beta)$ is the smallest positive root of

$$\alpha^{Fast}(\overline{G}(x-\sigma)+\overline{G}(x-\sigma))+(1-\alpha^{Fast})\lambda^s\overline{G}(x)x-\alpha^{Fast}(\overline{G}(x-\sigma)-\overline{G}(x+\sigma))=0, \quad (70)$$

where $\alpha^{Fast}(\alpha,\beta)=\frac{\alpha}{\alpha+(1-\alpha)(1-\beta)}$. When there is no slow market and the level of fast trading is α^{Fast}, the equilibrium spread solves the same equation (see Eq. (45)). Thus, $S^*(\alpha,\beta)=S^*(\alpha^{Fast})$. Hence $S^*(\alpha,\beta)$ increases in α and β because (i) $\alpha^{Fast}(\alpha,\beta)$ increases in α and (ii) β and $S^*(\alpha)$ increases in α.

Proof of Proposition 9. First, consider the case in which $\lambda \leq \widehat{\lambda}(\sigma,C,\pi)$. In this case, $\alpha^{SO}>0$. Now suppose that there exists a pair (α_0,β_0) such that $0<\alpha_0<1$ and $0<\beta_0<1$ with $W(\alpha_0,\beta_0)>W(\alpha^{SO},0)$ (to be contradicted). In this case, the FOCs of the optimization problem: $\text{Max}_{\alpha,\beta}W(\alpha,\beta)$ must be satisfied for (α_0,β_0). This implies:

$$\begin{aligned}\frac{\partial W}{\partial \beta} &= (1-\alpha_0)(\psi(0)-\psi(\alpha_0^{Fast}))\\ &+ \frac{\partial \alpha^{Fast}}{\partial \beta}\left(\alpha_0\frac{\partial \phi(\alpha^{Fast})}{\partial \alpha}+(1-\alpha_0)(1-\beta_0)\frac{\partial \psi(\alpha_0^{Fast})}{\partial \alpha}\right)=0, \end{aligned} \quad (71)$$

$$\begin{aligned}\frac{\partial W}{\partial \alpha} &= \phi(\alpha_0^{Fast})-C-(1-\beta_0)\psi(\alpha_0^{Fast})-\beta_0\psi(0)\\ &+ \frac{\partial \alpha^{Fast}}{\partial \alpha}\left(\alpha_0\frac{\partial \phi(\alpha^{Fast})}{\partial \alpha}+(1-\alpha_0)(1-\beta_0)\frac{\partial \psi(\alpha_0^{Fast})}{\partial \alpha}\right)=0. \end{aligned} \quad (72)$$

where $\alpha_0^{Fast}=\alpha^{Fast}(\alpha_0,\beta_0)$. At (α_0,β_0), $\frac{\partial \alpha^{Fast}}{\partial \alpha}=\frac{(1-\beta_0)}{(\alpha_0+(1-\alpha_0)(1-\beta_0))^2}$ and $\frac{\partial \alpha^{Fast}}{\partial \beta}=\frac{\alpha_0(1-\alpha_0)}{(\alpha_0+(1-\alpha_0)(1-\beta_0))^2}$. Using this and the fact that the expressions in large parentheses in Eq. (71) and Eq. (72) are equal, (α_0,β_0) must satisfy the following condition:

$$(\psi(0)-\psi(\alpha_0^{Fast}))(\alpha_0\beta_0+(1-\beta_0))=\alpha_0(\phi(\alpha_0^{Fast})-C-\psi(\alpha_0^{Fast})). \quad (73)$$

Moreover, if $W(\alpha_0, \beta_0) > W(\alpha^{SO}, 0)$ then $W(\alpha_0, \beta_0) > \psi(0)$. Thus, Eq. (41) implies

$$(1-\alpha_0)\beta_0(\psi(0) - \psi(\alpha_0^{Fast})) + \alpha_0(\phi(\alpha_0^{Fast}) - C - \psi(\alpha_0^{Fast})) > \psi(0) - \psi(\alpha_0^{Fast}). \tag{74}$$

Using Eq. (73), this implies: $(\psi(0) - \psi(\alpha_0^{Fast}) > \psi(0) - \psi(\alpha_0^{Fast})$, which is impossible.

The remaining possibility is that $(\alpha_0, \beta_0) = (0, 1)$ yields a larger social welfare than $(\alpha^{SO}, 0)$. However, as explained in the text, we have: $W(0,1) = W(0,0) = \psi(0)$. Moreover, if $\lambda \leq \hat{\lambda}(\sigma, C, \pi)$, we have $W(\alpha^{SO}, 0) > W(0,0)$. Hence, we deduce that $W(\alpha^{SO}, 0) > W(0,1)$.

Thus, we have shown that $W(\alpha, \beta) \leq W(\alpha^{SO}, 0)$ for all $0 \leq \alpha \leq 1$ and $\beta \geq 0$ when $\lambda \leq \widehat{\lambda}(\sigma, C, \pi)$. When $\lambda > \widehat{\lambda}(\sigma, C, \pi)$, $\alpha^{SO} = 0$. Thus, $W(\alpha^{SO}, 0) = W(0, \beta)$, $\forall\beta$. In sum, the allocation $(\alpha, \beta) = (\alpha^{SO}, 0)$ dominates (at least weakly and sometimes strictly) any other allocations (α, β). The allocation $(\alpha^{SO}, 0)$ can be implemented with (i) a tax larger than $\psi(0)$ imposed to institutions joining the slow market and (ii) a tax T^* on institutions investing in the fast trading technology chosen as described in Proposition 7. Indeed, the first tax is larger than institutions' expected profit, $\psi(0)$, on the slow market. Hence, it deters all institutions to join the slow market. Furthermore, as Proposition 7 shows, the second tax induces a level of fast trading just equal to α^{SO} when there is no slow market or equivalently when $\beta = 0$.

Bibliography

Baron, M., Brogaard, J., Kirilenko, A., 2014. Risk and return in high frequency trading. Unpublished working paper, Princeton University.

Bagnoli, M., Bergstrom, T., 2005. Log-concave probability and its applications. Economic Theory 26, 445-469.

Bessembinder, H., Hao, J., Zheng, K., 2013. Market Making Obligations and Firm Value. Unpublished working paper, David Eccles School of Business.

Breon–Drish, B., 2013. On Existence and Uniqueness of Equilibrium in Noisy Rational Expectations Economies. Unpublished working paper, Stanford University.

Brogaard, J., Hendershott, T., Riordan, R., 2014. High frequency trading and price discovery. Review of Financial Studies 27, 2267-2306.

Brogaard, J., Hagströmer, B. , Norden, L., Riordan, R., 2014. Trading fast and slow: colocation and market quality. Unpublished working paper, University of Stockholm.

Budish, E., Cramton, P., Shim, J., 2014. The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. Unpublished working paper, University of Chicago.

Chiyachantana, C., Jain, P., 2009. Institutional Trading Frictions. Unpublished working paper, University of Memphis.

Colombo, L., Femminis, G., Pavan, A., 2014. Information acquisition and welfare. Forthcoming Review of Economic Studies.

Dow, J., 2005. Is liquidity self-fulfilling. Journal of Business 77, 895-908.

Du, S., Zhu, H., 2013. Dynamic ex–post equilibrium, welfare and optimal trading frequency in double auctions. Unpublished working paper, MIT.

Duffie, D., Garleanu, N., Pedersen, L. H., 2005. Over-the-Counter Markets. Econometrica 73, 1815-1847.

Ganguli, J., Yang, L., 2009. Complementarities, multiplicity, and supply information. Journal of the European Economic Association 7, 90-115.

Glode, V., Green, R., Lowery, R., 2012. Financial Expertise as an Arms Race. Journal of Finance 67, 1723-1759.

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

Grossman, S., Stiglitz, J., 1980. On the Impossibility of Informationaly Efficient Markets. American Economic Review 70, 393-408.

Grossman, S., 1976. On the efficiency of competitive stock markets where traders have diverse information. Journal of Finance 31, 573-585.

Hasbrouck, J., 2005. Empirical market Microstructure. Oxford University Press, Inc. New-York.

Hendershott, T., Riordan R., 2013. Algorithmic Trading and the Market for Liquidity. Journal of Financial and Quantitative Analysis 48, 1001-1024.

Hirshleifer, J., 1971. The Private and Social Value of Information and the Reward to Inventive Activity. American Economic Review 61, 561-574.

Hu, X., Qin, C. 2013. Information acquisition in financial markets and welfare effects in a model of competitive financial markets. Economic Theory 54, 199-210.

Jovanovic, B., Menkveld, A., 2011. Middlemen in limit order markets.Unpublished working paper, New York University.

Kirilenko, A., Kyle, P., Samadi, M., Tuzun, T., 2011. The Flash Crash: The impact of high-frequency trading on an electronic market. Unpublished working paper, University of Maryland.

Llosa, L.G., Venkateswaran, V., 2012. Efficiency with endogenous information choice. Unpublished working paper. University of California at Los Angeles.

Manzano, C., Vives, X., 2012. Public and private learning from prices, strategic substitutability and complementarity, and equilibrium multiplicity. Journal of Mathematical Economics 47, 346-369.

O'Hara, M., Ye, M., 2011. Is market fragmentation harming market quality. Journal of Financial Economics 100, 459-474.

Pagnotta, E., Phillipon, T., 2013. Competing on Speed. Unpublished working paper, Stern Business School, New-York University and NBER.

Vayanos, D., 1999. Strategic trading and welfare in a dynamic market. Review of Economic Studies 66, 219-254.

Zhang, S., 2013. Need for speed: an empirical analysis of high and soft information in a high frequency world. Unpublished working paper, University of Manchester.

Zhu, H., 2014. Do dark pools harm price discovery? Review of Financial Studies 27, 748-789.

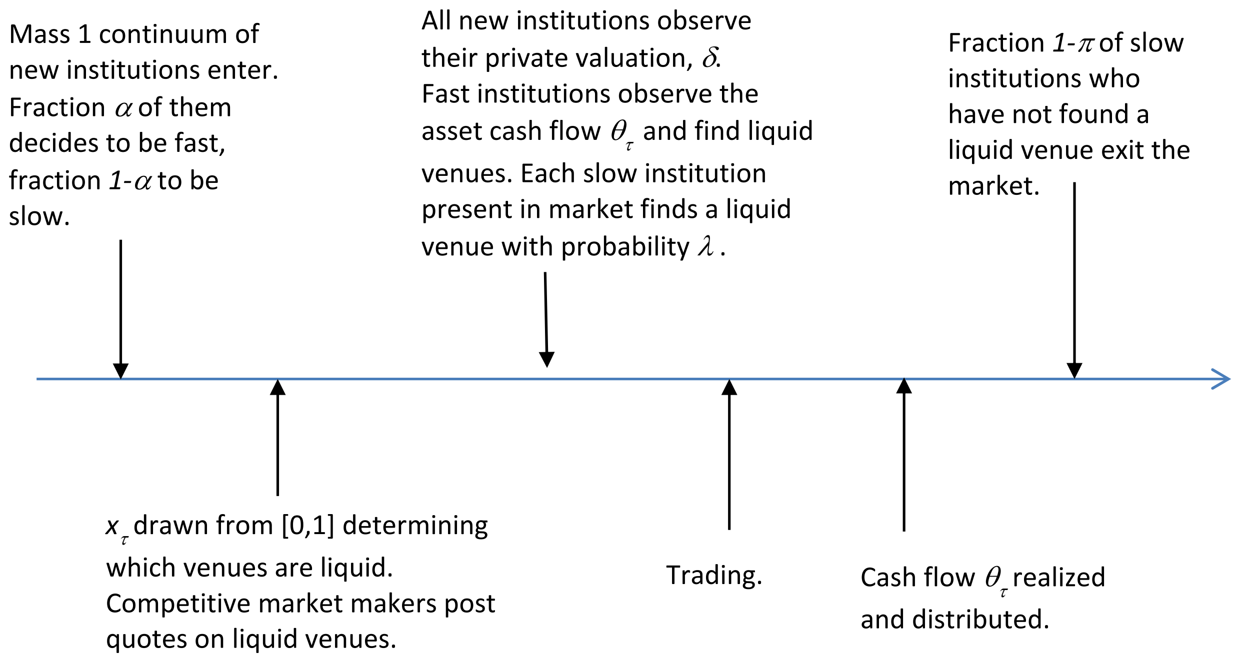

Figure 1: Timing of Institutions' decisions. This figure shows the sequence of play within one period, say, τ.

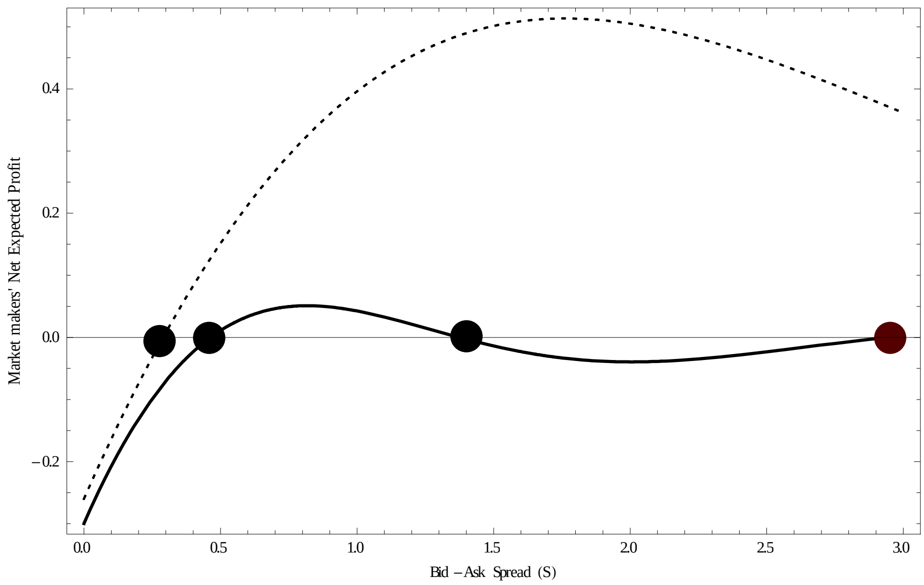

Figure 2: Equilibrium Uniqueness of the Bid-Ask Spread. This figure shows market makers' expected profit (the difference between the left hand side and the right hand side of Eq. (11)) as a function of their bid-ask spread (S) when the distribution of traders' private valuation is normal with standard deviation, σ_δ=0.9 (plain line) and σ_δ=2 (dashed line). Other parameter values are α=0.1, σ=3, λ = 0.8 and π=1. Dots are values of zero profits bid-ask spreads.

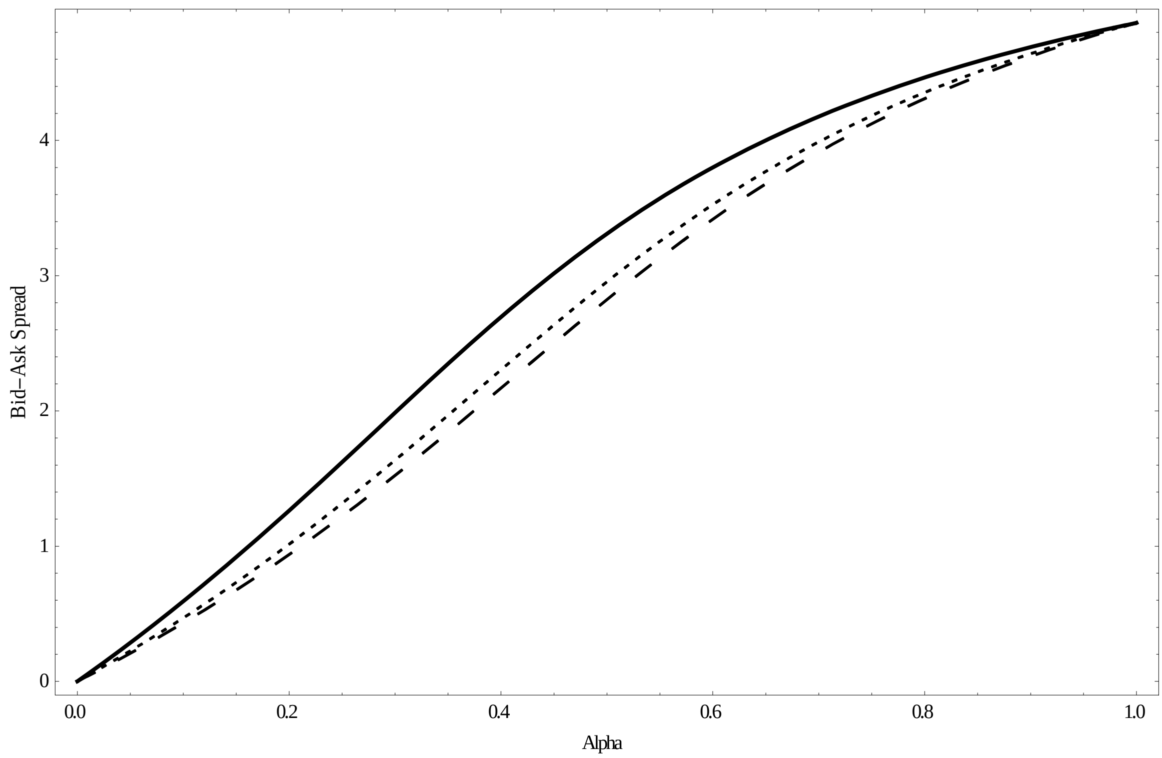

Figure 3: Bid-Ask Spreads and Fast Trading. This figure shows the equilibrium bid-ask spread (S^*) as a function of the level of fast trading (α) when the distribution of traders' private valuation is normal. Parameter values: λ=0.2 (plain line), λ =0.5 (small dashed line), λ =0.9 (large dashed line) for σ=5, σ_δ = 4, and π=0.9.

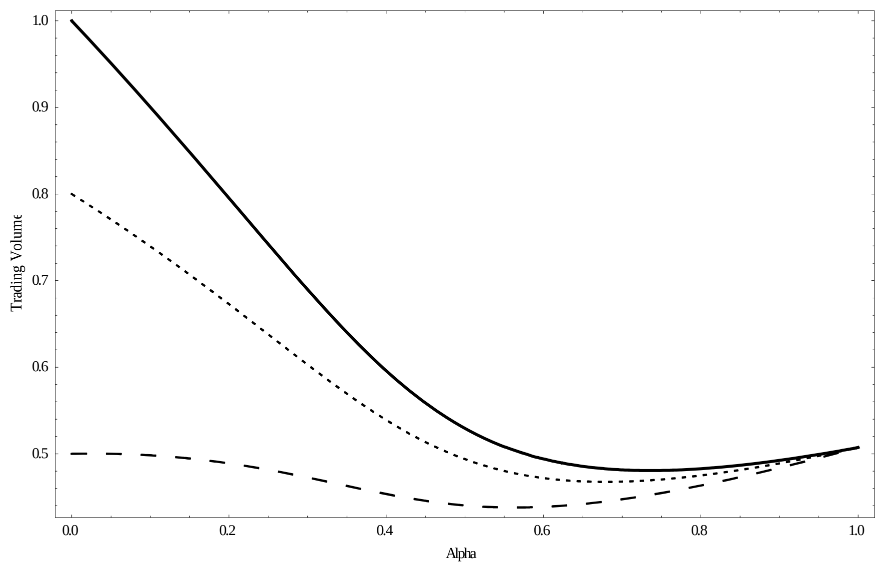

Figure 4: Trading Volume and Fast Trading. This figure shows equilibrium trading volume as a function of the level of fast trading (α) for different values of λ: 0.5 (large dashed line), 0.8 (dotted line), and 1 (plain line) when δ has a normal distribution with σ_δ=3.5, σ=5, and π=1.

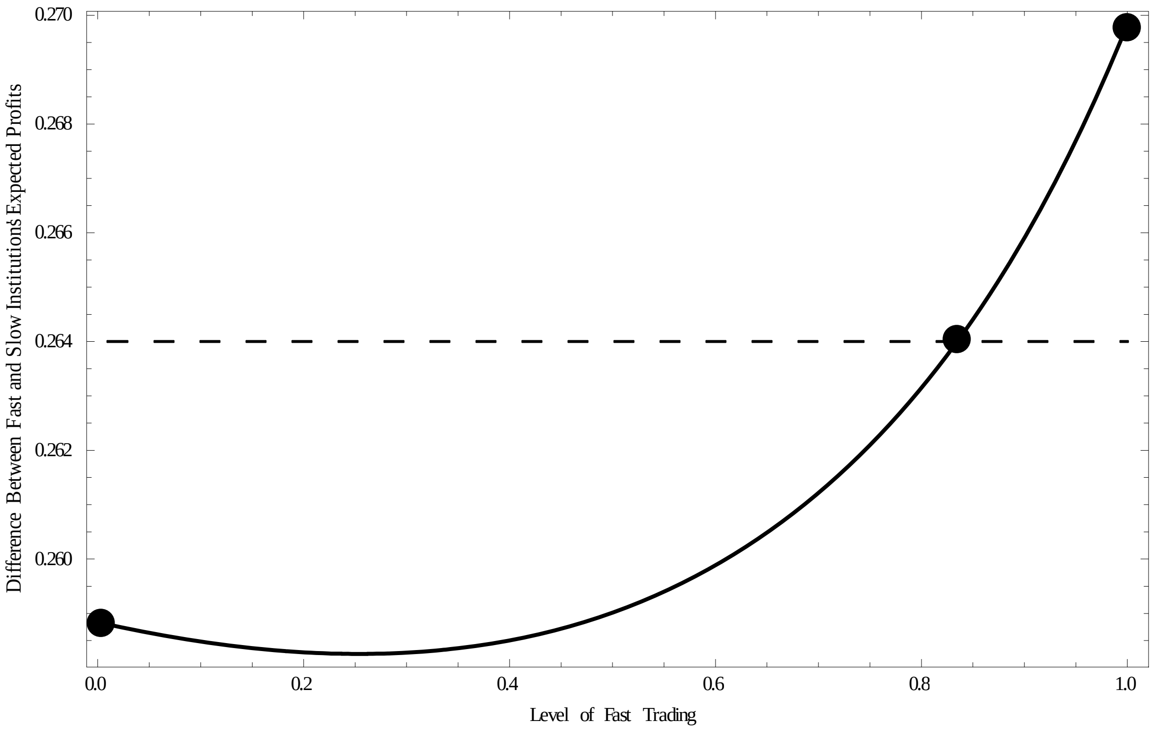

Figure 5: Equilibrium Fast Trading When Institutions' Decisions are Complements. The distribution of institutions' private valuation is as in Example 2, with φ=1.5, C=0.264, σ=3, δ=7, λ=0.5, π=0.99, and r=0. In this case, λ^s=0.99. Equilibrium levels of fast trading are indicated by large dots.

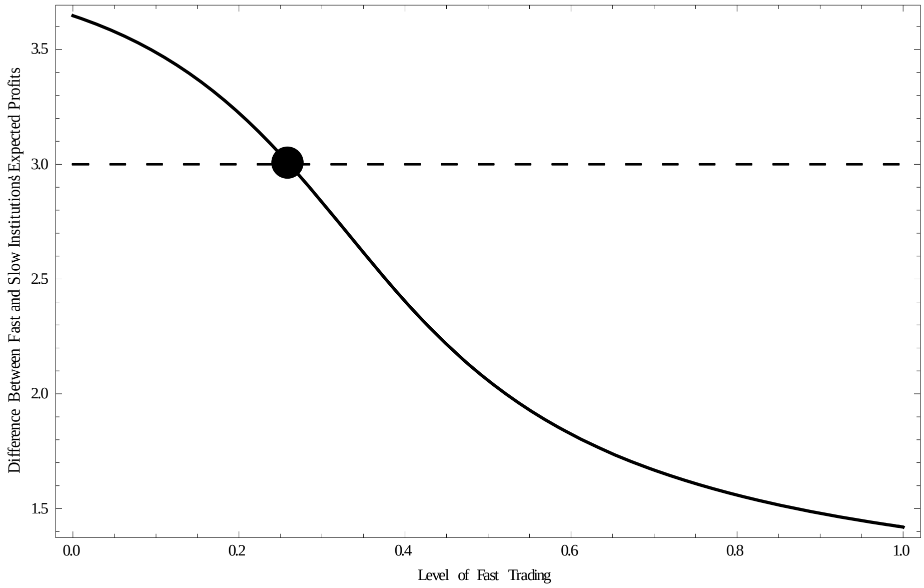

Figure 6: Equilibrium Fast Trading When Institutions' Decisions are Substitutes. Institutions' private valuations have a normal distribution with standard deviation σ_δ=4. Other parameters are λ=0.3, π=0.9, σ=6, r=0, and C=3. In this case, λ^s =0.8 and the equilibrium level of fast trading is indicated by the large dot (α^*=0.26).

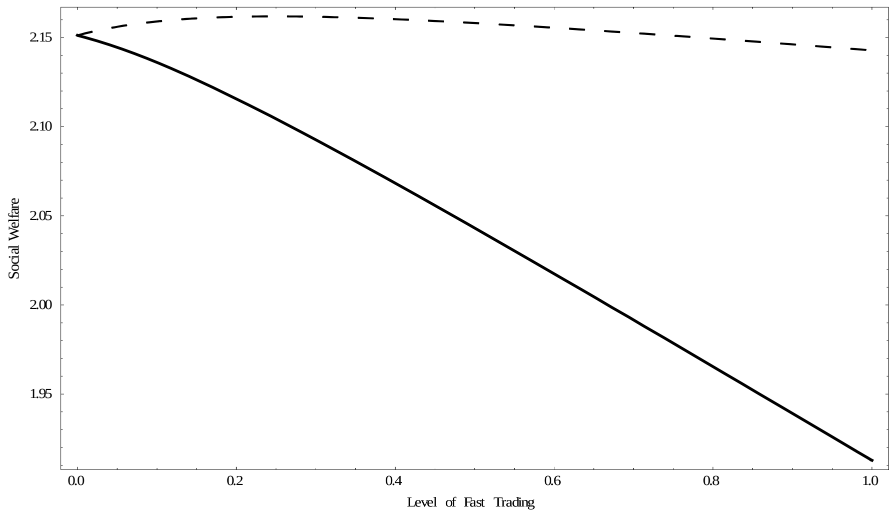

Figure 7: Social Welfare and Fast Trading. This figure shows social welfare as a function of the level of fast trading for C=4.77 (dashed line) and C =5 (plain line) when institutions' private valuations are normally distributed with σ_δ =10. Other parameters are σ=5, π=0.9, r=0, and λ=0.0356 (so that λ^s=0.27).