Equilibrium Fast Trading[1]

Bruno Biais[2] Thierry Foucault[3] Sophie Moinas[4]

March, 2014

[1]Many thanks for helpful comments to an anonymous referee, Alex Guembel, Terrence Hendershott, Andrei Kirilenko, Carolina Manzano, Albert Menkveld, Anya Obizhaeva, Rafael Repullo, Ailsa Roëll, Xavier Vives and participants in the conference on "High Frequency Trading: financial and regulatory implications" in Madrid, the 2012 SFS finance cavalcade, and the 2013 American Finance Association Meetings. Moinas acknowledges the support of ANR (ANR-09-JCJC-0139-01). Biais acknowledges the support of the ERC (Grant 295484 - TAP.)

[2]Toulouse School of Economics (CNRS-CRM and FBF-IDEI Chair on Investment Banking and Financial Markets); e.mail: bruno.biais@univ-tlse1.fr

[3]Department of Finance, HEC, Paris; e-mail: foucault@hec.fr

[4]Toulouse School of Economics (Université de Toulouse 1 Capitole and CRM); e.mail: sophie.moinas@iae-toulouse.fr

Abstract

High–speed market connections and information processing improve the ability to seize trading opportunities, raising gains from trade. They also enable fast traders to process 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 platforms: one accepting fast traders, the other banning them. Utilitarian welfare is maximized by having i) a single platform on which fast and slow traders coexist and ii) Pigovian taxes on investment in the fast–trading technology.

1 Introduction

Investors must process very large amounts of information, both about the fundamentals of the economy (earnings, growth, interest rates,...) and about the evolution of the market (prices, quotes, volume). The latter is particularly difficult to collect in today's fragmented markets (see, e.g., O'Hara ad Ye, 2011). For example, Wall Street institutions trading an NYSE listed stock must closely monitor the dynamics of the order book and order flow on the Exchange, and also other markets, such as, e.g., Nasdaq, Bats and Direct Edge, or derivatives markets.

To cope with this huge flow of information, financial institutions invest in fast connections to trading venues and high–speed information processing capacities. For example, trading firms can buy co–location rights, i.e., the placement of their computers next to the exchange's servers, to reduce latencies (the delay between emission and reception of a message) by a few milliseconds. Another example is fiber optic cables strung over the Atlantic or between Chicago and New York, enabling their users to get information from, and send orders to, markets 5 milliseconds before their competitors. Yet another example is the venture between the inter-dealer broker BCG Partners and the high–frequency trading group Tradeworx announced in August 2012 to explore data transmission through the air on microwave radio signals, as the latter travel nearly 50 percent faster than light through optical fiber. Other forms of such investments include the purchase of powerful computers and the development of smart programs collecting and comparing data from several markets and automatically firing orders based on this data.[1]

Investments in fast trading technologies have grown considerably in recent years. And they are expensive. 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. Against this cost, investments in fast trading technology have both positive and negative consequences for the functioning of markets.

[1]Investment also includes investment in human capital. For example, about a third of the employees of Renaissance Technologies (a hedge fund that is extremely active in fast computerized trading) have Ph.Ds.

On the one hand, they help traders cope with market fragmentation. Thus, financial institutions can seize trading opportunities before they vanish. This enhances market participants' ability to reap mutually beneficial gains from trade. On the other hand, investment in high–speed connections and information processing enable fast traders to access value relevant information before slow traders. Thus, fast traders are superiorly informed about future price changes. For example, Kirilenko et al. (2011) note that "*possibly due to their speed advantage or superior ability to predict price changes, high–frequency traders are able to buy right as the prices are about to increase.*" Consistent with this view, Hendershott and Riordan (2013) or Brogaard, Hagströmer, Norden, and Riordan (2014) find that fast investors' orders are more informative than slow ones (see for instance Table 5 in Brogaard, et al. (2014)). The informational advantage of fast institutions raises adverse selection costs. For example, Baron, Brogaard, and Kirilenko (2012) 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 concerns about high–frequency traders imposing adverse selection on other investors"*.[2]

These observations raise a number of issues: Overall, do fast traders enhance or deteriorate the functioning of markets? Do market forces lead to an optimal amount of investment in fast–trading technologies? Is policy intervention called for? And if so, which policy responses are most appropriate?

To examine these issues, we consider a simple model suitable for welfare and policy analysis. Our model features financial institutions with 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.[3]

Initially, institutions can decide to invest in a fast–trading technology, which brings two benefits: (a) advance information about the common value of the asset and (b) the

[2] They also point to the benefits of high–frequency traders with regards to price efficiency. They note, however, that such benefits might be limited, as they find that high–frequency traders' orders predict prices only on very short horizons, of less than 4 seconds.

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

ability to always secure a trade when perceiving an opportunity. If an institution decides not to invest in the fast–trading technology and remains slow hen (a) it does not observe advance information and (b) it may not be able to perfectly implement its trading plan. For example, suppose the institution wants to buy 500 shares with an executable limit order. It has to search for the best quotes across market venues. When the institution is fast, it is able to conduct this search efficiently, to rapidly locate counterparties, and to execute the purchase. In contrast, when the institution is slow, it takes time to search for quotes and counterparties. During this delay, attractive quotes may be cancelled or executed, preventing the institution from executing its order. This execution risk is increasing in the fragmentation of the market. To model this in the simplest possible way, we assume a slow institution finds a trading counterparty and executes its trade with probability $\lambda < 1$, only. Because, as discussed above, market fragmentation makes it difficult for slow institutions to execute their trades, λ decreases with the fragmentation of the market.

We first analyze equilibrium allocations and prices for a given fraction (α) of fast institutions. The larger α, the greater the information content, and hence price impact, of trades. Now, institutions prefer abstain from trading when the price impact cost 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.

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 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 the equilibrium level of investment in the fast trading technology is too high compared to the level maximizing utilitarian welfare. Indeed, when institutions decide to be fast, they account for the private benefit and cost of this decision but they ignore the negative externality they inflict on other traders. As a result, there is in general too much investment in fast trading. 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.

We analyze various possible policy interventions to mitigate 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 one is to let a "slow market" (on which fast trading is banned) coexist with the fast market. This always dominates a complete ban on fast trading or "laissez-faire". However, it can lead to *underinvestment* in the fast trading technology. Slow institutions migrate to the slow market where there is no adverse selection. This reduces the expected profits of fast institutions. In this context, there are only two possible equilibrium outcomes: either all institutions are slow, 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, there is too little investment in the fast trading technology relative to the utilitarian optimum. The second approach is Pigovian taxation of the fast trading technology. 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. Thus, Pigovian taxation yields better outcomes than having slow and fast platforms.

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 should raise the informational content of trades. This reduction has an ambiguous effect on trading volume, however. Indeed, it increases the chance that an institution is able to carry out its desired trades, but it also raises price impact costs. Consequently, trading volume is non monotonic in the level of fast trading. The model also implies that an increase in market fragmentation should lower the profitability of fast institutions because it increases the informativeness 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

Our analysis is in line with the seminal paper of Grossman and Stiglitz (1980). In both cases, the fraction of informed agents affects the outcome of the trading process and is determined in equilibrium. The two main differences between our model and theirs are the following: First, in our framework, all participants are rational and make optimal decisions, i.e., there are no noise traders. Second, investment in fast trading does not only generate advance information, it also enhances the ability to seize trading opportunities. These differences in modelling approaches yield differences in results.

First, as all traders are rational and have well defined preferences, we can perform a welfare analysis of investment in fast trading technologies, comparing equilibrium investment to its socially optimal counterpart. Welfare analyses in models of informed trading in financial markets are scarce, precisely because these models often rely on exogenous noise trading.

Second, in our model, there exist parameter values for which the socially optimal level of investment in the fast trading technology is strictly greater than zero, despite the fact that this technology is a source of adverse selection. The reason is that it also increases the likelihood for traders to realize gains from trade. Thus, in computing social welfare, one must trade-off this benefit with the negative externality associated to greater informational asymmetries, which lead to lower gains from trade. This benefit needs to be small for the socially optimal level of fast trading to be zero.

Third, in Grossman and Stiglitz (1980), investments in information acquisition are always strategic substitutes. In contrast, in our model they can also be strategic complements. In this case, institutions' investment decisions reinforce each other, which raises the possibility of equilibrium multiplicity and investment waves. In a CARA–Gaussian model, Ganguli and Yang (2009) analyze the case where traders observe imperfect private signals on aggregate supply as well as on common values. Breon–Drish (2013) extends Grossman and Stiglitz (1980) to the non Gaussian case. In both models, complementarity in information acquisition arises when prices become less informative as the number of informed investors increases.[4] This interesting mechanism is completely different from that at play in our model, whereby financial institutions 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.

Budish, Cramton, and Shim (2014) develop a model in which traders invest in speed to be first to react to and profit from information arrival. In their model, however, trading is a zero sum game and the social cost of fast trading is just the technological costs borne by investors. Furthermore, fast trading has no social benefit. Hence, trading slow (they advocate batch auctions every second) is always the socially optimal solution. In contrast, in our model, the socially optimal market structure calls for coexistence of fast and slow institutions in many cases.

Pagnotta and Phillipon (2013) analyze competition in speed between markets. A faster

[4]In contrast, in our model, the informativeness of prices always increases in the fraction of institutions that are fast.

market in their model means that investors can interact with the market, and therefore realize gains from trade, more frequently. Similarly, in our model, fast institutions are more likely to carry out their desired transactions than slow institutions. However, in addition, they can also obtain advance information. As previously explained, this is key for our findings. Furthermore, in our model, each investor chooses the speed at which it operates on a given market. In contrast, in Pagnotta and Phillipon (2013), each market chooses the speed at which all its participants operate. Excessive investment in speed can arise in their model because markets seek to relax competition through vertical differentiation. In contrast, in our model, excessive investment arises because investors do not internalize the adverse selection cost they inflict on others when they become fast. 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). In sum, Pagnotta and Phillipon (2013) focus on the evolution of trading technologies supplied by markets (platform's infrastructure) whereas we focus on the demand of fast trading technologies by investors (traders' infrastructure). Thus, the two models complement each other.

3 Model

Consider a population of risk–neutral financial institutions, indexed by the trading round at which they contact the market. Institution $\tau \in \{1, ..., T\}$ trades a short–lived asset at the beginning of trading round τ. At the end of the trading round, the asset pays off cash–flow, θ_τ, equal to $+\epsilon$ or $-\epsilon$ with probability $\frac{1}{2}$, where $\epsilon \geq 0$. Across periods, cash-flows are i.i.d.[5]

Investment in the fast–trading technology: At date $\tau = 0$, before any trading occurs, all institutions simultaneously decide whether to invest in infrastructures (computers, co-location, etc.) and intellectual capital (skilled traders, codes, etc.) to increase the speed with which they receive information from markets, process this information, and act

[5]The assumption that the asset is short–lived is just for simplicity. Qualitatively identical results would obtain if we considered patient agents, trading a long lived asset paying–off $\sum_{\tau=1}^{T} \theta_\tau$ at time T, where θ_τ is publicly observed at the end of each period. The only difference would be that, at round t, the unconditional expectation of the dividend would be $\sum_{\tau=1}^{t-1} \theta_\tau$ instead of 0.

upon it. The cost of this investment is C. The institutions who invest in the fast trading technology are "fast," the others are "slow." The fraction of fast institutions, α, is the *level of fast trading* in the market. It is endogenized in Section 5.

In reality, as explained in the introduction, investment in fast trading technologies provides financial institutions with a quicker access to two types of information. First, fast institutions can access and process information on common values before other market participants. To capture this, in the simplest possible way, we assume that if institution τ is fast then it observes θ_τ, just before trading.

Second, fast institutions are more likely to fully realize their private gains from trade because they have a more comprehensive information on market conditions (e.g., posted quotes) and can take advantage of good deals more efficiently. Regulations such as the MiFID in Europe or RegNMS in the U.S. led to fragmentation of trading among competing platforms.[6] As a result, investors have to compare trading opportunities among several markets. Fast access to and processing of market data help investors in locating attractive quotes before they are taken or withdrawn. It also helps traders to split their orders in the most profitable way across trading venues (see Foucault and Menkveld (2008)). Accordingly, we assume that slow institutions are less likely to realize gains from trade than fast institutions.[7] Namely, if institution τ wishes to trade, it can realize its desired trade with certainty if it is fast but with only probability λ if it is slow. This assumption formalizes the notion that less efficient search for the best price results in a utility loss for slow traders (e.g., less efficient hedging).

Investors' valuations: The valuation of institution τ is $v_\tau = \theta_\tau + \delta_\tau$ where δ_τ is its private valuation for the asset. At the beginning of round τ, institution τ observes its private valuation, δ_τ. If it is fast, the institution also observes the realization of the common value component, θ_τ. If it is slow, it observes θ_τ only at the end of period τ.

[6]For instance, in February 2014, the three most active competitors of the London Stock Exchange, namely Chi-X Europe, Turquoise, and BATS Europe reached a daily market share in FTSE 100 stocks of 19.6%, 11.9% and 6.0% respectively, while that of the London Stock Exchange was 62.3%. Source: http://http://fragmentation.fidessa.com/.

[7]Hendershott and Riordan (2013) write that "*Technological progress in the form of algorithmic trading [...] reduces monitoring frictions, which [...] facilitates gains from trade.*" They show for instance that algorithmic (i.e., fast) traders are more likely than human (slow) traders to hit quotes when liquidity is cheap (e.g., bid-ask spreads are narrow).

Differences in private values 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.[8] Differences in tax regimes can also induce differences in private values.

We assume private valuations are i.i.d. across institutions and continuously distributed on $[-\overline{\delta}, \overline{\delta}]$ with a cumulative probability distribution $G(\cdot)$ and density function $g(\cdot)$. The average private valuation is zero ($E(\delta_\tau) = 0$), and $G(\cdot)$ is symmetric around its mean: $G(\delta) = \Pr(\delta_\tau \leq \delta) = 1 - \Pr(\delta_\tau \geq -\delta)$. Thus, $G(0) = \frac{1}{2}$. In several examples, we consider the limit case where $\overline{\delta} \to \infty$ and private valuations are normally distributed with standard deviation σ_δ.

Trading: For simplicity, we assume trades can only take three values: 0, 1, or -1. Thus, in trading round τ, institution τ can seek to buy one share ($\omega_\tau = 1$), sell one share ($\omega_\tau = -1$), or abstain from trading ($\omega_\tau = 0$). If the institution is fast, it locates counterparties and executes ω_τ with probability 1. If the institution is slow, it locates counterparties and executes ω_τ with probability $\lambda \leq 1$. ω_τ, if executed, trades at price $E(\theta_\tau|\omega_\tau)$, the expectation of the common value of the payoff conditional on the order.

This is the simplest specification of the outcome of the trading process that is consistent with rationality and participation constraints: the transaction's price reflects the information content of the order and, correspondingly, the trading counterparties of the institution earn non–negative expected profits. A micro–foundation for this specification is to assume the institution sends a market order, executed against the quotes placed by competitive risk neutral market-makers, as in Glosten and Milgrom (1985) or Easley and O'Hara (1987). This is in line with the empirical findings of Brogaard, Hendershott, and Riordan (2014) that high–frequency traders trade in the direction of permanent price changes with market orders, and Baron, Brogaard, and Kirilenko (2012) that most of high–frequency traders' profits are generated by aggressive, liquidity–taking, trades. Our qualitative conclusions, however, should be robust to generalizing this specification: the driving force underlying

[8] For instance, by regulatory requirements, some institutional investors can only hold investment grades bonds. Thus, they value these bonds at a premium relative to other investors.

our results is the increased adverse selection generated by fast trading, which hurts all institutions. This effect arises independently of the specification of the trading game.

Timing: Figure 1 recaps the timing of the model. At $\tau = 0$, each institution decides whether to pay C, and become fast, or not. Then, each institution has one trading opportunity. At date τ, institution τ observes its private valuation δ_τ, and, if it is fast, it observes θ_τ. Then it optimally chooses its portfolio ($\omega_\tau = 1$, $\omega_\tau = -1$, or $\omega_\tau = 0$) and it trades at price $E(\theta_\tau|\omega_\tau)$, with certainty if the institution is fast, and with probability λ if it is slow. Then the asset pays off and the next institution lines up for the next round of trade.

[**Insert Figure 1 About Here**]

As θ_τ and δ_τ are i.i.d., each trading round is identical. For notational simplicity, we hereafter drop the subscript τ.

4 Trading with fast and slow investors

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

4.1 Equilibrium: Definition, Existence, and Uniqueness

Let a and b be the prices at which institutions buy and sell the asset (the "ask" and "bid" prices). Furthermore, let $v(\delta, i) = i\epsilon + \delta$ denote the expected valuation of institution τ at the beginning of trading round τ, with $i = 0$ if the institution is slow, $i = 1$ if the institution is fast and the asset cash-flow is high, and $i = -1$ if the institution is fast and the asset cash-flow is low. An institution with type i and private valuation δ buys the asset if $v(\delta, i) \geq a$, sells it if $v(\delta, i) \leq b$, and does not trade if its valuation falls within the bid-ask spread.[9] Thus, in equilibrium, the ask and bid prices, a^* and b^*, solve:

$$a^* = E(\theta \,|\omega = 1) = E(v \,|v(\delta, i) \geq a^*)\,, \tag{1}$$

$$b^* = E(\theta \,|\omega = -1) = E(v \,|v(\delta, i) \leq b^*)\,. \tag{2}$$

Equilibrium conditions (1) and (2) are similar to the equations defining equilibrium in the limit–order models of Glosten (1994) and Biais, Martimort, and Rochet (2000).

We have:

$$E(\theta \,|v(\delta, i) \geq a^*) = \Pr(\theta = \epsilon \,|v(\delta, i) \geq a^*)\,\epsilon + (1 - \Pr(\theta = \epsilon \,|v(\delta, i) \geq a^*))(-\epsilon).$$

That is,

$$E(\theta \,|v(\delta, i) \geq a^*) = (2 \Pr(\theta = \epsilon \,|v(\delta, i) \geq a^*) - 1)\epsilon. \tag{3}$$

Similarly:

$$E(\theta \,|v(\delta, i) \leq b^*) = (2 \Pr(\theta = \epsilon \,|v(\delta, i) \leq b^*) - 1)\epsilon. \tag{4}$$

For a given private valuation, fast institutions are more likely to buy the asset if they have good information ($i = 1$) than if they have bad information ($i = -1$) because $v(\delta, 1) > v(\delta, -1)$ for $\epsilon > 0$. Thus, $\Pr(\theta = \epsilon \,|v(\delta, i) \geq a^*) \geq \frac{1}{2} \geq \Pr(\theta = -\epsilon \,|v(\delta, i) \leq b^*)$. We deduce from (1), (2), (3), and (4) that $a^* \geq b^*$ with strict inequalities iff $\alpha > 0$ and $\epsilon > 0$. As the distributions of cash-flows and private valuations are symmetric, we focus on symmetric equilibria in which $a^* = -b^*$.

To study the properties of these equilibria, it is convenient to analyze how the expected profit of the market participants quoting the ask price vary with this price. Let

$$\Pi(a; \alpha, \lambda, \epsilon) = a - E(v \,|v(\delta, i) \geq a)$$

be expected profit of the sellers at price a. Equilibrium ask prices, a^*, are such that this profit is zero (see (1)), i.e., $\Pi(a^*; \alpha, \lambda, \epsilon) = 0$. Note that $\Pi(a; \alpha, \lambda, \epsilon)$ is not necessarily

[9] If an institution is indifferent between buying the asset or not (i.e., $v(\delta, i) = a$), we assume that it buys the asset. This tie-breaking rule is innocuous because this event has zero probability since δ has a continuous distribution.

increasing in a. Indeed, an increase in a has two effects. On the one hand, it raises the sellers' revenue per trade. On the other hand, it also changes the mix of institutions buying at this price since only institutions with a valuation larger than the ask price buy the asset. Correspondingly, trades convey a stronger informational signal. The latter effect can be stronger than the former (a \$1 increase in a can raise $E(v\,|v(\delta,i)\geq a)$ by more than \$1), especially when adverse selection is strong. When this happens, $\Pi(a;\alpha,\lambda,\epsilon)$ can increase and decrease with a so that there exist multiple prices for which $\Pi(a;\alpha,\lambda,\epsilon)=0$ and therefore (1) has multiple solutions a^*, as the next example illustrates.[10]

Example 1. Suppose that institutions' private valuations are normally distributed with standard deviation σ_δ. Figure 2 plots $\Pi(a;\alpha,\lambda,\epsilon)$ when $\lambda=0.8$, $\alpha=0.1$, and $\epsilon=3$, for $\sigma_\delta=1$ or $\sigma_\delta=2$. When $\sigma_\delta=1$, $\Pi(a;\alpha,\lambda,\epsilon)$ is non monotonic in a. For this reason, there are three ask prices such that (1) holds: $a^*=0.60$, $a^*=1.56$, and $a^*=2.83$. When $\sigma_\delta=2$, the adverse selection problem is less acute. In this case, sellers' expected profit decreases in a everywhere and, as a result, there is a unique equilibrium ask price, $a^*=0.365$.$\blacksquare$

[Insert Figure 2 about here]

Lemma 1 *: An equilibrium price always exists because (1) always has at least one solution, $0\leq a^*\leq\epsilon$. When $\alpha=0$ or $\epsilon=0$, the unique equilibrium is $a^*=0$. Otherwise, equilibrium is not necessarily unique.*

When there exist multiple solutions to (1), economic reasoning suggests to select prices that cannot be profitably undercut. Consider Figure 2 again. Ask prices $a^*=1.56$ and $a^*=2.83$ satisfy the zero profit condition (i.e., they solve (1)) but they can be profitably undercut because any price sufficiently close to and above $a^*=0.60$ yields a strictly positive expected profit to a seller. This is a more general principle. When (1) has several solutions, only the smallest solution cannot be profitably undercut as the next lemma states.

[10]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. This phenomenon is pervasive and can arise whether or not $\overline{\delta}<\epsilon$. In fact, when $\overline{\delta}<\epsilon$, for each value of α, (1) has always at least two solutions, one in which $a^*=\epsilon$ and one in which $a^*<\overline{\delta}$.

Lemma 2 *: Let $a^*_{\min}(\alpha)$ be the smallest solution to (1). This equilibrium price is the only one that cannot be profitably undercut.*

Hence, if one adds the natural economic requirement that an equilibrium price should not be profitably undercut then equilibrium is always unique. We therefore focus on the equilibrium in which the ask price is $a^*_{\min}(\alpha)$ when there are multiple equilibrium prices. For our purposes, this assumption is conservative because the negative externality due to fast trading increases with the equilibrium ask price. Furthermore, the equilibrium in which the ask price is $a^*_{\min}(\alpha)$ is Pareto dominant because institutions' expected profits decline with the ask price (see (9) and (10) in Section 5).

4.2 Price Impacts, Trading Volume, and Fast Trading

We now study how the level of fast trading, α, affects the price impact of trades: a^* (a measure of market illiquidity), as well as trading volume.

Proposition 1 *(price impacts): The equilibrium ask price increases in the level of fast trading, α, and the volatility of the asset fundamental value, ϵ. It decreases with the likelihood that a slow institution finds a trading opportunity, λ.*

When α increases or λ decreases, orders are more likely to stem from fast institutions. Hence, the order flow is more informative and, for this reason, trades move prices more. Thus, the model implies that the informational impact of trades should increase with the level of fast trading.[11] Figure 3 illustrates this testable implication when the distribution of traders' private valuation is normal. It also shows that the equilibrium ask price decreases with the dispersion of traders' private valuations (i.e., when σ_δ increases), because heterogeneity in these valuations increases the variance of the non–informational component of the order flow, which lowers the informational content of trades.

[Insert Figure 3 about here]

[11] Hendershott, Jones, and Menkveld (2011) find that the informational impact of trades has declined on the NYSE after a change in market structure that made algorithmic trading easier on this market. However, it is not clear whether the change in market structure considered in Hendershott, Jones, and Menkveld (2011) corresponds to an increase in α or an increase in λ (the possibility for slow traders to better identify good prices in the market). For a fixed value of α, Proposition 1 implies that the informational impact of trades declines in λ, in line with their findings.

The difference between the likelihood of a trade by fast and by slow institutions at equilibrium plays an important role for the analysis of trading volume and the decisions to invest in the fast–trading technology (see next section). Let $\Delta\text{Vol}(a^*(\alpha), \alpha)$ be this difference when the level of fast trading is α. As institutions buy the asset if their valuation is higher than a^* and sell it if their valuation is lower than $b^* = -a^*$, we have:

$$\Delta\text{Vol}(a^*(\alpha), \alpha) = 2(\Pr(v(\delta, i) \geq a^* \,|i \neq 0) - \lambda \Pr(v(\delta, i) \geq a^* \,|i = 0)), \tag{5}$$

where the factor 2 comes from the symmetry of institutions' private valuations. Using this symmetry, straightforward manipulations of (5) yield

$$\Delta\text{Vol}(a^*(\alpha), \alpha) = 2\left[(1-\lambda)(1-G(a^*)) + \frac{1}{2}(G(a^*) - G(a^* - \epsilon)) + \frac{1}{2}(G(a^*) - G(a^* + \epsilon))\right],$$

which is equivalent to:

$$\Delta\text{Vol}(a^*(\alpha), \alpha) = 2\left[(1-\lambda)(1-G(a^*)) + \frac{1}{2}(G(a^*) - G(a^{*+})) - \frac{1}{2}(G(a^* + \epsilon) - G(a^{*-}))\right], \tag{6}$$

where $a^{*+} = Max\{a^* - \epsilon, -a^*\}$ and $a^{*-} = Max\{a^*, \epsilon - a^*\}$.

The two first terms in the brackets in (6) are positive while the latter is negative. Thus, the sign of ΔVol is ambiguous. To explain why, we now discuss the economic interpretation of the terms in the brackets in (6).

Consider a fast and a slow institution with the same private valuation, δ. In the absence of advance information on the asset payoff ($\epsilon = 0$), both institutions wish to trade at a^*, provided that $\delta > a^*$. The fast institution is more likely to do so because of its technological investment. This effect raises the likelihood of a trade for the fast institution relative to the slow one by $(1-\lambda)(1-G(a^*))$, the first term in the brackets in (6).

The fast trading technology also provides advanced access to cash-flow information and thereby it affects an institution's valuation for the asset. This effect cuts both ways in term of incentives to trade for a fast institution. On the one hand, good news about the asset cash-flow (that occurs with probability $\frac{1}{2}$) might induce an institution to buy the asset, whereas it would not trade without information (if it were slow). This happens when an institution's private valuation is in $[a^{*+}, a^*]$. This effect also raises the likelihood of a trade by fast institutions, relative to slow ones, by $\frac{1}{2}(G(a^*) - G(a^{*+}))$ (the second term in brackets in (6)). On the other hand, bad news about the asset cash flow might induce a fast institution to abstain from trading whereas it would buy the asset if slow. This happens if its private valuation is in $[Max\{b^* + \epsilon, a^*\}, a^* + \epsilon] = [Max\{\epsilon - a^*, a^*\}, a^* + \epsilon]$. This effect *reduces* the likelihood of a trade by a fast institution relative to an otherwise identical slow institution by $\frac{1}{2}(G(a^* + \epsilon) - G(a^{*-}))$ (the last term in (6)).

There exist specifications for the distribution of institutions' private valuations (see Example 2 in the next section) such that, for some values of α, the last effect dominates. In these cases, $\Delta\text{Vol}(a^*(\alpha), \alpha) < 0$: fast institutions trade less than slow institutions in equilibrium. At first glance, this possibility seems paradoxical because the fast trading technology allows an institution to lock in a trade with certainty. However, this will occur only if the institution finds it optimal to trade. By providing advanced information on future cash-flows, the fast trading technology also reduces the mass of states in which fast institutions wish to trade and can thereby reduce their trading likelihood.

The next lemma provides a sufficient condition on the distribution of institutions' private valuations such that fast institutions always trade more than slow institutions ($\Delta\text{Vol}(a^*(\alpha), \alpha) > 0$) 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 *: Fast institutions trade more frequently than slow institutions in equilibrium ($\Delta\,Vol(a^*(\alpha), \alpha) > 0, \forall \alpha$) 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) $\lambda \le \frac{1}{2}$, or (ii) $\lambda > \frac{1}{2}$ and*

$$2G(\epsilon) + \left(\frac{1 - G(2\epsilon)}{1 - G(\epsilon)}\right) \ge 2\lambda. \tag{7}$$

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.[12] In this case, for $\lambda \leq 1/2$, fast institutions always trade more than slow institutions simply because, conditional on wishing to trade, slow institutions are much less likely to realize their gains from trade (the first effect in (6)). When $\lambda > \frac{1}{2}$, this effect is not strong enough to guarantee that fast institutions always trade more than slow institutions and Condition (7) is required. This condition is satisfied when $\epsilon \geq Max\{G^{-1}(\lambda), 0\}$, i.e., when the asset volatility is sufficiently high and/or λ sufficiently low.

Equilibrium trading volume is

$$
\begin{aligned}
Vol(a^*(\alpha), \alpha) &= \Pr(v(\delta, i) \geq a^*) + \Pr(v(\delta, i) \leq -a^*) \\
&= \alpha(2 - (G(a^* - \epsilon) + G(a^* + \epsilon)) + 2(1 - \alpha)\lambda(1 - G(a^*)).
\end{aligned}
$$

We deduce that:

$$
\frac{d\mathrm{Vol}(a^*, \alpha)}{d\alpha} = \frac{\partial \mathrm{Vol}(a^*, \alpha)}{\partial \alpha} + \frac{\partial \mathrm{Vol}(a^*, \alpha)}{\partial a^*}\frac{\partial a^*}{\partial \alpha} = \Delta \mathrm{Vol}(a^*(\alpha), \alpha) + \underbrace{\frac{\partial \mathrm{Vol}(a^*, \alpha)}{\partial a^*}\frac{\partial a^*}{\partial \alpha}}_{<0}. \quad (8)
$$

Thus, when $\Delta\mathrm{Vol}(a^*(\alpha), \alpha) > 0$, the effect of an increase in the level of fast trading on trading volume, $\mathrm{Vol}(a^*(\alpha), \alpha)$, is ambiguous (otherwise it is clearly negative). Indeed, a small 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. If fast institutions are more likely to trade than slow institutions (i.e., $\Delta\mathrm{Vol}(a^*(\alpha), \alpha) > 0$), this effect increases trading volume in equilibrium. Second, the increase in the level of fast trading raises the price impact of trades ($\frac{\partial a^*}{\partial \alpha} > 0$), which leads all institutions to abstain from trading more frequently. This effect (second term in (8)), always reduces trading volume.[13] For these reasons, the effect of an increase in the level of fast trading on trading volume can be non monotonic.

[Figure 4 about here]

Figure 4 illustrates this point when investors' private valuations are normally distributed. It depicts equilibrium trading volume ($\mathrm{Vol}(a^*(\alpha), \alpha)$) as a function of α for

[12] See Bagnoli and Bergstrom (2005).

[13] Formally: $\frac{\partial Vol(a^*, \alpha)}{\partial a^*} = -\alpha(g(a^* - \epsilon) + g(a^* + \epsilon)) - 2(1 - \alpha)\lambda g(a^*) < 0$.

various values of λ. Observe that trading volume is non monotonic in α, even when $\lambda = 1$. 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 at $\tau = 0$. To do so, we first analyze the gains of fast and slow institutions, which are then compared to determine investment decisions.

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. Let $\omega(\delta, i)$ be the trading decision of an institution with private valuation δ and type i, in equilibrium. Recall that $\omega(\delta, i) = 1$ if $v(\delta, i) \geq a^*$, $\omega(\delta, i) = -1$ if $v(\delta, i) \leq b^*$, and $\omega(\delta, i) = 0$, otherwise. We deduce that:

$$\phi(\alpha) = 2E(\delta + i\epsilon - a^*(\alpha))\omega(\delta, i)\,|i \neq 0)\,,$$

where the factor 2 comes from the fact that $b^* = -a^*$ and the symmetry of institutions' private valuations. This implies:

$$\phi(\alpha) = \int_{a^*(\alpha)-\epsilon}^{\overline{\delta}} (\delta + \epsilon - a^*(\alpha))g(\delta)d\delta + \text{Max}\{\int_{a^*(\alpha)+\epsilon}^{\overline{\delta}} (\delta - \epsilon - a^*(\alpha))g(\delta)d\delta, 0\}. \quad (9)$$

The first term in (9) is the gain of fast institutions when they buy (resp. sell) the asset and the asset cash-flow is high (resp. low). The second term is their gain when they buy (resp. sell) the asset and its cash-flow is low (resp. high), which can occur only if $\overline{\delta} - \epsilon \geq a^*(\alpha)$. Similarly, for a slow institution, the ex–ante expected gain is:

$$\psi(\alpha) = 2\lambda \times \text{Max}\{\int_{a^*(\alpha)}^{\overline{\delta}} (\delta - a^*(\alpha))g(\delta)d\delta, 0\}, \quad (10)$$

[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.

because a slow institution with private valuation δ wishes to buy (sell) the asset only if $\delta \geq a^*(\alpha)$ and is actually able to carry out this trade with probability λ. The ability to trade fast raises the gains of fast institutions above the gains of slow ones. However, the expected gains of both types decline with $a^*(\alpha)$, and therefore α.[15] This is stated in the next proposition.

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 profits of slow and fast institutions.*

Proposition 2 implies that the entry of a new fast institution exerts a *negative externality* on all institutions, fast or slow, because it increases the price impact of their trades ($a^*(\alpha)$).

Henceforth, we assume that $\overline{\delta} \geq 2\epsilon$. This assumption is innocuous. It simply reduces the number of cases to analyze because, under this assumption, the expressions under the $Max\{\cdot\}$ in (9) and (10) are always strictly positive ($a^*(\alpha) + \epsilon \leq \overline{\delta}$ since $a^*(\alpha) \leq \epsilon$). When $\overline{\delta} < 2\epsilon$, this is not necessarily true and one must therefore consider several cases when manipulating $\phi(\alpha)$ and $\psi(\alpha)$. This lengthens the analysis without providing additional insights.

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

$$\phi(\alpha) - \psi(\alpha) \geq C.$$

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 in fast trading technologies are strategic substitutes: the greater the level of fast trading, the lower the relative value of fast trading. In contrast, if

[15] Hence, when there are multiple equilibria at date $\tau \geq 0$, the equilibrium with the smallest equilibrium price is Pareto dominant, as mentioned in the previous section.

$\phi - \psi$ 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 (9) and (10), we obtain after some algebra that:

$$\frac{\partial \Delta(\alpha)}{\partial \alpha} = -\frac{\partial a^*}{\partial \alpha} \times \Delta \mathrm{Vol}(a^*(\alpha), \alpha), \tag{11}$$

where $\Delta\mathrm{Vol}(a^*(\alpha), \alpha)$ (given in (6)) is the difference between the likelihood of a trade for a fast and a slow institution in equilibrium. The equilibrium ask price increases with α (Proposition 1). Hence, whether institutions' decisions are locally substitutes or complements is determined by the *sign* of $\Delta\mathrm{Vol}(a^*(\alpha), \alpha)$. This is intuitive: the increase in the cost of trading $(a^*(\alpha))$ due to an increase in the level of fast trading hurts more those institutions that trade more. As explained in the previous section, fast institutions trade more than slow ones in equilibrium iff $\Delta\mathrm{Vol}(a^*(\alpha), \alpha) > 0$. Thus, an increase in α reduces the relative value of fast trading iff $\Delta\mathrm{Vol}(a^*(\alpha), \alpha)$ is positive. Using (6) and $a^*(0) = 0$,

$$\Delta Vol(a^*(0), 0) = (1 - \lambda) > 0, \tag{12}$$

which implies $\frac{\partial \Delta(0)}{\partial \alpha} < 0$ (see (11)). 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 value of being fast. This is not necessarily the case for larger values of α, however, unless $\Delta\mathrm{Vol}(a^*(\alpha), \alpha) > 0$ for all levels of fast trading. This is the case if the distribution of institutions' private valuations satisfies one of the conditions in Lemma 3.

Corollary 1 *: Under the conditions of Lemma 3,* $\Delta Vol(a^*(\alpha), \alpha) > 0$, $\forall \alpha$. *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 substitutes at least for α close to zero (see (12)). 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. Assume $\overline{\delta} > 2\epsilon$. Define $\gamma = \left(\overline{\delta} - \varphi\left(\overline{\delta} - \epsilon\right)\right)/\epsilon$ with $\varphi \in [1, \frac{\overline{\delta}}{\overline{\delta}-\epsilon}]$. Assume $g(\delta) = \varphi(2\overline{\delta})^{-1}$ if $-\overline{\delta} \leq \delta \leq -\epsilon$, $g(\delta) = \gamma(2\overline{\delta})^{-1}$ if $-\epsilon \leq \delta \leq \epsilon$, and $g(\delta) = \varphi(2\overline{\delta})^{-1}$ if $\epsilon \leq \delta \leq \overline{\delta}$. 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 $[-\overline{\delta}, -\epsilon]$ or $[\epsilon, \overline{\delta}]$) is greater than the mass of traders with intermediate private valuations (in $[-\epsilon, \epsilon]$).

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

Figure 5 illustrates Corollary 2 when for $\varphi = 1.5$, $\epsilon = 3$, $\overline{\delta} = 7$ and $\lambda = 0.99$. 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 Figure 5 about here]

Equilibrium fast trading: We now study the equilibrium determination of the level of fast trading, α. First, consider corner equilibria. If

$$\phi(1) - \psi(1) > C, \tag{13}$$

then institutions prefer to invest when they expect all the others to do so. Hence, $\alpha^* = 1$ is an equilibrium if Condition (13) holds. Symmetrically, if

$$\phi(0) - \psi(0) < C, \tag{14}$$

then institutions prefer not to invest when they expect the others also won't. Hence, $\alpha^* = 0$ is an equilibrium if Condition (14) 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:[16]

$$\phi(\alpha^*) - \psi(\alpha^*) = C. \tag{15}$$

As $\phi(\alpha) - \psi(\alpha)$ is continuous in α, at least one of these three equilibrium conditions must hold. Thus, an equilibrium always exists. Furthermore, if $\Delta(0) = \phi(0) - \psi(0) > C$ then, in equilibrium, some institutions invest in the fast trading technology. This yields the next proposition.

Proposition 3 *: We have:*

$$\Delta(0) = \underbrace{(1-\lambda)E(|\delta|)}_{\textit{Search Value}} + \underbrace{(2G(\epsilon)-1)(\epsilon - E(|\delta|\,||\delta| \leq \epsilon))}_{\textit{Speculative Value}}. \tag{16}$$

Thus, $\Delta(0)$ *decreases with* λ*, increases with* ϵ*, and there is an equilibrium with fast trading* ($\alpha^* > 0$) *if*

$$(1-\lambda)E(|\delta|) + (2G(\epsilon)-1)(\epsilon - E(|\delta|\,||\delta| \leq \epsilon)) > C. \tag{17}$$

The value of fast trading at $\alpha = 0$, $\Delta(0)$, measures the increase in expected profit for an institution that becomes fast when all others are slow. To gain insight on the determinants of $\Delta(0)$, suppose first that $\epsilon = 0$. In this case, the institution that becomes fast increases by $(1-\lambda)$ its chance of realizing its gain from trade. Ex-ante, expected gains from trade are equal to $E(|\delta|)$. Thus, adoption of the fast trading technology generates an increase in expected profit equal to $(1-\lambda)E(|\delta|)$ for the first adopter. This effect is captured by the first term in (16). We call this the "*search value*" of the trading technology.

When $\epsilon > 0$, in addition to the previous effect, adoption of the fast trading technology brings a speculative gain: a fast institution can exploit private information about the asset common value, θ. The average "*speculative value*" of the fast trading technology is given

[16] Because we consider an integer number (T) of institutions instead of a continuum, there might not always exist a number N of institutions such that $\alpha^* = N/T$. For simplicity, we neglect this integer problem, which vanishes as T becomes large.

by the second term in (16). To see why, 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 ($\theta = \epsilon$). As $\alpha = 0$, it buys the asset iff $\delta + \epsilon - a^*(0) > 0$, i.e., only if $\delta \geq -\epsilon$. However, the institution would have purchased the asset anyway if slow when $\delta \geq 0$. Thus, the technology has speculative value only when $-\epsilon \leq \delta < 0$. In this case, if fast, the institution buys the asset and earns $\delta + \epsilon$ whereas if slow it *sells* the asset and earns $-(\delta + \epsilon)$. Thus, the net speculative gain of the technology is $\delta + \epsilon - (-(\delta + \epsilon)) = 2(\delta + \epsilon)$, conditional on $-\epsilon \leq \delta < 0$ and $\theta > 0$. This generates an average speculative gain of $2(G(0) - G(-\epsilon))\,(\epsilon - E(|\delta|\,||\delta| \leq \epsilon))$ when $\theta > 0$.[17] By symmetry, this is also the average speculative gain when $\theta < 0$. Thus, the total average speculative value of the fast trading technology is $2(G(0) - G(-\epsilon))\,(\epsilon - E(|\delta|\,||\delta| \leq \epsilon)) = (2G(\epsilon) - 1)(\epsilon - E(|\delta|\,||\delta| \leq \epsilon))$ because $G(.)$ is symmetric around 0.

The previous calculation holds 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, $\Delta(\alpha)$, is the sum of the search value and the speculative value of the fast trading technology. Closed-form expressions for these components, however, cannot be obtained for $\alpha > 0$ as they depend on the equilibrium price, $a^*(\alpha)$, 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, that 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.*

Figure 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 $\Delta(\alpha) = \phi(\alpha) - \psi(\alpha)$. In this example, $\Delta(0) = 4.57$ and $\Delta(1) = 1.49$. Thus, for $C = 3$, there is an interior equilibrium,

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

$\alpha^* \approx 0.305$. Clearly, as the cost of fast trading increases (the horizontal line shifts up in Figure 6), the level of fast trading declines.

[Insert Figure 6 about here]

Now consider the case in which, at least for some ranges of α, institutions' decisions are complements. In this 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 case, 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.

Institutions can be trapped in the "All-Fast" equilibrium for the following reason. If an institution expects a large number of institutions to be fast, it anticipates large price impacts. When institutions' decisions are complements, slow institutions are *more* affected because, as explained previously, they are more likely to trade in equilibrium. Hence, the loss in profits of not investing is relatively high (which explains why $\Delta(1) \geq \Delta(0)$ and yet $\psi(0) > \phi(1) - C$). This prospect pushes each institution to acquire the fast trading technology, confirming thereby institutions' beliefs about other institutions' decisions.

More generally, when $\Delta(\alpha)$ increases in α, each new fast institution reinforces other institutions' incentives to be fast. This self-reinforcing mechanism for institutions' investment decisions can be interpreted as an arms race, similar to Glode, Green and Lowery (2012), and it implies the possibility of investments waves.

Figure 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 $\alpha_{\max} = 1$. 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 the figure shows, when $C = 0.264$, the possible equilibrium levels of fast trading: (i) All Slow ($\alpha^* = 0$), (ii) All Fast ($\alpha^* = 1$), and (iii) $\alpha_3^* = 80\%$. In this interior equilibrium, the level of fast trading increases in C, counterintuitively, because the equilibrium level of investment now critically depends on institutions' beliefs about other institutions' choices. These beliefs can be disconnected from technological costs and yet be self-fulfilling.

This interior equilibrium is not stable, however. 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 Figure 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]

[18]When an interior equilibrium, α^*, is not stable, $\Delta(\alpha)$ must necessarily be increasing at $\alpha = \alpha^*$, in line 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 Figure 6, the equilibrium with $\alpha^* = 1$ also arises from complementarity in information acquisition decisions by institutions ($\Delta(\alpha)$ is increasing at $\alpha = 1$). Yet this equilibrium is stable.

[19]All institutions optimally decide to trade or not by comparing their valuation for the asset to the equilibrium price (see Lemma 1). Hence, there are no noise traders in our setting. This feature enables on to conduct welfare analysis because all investors' welfare is well defined. It also implies that all investors optimally adjust their trading strategies when the market structure (e.g., the fraction of fast institutions; see below) changes. Thus, the model can be used for policy analysis.

6.1 Excessive Fast Trading

Utilitarian welfare is equal to[20]

$$W(\alpha) = \alpha\left(\phi(\alpha) - C\right) + (1-\alpha)\psi(\alpha). \tag{18}$$

Let α^{SO} be the socially optimal level of fast trading, that is, the level of α that maximizes $W(\alpha)$. (18) yields

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

The term within brackets is positive because an increase in α reduces fast and slow institutions' expected profits ($\frac{\partial\phi(\alpha)}{\partial\alpha} \leq 0$ 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

$$\frac{\partial W(\alpha)}{\partial \alpha} = \underbrace{\Delta(\alpha)}_{\text{Social Value of Fast Trading}} - \underbrace{(C + C_{ext}(\alpha))}_{\text{Social Cost of Fast Trading}}. \tag{19}$$

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 (19). 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 is a balancing act between the social cost and benefit of fast trading. This level is not necessarily zero because, at $\alpha = 0$, the search value of the fast trading technology can be large relative to its social cost (see Proposition 6 below). Yet, when $\epsilon > 0$, the socially optimal level of fast trading is always smaller (and in most cases strictly smaller) than the equilibrium level of fast trading, as the next proposition shows.

Proposition 5 *: When $\epsilon > 0$, the socially optimal level of fast trading is smaller than the equilibrium one ($\alpha^{SO} \leq \alpha^*$), with a strict inequality when there is at least one interior equilibrium, i.e., when $\alpha^* \in (0,1)$ (e.g., when $\Delta(1) < C < \Delta(0)$). When $C \geq \Delta(0)$, the*

[20]The counterparties of the financial institutions are risk neutral and obtain zero–expected profits. Hence, their contribution to utilitarian welfare is equal to 0.

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. Overinvestment arises because, when making their investment decisions, institutions do not internalize the negative externality they impose on others (i.e., they ignore $C_{ext}(\alpha)$). In contrast, underinvestment never arises in our model, where investment in fast trading technologies does not generate any positive externality.[21] Our overinvestment result is illustrated in Figure 7, which depicts social welfare when $\epsilon = 5$, institutions' private valuations are normally distributed with $\sigma_\delta = 10$, $\lambda = 0.27$, and $C = 4.77$ or 5. For $C = 4.77$, the social optimum is strictly positive and equal to $\alpha^{SO} \approx 23\%$ 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 Figure 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, tends to amplify the overinvestment problem because it tends to disconnect investment decisions from the technological cost. For instance, suppose that $\Delta(0) \leq C < \Delta(1)$, a case that arises when there is complementarity in institutions' decisions (see the previous section). 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

[21] A positive externality could arise if λ, the likelihood of finding a counterparty, increased in α. The effect of α on λ in reality is unclear, however. In particular, an increase in the level of fast trading could reduce the chance for slow institutions to obtain good quotes. In fact, institutional investors complain that in the new trading environment, good quotes disappear before they hit them. Participants refer to this phenomenon as "ghost liquidity" (see "*In push for HFT rules, European Commision frets about 'ghost liquidity', not loss of some players*", HFT Review, November 2012; available at http://www.hftreview.com/). Ghost liquidity is consistent with faster traders being able to take advantage of good quotes more quickly than slow traders. This is another negative externality, that would reinforce our results.

value of being fast relatively high, despite the fact that investment is highly inefficient because the technological cost is relatively large ($\Delta(0) \leq C$) so that $\alpha = 0$ would maximize utilitarian welfare.

Overinvestment in fast trading technologies does not mean that one should necessarily bar institutions from using them. Indeed, fast trading increases the likelihood that an institution will realize gains from trade, which generates a social benefit. When this benefit is large enough (i.e., when λ is small and/or average gains from trade are large enough), the socially optimal level of fast trading is strictly positive. In fact a necessary and sufficient condition for $\alpha^{SO} > 0$ is:

$$\frac{\partial W}{\partial \alpha}(0) = \Delta(0) - C_{soc}(0) > 0. \tag{20}$$

This leads to our next proposition.

Proposition 6 *: The socially optimal level of investment in fast trading technologies is strictly larger than zero if and only if* $\lambda < \widehat{\lambda}(\epsilon, C)$ *where*

$$\widehat{\lambda}(\epsilon, C) = Max\{1 - \frac{(2G(\epsilon) - 1)(E(|\delta| \, ||\delta| \leq \epsilon)) + C}{E(|\delta|)}, 0\}. \tag{21}$$

Furthermore, $\widehat{\lambda}(\epsilon, C)$ *decreases with* C *and* ϵ*.*

The search value of the technology at $\alpha > 0$ must be at least as large as the cost of the technology for investment to be socially optimal. For this reason, $\widehat{\lambda}$ decreases in the technological cost C. Furthermore, $\widehat{\lambda}(\epsilon, 0) < 1$. Hence, even when $C = 0$, there can be overinvestment in the fast trading technology. Indeed, when $C = 0$, the only equilibrium is $\alpha^* = 1$. Yet, if $\lambda > \widehat{\lambda}(\epsilon, 0)$ then $\alpha^{SO} = 0$ and *even when* $\lambda < \widehat{\lambda}(\epsilon, 0)$, α^{SO} will be positive but strictly less than one. The reason is that the social cost of fast trading includes not only real resources invested in the technology but also the negative externality generated by fast trading. [22]

[22] 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."*

The threshold $\widehat{\lambda}$ also decreases with ϵ because a larger ϵ raises the range of private valuations for which institutions make socially inefficient trading decisions (i.e., do not trade, sell when they should buy or vice versa). Proposition 5 focuses on $\epsilon > 0$. For completeness, the next corollary considers the particular case in which $\epsilon = 0$ and the fast trading technology has no speculative value because there is no uncertainty about the cash flow of the asset.

Corollary 3 *(benchmark: no adverse selection): When $\epsilon = 0$, the socially optimal level of fast trading is $\alpha^{SO} = 1$ if $\lambda < \widehat{\lambda}(0, C)$ and $\alpha^{SO} = 0$ if $\lambda \geq \widehat{\lambda}(0, C)$ where $\widehat{\lambda}(0, C) = Max\{\frac{E(|\delta|)-C}{E(|\delta|)}, 0\}$. 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 trading is either zero or one, depending on whether the social benefit of fast trading (i.e., $(1 - \lambda)E(|\delta|)$) is less than or higher than the technological cost. This calculation 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

While investment in fast trading is in general too high, an outright ban of fast trading technologies is not desirable because when $\lambda \leq \widehat{\lambda}(\epsilon, C)$, the socially optimal level of fast trading is greater than zero. In this section, we analyze two possible responses: (i) Pigovian taxes, and (ii) "slow markets".

6.2.1 Pigovian taxation

Suppose that the social planner can levy a lump sum tax T on fast institutions. This raises the total cost of being fast to $C + T$. Let $\alpha^{**}(T)$ be the equilibrium level of fast trading for a tax T, 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:

$$\frac{\partial W(\alpha^{SO})}{\partial \alpha^{SO}} = \phi(\alpha^{SO}) - \psi(\alpha^{SO}) - C + \underbrace{(\alpha^{SO}\frac{\partial \phi(\alpha^{SO})}{\partial \alpha^{SO}} + (1-\alpha^{SO})\frac{\partial \psi(\alpha^{SO})}{\partial \alpha^{SO}})}_{-C_{ext}(\alpha^{SO})} = 0. \quad (22)$$

Set $T^* = C_{ext}(\alpha^{SO})$. Using (22), we have:

$$\phi(\alpha^{SO}) - \psi(\alpha^{SO}) = C + T^*. \quad (23)$$

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 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 the public good.

When $\alpha^{SO} = 1$, $\frac{\partial W(1)}{\partial \alpha^{SO}} > 1$. Hence, using (22), $\phi(1) - \psi(1) > C + C_{ext}(1)$. For this reason, if $\alpha^{SO} = 1$ then $\alpha^{**}(T^*) = 1$ is the unique equilibrium when the tax is $T^* = C_{ext}(1)$. Thus, the tax $T^* = C_{ext}(\alpha^{SO})$ also implements the socially optimal level of fast trading when $\alpha^{SO} = 1$. This is not necessarily the case when $\alpha^{SO} = 0$. Indeed, because of complementarities in institutions' decisions, there might exist αs for which $\Delta(\alpha) > C + C_{ext}(0)$ for some parametrization of the model. However, in this case, the social planner can prevent investment in the fast trading technology with a tax that exceeds the largest possible value of $\Delta(\alpha)$ or by simply banning fast trading. The next proposition summarizes these results.

Proposition 7 : *When $\lambda < \widehat{\lambda}(\epsilon, C)$, 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 $\lambda \geq \widehat{\lambda}(\epsilon, C)$, a tax that exceeds the largest possible value of $\Delta(\alpha)$ implements the socially optimal*

[23] If institutions' decisions are substitutes everywhere, this is the unique equilibrium whereas if institutions' 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 $\alpha^{SO} < 1$.

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 (ϵ 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.

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), which is avoided with a lump sum tax. This suggests that taxing investment in fast trading technology rather than fast institutions' trades is more efficient.

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:

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

where the second equality follows from (23) and, 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 after redistribution. They therefore equally benefit from the improvement in welfare relative to the case in which fast trading is forbidden.

[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.

6.2.2 Slow and Fast Markets

Another way to alleviate the negative externality of fast trading is to create "slow–only" markets. In this section, we show this can lead to underinvestment in fast trading. Consider the following extension of our baseline model. There is a fast market (F) and a slow one (S). The slow market imposes a "speed limit"on all traders. For instance, S might have high latency (slowing messages between platform and traders), could batch incoming orders (see "*High-frequency traders face speed limits*," Financial Times, April 28, 2013), or, more directly, could deny entry to traders known to be fast. In the context of our model this means that, on S, all participants have the characteristics of our slow traders: at round τ they are uninformed about the realization of θ_τ, and they carry out their desired transaction with probability λ. In contrast, on F, fast institutions can fully exploit the power of the fast technology, as in our baseline model.

At date $\tau = 0$, institutions decide: (i) whether to be fast or slow and (ii) if they are slow, whether to trade on the fast market or the slow one. Let β denote the fraction of slow institutions trading on the slow market. Thus, the mass of slow institutions on the fast market is $(1-\alpha)(1-\beta)$. Before analyzing the equilibrium values of α and β, we first study equilibrium price and trade decisions at dates $\tau > 0$, similarly to the baseline model. The bid-ask spread on S is zero because all market participants are uninformed. Let $a_F^*(\alpha, \beta)$ be the equilibrium ask price on F. For given values of α and β, conditional on receiving a buy order, sellers on the fast market expect the asset cash-flow to be $\mathrm{E}(v\,|v(\delta, i) \geq a_F^*)$ where

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

Thus, the determination of the equilibrium price on the fast market is identical to that in the baseline model, except that $\alpha^F(\alpha, \beta)$ plays the role of α. Hence: $a_F^*(\alpha, \beta) = a^*(\alpha^F(\alpha, \beta))$, where $a^*(\cdot)$ is the equilibrium ask price in the baseline model when the level of fast trading is α^F. As (i) α^F increases with β or α and (ii) $a^*(\cdot)$ increases with α, the bid-ask spread on the fast market ($a_F^* - b_F^* = 2a_F^*$) increases with α and β.

Institutions' expected gain on the fast market are obtained by replacing $a^*(\alpha)$ by $a^*(\alpha^F)$

in (9) and (10). Thus:

$$\phi^F(\alpha, \beta) = \phi(\alpha^F) \text{ and } \psi^F(\alpha, \beta) = \psi(\alpha^F),$$

where $\phi^F(\alpha, \beta)$ and $\psi^F(\alpha, \beta)$ are respectively fast and slow institutions' expected gains in the fast market. Institutions' expected gain on the slow market, ψ^S, is identical to that obtained on the fast market when all investors are slow in the baseline model. That is:

$$\psi^S(\alpha, \beta) = \psi(0).$$

Institutions' expected gains on the fast market, $\phi(\alpha^F)$ or $\psi(\alpha^F)$, decrease with $\alpha^F$and therefore β. Hence, *slow* institutions who migrate 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^F) that trades on this market stem from fast traders. This is stated in the next corollary.

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

Now, consider institutions' decisions at date $\tau = 0$. For simplicity, and because this is not key for our conclusions, we assume that there is no cost of joining one or the other market.[25] Slow institutions trade at the same speed on S and F (their likelihood of executing a trade is λ in each case) but there is no adverse selection in the slow market. Thus, trading exclusively on the slow market is a dominant strategy for slow institutions. This implies that $\beta = 1$ in equilibrium.[26] Hence, in any equilibrium with $\alpha^* > 0$, fast institutions cannot make speculative profits at the expense of slow institutions, which

[25]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.

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

considerably reduces their expected profit. Indeed, for any α, they obtain:

$$\phi^F(\alpha, 1) - C = \phi(1) - C,$$

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

Thus, the choice between being fast and slow boils down to a comparison between $\phi^F(\alpha, 1)$ and $\psi^S(\alpha, 1)$, 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 $\phi^F(\alpha, \beta) > \phi^F(\alpha, 1) > \psi(0)$). Thus, all institutions choose to be fast and trade on F in equilibrium. If $\phi(1) - C \leq \psi(0)$, all institutions are better off being slow and trading on S, for all values of α. Thus, in equilibrium, all institutions choose to be slow and only trade on the slow market.

Proposition 8 *: When institutions can join a slow market or a fast market, if $\phi(1) - \psi(0) \leq C$ no institution becomes fast and all trade on the slow market ($\alpha^* = 0$ and $\beta^* = 1$), while if $C < \phi(1) - \psi(0)$ all institutions are fast and only trade on the fast market ($\alpha^* = 1$ and $\beta^* = 0$).*

Thus, the introduction of a slow market 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 on the slow market or (ii) the "All Slow" equilibrium with no activity on the fast market.

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

$$\begin{aligned} W(\alpha, \beta) &= \beta(1-\alpha)\psi^S(\alpha, \beta) + (1-\beta)(1-\alpha)\psi^F(\alpha, \beta) + \alpha(\phi^F(\alpha, \beta) - C) \\ &= \beta(1-\alpha)\psi(0) + (1-\beta)(1-\alpha)\psi(\alpha^F) + \alpha(\phi(\alpha^F) - C). \end{aligned} \tag{24}$$

Using Proposition 8, with the slow market, utilitarian welfare in equilibrium is either:

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

or

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

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 the slow market is always less than or equal to equilibrium social welfare with the slow market ($\psi(0)$) 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 a slow market, 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 a slow and a fast market is greater than social welfare with only a fast market.*

Thus, if one cannot tax fast institutions, opening a slow market *improves* welfare. However, this 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 $\lambda \leq \widehat{\lambda}(\epsilon, C)$ and $\phi(1) - \psi(0) < C$.[27] In this case, in the absence of a slow market, the socially optimal level of fast trading, α^{SO}, is strictly between zero and one (Proposition 6). For this level, we therefore have:

$$W(0,0) < W(\alpha^{SO}, 0). \tag{25}$$

If the slow market opens, the equilibrium obtained in this case is such that all institutions remain slow and trade on the slow market. Social welfare in equilibrium is therefore $W(0,1) = W(0,0)$. (25) implies 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 the slow market ($\beta = 1$) because institutions

[27] In other cases ($\lambda > \widehat{\lambda}(\epsilon, C)$ or $C \leq \phi(1) - \psi(0)$), the equilibrium outcome with the slow market coincides with the outcome maximizing social welfare in the absence of the slow market.

joining the slow market exert a negative externality on those on the fast market (Corollary 4).

Thus, regulators are between a rock and a hard place: with only a fast market, there is overinvestment in the fast trading technology in equilibrium, whereas with a slow market, 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.

Proposition 9 *: If the regulator can use Pigovian taxes, it should choose a tax larger than $\psi(0)$ for institutions trading on the slow market (to preclude trading on this market) and 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 Figure 7 with $C = 4.77$. Without a slow market, all institutions choose to be fast. Utilitarian welfare is $W(1, 0) = \phi(1) - C = 2.142$. With a slow market, this equilibrium unravels and all institutions choose to be slow. Investors' welfare improves (Corollary 5) 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 on the slow market (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, $\alpha^* = \alpha^{SO} = 23\%$ 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).$\blacksquare$

7 Empirical implications

In this section, we briefly discuss the testable implications of the model. The informational content of trades should be inversely related to the cost of fast trading (e.g., co-location fees). Indeed, at any stable equilibrium, an increase in C triggers a drop in the level of fast trading (α) and the informational content of trades increases in this level (Proposition 1). In contrast, an increase in the cost of fast trading should have an ambiguous effect on trading volume (see the analysis in Section 4.2).

Anecdotal evidence suggests 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 (Corollary 2). 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). The model suggests two other, less obvious, explanations, highligted in Implications 1 and 2 below.

Implication 1: *Holding α constant, the expected profits of fast institutions should decrease when market fragmentation or the fraction of slow institutions trading on slow markets increases.*

As discussed in the introduction, as market fragmentation increases, λ goes down. This hurts fast traders in our model. Indeed, a reduction in λ means that the informational impacts of trades increases. Accordingly, in equilibrium, fast and slow institutions' expected profits decline (see (9) and (10)). Similarly, an increase in the fraction of slow institutions trading on slow markets (β) raises the price impact of trades on the fast market (Corollary 4) and thereby lowers the profitability of fast trading, other things equal.

OTC 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 variance of the asset payoff is captured by ϵ, which thus measures fundamental volatility (as opposed to volatility driven by microstructure effects.) The effect of ϵ on the profitability of fast trading is ambiguous.[28] An increase in volatility has two opposite effects on the profitability of fast institutions. On the one hand, holding the bid and ask prices constant, it raises the speculative value of the fast trading technology. On the other hand, it raises the price impacts of trades because $a^*(\alpha)$ increases in the volatility, ϵ. The first effect raises fast institutions' expected profits whereas the second lowers it. The former effect dominates the latter when α is small but not necessarily when α is large.

Now consider the effects of variations in λ and ϵ on the equilibrium level of fast trading, α^*. The model suggests that analyzing the effects of these parameters 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 it negatively affects slow institutions even more. A good example is an increase in market fragmentation. For low levels of fast trading, an increase in market fragmentation decreases both the expected profit of fast and slow institutions but relatively more so for slow institutions. This follows directly from Proposition 3 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. 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 C, equilibrium investment in fast trading should*

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

increase when (a) trading becomes more fragmented or (b) the volatility of the fundamental value 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 (due to a decline in C, for instance), increased market fragmentation or volatility might instead lower the profitability of fast trading and force some fast trading firms to exit. This highlights the importance of controlling for the cost of fast trading, and its variations, 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 $\sigma_\delta = 4$ and $\epsilon = 7$. For a large cost of fast trading, $C = 4.5$, an increase in fragmentation (from $\lambda = 0.8$ to $\lambda = 0.6$), generates an *increase* in α^*, from 4.1% to 11.9%. In contrast, for a low cost of fast trading, $C = 3$, the same increase in fragmentation generates a *decrease* in α^*, from 30.6% to 26.2%.

8 Conclusion

Investment in fast trading technology helps financial institutions cope with market fragmentation. To the extent that this enhances their ability to reap mutual gains from trade, it improves social welfare. On the other hand, fast institutions observe value relevant information before slow ones, 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 that 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).

Fast trading would unambiguously increase welfare if it could improve traders' ability to seize gains from trade, without providing them advance information. The two are impossible to disentangle, however. Getting fast access to quotes (which is required for efficient search) is also a way to get fast access to information (because quotes and trades contain information in financial markets). Another way to mitigate the adverse consequences of fast traders is to let slow institutions trade on "slow–only" platforms. Yet, we show that this can lower equilibrium investment in fast trading below its socially optimal counterpart. On the other hand, if the regulator can impose Pigovian taxes on investments in fast technology (equal to the externalities they generate), the socially optimal level of investment in fast trading can be implemented.

This raises the issue of the political economy and implementation of the taxation of investments in fast trading. Even if, before investment decisions are made, Pigovian taxes are Pareto optimal, after the fact, once investment is sunk, they no longer are. And they would attract political opposition from institutions having already invested in fast trading. An adequate policy response could be to tax new investment in fast trading, while leaving past investment untaxed. For example, given the existing fiber optic cables between Chicago and New York, it could make sense to tax additional investment in microwave signal transmission.

Proofs

Remark. The proofs of Corollary 1, Proposition 7, Corollary 4, Proposition 8, and Corollary 5 stem directly from the arguments in the text.

Proof of Lemma 1. Using Bayes rule, we obtain that for any $a \leq \epsilon$:[29]

$$\mathrm{E}(v\,|v(\delta,i)\geq a)=\left(\frac{\alpha(G(a+\epsilon)-G(a-\epsilon))}{\alpha(2-(G(a+\epsilon)+G(a-\epsilon))+2(1-\alpha)\lambda(1-G(a))}\right)\epsilon,$$

Using (3), we have:

$$\Pi(a;\alpha,\lambda,\epsilon)=a-\left(\frac{\alpha(G(a+\epsilon)-G(a-\epsilon))}{\alpha(2-(G(a-\epsilon)+G(a+\epsilon))+2(1-\alpha)\lambda(1-G(a))}\right)\epsilon. \tag{26}$$

Thus, using the symmetry of $g(.)$, we deduce that:

$$\Pi(0;\alpha,\lambda,\epsilon)=-\frac{\alpha(2G(\epsilon)-1)}{\alpha+(1-\alpha)\lambda}\epsilon, \tag{27}$$

and that:

$$\Pi(\epsilon;\alpha,\lambda,\epsilon)=\epsilon-\left(\frac{\alpha(2G(2\epsilon)-1)}{\alpha(3-2G(2\epsilon))+4(1-\alpha)\lambda(1-G(\epsilon))}\right)\epsilon. \tag{28}$$

Thus, when $\alpha > 0$ and $\epsilon > 0$, $\Pi(0;\alpha,\lambda,\epsilon) < 0$ and $\Pi(\epsilon;\alpha,\lambda,\epsilon) \geq 0$. As $\Pi(\cdot)$ is continuous in a, there is therefore at least one equilibrium ask price, a^* such that $a^* > 0$ and $a^* \leq \epsilon$.[30] If $\alpha = 0$ or $\epsilon = 0$, $\Pi(0;\alpha,\lambda,\epsilon) = 0$ and $\Pi(a;\alpha,\lambda,\epsilon) > 0$ for all $a > 0$. Thus, in this case, $a^* = 0$ is the unique equilibrium ask price. Thus, for all values of the parameters, there exists at least one equilibrium price and this price is unique when $\alpha = 0$ or $\epsilon = 0$.

Proof of Lemma 2. By definition, $a^*_{\min}(\alpha)$ is the first price $a > 0$ such that $\Pi(a^*_{\min}(\alpha);\alpha,\lambda,\epsilon) = 0$. As $\Pi(0;\alpha,\lambda,\epsilon) < 0$ and $\Pi(a;\alpha,\lambda,\epsilon)$ is continuous, we deduce that $\Pi(a;\alpha,\lambda,\epsilon) < 0$ for all $a < a^*_{\min}(\alpha)$. Hence, equilibrium price $a^*_{\min}(\alpha)$ cannot be profitably undercut. This also implies that:

$$\frac{\partial\Pi}{\partial a}\Big|_{a=a^*_{\min}(\alpha)} > 0, \tag{29}$$

as otherwise there would exist a price slightly smaller than $a^*_{\min}(\alpha)$ such that $\Pi(a;\alpha,\lambda,\epsilon) >$

[29] For $a \leq \epsilon$, the event $v(\delta,i) \geq a$ has always a strictly positive probability because fast institutions with a positive private valuation and good news always buy the asset ($v(\delta,1) > \epsilon$ for $\delta > 0$). Hence, $\mathrm{E}(v\,|v(\delta,i) \geq a)$ is well defined for $a \leq \epsilon$.

[30] Dealers' expected profit, $\Pi(.;\alpha,\lambda,\epsilon)$, is continuous in a because $G(.)$ is continuous.

0, which, as we just observed, is impossible. Using (29) and the fact that $\Pi(a;\alpha,\lambda,\epsilon)$ is continuous, we deduce that there is always an ask price a_0 arbitrarily close to but larger than $a^*_{\min}(\alpha)$ such that $\Pi(a_0;\alpha,\lambda,\epsilon) > 0$. Thus, any equilibrium price above $a^*_{\min}(\alpha)$ can be profitably undercut.

Proof of Proposition 1: Remember that a^* is such that $\Pi(a^*;\alpha,\lambda,\epsilon) = 0$. Hence, using the implicit function theorem, we obtain that:

$$\frac{\partial a^*}{\partial \alpha} = -\frac{\frac{\partial \Pi}{\partial \alpha}\big|_{a=a^*(\alpha)}}{\frac{\partial \Pi}{\partial a}\big|_{a=a^*(\alpha)}}. \tag{30}$$

We know from (29) that $\frac{\partial \Pi}{\partial a}\big|_{a=a^*(\alpha)} > 0$. Using (26), we deduce:

$$\frac{\partial \Pi}{\partial \alpha}\big|_{a=a^*(\alpha)} = -\frac{2\lambda(G(a^*+\epsilon) - G(a^*-\epsilon))(1-G(a^*))\epsilon}{(\alpha(2-(G(a^*+\epsilon)+G(a^*-\epsilon)) + 2(1-\alpha)\lambda(1-G(a^*)))^2} < 0. \tag{31}$$

Hence, using (30), we deduce that $\frac{\partial a^*}{\partial \alpha} > 0$. Using a similar reasoning, we deduce that (i) $\frac{\partial a^*}{\partial \lambda} < 0$ because $\frac{\partial \Pi}{\partial \lambda}\big|_{a=a^*(\alpha)} > 0$ and (ii) $\frac{\partial a^*}{\partial \epsilon} > 0$ because $\frac{\partial \Pi}{\partial \epsilon}\big|_{a=a^*(\alpha)} < 0$.

Proof of Lemma 3: Using (**??**), we deduce that $\Delta\text{Vol}(a^*(\alpha),\alpha) > 0$ iff:

$$\frac{1-G(a^*-\epsilon)}{1-G(a^*)} + \frac{1-G(a^*+\epsilon)}{1-G(a^*)} \geq 2\lambda. \tag{32}$$

Let define $f(x,y) = \frac{1-G(x+y)}{1-G(x)}$ where y is a constant. Condition (32) is:

$$f(a^*,-\epsilon) + f(a^*,\epsilon) \geq 2\lambda. \tag{33}$$

We have:

$$\frac{df(x,y)}{dx} = \frac{-g(x+y)(1-G(x)) + g(x)(1-G(x+y))}{(1-G(x))^2}.$$

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, $h_g(x+\epsilon) < h_g(x) < h_g(x-\epsilon)$. Hence, setting $y = -\epsilon$, we deduce that $f(a^*,-\epsilon)$ decreases with a and is therefore minimal in $a = \epsilon$. Symmetrically for $y = +\epsilon$, $f(a^*,\epsilon)$ increases with a and is therefore minimal in

$a = 0$. Thus, Condition (33) is satisfied if $f(\epsilon, -\epsilon) + f(0, \epsilon) \geq 2\lambda$, that is:

$$\frac{1 - G(0)}{1 - G(\epsilon)} + \left(\frac{1 - G(0)}{1 - G(\epsilon)}\right)^{-1} \geq 2\lambda.$$

This condition is always satisfied if $\lambda \leq \frac{1}{2}$ 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 $\lambda > \frac{1}{2}$ because the L.H.S of the previous equation reaches its minimum for $\epsilon = 0$, for which it is equal to 2. This proves the first part of the proposition.

Now suppose that $h_g(.)$ increases with x. Thus, $h_g(x + \epsilon) > h_g(x) > h_g(x - \epsilon)$. Hence, setting $y = -\epsilon$, we deduce that $f(a^*, -\epsilon)$ increases with a and is therefore minimal in $a = 0$. Symmetrically, $f(a^*, \epsilon)$ decreases with a and is therefore minimal in $a = \epsilon$. Thus, Condition (33) is satisfied if $f(0, -\epsilon) + f(\epsilon, \epsilon) \geq 2\lambda$, that is, using the symmetry of $g(.)$:

$$2G(\epsilon) + \left(\frac{1 - G(2\epsilon)}{1 - G(\epsilon)}\right) \geq 2\lambda.$$

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

Proof of Proposition 2. As $a^*(\alpha)$ increases with α, it is immediate from (9) and (10) that fast and slow institutions' expected profits (weakly) decrease with α. Now, we have:

$$\begin{aligned}\phi - \psi &= \int_{a^*(\alpha)-\epsilon}^{\overline{\delta}} (\delta + \epsilon - a^*(\alpha))g(\delta)d\delta + \text{Max}\{\int_{a^*(\alpha)+\epsilon}^{\overline{\delta}} (\delta - \epsilon - a^*(\alpha))g(\delta)d\delta, 0\} \\ &\quad -2\lambda\text{Max}\{\int_{a^*(\alpha)}^{\overline{\delta}} (\delta - a^*(\alpha))g(\delta)d\delta, 0\}.\end{aligned}$$

Suppose first that $\overline{\delta} \geq a^*(\alpha) + \epsilon$. Then, we have:

$$\begin{aligned}\phi - \psi &= \int_{a^*(\alpha)-\epsilon}^{\overline{\delta}} (\delta + \epsilon - a^*(\alpha))g(\delta)d\delta + \int_{a^*(\alpha)+\epsilon}^{\overline{\delta}} (\delta - \epsilon - a^*(\alpha))g(\delta)d\delta \\ &\quad -2\lambda\left[\frac{1}{2}\int_{a^*(\alpha)}^{\overline{\delta}} (\delta + \epsilon - a^*(\alpha))g(\delta)d\delta + \frac{1}{2}\int_{a^*(\alpha)}^{\overline{\delta}} (\delta - \epsilon - a^*(\alpha))g(\delta)d\delta\right].\end{aligned}$$

Thus:

$$\begin{aligned}\phi-\psi &= (1-\lambda)\int_{a^*(\alpha)-\epsilon}^{\overline{\delta}}(\delta+\epsilon-a^*(\alpha))g(\delta)d\delta+(1-\lambda)\int_{a^*(\alpha)+\epsilon}^{\overline{\delta}}(\delta-\epsilon-a^*(\alpha))g(\delta)d\delta\\ &\quad +\lambda\int_{a^*(\alpha)-\epsilon}^{a^*(\alpha)}(\delta+\epsilon-a^*(\alpha))g(\delta)d\delta-\lambda\int_{a^*(\alpha)}^{a^*(\alpha)+\epsilon}(\delta-\epsilon-a^*(\alpha))g(\delta)d\delta.\end{aligned}$$

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 $\int_{a^*(\alpha)}^{a^*(\alpha)+\epsilon}(\delta-\epsilon-a^*(\alpha))g(\delta)d\delta<0$ and $\int_{a^*(\alpha)-\epsilon}^{a^*(\alpha)}(\delta+\epsilon-a^*(\alpha))g(\delta)d\delta>0$. Thus, $\phi-\psi>0$ for $\overline{\delta}\geq a^*(\alpha)+\epsilon$. Similar steps can be followed when $a^*(\alpha)\leq\overline{\delta}<a^*(\alpha)+\epsilon$ or $\overline{\delta}<a^*(\alpha)$.

Proof of Corollary 2: Using the definition of $g(.)$ in Example 2, we deduce that in this example:

$$\begin{aligned}G(a^*+\epsilon) &= \left(\overline{\delta}+\gamma\epsilon+\varphi a^*\right)/(2\overline{\delta}),\\ G(a^*-\epsilon) &= \left(\gamma a^*+\varphi(\overline{\delta}-\epsilon)\right)/(2\overline{\delta}),\\ G(a^*) &= \frac{1}{2}+\frac{\gamma}{2\overline{\delta}}a^*.\end{aligned}$$

Hence, using (6) and $\left(\gamma\epsilon+\varphi\left(\overline{\delta}-\epsilon\right)\right)/2\overline{\delta}=1/2$, we obtain:

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

We have $\Delta\text{Vol}(a^*(0),0)=1-\lambda>0$ and $\Delta\text{Vol}(a^*(1),1)=(1-\lambda)-\left(\varphi+\gamma(1-2\lambda)\right)a^*(1)/2\overline{\delta}$. Furthermore $\Delta\text{Vol}(a^*(\alpha),\alpha)$ decreases with α because a^* increases with α and $\varphi\geq\gamma$. Thus, there are two cases to consider. If $\lambda<\text{Min}\{1,\frac{2\overline{\delta}-(\gamma+\varphi)a^*(1)}{\overline{\delta}-\gamma a^*(1)}\}$ then $\Delta\text{Vol}(a^*(1),1)>0$. In this case, $\Delta\text{Vol}(a^*(\alpha),\alpha)>0$ for all α and institutions' decisions are globally complements ($\frac{\partial\Delta(\alpha)}{\partial\alpha}<0,\ \forall\alpha$).[31] If instead, $\lambda>\text{Min}\{1,\frac{2\overline{\delta}-(\gamma+\varphi)a^*(1)}{\overline{\delta}-\gamma a^*(1)}\}$ then $\Delta\text{Vol}(a^*(1),1)<0$. By continuity of $\Delta\text{Vol}(a^*(\alpha),\alpha)$, we deduce that there is one value of α, denoted α_0, such that $\Delta\text{Vol}(a^*(\alpha),\alpha)<0$ 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 3: Observe first that:

$$E(|\delta|\,||\delta|\leq x)=\frac{(2G(x)-1)}{2}\int_0^x \delta g(\delta)d\delta. \tag{34}$$

Using (9) and (10), the symmetry of $g(.)$, and the fact that $a^*(0)=0$, we obtain:

$$\phi(0)=\epsilon(G(\epsilon)-G(-\epsilon))+2\int_0^{\overline{\delta}}\delta g(\delta)d\delta-2\int_0^{\epsilon}\delta g(\delta)d\delta,$$

and

$$\psi(0)=2\lambda\int_0^{\overline{\delta}}\delta g(\delta)d\delta.$$

Hence, using (34), the symmetry of $G(.)$, and $\Delta(0)=\phi(0)-\psi(0)$,

$$\Delta(0)=(1-\lambda)E(|\delta|)+(2G(\epsilon)-1)(\epsilon-E(|\delta|\,||\delta|\leq\epsilon)).$$

This is clearly decreasing with λ. Moreover $\frac{\partial\Delta(0)}{\partial\epsilon}=2G(\epsilon)-1\geq 0$. Last, if $\Delta(0)>C$, some investors find optimal to become fast traders when $\alpha=0$. As there always exists at least one equilibrium fraction of fast investors, we deduce that if $(1-\lambda)E(|\delta|)+(2G(\epsilon)-1)(\epsilon-E(|\delta|\,||\delta|\leq\epsilon))>C$ then all equilibria are such that $\alpha^*>0$.

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

Proof of Proposition 5. Consider first an interior equilibrium ($0<\alpha^*<1$). If $\epsilon>0$, 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 (15)). Thus, $\Delta(\alpha^*)<C_{soc}(\alpha^*)$ when $\epsilon>0$, because, under the latter condition, $C_{ext}(\alpha^*)>0$. Hence, at equilibrium, the social cost of fast trading strictly exceeds the social benefit, which implies $\alpha^{SO}\neq\alpha^*$.[32] Furthermore, as $\phi(\alpha^*)-\psi(\alpha^*)=C$,

[32]When $\Delta(\alpha)<C_{soc}(\alpha)$, a small decrease in α makes social welfare larger because $\frac{\partial W(\alpha)}{\partial\alpha}<0$ (see (19)).

(18) yields:

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

where the inequality is strict because $\alpha^{SO} \neq \alpha^*$. Hence, the L.H.S of this inequality must be strictly positive. 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 less than zero if $\alpha^{SO} > 0$. Thus, the first inequality implies that $\alpha^{SO} = 0$. In other words, if there is at least one equilibrium such that $\alpha^* = 0$ then $\alpha^{SO} = 0$. Finally, if there is an equilibrium in which $\alpha^* = 1$ (which requires $C \leq \Delta(1)$), we necessarily have $\alpha^{SO} \leq \alpha^*$.

Proof of Proposition 6. We first show that $C_{ext}(0) = (2G(\epsilon) - 1)$. By definition $C_{ext}(0) = \frac{\partial \psi(0)}{\partial \alpha}$. Using (10)

$$
C_{ext}(0) = -2\lambda \frac{\partial a^*(0)}{\partial \alpha}(1 - G(a^*(0)) = -\lambda \frac{\partial a^*(0)}{\partial \alpha}, \tag{35}
$$

because $a^*(0) = 0$. Furthermore, using (31), we have:

$$
\frac{\partial \Pi}{\partial \alpha}\Big|_{a=a^*(0)} = -\frac{(G(\epsilon) - G(-\epsilon))\epsilon}{\lambda},
$$

and

$$
\frac{\partial \Pi}{\partial a}\Big|_{a=a^*(0)} = 1.
$$

Thus, using (30),

$$
\frac{\partial a^*(\alpha)}{\partial \alpha}|\alpha = 0 = -\frac{\frac{\partial \Pi}{\partial \alpha}\big|_{a=a^*(0)}}{\frac{\partial \Pi}{\partial a}\big|_{a=a^*(0)}} = \frac{(2G(\epsilon) - 1)\epsilon}{\lambda}.
$$

Finally, using (35),

$$C_{ext}(0) = (2G(\epsilon) - 1)\epsilon. \tag{36}$$

Using (36) and the expression of $\Delta(0)$ in (16), (20) is equivalent to:

$$(1-\lambda)E(|\delta|) - (2G(\epsilon) - 1)E(|\delta|\,||\delta| \leq \epsilon)) > C,$$

that is: $\lambda \leq \widehat{\lambda}(\epsilon, C)$. Thus, $\alpha^{SO} > 0$ iff $\lambda < \widehat{\lambda}(\epsilon, C)$. Clearly, $\widehat{\lambda}(\epsilon, C)$ decreases with C. Furthermore, it decreases with ϵ because $(2G(\epsilon) - 1)E(|\delta|\,||\delta| \leq \epsilon)$ increase with ϵ.

Proof of Corollary 3. Using (21) for $\epsilon = 0$,

$$\widehat{\lambda}(0, C), = Max\{\frac{E(|\delta|) - C}{E(|\delta|)}, 0\}.$$

When $\epsilon = 0$, the bid-ask spread is zero fo all values of α. Hence, $\frac{\partial a^*}{\partial \alpha} = 0$ and $C_{ext}(\alpha) = 0$. Moreover, (11) and (16) yield $\Delta(\alpha) = \Delta(0) = (1-\lambda)E(|\delta|)$, for all α. Hence, $\Delta(\alpha) - C_{soc}(\alpha) = \Delta(0) - C = (1-\lambda)E(|\delta|) - C, \forall \alpha$. Thus, using (19), we deduce that:

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

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

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

Proof of Proposition 9.

First, consider the case in which $\lambda \leq \widehat{\lambda}(\epsilon, C)$. In this case, $\alpha^{SO} > 0$ (Proposition 6). 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:

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

By definition of α^F, $\frac{\partial \alpha^F}{\partial \alpha} = \frac{(1-\beta)}{(\alpha+(1-\alpha)(1-\beta))^2}$ and $\frac{\partial \alpha^F}{\partial \beta} = \frac{\alpha(1-\alpha)}{(\alpha+(1-\alpha)(1-\beta))^2}$. Using this and the fact that the expressions in large parentheses in the previous FOCs ($\frac{\partial W}{\partial \beta}=0$ and $\frac{\partial W}{\partial \alpha}=0$) are equal, we deduce that (α_0, β_0) must satisfy the following condition:

$$(\psi(0)-\psi(\alpha_0^F))(\alpha_0\beta_0 + (1-\beta_0)) = \alpha_0(\phi(\alpha_0^F) - C - \psi(\alpha_0^F)). \tag{37}$$

Moreover, if $W(\alpha_0, \beta_0) > W(\alpha^{SO}, 0) > \psi(0)$, we deduce from (24) that:

$$(1-\alpha_0)\beta_0(\psi(0)-\psi(\alpha_0^F)) + \alpha_0(\phi(\alpha_0^F) - C - \psi(\alpha_0^F)) > \psi(0) - \psi(\alpha_0^F)$$

Using (37), this implies: $(1-(1-\alpha_0)\beta)(\psi(0)-\psi(\alpha_0^F)) > \psi(0)-\psi(\alpha_0^F)$, which is impossible since $\psi(0) > \psi(\alpha_0^F)$.

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 $\lambda \leq \widehat{\lambda}(\epsilon, C)$, 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 $\lambda \leq \widehat{\lambda}(\epsilon, C)$, with a strict inequality if $\beta > 0$. When $\lambda > \widehat{\lambda}(\epsilon, C)$, $\alpha^{SO} = 0$. 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, Matthew, Jonathan Brogaard, and Andrei Kirilenko (2012), "The trading profits of high–frequency traders," Working paper, University of Washington.

Bagnoli, Mark, and Ted Bergstrom (2005), "Log-concave probability and its applications," *Economic Theory*, 445-469.

Bessembinder, Hendrik , Jia Hao and Kuncheng Zheng (2013), "Market Making Obligations and Firm Value," Working paper, David Eccles School of Business.

Biais, Bruno, David Martimort, and Jean-Charles Rochet, 2000, "Competing Mechanisms in a Common Value Environment," *Econometrica*, 68, 799-838.

Breon–Drish, Bradyn (2013), "On Existence and Uniqueness of Equilibrium in Noisy Rational Expectations Economies," Working paper, Stanford University.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014), "High frequency trading and price discovery,"forthcoming *Review of Financial Studies*.

Brogaard, Jonathan, Björn Hagströmer, Lars Nordén, and Ryan Riordan (2014), "Trading fast and slow: colocation and market quality," Working paper, University of Stockholm.

Budish, Eric, Peter Cramton, and John Shim (2014) "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," Working paper, University of Chicago.

Dow, James (2005), "Is liquidity self-fulfilling," *Journal of Business*, 77, 895-908.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen (2005), "Over-the-Counter Markets," *Econometrica*, 73, 1815-1847.

Easley, David and Maureen O'Hara (1987), "Price, trade size and information is securities markets," *Journal of Financial Economics*, 19, 69-90.

Foucault, Thierry and Albert Menkveld (2008), "Competition for Order Flow and Smart Order Routing Systems," *Journal of Finance*, 63, 119-158.

Ganguli, Jayant, and Liyan Yang (2009), "Complementarities, multiplicity, and supply information," *Journal of the European Economic Association*, 7, 90-115.

Glode, Vincent, Richard Green, and Richard Lowery (2012), "Financial Expertise as an Arms Race," *Journal of Finance*, 67, 1723-1759.

Glosten, L., 1994, "Is the Electronic Open Limit Order Book Inevitable?" *Journal of Finance*, 49, 1127-1161.

Glosten, Larry, and Paul Milgrom (1985), "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14, 71–100.

Grossman, Sandford, and Joseph Stiglitz (1980), "On the Impossibility of Informationaly Efficient Markets," *American Economic Review*, 70, 393-408.

Hasbrouck, Joel, 2005, Empirical market Microstructure, *Oxford University Press*.

Hendershott, Terrence, Charles Jones, and Albert Menkveld (2011), "Does Algorithmic trading improve liquidity," *Journal of Finance*, 66, 1-33.

Hendershott, Terrence, and Ryan Riordan (2013), "Algorithmic Trading and the Market for Liquidity,"*Journal of Financial and Quantitative Analysis*, 48, 1001-1024.

Hirshleifer, Jack (1971), "The Private and Social Value of Information and the Reward to Inventive Activity," *American Economic Review*, 61, 561-574.

Jovanovic, Boyan, and Albert Menkveld (2011), "Middlemen in limit order markets," Working paper, New York University.

Kirilenko, Andrei, Pete Kyle, Mehrdad Samadi, and Tugkan Tuzun (2011), "The Flash Crash: The impact of high-frequency trading on an electronic market," Working paper, University of Maryland.

Manzano, Carolina, and Xavier Vives (2012), "Public and private learning from prices, strategic substitutability and complementarity, and equilibrium multiplicity," *Journal of Mathematical Economics*, 47, 346-369.

O'Hara, Maureen, and Mao Ye (2011), "Is market fragmentation harming market quality," *Journal of Financial Economics*, 100, 459-474.

Pagnotta, Emiliano, and Thomas Phillipon (2013), "Competing on Speed," Working paper, Stern Business School, New-York University.

.

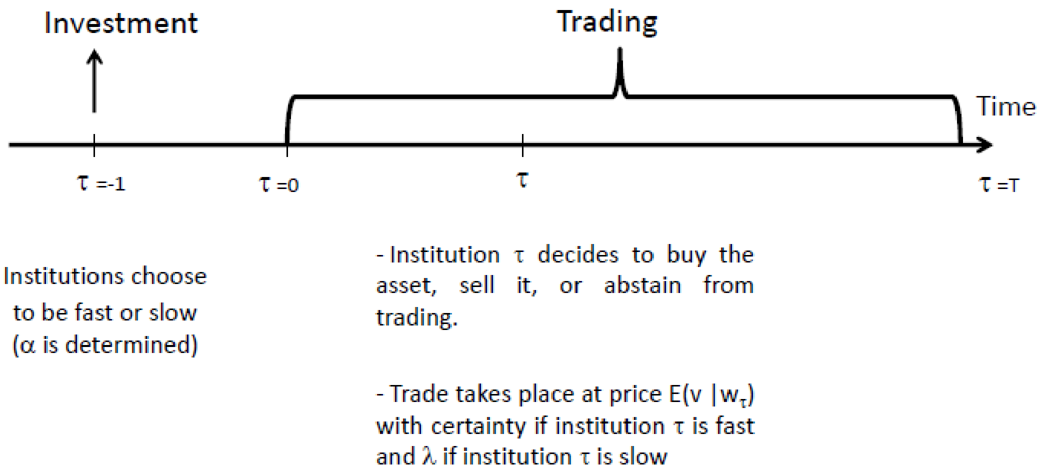

Figure 1: Timing of Institutions' decisions.

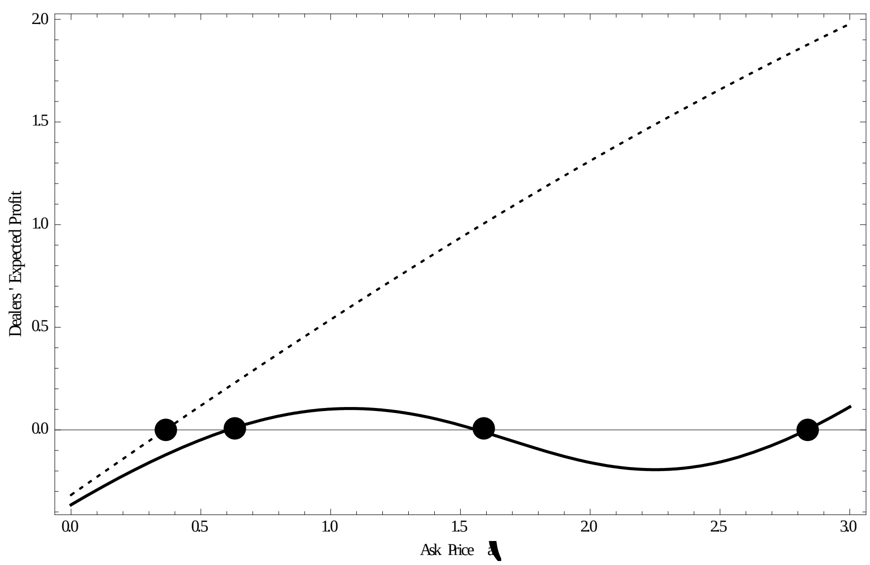

Figure 2: Equilibrium Uniqueness. This figure shows sellers' expected profit (Π) as a function of their ask price (a) when the distribution of traders' private valuation is normal with standard deviation, σ_δ=1 (plain line) and σ_δ=2 (dashed line). Other parameter values are α=0.1, ε=3 and λ = 0.8. Dots are values of the ask price for which equilibrium condition (1) is satisfied.

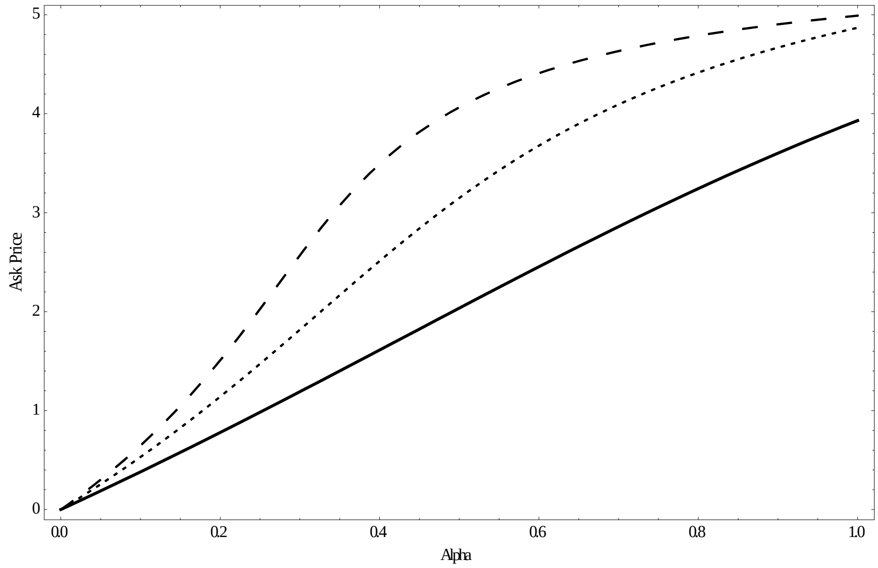

Figure 3: Price Impacts and Fast Trading. This figure shows the ask price (a^*) as a function of the level of fast trading (α) when the distribution of traders' private valuation is normal. Parameter values: σ_δ=6 (plain line), σ_δ =5 (small dashed line), σ_δ =4 (large dashed line) for ε=5 and λ = 0.8.

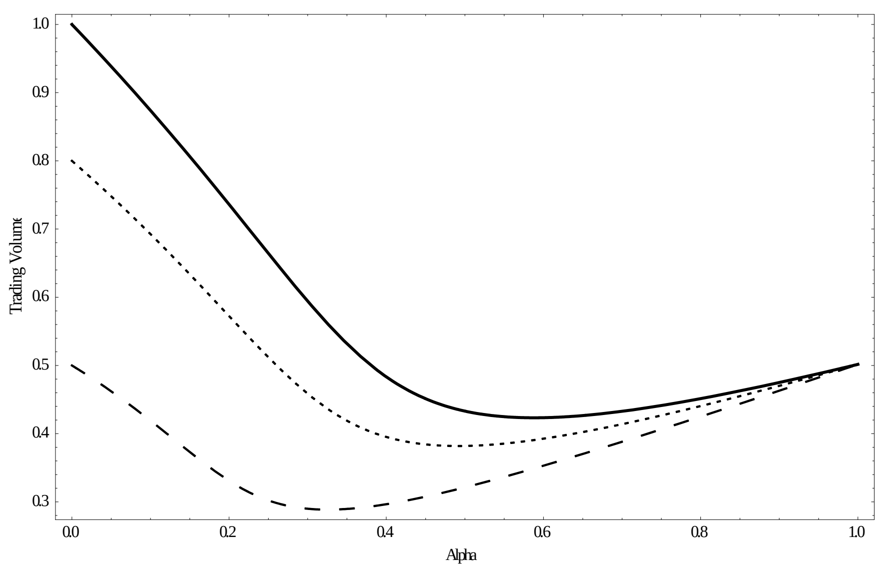

Figure 4: Trading Volume and Fast Trading. This figure shows equilibrium trading volume as a function of the level of fast trading (α) for different values of λ: 0.5 (large dashed line), 0.8 (dotted line), 0.9 (dashed line) and 1 (plain line) when δ has a normal distribution with σ_δ=3 and ε=5.

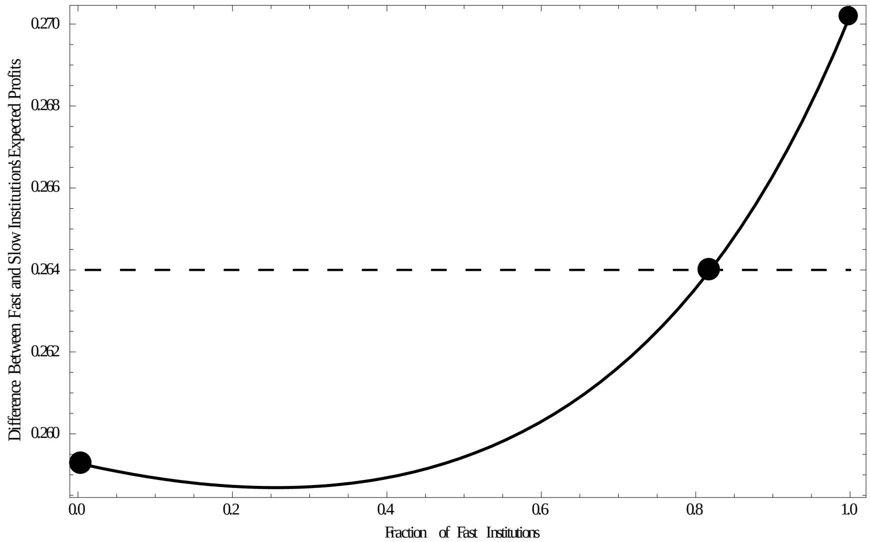

Figure 5: Equilibrium Fast Trading When Institutions' Decisions are Complements. The distribution of institutions' private valuation is as in Example 2, with φ=1.5 and C=0.264. Equilibrium levels of fast trading are indicated by large dots.

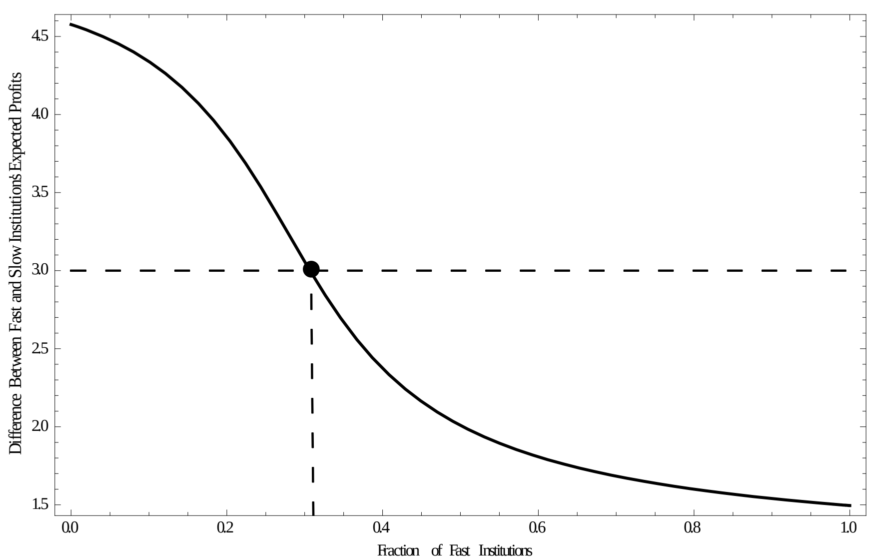

Figure 6: Equilibrium Fast Trading When Institutions' Decisions are Substitutes. Institutions' private valuations have a normal distribution with mean zero and standard deviation σ_δ=4. Other parameters are λ=0.8, ε=7, and C=3. In this case, the equilibrium level of fast trading is indicated by the large dot (α^*=0.305).

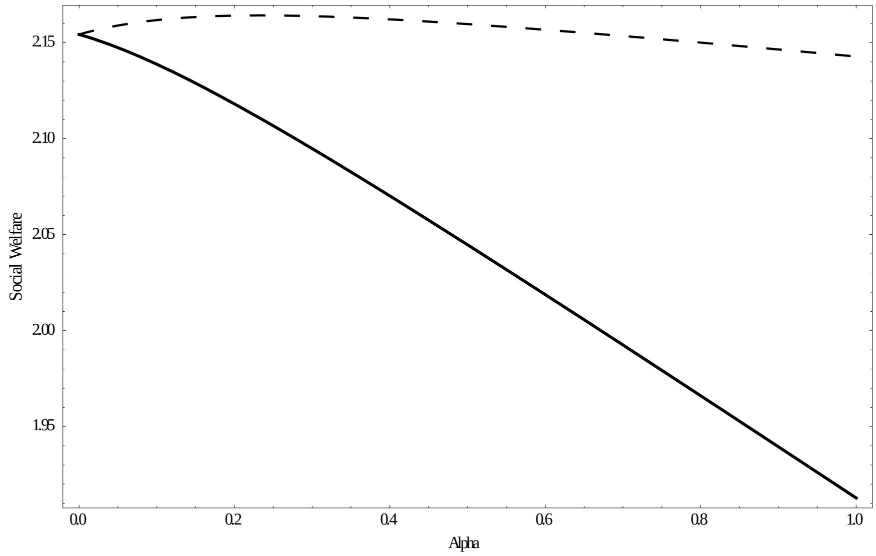

Figure 7: Social Welfare and Fast Trading. This figure shows social welfare as a function of the level of fast trading for C=4.77 (dashed line) and C =5 (plain line) when institutions' private valuations are normally distributed with σ_δ =10, ε=5, and λ=0.27.