Equilibrium High Frequency Trading[1]

Bruno Biais
Toulouse School of Economics
21 Allée de Brienne
31000 Toulouse, France
bruno.biais@univ-tlse1.fr

Thierry Foucault
HEC, Paris
1 rue de la Liberation
78351 Jouy en Josas, France
foucault@hec.fr

Sophie Moinas
Toulouse School of Economics
21 Allée de Brienne
31000 Toulouse, France
sophie.moinas@univ-tlse1.fr

November 18, 2011

[1]We are grateful to Terrence Hendershott, Carolina Manzano, Albert Menkveld, Rafael Repullo and participants to the conference on "High Frequency Trading: financial and regulatory implications" in Madrid for their suggestions. This research benefited from the support of the " Financial Markets and Investment Banking Value Chain" Chair sponsored by the Fédération Bancaire Française and of the ANR (ANR-09-JCJC-0139-01). Many thanks to Vivien Lévy-Garboua, Charles-Albert Lehalle, Eric Roussel and Jean Tricou for stimulating discussions. All errors are ours.

Abstract

High Frequency Trading (HFT) improves investors' ability to seize trading opportunities, which raises gains from trade. It also enables fast traders to process information before slow traders, which generates adverse selection. We first analyze trading equilibria for a given level of HFT and then endogenize investment in HFT. When some traders become fast, it increases adverse selection costs for the others, thus HFT generates negative externalities. Therefore equilibrium investment in HFT exceeds its utilitarian welfare–maximizing counterpart. Furthermore, since it involves fixed costs, investment in HFT is more profitable for large institutions than for small ones. Hence, in equilibrium, small institutions are less informed than large ones and exit the market when HFT becomes prevalent.

1 Introduction

'Well in *our* country' said Alice, still panting a little, 'you'd generally get to somewhere else – if you ran very fast for a long time, as we have been doing.'

'A slow sort of country!' said the Queen. 'Now, *here*, you see, it takes all the running *you* can do, to keep in the same place. If you want to go somewhere else, you must run at least twice as fast as that!'

Lewis Carroll, *Through the Looking Glass*

Traders must collect and process vast amounts of information about fundamentals, quotes, transaction prices, etc... Computers can complete these tasks faster and sometimes better than humans. Algorithmic trading thus relies on computers to collect and process information and submit and manage orders. It is widely used by proprietary trading desks, market makers, investment managers, brokers and hedge funds for a variety of trading tasks. Algorithmic trading is not new. For instance, traders have automated portfolio insurance strategies, that require simultaneous trades in a large number of securities, since the 80s. The recent years however have seen the rapid emergence of a new form of algorithmic trading, known as high frequency trading (hereafter HFT). HFT strategies are quite diverse but they all share one common feature: they rely on obtaining market data, processing these data, and trading accordingly at very high speed. As a result of this evolution, the pace of trading has considerably accelerated in recent years, so that multiple trades and orders within a second are now common. And this increase in the pace of the market seems to compel all financial institutions to catch up with speed, as the Red Queen in the quote above.

Now this trend is not anecdotal, it does not affect a small fringe of the market. Quite to the contrary, it is very pervasive. As pointed out by the SEC, "*HFT is a dominant component of the current market structure and is likely to affect all aspects of its performance.*" (See SEC 2010, p.45). For example, Brogaard (2010) finds that 26 high frequency traders participate in 68% of dollar volume for 120 Nasdaq stocks, while the pure play high frequency firms in Kirilenko et al (2010) account for 34% of the trading volume in the E.mini S&P 500 futures.

Not only is HFT very active, it is also quite profitable. Menkveld (2010) and Brogaard (2010) estimate that high frequency traders earn significant risk-adjusted returns. Saraiya and Mittal (2009) report that the annual aggregate profits of HFT are estimated to $21 billion. Furthermore, the bulk of these returns accrues to a small group of players. A report of the TABB group states that 2% of the 20,000 proprietary trading firms in the

U.S. account for more than 70% of the trading volume (see Iati, 2009). One possible reason for this concentration is that HFT requires significant fixed investments.[1] To be among the fastest in the market, proprietary trading firms must acquire hardware, develop and maintain codes, hire highly qualified personnel (e.g., Ph.Ds in mathematics, physics, computer science etc...), subscribe to expensive real time data feed, and obtain ultra-fast ("low latency") connections to exchanges' trading systems (e.g., by paying a fee to locate their computers just next to exchanges' servers, a practice known as co-location).

These market developments beg the following questions. Does the observed growth in HFT improve or impair the workings of markets? Do the profits of HFT reflect the value it creates or of the losses of lower frequency investors? What is the social value of the significant amount of investment in HFT technology? Is this allocation of resources efficient? Is policy intervention called for?

The goal of this paper is to offer a theoretical framework to shed light on these issues. Our model features a continuum of financial institutions (proprietary trading firms, hedge funds, banks' prop trading desks), differing in sizes, and who can invest or not in HFT, i.e., a technology giving them faster and better access to market information and trading opportunities. The fraction of financial institutions investing in this technology determines the overall level of HFT activity. We assume that fast access to markets gives an edge to high frequency traders in two ways:

- High frequency traders' "search capacity" for liquidity is higher. That is, they are more likely than other institutions to locate beneficial trading opportunities. This is because computers expand investors' cognition capacity and fast connections to trading platforms enable high frequency traders to react faster to fleeting trading opportunities and cope with market fragmentation.[2] For instance, a computer can identify more quickly than a human the package of trades at the current quotes which would results in a good hedge for an institution's overall position. Being super fast in achieving these trades is important as quotes can change very quickly as well.[3]

- Ultra–fast connections to trading platforms and data providers enable high–frequency traders to react faster than others to new information relevant for the value of an

[1] See for instance "*Citigroup to expand electronic trading capabilities by buying Automated Trading Desk*," International Herald Tribune, July 4, 2007. The author of this article notes that "Goldman spends tens of millions of dollars on this stuff. They have more people working in their technology area than people on the trading desk...The nature of the markets has changed dramatically."

[2] On this point, see the theoretical analyses of Biais, Hombert and Weill (2007) and Foucault, Kadan and Kandel (2010) and the empirical findngs of Hendershott, Jones and Menkveld (2010) and Hendershott and Riordan (2010).

[3] For instance, consider a high frequency firm engaged in market-making in a call option. It can charge a very competitive spread in this option if it can hedge its position at times in which the liquidity of the underlying security is plentiful, thereby reducing the cost of its hedge.

asset (be it macro news, corporate announcements, changes in prices of assets with correlated payoffs, or changes in supply and demand expressed in the order book.) Consequently, HFT orders are more informed than slower orders, as shown empirically by Hendershott and Riordan (2010), Brogaard (2010), Kirilenko et al (2010), and Hendershott and Riordan (2011). As noted by Kirilenko et al (2010): "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." But the flip–side of this superior ability is the adverse selection problem it creates for other traders. This in line with concern expressed by James A. Brigagliano, Co-Acting Director, Division of Trading and Markets, at the SEC: "The Commission recognizes concerns have been raised that high frequency traders have the ability to access markets more quickly through high-speed trading algorithms and co-location arrangements. This ability may allow them to submit or cancel their orders faster than long-term investors, which may result in less favorable trading conditions for these investors. (Brigagliano, 2009, page 5)"

As a result, in our model, an increase in the level of HFT has two effects. On one hand, it can increase the likelihood that assets are transferred to investors who value them the most for non informational reasons (e.g., hedging needs or tax purposes). Such transfers are mutually beneficial for buyers and sellers and therefore enhance aggregate welfare. On the other hand, as high frequency traders also have access to advance information on asset cash-flows, an increase in the level of HFT raises adverse selection costs.[4] Hence, the entry of a new high frequency trader exerts a negative externality on the other traders. This adverse–selection negative externality can reduce market participation and gains from trade. In particular it can crowd out slow investors from the market. This echoes the concerns raised in the IOSCO consultation report on "Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency" (July 2011, page 10): "some market participants have also commented that the presence of high frequency traders discourages them from participating as they feel at an inherent disadvantage to these traders' superior technology."

Now, HFT is not a given. Its level results from the choice of market participants whether to invest in hardware, code, personnel, connections and data–feed. This choice, in

[4]In Foucault, Röell, and Sandas (2003), liquidity suppliers monitor the market to alleviate adverse selection. By facilitating such monitoring, algorithms can reduce adverse selection for these traders. But, in electronic limit order markets, the distinction between liquidity suppliers and liquidity providers is blurring. Indeed, high frequency traders use both limit and market orders (as documented by Brogaard, 2010). If these order placement strategies are based on superior information, they will generate adverse selection costs for slow traders, irrespective of whether the latter use market or limit orders. Jovanovic and Menkveld (2010) offer an insightful theoretical and empirical analysis of these issues.

turn, reflects the above mentioned positive and negative consequences of HFT. Thus, we study the equilibrium determination of the endogenous level of HFT. For a given level of HFT, each institution trades more profitably when it is fast. However, financial institutions with a larger size can trade a wider array of asset classes and can therefore better amortize the fixed technological costs required for HFT.[5] For this reason, there is a critical size below which financial institutions choose to remain slow. Thus, in equilibrium there is a non–level playing field: a few actively trading fast institutions coexist with smaller, slower and less active institutions who bear the brunt of adverse selection costs.

While an increase in HFT reduces the trading profits of slow and fast traders, it may affect them differently. We identify conditions under which the profits of slow institutions decline faster with the level of HFT than the profits of fast institutions. Therefore, because the decision to invest in the HFT technology depends on the comparison between the profits of slow institutions and those of fast ones, the attractiveness of HFT can increase with its prevalence. Consequently, for some parameter values, institutions' investment decisions in HFT reinforce each other, i.e., they are strategic complements. Thus, there is an element of coordination in the institutions' decisions to become an HFT. This can generate multiple equilibria for the level of HFT. For instance, if it is expected that only a few institutions will be fast, then the gain of becoming fast is relatively small. This leads most institutions to remain slow, thus vindicating the initial expectation. Conversely, if it is expected that many institutions will invest in HFT, the cost of remaining slow is high. Against this backdrop, many institutions will choose to invest in HFT, again confirming the initial expectations. Thus HFT can be contagious: financial institutions may suddenly decide to invest massively in HFT simply because they expect others to be fast and therefore need to be fast also, lest they should be side–lined, very much like the red queen in the epigraph to this paper.[6]

Finally, we study whether the equilibrium level of HFT is socially optimal. We show that this level is in general excessive relative to the social optimum. The reason is that, in making their investment decisions in HFT, financial institutions do not internalize the impact of their decision on other investors' welfare. As this decision generates a negative externality by raising adverse selection, the equilibrium level of HFT can be too high relative to the level maximizing the aggregate welfare of slow and fast institutions. It is worth stressing however that this result does not mean that HFT should be banned. Indeed, by expanding institutions' capacity to search for trading opportunities, HFT has

[5] The business lines relevant to measure size in our context include prop–trading and market–making, but exclude such activities as commercial banking and corporate financing, which are unrelated to trading.

[6] This is in line with the analysis of financial expertise as an arms race by Glode, Green and Lowery (2011).

a positive effect on aggregate welfare. Thus, in many cases, the socially optimal level of HFT is strictly positive, but smaller than that set by institutions' equilibrium decisions.

Our analysis has several empirical implications about the consequences of HFT.

- First, it implies that the informational content of trades should increase with the level of HFT.[7]

- Second, in our theoretical model an increase in HFT raises short–term volatility. In a sense, this is the flip–side of the information content of HFT orders. By trading on advance information, high–frequency traders move prices rapidly.

- Third, our analysis implies that an increase in the level of HFT can increase or decrease trading volume (as found in Jovanovic and Menkveld (2010)). This reflects the fact that HFT increases the likelihood that traders identify trading opportunities (which tends to increase trading volume), as well as adverse selection costs (which tend to reduce market participation.)

- Fourth, our analysis implies that the decision to invest in the HFT technology can be contagious. Chaboud et al.(2010) track over time the number of trading desks equipped for algorithmic trading on EBS (a foreign–exchange trading–platform) and look at the effect of this change on volatility. Such data could also be useful to study wether there is complementarity in financial institutions' decisions to invest in algorithmic trading. For instance, one could study whether there is evidence of an acceleration in the number of new trading desks equipped for algorithmic trading as this number increases.

The econometric challenge raised by these implications is that, as shown in the present paper, the level of HFT as well as the informational impact of prices and the volume of trade are jointly endogenous. Thus, it will be necessary to identify natural experiments or instruments to identify the impact of HFT and disentangle it from that of other variations in the environment.

The paper is organized as follows. The next section discusses the link between our paper and related theoretical literature. Section 3 presents the model. In Section 4, we analyze equilibrium trades and prices, holding the level of HFT constant. We endogenize this level in Section 5 and show that the equilibrium level of HFT can be excessive in Section 6. Section 7 discusses the policy implications of our analysis. Section 8 briefly concludes. Proofs not given in the text are in the appendix.

[7] This is consistent with the empirical finding by Hendershott and Riordan (2010), Brogaard (2010), Kirilenko et al (2010), and Hendershott and Riordan (2011) that high frequency traders' orders are more information than slow orders.

2 Related literature

For tractability, we consider a simple one–period trading game. Yet, in practice, HFT strategies are highly dynamic. Martinez and Rosu (2011) analyze a dynamic model in which high–frequency traders are assumed to hold private information (as in the present paper) and to prefer taking bets on changes in the asset value rather than its level. In this context they show that HFT can generate a large fraction of volatility and trading activity. Hoffman (2010) considers a dynamic order–driven market, where limit orders are exposed to the risk of being picked off. High–frequency and human traders randomly access the market. Human traders are picked off when the value of the asset has moved against them. High–frequency traders are assumed to be picked off only if their successor also engages in HFT. Thus, as in the present paper, when a trader engages in HFT it is profitable for him, but it exerts a negative externality on the others.

While these papers take the fraction of high–frequency traders as given, we endogenize it. Indeed, one the main theoretical contributions of the present paper is to shed new light on equilibrium information acquisition in financial markets. As explained above, investing in HFT technology is a way to obtain value–relevant information before slower traders. Seminal analyses of the value of information in financial markets are offered by Grossman and Stiglitz (1980), Admati and Pfleiderer (1984) and Admati and Pfleiderer (1986). In particular, Admati and Pfleiderer (1984) observe that the value of information for one trader depends on the amount and quality of information other agents possess. In these models, the greater the number of traders that are privately informed, the more information is impounded in the price, the less profitable it is to be informed. Thus, information acquisition decisions are strategic substitutes. Because of such substitutability, there is a unique equilibrium in the fraction of informed agents in Grossman and Stiglitz (1980), in contrast with the multiplicity arising in our model.

In a dynamic context, strategic complementarities can arise between successive traders, as shown by Dow and Gorton (1994) and Chamley (2007.) When traders have short–horizons, they must unwind their holdings before all information has been made public. Once they have established their position, based on their private information, they benefit from the arrival of further informed agents. Indeed, the latter drive the price towards the true value of the asset, at which the short–horizon traders can profitably unwind their inventory. Our strategic complementarity result reflects different economic forces since there is only one round of trade in our model.

Ganguli and Yang (2009) extend the Grossman and Stiglitz (1980) setup to the case where traders can acquire information on the asset's payoff and also on its (random) supply. They show that, the additional dimension of supply information can lead to strategic complementarities and equilibrium multiplicity. Again, our strategic complementarity result is different from theirs, since there is only one dimension of private information in our model.

In our analysis, strategic complementarity in information acquisition is rooted in the fact that all traders face potential gains from trade and their ability to realize these gains varies according to whether they are informed or not. Complementarity arises when this ability declines faster for uninformed traders than for informed ones, as the fraction of informed agents increases.[8] In addition, the bulk of the literature on information acquisition in financial markets did not consider the welfare effects of this decision. In our paper all trading decisions are made by rational agents considering gains from trade. In this context we examine whether the equilibrium level of HFT is socially optimal.

3 Model

Consider a unit mass continuum of risk–neutral, profit maximizing financial institutions. Until Section 5 we focus on trading in one asset only.

Values: The asset's payoff at date $\tau = 2$ is v, which can be equal to $\mu + \epsilon$ or $\mu - \epsilon$ with equal probability. Institutions have no endowment in the asset, and can buy or sell up to one share. They can trade at date $\tau = 1$, just after learning their private value for the asset. This private value adds to the payoff of the asset and can be equal to $\delta > 0$ or $-\delta$ with equal probability. That is, each institution values the asset at $v + \delta$ or at $v - \delta$. Private values are i.i.d across institutions. They capture in a simple way that other considerations than expected cash–flows affect the willingness of investors to hold assets. For example, regulation can make it costly or attractive for certain investors, such as insurance companies or pension funds or banks to hold certain asset classes.[9] Differences in tax regimes can also induce differences in private values. Differences in private values generate trading without noise traders, hence welfare is well defined.

High Frequency Trading: At time $\tau = 0$, institutions simultaneously decide whether to invest in the infrastructures (computers, colocation, ...) and intellectual capital (skilled traders, codes, ...) necessary to engage in HFT, at cost C.[10] We refer to these players as *fast* institutions and denote the fraction of institutions that are fast by α. The remaining fraction is referred to as *slow* institutions. HFT technology helps fast institutions in two ways.

[8] Subsection 4.1.4 in Chamley (2007) points to this effect, but does not model it.

[9] E.g., some institutional investors can only hold investment grades bonds, which they will value at a premium relative to other investors (see, Chen, Lookman, Schürhoff, and Seppi, 2011).

[10] This decision is endogenized in Section 5.

- First, fast institutions access and process information flows before slow institutions. To capture this, we assume that, just before trading, at the same time as they learn their private value, fast institutions observe whether $v = \mu + \epsilon$ or $\mu - \epsilon$. This assumption is consistent with empirical evidence. For example, Hendershott and Riordan (2010) and Brogaard (2010) find that market orders placed by high frequency traders have a greater permanent impact on prices than market orders placed by humans. Similarly, Hendershott and Riordan (2011) find that HFT's marketable orders' informational advantage is sufficient to overcome the bid–ask spread and trading fees.

- Second, fast institutions are more likely to find trading opportunities. Regulations such as the MiFID in Europe or RegNMS in the U.S. led to competition and, in turn, fragmentation between trading platforms. This implies that quotes for the same security are posted in various trading venues.[11] Thus, investors have to search for the best price among multiple trading venues and to compare trading opportunities among several markets. HFT technologies improve search efficiency and help investors locating attractive quotes before they have been hit or withdrawn.[12] To capture this, we assume that slow institutions are less likely to find a trading opportunity than fast institutions. Namely, slow institutions find a trading counterparty with probability $\rho < 1$, while fast institutions find it with probability 1.

Trading: Our modeling of the trading process is intended to capture, in the simplest possible way, the consequences of traders' heterogeneity. When institutions find a trading opportunity, they are matched with rational competitive liquidity suppliers. At this point they decide whether to buy one share, sell one share or abstain from trading. The transaction price equals the expectation of the asset payoff, v, conditional on the institution's order, as in Glosten and Milgrom (1985). While it offers a tractable setup to model equilibrium prices and gains from trade, this stylized market mechanism abstracts from the richness of trading strategies available to high–frequency traders in limit order markets.[13]

Timing: Summing up the above discussion, timing in our model is as follows:

[11] For instance, in May 2011, the three most active competitors of the London Stock Exchange, namely Chi-X, BATS Europe and Turquoise, reached a daily market share in FTSE 100 stocks of 27.5%, 7.4% and 5.2%, respectively while that of the London Stock Exchange was 51%. Source: http://www.ft.com/trading-room.

[12] In line with this hypothesis, Garvey and Wu (2010) find that traders who get quicker access to the NYSE because of their geographical proximity pay smaller average effective spreads.

[13] Biais, Hombert and Weill (2010) offer an analysis of such aspects, but their model does not incorporate adverse selection (see also Parlour (1998), Foucault (1999) and Goettler et al (2005)).

- At $\tau = 0$, institutions decide whether to pay C, and become fast, or not.
- At $\tau = 1$,
  1. Each institution observes its private valuation δ or $-\delta$, and, if it is fast, observes the realization of v: $\mu + \epsilon$ or $\mu - \epsilon$.
  2. Each institution finds a trading opportunity or not and, if it does, optimally chooses whether to buy one unit, sell one unit or do nothing.
  3. Liquidity providers execute order ω at price $E(v|\omega)$.
- At $\tau = 2$, v is realized.

At $\tau = 1$, there are six types of institutions: (i) fast institutions with good news and high private valuations (which we denote by GH), (ii) fast institutions with good news and low private valuations (GL), (iii) fast institutions with bad news and high private valuations (BH), (iv) fast institutions with bad news and low private valuations (BL), (v) slow institutions with high private valuation (H), and (vi) slow institutions with low private valuation (L).

4 Price formation and trading with fast and slow investors

This section analyzes equilibrium transaction prices and trading volume, for a given level of α. We focus on the case where the institution decides to buy the asset (or abstain from trading). The corresponding price is denoted by a. The case of sales (at price b) is symmetric, e.g., the markup at which institutions buy $(a - \mu)$ is equal to the discount at which they sell $(\mu - b)$. Since there is a unit–mass continuum of institutions, trading volume is the unconditional probability that an institution trades.

4.1 Equilibrium price formation and trading

As a benchmark, first consider the case in which all institutions are slow ($\alpha = 0$). Their orders do not convey any information and execute at μ. Institutions with a high private valuation buy while those with a low valuation sell. Trading volume (denoted by Vol) is equal to the fraction of institutions that find a counterparty, ρ.

When $\alpha > 0$, the analysis is more complex. As will be clear below, equilibria can involve pure or mixed strategies. To characterize these, denote by β_j^F the probability that a fast institution $j \in \{GH, GL, BH, BL\}$ buys, and by β_j^S the probability that, conditional on finding a counterparty, a slow institution $j \in \{H, L\}$ buys.

Transaction prices cannot exceed the highest possible payoff for the security, $(\mu + \epsilon)$, and cannot be smaller than the smallest possible payoff for the security, $(\mu - \epsilon)$. Hence, fast institutions with good news on v and high private valuation always buy, i.e., $\beta_{GH}^F = 1$. Symmetrically, fast institutions with bad news and low private valuation never buy, i.e., $\beta_{BL}^F = 0$. Applying Bayes law, one obtains the following lemma.

Lemma 1 *Buy orders execute at price*

$$a = E(v\,|buy) = \mu + \frac{\alpha(1 + \beta_{GL}^F - \beta_{BH}^F)}{2\left((1-\alpha)\rho(\beta_H^S + \beta_L^S) + \frac{\alpha}{2}(1 + \beta_{GL}^F + \beta_{BH}^F)\right)}\epsilon. \tag{1}$$

First suppose that $\delta > \epsilon$. In this case the reservation price of a fast institution with good news but a low private valuation, $\mu + \epsilon - \delta$ is smaller than μ. Since $a \geq \mu$, institutions with a low private valuation never buy, even if they are fast and receive a good signal, i.e., $\beta_{GL}^F = \beta_L^S = 0$. This yields our first proposition.

Proposition 1 *When $\delta > \epsilon$ there exists an equilibrium in which: (i) all institutions buy if and only if they have a high private valuation, (ii) all trades take place at price μ, and (iii) trading volume equals $\alpha + (1-\alpha)\rho$.*

When $\delta > \epsilon$, news about v are small relative to private valuation shocks. Hence, prices and allocations are identical to those that would prevail without private information on v. Thus, prices are as without HFT, but trading volume is higher since some institutions are more likely to find a counterparty. Therefore, if $C = 0$, HFT ($\alpha > 0$) Pareto dominates the benchmark case ($\alpha = 0$). There are other equilibria however, in which algorithms have negative consequences. To see why, consider fast institutions with a high private valuation but bad news. If they expect the ask price to be higher than their reservation price, $\mu - \epsilon + \delta$, then they do not trade (i.e., $\beta_{BH}^F = 0$). This expectation can be self-fulfilling since the ask price is inversely related to the likelihood of a trade by this type of institution (see equation (1)). This is in line with the analyses of Glosten and Milgrom (1985) and Dow (2005), which underscore the possibility 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). It is also in line with the analysis of Admati and Pfleiderer (1988), who emphasize that investors will choose to trade where and when they expect liquidity, thus providing liquidity themselves, and participating in a virtuous cycle. In the supplementary appendix, however, we show that that the equilibrium in

Proposition 1 Pareto dominates those with low liquidity. And, hereafter, when multiple equilibria arise, we will focus on the Pareto dominating one, if it exists.

In the rest of the paper, we assume $\delta < \epsilon$. In this case adverse selection problems are more severe because private information on v is large relative to gains from trade. To simplify the analysis and reduce the number of possible cases, we hereafter assume

$$\frac{\epsilon}{2} < \delta < \epsilon. \tag{2}$$

That is, the volatility of the fundamental value is higher than the dispersion in private valuations ($\delta < \epsilon$), but the latter is still significant ($\frac{\epsilon}{2} < \delta$).[14] Equation (2) implies

$$\mu < \mu + \epsilon - \delta < \mu + \delta < \mu + \epsilon < \mu + \epsilon + \delta. \tag{3}$$

The first term from the left is the unconditional expectation of v. The second one is the valuation of the security for a fast investor with good news but negative private value. The third term is the valuation of the security for a slow investor with positive private value. The fourth term is the valuation of the security for the liquidity suppliers given good news on the fundamental. The fifth and last term is the valuation of the security for a fast investor with good news and positive private value. The ranking of institutions' possible valuations for the asset in equation (3) implies there are 5 candidate equilibria, spelled out below. They correspond to increasingly high ask prices. In all candidate equilibria, $\beta^F_{GH} = 1$ and $\beta^F_{BL} = \beta^S_L = 0$, as mentioned above. Furthermore, $\beta^F_{BH} = 0$ since $a > \mu > \mu - \epsilon + \delta$.

- **P1:** If $\mu \leq a < \mu + \epsilon - \delta$, fast institutions with good news buy, whatever their private value, while slow institutions buy if and only if their private value is high. Hence, $\beta^F_{GL} = 1$ and $\beta^S_H = 1$.

- **M1:** If $a = \mu + \epsilon - \delta$, fast institutions with good news and high private value buy. So do slow institutions with high private value, i.e., $\beta^S_H = 1$. Fast institutions with good news but low private value are indifferent between buying and not trading. They play mixed strategies, buying with probability $0 \leq \beta^F_{GL} \leq 1$.

- **P2:** If $\mu + \epsilon - \delta < a < \mu + \delta$, fast institutions buy if they have good news and high private value, but they do not trade if their private value and their information on v conflict, i.e., $\beta^F_{GL} = 0$. Slow institutions with high private value buy, i.e., $\beta^S_H = 1$.

[14] The case where $\frac{\epsilon}{2} \geq \delta$ is analyzed in the supplementary appendix to this paper. The qualitative results are similar to those presented here.

- **M2:** If $a = \mu + \delta$, fast institutions with good news and high private value buy, but they do not trade if their private value and their information on v conflict, i.e., $\beta^F_{GL} = 0$. Slow institutions with high private value are indifferent between buying or not trading. They play a mixed strategy, buying with probability $\beta^S_H \in [0, 1]$.

- **P3:** If $a = \mu + \epsilon$, fast institutions with good news and high private value buy. Other types choose not to trade. Hence, $\beta^F_{GL} = 0$, and $\beta^S_H = 0$.

P3 generates "crowding out" since slow institutions are sidelined and only fast institutions trade. This implies that only a small fraction of the potential gains from trade can be reaped. Unfortunately, such equilibrium can be pervasive. Suppose liquidity suppliers anticipate that only fast institutions with good news buy. Correspondingly, they set $a = \mu + \epsilon$. As a result, slow institutions choose not to trade. So do fast institutions whose private value and signal on v conflict. Hence, the expectations of the liquidity suppliers are self-fulfilling. Under (2), this holds for all parameter values. Hence we can state our next proposition.

Proposition 2 *There always exists a crowding out equilibrium (P3).*

To spell out the conditions under which other equilibria than P3 exist, denote:

$$\alpha_{P1} = \frac{\rho(\epsilon - \delta)}{\rho(\epsilon - \delta) + \delta}, \ \alpha_{P2} = \frac{\rho(\epsilon - \delta)}{\rho(\epsilon - \delta) + \frac{\delta}{2}}, \alpha_{P3} = \frac{2\rho\delta}{2\rho\delta + \epsilon - \delta}.$$

Relying on these notations, and noting that $\alpha_{P1} < \alpha_{P2} < \alpha_{P3}$, we state our next proposition.

Proposition 3 *1. If $0 < \alpha \leq \alpha_{P3}$ there exists an equilibrium of type M2, in which $\beta^S_H = \frac{\alpha}{2(1-\alpha)\rho\delta}(\epsilon - \delta)$.*

2. If $\alpha < \alpha_{P1}$, there exists an equilibrium of type P1, in which $a = \mu + \frac{\alpha}{\alpha + (1-\alpha)\rho}\epsilon$.

3. If $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$, there exists an equilibrium of type M1, in which $\beta^F_{GL} = \frac{2(1-\alpha)\rho}{\alpha\delta}(\epsilon - \delta) - 1$.

4. If $\alpha_{P2} < \alpha < \alpha_{P3}$, there exists an equilibrium of type P2, in which $a = \mu + \frac{\alpha}{\alpha + 2(1-\alpha)\rho}\epsilon$.

Figure 1 illustrates these results, highlighting that when $0 < \alpha < \alpha_{P3}$ there are three equilibria. However, as claimed in the next lemma, those with low trading volume (P3 and M2) are Pareto dominated by the others (P1, M1, or P2).

Lemma 2 *For each value of α, there is a unique Pareto dominant equilibrium: P1 when $0 \leq \alpha < \alpha_{P1}$, M1 when $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$, P2 when $\alpha_{P2} < \alpha < \alpha_{P3}$, M2 when $\alpha = \alpha_{P3}$, and P3 when $\alpha > \alpha_{P3}$.*

Hereafter, for each value of α, we focus on the Pareto dominant equilibrium. Figure 2 shows the evolution of the price impact of buy orders, $a - \mu$, as a function of α. In our framework, this is a measure of the informativeness of trades. It weakly increases in α because: (i) the fraction of investors with news increases and (ii) trading strategies become increasingly dependent on news (e.g., slow institutions stop trading when $\alpha > \alpha_{P3}$).[15] Let $\psi(\alpha)$ and $\phi(\alpha)$ be the expected gain for slow and fast institutions respectively. Using Proposition 3, we obtain[16]

$$\psi(\alpha) = \begin{cases} (\delta - \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon)\rho & \text{for } 0 \leq \alpha < \alpha_{P1}, \\ (2\delta - \epsilon)\rho & \text{for } \alpha_{P1} \leq \alpha \leq \alpha_{P2}, \\ (\delta - \frac{\alpha/2}{\alpha/2+(1-\alpha)\rho}\epsilon)\rho & \text{for } \alpha_{P2} < \alpha \leq \alpha_{P3}, \\ 0 & \text{for } \alpha > \alpha_{P3}. \end{cases} \tag{4}$$

and

$$\phi(\alpha) = \begin{cases} \frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho}\epsilon & \text{for } 0 \leq \alpha < \alpha_{P1}, \\ \delta & \text{for } \alpha_{P1} \leq \alpha \leq \alpha_{P2}, \\ \frac{1}{2}(\delta + \frac{(1-\alpha)\rho}{\alpha/2+(1-\alpha)\rho}\epsilon) & \text{for } \alpha_{P2} < \alpha \leq \alpha_{P3}, \\ \delta/2 & \text{for } \alpha > \alpha_{P3}. \end{cases} \tag{5}$$

which yields the following corollary.

Corollary 1 *The expected gains from trades of each fast or slow institution (weakly) decrease in the fraction of fast institutions.*

Figure 3 illustrates that the expected gains of slow *and* fast institutions declines with α. This arises for two reasons. First, as trades become more informative, institutions buy at a higher markup (or sell at more discounted prices). Second, as price impact increases,

[15] At first glance, the empirical findings in Hendershott et al.(2011) do not support this implication of our model. They find empirically that the informational impact of trades has declined on the NYSE after a change in market structure that made algorithmic trading easier on the NYSE. However, it is not clear whether Hendershott et al.(2011)'s proxy for algortithmic trading captures the effect of high frequency traders or other forms of algorithmic trading. Moreover the market structure change considered in Hendershott et al.(2011) may also have helped slow traders to better find trading opportunities (an increase in ρ). Our model predicts that for a fixed value of α, the informational impact of trades declines in ρ.

[16] A derivation of $\phi(\alpha)$ and $\psi(\alpha)$ is given in the proof of Lemma 2. The case where $\alpha = \alpha_{P_3}$ and where the Pareto-dominant equilibrium is M2 is actually a limit case of $\alpha_{P2} < \alpha < \alpha_{P3}$.

institutions trade less. For instance, fast institutions with low private valuations but good news trade less or stop trading when $\alpha > \alpha_{P1}$ because their impact on prices is too high. Similarly, slow institutions pull out from the market when $\alpha > \alpha_{P3}$. For these reasons, the entry of a new fast institution exerts a *negative externality* on all other institutions, fast or slow. Fast institutions however always get greater expected gains than slow ones because (i) they trade more and (ii) they profit from their private information. The latter source of gain is obtained at the expense of slow institutions and does not increase aggregate welfare (the weighted average of slow and fast institutions' gains). It may even decrease aggregate welfare if it leads to a situation in which institutions stop trading in some states. This reflects the above mentioned negative externality.

To build further intuition on this externality and the welfare effects of HFT, it is useful to contrast two polar cases: the benchmark case where all institutions are slow ($\alpha = 0$) and that in which all institutions are fast ($\alpha = 1$). In the former, institutions' expected gains are: $\psi(0) = \rho\delta$ whereas in the latter their expected gains are $\phi(1) = \delta/2$. Thus, if $\rho > 1/2$, even if $C = 0$, all institutions are better off with $\alpha = 0$ than with $\alpha = 1$. Yet, under our assumption that $\epsilon > \delta$, $\alpha = 0$ is not individually optimal, since $\phi(0) = \epsilon > \psi(0) = \rho\delta$. This is akin to the Prisoner's dilemma and reflects the negative externality generated by HFT. We come back to this point in Section 6.

Turning back to the general case, our next corollary states the effect of α on trading volume.

Corollary 2 *Equilibrium trading volume is:*

$$Vol(\alpha) = \begin{cases} \alpha + (1-\alpha)\rho & \text{for } 0 \leq \alpha < \alpha_{P1}, \\ (1-\alpha)\rho\epsilon/\delta & \text{for } \alpha_{P1} \leq \alpha \leq \alpha_{P2}, \\ \alpha/2 + (1-\alpha)\rho & \text{for } \alpha_{P2} < \alpha \leq \alpha_{P3}, \\ \alpha/2 & \text{for } \alpha > \alpha_{P3}. \end{cases} \tag{6}$$

Thus, trading volume is non monotonic in the level of high frequency trading, α (see Figure 4).

HFT increases the probability of finding a counterparty, but because it generates adverse selection it can reduce trading for institutions finding a counterparty. Hence, trading volume is not monotonic in α, as illustrated in Figure 4. When α is very low, adverse selection is limited and the main effect of an increase in α is to increase the probability that an institution finds a counterparty. Furthermore, when α is very large, most institutions participating in trading are fast, and an increase in α increases total trading volume. Therefore, when there is either little HFT or a lot of it, trading volume is increasing in α.

In contrast, for intermediate values of α, trading volume can decrease in the level of HFT. Indeed, an increase in this level leads fast institutions to trade less intensively because their price impact is higher (specifically, fast institutions do not trade when their signal and private valuations conflict while they would trade for sufficiently low levels of HFTs).[17] There is a discrete drop in trading when α increases beyond α_{P3}, due to the fact that at this point slow institutions stop trading. More precisely, a small increase in α at $\alpha = \alpha_{P3}$ implies that trading volume drops from $\frac{\alpha_{P3}}{2} + (1-\alpha_{P3})\rho$ to $\frac{\alpha_{P3}}{2}$. Thus, an increase in HFT can be associated with a drop in trading volume in some cases. This is in line with the finding by Jovanovic and Menkveld (2010) that for Dutch stocks the entry of a fast trader on Chi-X led to a drop in volume.[18]

5 Scope of High Frequency Trading

In the previous section, the "scope of High Frequency Trading", α, was exogenous. We now study its equilibrium determination.

5.1 Heterogeneity in institutions' size

While C is the same for all market participants, institutions are heterogeneous in size. Large institutions can take advantage of their investment in HFT facilities on a greater scale than smaller ones. To model heterogeneity in the scale of institutions, while preserving the tractability of the model presented in the previous section, we proceed as follows.

We assume that investors have potential access to a continuum of markets of size N. Each market is as in the previous section, and, for simplicity, the random variables are i.i.d across markets. The scale of an institution is defined by the number of markets to which it can participate. Namely, an institution of type t can participate in $n(t) \leq N$ markets. $n(t)$ increases in t, i.e., a higher value of t corresponds to a bigger institution. Thus, we refer to t as the *size* of an institution.

Our key assumption is that institutions' sizes are distributed over $[\underline{t}, \bar{t}]$ with density

[17] More precisely, when $\rho > 1/2$, a small increase in the fraction of fast institutions increases the trading volume when $\alpha < \alpha_{P1}$ or $\alpha > \alpha_{P3}$ and it decreases trading volume when $\alpha_{P1} \leq \alpha \leq \alpha_{P3}$. When $\rho \leq 1/2$, a small increase in the fraction of fast institutions increases the trading volume when $\alpha < \alpha_{P1}$ or $\alpha > \alpha_{P2}$ and it decreases trading volume when $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$.

[18] Anecdotal evidence also suggests that, as High Frequency 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.

$f(t)$ such that:

$$f(t) = \frac{N}{n(t)}. \tag{7}$$

As $n(t)$ increases with an institution size, the mass of institutions of a given size, $f(t)$, decreases with size. This assumption captures the notion that there are a few big institutions with access to many markets and many small institutions with access to only a few markets. (7) is in the spirit of Zipf's law, as the density of type t is inversely proportional to its rank, $n(t)$.[19]

While larger institutions are active in a greater number of markets, smaller institutions are more numerous. Equation (7) implies that the two effects offset one another, so that the mere fact that an institution is present in a market does not convey any information about its size. More formally, (7) implies that the total number of markets in which the population of type–t institutions are active is $n(t)f(t) = N$, i.e., it is constant across types t. Hence, *within each market*, investors' types have a uniform distribution. This enables us to keep the framework presented in Section 2, and in particular the updating rules underlying equation (1), while allowing for heterogeneity among institutions.

To illustrate and clarify the intuition, consider the following discrete example, depicted in Figure 5: There are four types of institutions and $N = 6$ markets. There is one institution of type t_6, which has access to 6 markets ($n(t_6) = 6$), two institutions of type t_3 which have access to three markets ($n(t_3) = 3$), three institutions of type t_2 who have access to two markets ($n(t_2) = 2$), and six institutions of type t_1 who have access to only one market ($n(t_1) = 1$).[20] Thus, in line with (7), the number of institutions of a given type multiplied by the number of markets to which it has access is constant across types t. And there is exactly one trader of each type in each market. Consequently, as mentioned above, that one trader is active in a market does not convey any information about its size.

5.2 Investing in the HFT technology

For a given level of HFT, α, the expected profit of a type t institution if it chooses to pay the investment cost C is:

$$\phi(\alpha)n(t) - C,$$

while if it does not invest in HFT, its expected profit is:

$$\psi(\alpha)n(t),$$

[19] For an application of Zipf's law in finance and economics see Gabaix and Landier (2003).

[20] Also, in keeping with Zipf's law, the most frequent type of trader (t_1) "occurs" twice as often as the second most frequent type (t_2), three times as often as the third most frequent type (t_3) and six times most often as the less frequent type (t_6).

where ϕ and ψ are defined in (5) and (4) respectively. Thus, an institution with size t is better off investing in algorithmic trading if and only if:

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

Since, $\phi(\alpha) > \psi(\alpha)\ \forall\alpha$, the above inequality is equivalent to

$$n(t) \geq \frac{C}{\phi(\alpha) - \psi(\alpha)},$$

or, as $n(.)$ is increasing,

$$t \geq n^{-1}(\frac{C}{\phi(\alpha) - \psi(\alpha)}).$$

Thus, defining the function $t^*(.)$ as

$$t^*(\alpha) = n^{-1}(\frac{C}{\phi(\alpha) - \psi(\alpha)}), \tag{8}$$

we obtain the following result.

Lemma 3 *For any given $\alpha \in [0, 1]$, an institution is better off investing in High Frequency Trading if and only if its size t is greater than $t^*(\alpha)$.*

Investment in the HFT technology is more profitable for large institutions, since they have more trading opportunities and therefore can better amortize the fixed cost C. But an institution's decision to invest in the HFT technology does not only depend on its own size. It also depends on the overall level of investment in this technology, α. As this level depends on other institutions' choices, institutions' technological choices are interdependent. The logic is the following: An increase in α raises the price impact of trades. This reduces both the profits earned by fast investors ($\phi(\alpha)$) and those earned by slow investors ($\psi(\alpha)$). But, as can be seen in Lemma 3 and equation (8), what matters for the decision to invest in HFT or not, is the difference between the profits of fast investors and those of slow investors, $\phi(\alpha) - \psi(\alpha)$. If $\phi(\alpha) - \psi(\alpha)$ is decreasing in α, then fast investors loose more than slow ones when α goes up. Hence $t^*(\alpha)$ increases in α, and the decisions to invest in HFT are strategic substitutes: the greater the fraction of institutions which have decided to invest in HFT, the higher the size threshold above which institutions decide to invest in HFT. In contrast, if $\phi(\alpha) - \psi(\alpha)$ is increasing in α, that is, slow investors are hurt more than fast investors by an increase in α. Consequently, $t^*(\alpha)$ is decreasing in α, and investments in HFT are strategic complements: the greater the fraction of institutions which have decided to invest in HFT, the lower the size threshold above which institutions decide to invest in HFT. Otherwise stated, the institutions' decisions to invest in HFT are mutually reinforcing.

The next proposition states the condition under which the decisions to invest in HFT are strategic substitutes or complements.

Proposition 4 *When $0 \leq \alpha < \alpha_{P1}$, $\phi(\alpha) - \psi(\alpha)$ is decreasing in α and the decisions to invest in HFT are strategic substitutes. If $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$ or $\alpha > \alpha_{P3}$, then $\phi(\alpha) - \psi(\alpha)$ is constant with α. When $\alpha_{P2} < \alpha < \alpha_{P3}$, $\phi(\alpha) - \psi(\alpha)$ is increasing in α and the decisions to invest in HFT are strategic substitutes if $\rho \leq \frac{1}{2}$ and strategic complements otherwise.*

To see why this result obtains, it is useful to compare $\phi(\alpha)$ and $\psi(\alpha)$ in equilibria *P1* and *P2*. When $0 \leq \alpha < \alpha_{P1}$ and *P1* prevails, then slow institutions buy if and only if they have located a trading opportunity and their private valuation is high so that their expected profit is:

$$\psi(\alpha) = \Pr(+\delta)\rho(\mu + \delta - a(\alpha)), \tag{9}$$

where $\Pr(+\delta)$ is the likelihood that a slow institution has a high valuation and $a(\alpha)$ is the price at which buy orders execute when the level of HFT is α. In this equilibrium, fast traders buy if and only if they have observed good news about v and their expected profit is:

$$\phi(\alpha) = \Pr(+\epsilon)(\mu + \epsilon - a(\alpha)),$$

where $\Pr(+\epsilon)$ is the likelihood that the asset value is large. Thus when $0 \leq \alpha < \alpha_{P1}$,

$$\frac{\partial}{\partial \alpha}(\phi - \psi)(\alpha) = -\frac{1}{2}(1 - \rho)\frac{\partial}{\partial \alpha}a(\alpha).$$

This is negative because slow investors trade less often than fast ones, and are therefore less affected by the increase in price impact. Thus, in this case, the decisions to invest in HFT are strategic substitutes.

When $\alpha_{P2} < \alpha < \alpha_{P3}$, and equilibrium *P2* prevails, slow traders still buy if and only if they have located a trading opportunity and their private valuation is high so that their profit is as in (9). But fast traders buy only if they have good news and high private valuations, so their profits are

$$\phi(\alpha) = \Pr(+\epsilon)\Pr(+\delta)(\mu + \delta + \epsilon - a(\alpha)).$$

Thus when $\alpha_{P2} < \alpha < \alpha_{P3}$,

$$\frac{\partial}{\partial \alpha}(\phi - \psi)(\alpha) = -\frac{1}{2}(\frac{1}{2} - \rho)\frac{\partial}{\partial \alpha}a(\alpha).$$

This is positive if and only if $\rho > \frac{1}{2}$. In that case, as ρ is relatively high, in *P2* slow investors trade more often than fast ones, and are therefore more affected by the increase in price impact. Hence the decisions to invest in HFT are strategic complements.

Building on this analysis, we now study the equilibrium determination of α, and show that when the decisions to invest in HFT are strategic complement equilibrium multiplicity can arise.

5.3 Corner equilibria

Denote by α^* the equilibrium fraction of investors who decide to invest in HFT. If $t^*(0) > \bar{t}$ there exists an equilibrium in which no institution invests in HFT, i.e., $\alpha^* = 0$. Indeed, $t^*(0) > \bar{t}$ implies that, even for the largest institution, incurring cost C is non profitable, when it is expected that no one will invest in algorithmic trading. From equation (8), the condition $t^*(0) > \bar{t}$ is equivalent to:

$$t^*(0) = n^{-1}(\frac{C}{\phi(0) - \psi(0)}) > \bar{t}. \tag{10}$$

Substituting $\alpha = 0$ in (4) and (5), $\psi(0) = \delta\rho$ and $\phi(0) = \epsilon$. Thus (10) becomes:

$$t^*(0) = n^{-1}(\frac{C}{\epsilon - \delta\rho}) > \bar{t}. \tag{11}$$

As $n(.)$ is increasing and $\epsilon > \delta\rho$ (under Condition (2)), equation (11) leads to the next proposition.

Proposition 5 *Denote $C_{\max} = n(\bar{t})(\epsilon - \delta\rho)$. If $C > C_{\max}$, there exists an equilibrium in which there is no investment in High Frequency Trading, i.e., $\alpha^* = 0$.*

Conversely, if $t^*(1) \leq \underline{t}$, even the smallest institution finds it optimal to invest in HFT when it expects all the others to do so. Thus, following a similar logic as for Proposition 5, we obtain our next result.

Proposition 6 *Denote $C_{\min} = n(\underline{t})\delta/2$. If $C < C_{\min}$, there exists an equilibrium in which all institutions invest in High Frequency Trading, i.e., $\alpha^* = 1$.*

It is not always the case that $C_{\max}$ is above $C_{\min}$. Indeed,

$$C_{\max} > C_{\min} \iff n(\bar{t})(\epsilon - \delta\rho) > n(\underline{t})\delta/2. \tag{12}$$

This is equivalent to $\rho < \rho^*$ where

$$\rho^* = \frac{\epsilon}{\delta} - \frac{1}{2}\frac{n(\underline{t})}{n(\bar{t})} > \frac{1}{2}. \tag{13}$$

Thus, building on Propositions 5 and 6, we have the following result.

Proposition 7 *If $\rho > \rho^*$, then $C_{\max} < C_{\min}$ and if $C \in (C_{\max}, C_{\min})$ there are at least two possible levels of High Frequency Trading in equilibrium: $\alpha^* = 1$ and $\alpha^* = 0$.*

The multiplicity of equilibria reflects the strategic complementarity discussed above. Indeed, as stated in Proposition 4, when $\rho > \frac{1}{2}$, investments in HFT are strategic complements, and, when $\rho > \rho^* > \frac{1}{2}$, this complementarity is very strong. Thus, the prevalence of HFT can be a self–fulfilling prophecy: if institutions expect the others to be fast, then they have an incentive to be fast too. In this sense, there is a form of herding or contagion in institutions' decisions to be fast.

5.4 Interior equilibria

So far we derived conditions for corner equilibria, we now consider interior equilibria. From Lemma 3, when market participants expect that a fraction α of institutions will invest in algorithmic trading, institutions with size greater than $t^*(\alpha)$ invest themselves. Since institutions' types within each market are uniformly distributed, in each market the mass of institutions with $t > t^*(\alpha)$ is

$$\frac{\bar{t} - t^*(\alpha)}{\bar{t} - \underline{t}}.$$

Therefore, if there exists a fixed–point $\alpha^* \in (0, 1)$ solving

$$\alpha^* = \frac{\bar{t} - t^*(\alpha^*)}{\bar{t} - \underline{t}}, \tag{14}$$

there exists an interior equilibrium. To encompass the case in which one or two corner equilibria exist, we can more generally state that α^* is an equilibrium fraction of HFT if it solves:

$$\alpha^* = Min\{Max\{\frac{\bar{t} - t^*(\alpha)}{\bar{t} - \underline{t}}, 0\}, 1\}. \tag{15}$$

Condition (15) states that, when institutions expect a fraction α^* to invest in HFT, then α^* is precisely the fraction of institutions which do so. This is similar to the endogenous determination of the fraction of informed agents in Grossman and Stiglitz (1980). Beyond the technical differences (two point distributions instead of normality, different price setting mechanism) the substantive economic differences between the two analyses include the following:

- In the present model there is no noise trading, and the level of non–informational trading endogenously adjusts to the level of adverse selection. Thus trading volume is endogenous and the social cost of adverse selection can be analyzed.

- To capture the salient features of high frequency trading we assume that when institutions incur cost C this investment both increases their real gains from trade (they are more likely to find a counterparty) and gives them an informational edge ($\rho = 1$ is the special case where there are only informational effects.)

- We consider financial institutions that are heterogeneous in size, which enables us to contrast the information acquisition decisions of small and large players.

The next proposition shows that in general the model has several possible equilibria when $\rho > 1/2$ (the exact number depending on the value of C, the exogenous parameters and the specification of $f(.)$) whereas, if it exists, the equilibrium is unique when $\rho \leq 1/2$. The difference between the cases $\rho > 1/2$ and $\rho \leq \frac{1}{2}$ follows from the fact that in the former case, investment decisions in HFT can be complements whereas they are substitutes in the former case (Proposition 4). For the next proposition, we define $H(\alpha, C) = Min\{Max\{\frac{\bar{t}-t^*(\alpha)}{\bar{t}-\underline{t}}, 0\}, 1\}$.

Proposition 8 *(Equilibrium high frequency trading)*

1. *If $\rho \leq 1/2$, $H(\alpha_{P3}, C) > \alpha_{P3}$ and $H(1, C) < \alpha_{P3}$, there exists no equilibrium level of high frequency trading. Otherwise, when $\rho \leq 1/2$, the equilibrium level of high frequency trading is unique and it decreases with C. It is equal to zero for $C \geq C_{\max}$ and one for $C \leq C_{\min}$.*

2. *If $\frac{1}{2} < \rho \leq \rho^*$, $C_{\min} < C < C_{\max}$, $H(\alpha_{P1}, C) > \alpha_{P1}$ and $H(1, C) < \alpha_{P3}$ then there exists no level of high frequency trading in equilibrium. Otherwise there can be one, two or three possible levels of high frequency trading in equilibrium and these equilibrium levels are all strictly positive. When $C < C_{\min}$, depending on parameter values, there can be one, two or three levels of high frequency trading. In all cases, this level is strictly positive. When $C > C_{\max}$, depending on parameter values, there can be one, two or three equilibria, which include no investment in high frequency trading.*

3. *If $\rho > \rho^*$ and $C_{\max} < C < C_{\min}$ then there are three possible levels of high frequency trading, $\alpha^* = 0$, $\alpha^* = 1$ and one interior level, $\alpha^* \in (\alpha_{P2}, \alpha_{P3}]$. When $C < C_{\max}$, depending on parameter values, there can be one, two or three levels of high frequency trading. In all cases, this level is strictly positive. When $C > C_{\min}$, depending*

on parameter values, there can be one, two or three equilibria, which include no investment in high frequency trading.

When $\rho < \rho^*$ and $C_{\min} < C < C_{\max}$, there are cases in which equation 15 has no solution, that is, there is no equilibrium level of high frequency trading. Indeed, when the level of high frequency trading crosses the threshold α_{P3}, slow institutions stop trading, which generates a discontinuous drop in liquidity and fast institutions' expected profits (see Figure 3). As a result, the size of the institution which is just indifferent between being fast or slow jumps upward at $\alpha = \alpha_{P3}$, creating a discontinuous drop in the fraction of fast institutions, i.e., $H(\alpha, C)$ at α_{P3}. This discontinuity can preclude the existence of a solution to (15). for instance, if $\rho \leq \frac{1}{2}$, $H(\alpha, C)$ is weakly decreasing in α and, for $\alpha > \alpha_{P3}$, it is constant and therefore equal to $H(1, C)$. Thus, if $H(\alpha_{P3}, C) > \alpha_{P3} > H(1, C)$, the function $H(\alpha, C)$ does not cut 45° line and there is no equilibrium fraction of fast institutions.

Inexistence is driven by the simultaneity of investment decisions. To see this, suppose again that $\rho \leq \frac{1}{2}$ and that parameters are such that an equilibrium does not exist if institutions make their investment decisions simultaneously, as assumed here. Now suppose instead institutions reach decisions one after another, starting with the highest type ($\bar{t}$) and ending with the lowest one ($\underline{t}$). When making that choice, each institution observes previous decisions and rationally anticipates the decisions that will be taken afterwards. In that case, the only subgame perfect equilibrium is such that all institutions with $t \geq t^*(\alpha_{P3})$ invest in HFT whereas smaller institutions don't.[21] Indeed, institutions larger than $t^*(\alpha_{P3})$ are strictly better off investing in HFT when they anticipate that institutions smaller than $t^*(\alpha_{P3})$ don't invest since $H(\alpha_{P3}, C) > \alpha_{P3}$. Once all institutions larger than $t^*(\alpha_{P3})$ have invested, an institution smaller than $t^*(\alpha_{P3})$ knows that if it invests, it will trigger a discontinuous drop in the expected profit of fast institutions and will not be able to cover its cost since $t^*(\alpha) > t^*(\alpha_{P3})$ for $\alpha > \alpha_{P3}$ (as $\alpha_{P3} > H(1, C)$). Hence institutions smaller than $t^*(\alpha_{P3})$ choose not to invest in HFT.

When $\rho > 1/2$, there can be up to three equilibria because for $\alpha \in (\alpha_{P2}, \alpha_{P3}]$, institutions' decisions are complements. Thus, for the same parameter values one can have an equilibrium with little HFT, an equilibrium with a medium level of HFT and an equilibrium with lots of HFT. In the latter, HFT creates its "own space" by reducing the trading gains of slow investors, so that paying the cost of being fast appears relatively more attractive when many other institutions pay this cost. Depending on which of these two equilibria one focuses, HFT can increase or decrease as C decreases.

[21] When the equilibrium exists and is unique, the unique subgame perfect equilibrium of this sequential investment game is identical to that obtained when institutions make their investment decision simultaneously.

As an example, suppose that the density of institutions' size is given by

$$f(t) = 1 + b(\bar{t} - t), \text{with } b = \frac{2(1-\Delta t)}{(\Delta t)^2} \text{ and } \Delta t = \bar{t} - \underline{t} < 1. \tag{16}$$

This specification, which is consistent with (7), guarantees that $f(t)$ integrates to one and implies that $n(\bar{t}) = N$, i.e., the maximum number of markets to which the largest institutions have access is N. Figure 6 shows plots function $H(\alpha, C)$ when $\epsilon = 1, \delta = 0.9$, $\Delta t = 0.9$ and $N = 10$. In Panel A, we set $\rho = 0.9$ and consider two values of C: $C = 2$ and $C = 2.5$. In this case, for each value of C, there are three equilibria: $\alpha^* = 0$, $\alpha^* = 1$ and $\alpha^* = 28.5\%$ when $C = 2$ and $\alpha^* = 0$, $\alpha^* = 1$ and $\alpha^* = 48.3\%$ when $C = 2.5$. This example shows that the effect of an increase in the cost of HFT, C, on the equilibrium level of HFT is ambiguous when there are multiple equilibria. If $\alpha^* = 28.5\%$ when $C = 2$ then an increase in the cost of HFT from $C = 2$ to $C = 2.5$ can trigger an increase in the level of HFT or a decrease, depending on which equilibrium obtains in the second case. In Panel B, we set $\rho = 0.6$ and $C = 4.145$. In this case, there are only two equilibrium levels of HFT: $\alpha^* = 6.21\%$ and $\alpha^* = 86\%$.

6 High Frequency Trading and social welfare

We now study whether the fraction of institutions engaging in HFT in equilibrium is socially optimal. Suppose all traders with types above t^* choose to invest in the HFT technology. As a result the fraction of fast investors in each market is

$$\alpha = \frac{\bar{t} - t^*}{\bar{t} - \underline{t}}. \tag{17}$$

Utilitarian welfare is:

$$W(\alpha) = \int_{\underline{t}}^{t^*} \psi(\alpha) n(t) f(t) dt + \int_{t^*}^{\bar{t}} [\phi(\alpha) n(t) - C] f(t) dt.$$

Because we assume that $n(t)f(t) = N$, this simplifies to

$$W(\alpha) = N[\int_{\underline{t}}^{t^*} \psi(\alpha) dt + \int_{t^*}^{\bar{t}} \phi(\alpha) dt] - C(1 - F(t^*)),$$

where $F(\cdot)$ is the cumulative probability distribution of financial institutions' size. Hence:

$$W(\alpha) = N[\psi(\alpha)(t^* - \underline{t}) + \phi(\alpha)(\bar{t} - t^*)] - C(1 - F(t^*)). \tag{18}$$

Since, by (17), $t^* = \bar{t} - \alpha(\bar{t} - \underline{t})$, (18) rewrites as:

$$W(\alpha) = N(\bar{t} - \underline{t})[(1-\alpha)\psi(\alpha) + \alpha\phi(\alpha)] - C(1 - F(\bar{t} - \alpha(\bar{t} - \underline{t}))). \tag{19}$$

Maximizing $W(\alpha)$ and comparing the solution to the equilibrium fraction of fast investors, α^* given in Proposition 8, one obtains the following proposition.

Proposition 9 *If there is an interior equilibrium fraction of high–frequency traders $\alpha^* \in (0,1)$, then, evaluated at $\alpha = \alpha^*$, utilitarian welfare is decreasing in the level of HFT, i.e., $\frac{\partial W}{\partial \alpha}|_{\alpha=\alpha^*} \leq 0$, with a strict inequality for some parameter values.*

Consider an interior equilibrium $\alpha^* \in (0,1)$.[22] Starting from this point, a small reduction in the level of investment in HFT would increase utilitarian welfare. This discrepancy between equilibrium and optimality arises because of the negative externality generated by HFT, factored in the calculation of the social optimum, but ignored by institutions when they make their investment decisions. While Proposition 9 offers a local result, the next proposition offers a stronger global result.

Proposition 10 *If $\rho > 1/2$, the level of investment in High Frequency Trading maximizing utilitarian welfare is $\alpha = 0$. If $\rho \leq 1/2$, there exists a threshold C^* such that*

- *If $C \geq C_{\max}$, then the equilibrium level of investment in High Frequency Trading is equal to its utilitarian–welfare maximizing counterpart, which is zero.*
- *If $C^* < C < C_{\max}$, then the equilibrium level of investment in High Frequency Trading is positive, it is strictly above its utilitarian–welfare maximizing counterpart.*
- *If $C \leq C^*$, then if the equilibrium level of investment in High Frequency Trading is positive, it is above or equal to its utilitarian–welfare maximizing counterpart.*

The social benefit of HFT is that it improves investors' ability to find counterparties. When the probability that slow investors find a counterparty is relatively large (as $\rho > 1/2$), these benefits are small. Hence the social benefits of HFT are lower than their social cost, reflecting the negative externality due to adverse selection. Consequently, utilitarian optimality rules out investment in HFT. And yet, in this case, as stated in Proposition 8, the equilibrium level of investment in HFT is bounded away from zero for $C < C_{\max}$. Hence, for $C < C_{\max}$ and $\rho > 1/2$, equilibrium HFT is always excessive.

[22]Proposition 9 does not cover the case $\alpha^* = 1$. When $\rho < 1/2$, $\alpha^* = 1$ if and only if $C < C_{\min}$. In general, the socially optimal level of High Frequency Trading will be smaller than 1 unless C is sufficiently small relative to $C_{\min}$.

In contrast, when $\rho \leq 1/2$, some investment in HFT can be socially optimal.[23] Yet, as stated in the proposition, unless C is low equilibrium will typically involve excessive investment in HFT. Again, such excess–investment arises because of the adverse–selection negative externality generated by HFT.

7 Policy implications

Short–term volatility: An issue that is often raised is whether HFT increases short–term volatility, and whether this is harmful for slower traders. For example, on pages 36 and 37 of the SEC Concept Release on Market Structure (SEC, 2010) one can read:

> "short–term price volatility may harm individual investors if they are persistently unable to react to changing prices as fast as high frequency traders."

Our theoretical analysis, takes a different perspective than the SEC, and thus sheds new light on the relationship between HFT and short–term volatility. In our model, long term volatility is the standard deviation of $\tilde{v}$, which is equal to ϵ, and is unaffected by HFT. Short term volatility can be measured as the standard deviation of prices, which affected by HFT. More precisely,[24]

$$\begin{aligned} Var(\tilde{P}) &= Var(\tilde{P}-\mu) = E[(\tilde{P}-\mu)^2] \\ &= \Pr(P=a)(a-\mu)^2 + \Pr(P=b)(b-\mu)^2 \\ &= (a-\mu)^2. \end{aligned}$$

Hence, the volatility of prices is simply $a-\mu$, the price impact of trades, which is increasing in the fraction α of institutions that engage in HFT.

Thus, in our theoretical setup, HFT does increase short–term volatility. In a sense, this is the flip–side of the information content of HFT orders. By trading on advance information, high–frequency traders move prices rapidly. But it is not the increase short–term price volatility in itself which directly hurts slower traders. Rather it is the adverse–selection induced by HFT, of which short–term price volatility is just a consequence.

[23] For instance, suppose that $f(t)$ is specified as in (16) and consider the following numerical example: $\epsilon = 1$, $\delta = 0.9$, $\Delta t = 0.9$, $\rho = 0.3$ and $N = 10$. In this case $C_{\min} = 3.68$. Thus, if $C = 1.6$, the unique equilibrium level of HFT is $\alpha^* = 1$ (see Proposition 8) while the socially optimal level of HFT is $\alpha^{\max} = 0.562$.

[24] This variance is computed conditional on a trade occurring, establishing transaction price P. This is the theoretical counterpart of the statistical approach taken when one computes the variance of prices or returns, based on the time–series of observed transactions prices.

Level playing–field: Another issue that is often raised is whether some aspects of HFT, such as co–location, prevent the market from being a level playing–field. For example, on page 59 of the SEC Concept Release on Market Structure (SEC, 2010) one can read:

> "does co–location provide proprietary firms an unfair advantage because they generally will have greater resources and sophistication to take advantage of co–location services than other market participants, including long term investors? ...
>
> Is it fair for some market participants to pay to obtain better access to the markets than is available to those not in a position to pay for or otherwise obtain co–location services? ...
>
> Are co–location fees so high that they effectively create a barrier for smaller firms?"

In line with this concern is the observation that a small group of HFT players amount for a very large fraction of equity trading volume (as mentioned above a report of the TABB group states that 2% of the 20,000 proprietary trading firms in the U.S. account for more than 70% of the trading volume.) In response to these concerns, advocates of HFT and co–location have answered that it is only necessary to ensure that access to co–location be fair, i.e., open to all at the same price. Once this condition is met, market forces should ensure optimal outcomes.

Our analysis offers a theoretical counter–argument to these claims. Even if co–location services can be purchased by all at the same price, market forces will not yield an optimal outcome. This is because, as shown above, i) only the most active traders will incur the fixed cost of investment in the HFT technology, and ii) when doing so they will impose a negative externality on the other traders. Hence there is excessive investment in HFT, i.e., the market outcome is not optimal.

Our analysis does not imply that HFT or colocation should be banned, however. As usual when there are negative externalities, Pigovian taxes can improve efficiency. To implement such taxes, one would need to observe HFT or proxies for it. This could be obtained by requesting that HFT market participants should register as such, or by using such variables as co–location or the ratio of cancelled to executed orders as proxies for HFT.

Separating the wheat from the chaff: Within the context of our model, if there was a market venue barring access to HFT, slow traders would optimally choose to trade in that market. Thus they would avoid adverse–selection and optimality would be restored.

And, if the presence of HFT did bring benefits to slow traders, then they would still have the option to trade on the open–market, from which HFT is not excluded.

The search for protection from HFT predation may indeed be part of the motivation for the routing of orders to internalizers, brokers and dark–pools. Yet, such protection has become less available. As noted by Saraiya and Mittal (2009, page 1):

> "Pools that once excluded high–frequency trading participants have now opened their gates... high–frequency trading firms ... are often the cause of short–term adverse selection in dark pools."

Saraiya and Mittal (2009, page 19) also suggest that investors with a strong bargaining position should request dark pool operators to protect them from HFT predation:

> "Depending on his influence with a dark pool operator ... a trader may be able to keep his flow from interacting with high–frequency traders."

But it's not clear that investors with limited bargaining power would be able to yield such influence on dark pool operators. What's more, it might be damaging for the market if a large fraction of low–frequency investors migrated away from lit markets. A more efficient solution could be to offer protection from HFT within lit markets. Thus, we suggest the possibility that exchanges (such as NYSE–Euronext, the LSE or the Deutsche Börse) offer slow traders the option to exclude execution against HFT orders. This would remain optional and could thus offer a market–based response to the potential adverse–selection problems created by HFT.

8 Conclusion

This paper offers an analysis of equilibrium investment in HFT technology and the ensuing trading and welfare. It underscores that, while HFT can help market participants locate trading opportunities, it also generates, adverse–selection related, negative externalities. Our analysis yields the following policy implications:

- HFT orders impound information into prices faster. This may increase short–term price volatility. Such volatility in itself is not dysfunctional. Yet, since it reflects the advance information of high–frequency traders, it is the symptom of an adverse–selection problem.

- To the extent that such adverse–selection generates negative externality for low–frequency traders, investment in HFT is excessive. This suggests that Pigovian taxes, such as, for example, taxes on co–location, could improve utilitarian welfare.

- Our analysis also suggests exchanges should offer low–frequency traders the option to exclude execution of their orders against HFT orders. This could provide an efficient market–based response to the adverse–selection problem generated by HFT.

While the present paper focuses on adverse–selection costs, HFT could generate other costs, related to operational and systemic risk, which might be quite significant, as suggested by the recent report of the SEC and CFTC on the flash–crash of May 6, 2010 and the analysis of Kirilenko et al (2010). This underscores the need to further study HFT, to examine the conditions under which this innovation can improve, rather than impair, the workings of markets.

Appendix: Proofs

Proof of Lemma 1.

In the market structure we consider, we have

$$a = E(v|buy) = \mu + \Pr(+\epsilon|buy)(\epsilon) + \Pr(-\epsilon|buy)(-\epsilon) = \mu + (2\Pr(+\epsilon|buy) - 1)\epsilon$$

Now

$$\Pr(+\epsilon|buy) = \frac{\Pr(+\epsilon \ \& \ \text{buy})}{\Pr(\text{buy})} = \frac{\Pr(\text{buy}|+\epsilon)\Pr(+\epsilon)}{\Pr(\text{buy}|+\epsilon)\Pr(+\epsilon) + \Pr(\text{buy}|-\epsilon)\Pr(-\epsilon)}.$$

Because we assume $\Pr(+\epsilon) = \Pr(-\epsilon)$, and with our notation for β, this yields

$$\Pr(+\epsilon|buy) = \frac{\alpha(\frac{1}{2} + \frac{\beta^F_{GL}}{2}) + (1-\alpha)(\frac{\beta^S_H}{2} + \frac{\beta^S_L}{2})}{[\alpha(\frac{1}{2} + \frac{\beta^F_{GL}}{2}) + (1-\alpha)\frac{\beta^S_H + \beta^S_L}{2}] + [\alpha\frac{\beta^F_{BH}}{2} + (1-\alpha)\frac{\beta^S_H+\beta^S_L}{2}]}.$$

That is

$$\Pr(+\epsilon|buy) = \frac{\alpha(1 + \beta^F_{GL} - \beta^S_H - \beta^S_L) + \beta^S_H + \beta^S_L}{\alpha(1 + \beta^F_{GL} + \beta^F_{BH}) + 2(1-\alpha)(\beta^S_H + \beta^S_L)}.$$

Consequently:

$$a = \mu + \frac{\alpha(1 + \beta^F_{GL} - \beta^F_{BH})}{\alpha(1 + \beta^F_{GL} + \beta^F_{BH}) + 2(1-\alpha)(\beta^S_H + \beta^S_L)}\epsilon.$$

■

Preliminary remarks for the proofs of Propositions 1, 2 and 3: For the proofs of this propositions, it useful first to write the expected profit $\overline{\Pi}^F_j$ of a fast institution of type $j \in \{GH, GL, BH, BL\}$ when it buys one share of the security. We obtain:

$$\begin{aligned}
\overline{\Pi}^F_{GH} &= (\mu + \epsilon + \delta - a), \\
\overline{\Pi}^F_{BH} &= (\mu - \epsilon + \delta - a), \\
\overline{\Pi}^F_{GL} &= (\mu + \epsilon - \delta - a), \\
\overline{\Pi}^F_{BL} &= (\mu - \epsilon - \delta - a).
\end{aligned}$$

Similarly, the expected profits $\overline{\Pi}^S_j$ of a slow institution of type $j \in \{H, L\}$ when it buys one share of the security is

$$\begin{aligned}
\overline{\Pi}^S_L &= (\mu - \delta - a) \\
\overline{\Pi}^S_H &= (\mu + \delta - a)
\end{aligned}$$

We have already observed that fast institutions with good (bad) news and a high (low) private valuations always buy (sell) in any equilibrium. In all the cases considered in

Propositions 1, 2 and 3, $\mu \leq a$. In a symmetric way the price at which the asset can be sold is less than μ. Hence, when $\epsilon > \delta$, it is immediate that when fast institutions expect to buy the asset at the price $a \geq \mu$, they never sell if they have a good signal and a low private valuation and they never buy if they have a bad signal and a high private valuation. Moreover, slow institutions with high private valuations never find optimal to sell and slow institutions with low private valuations never find optimal to buy. This implies that in all cases, $\beta^S_L = 0$, $\beta^F_{GH} = 1$ and $\beta^F_{BL} = 0$. Hence we just need to check that fast institutions with good news and low private valuations on the one hand and slow institutions with high private valuations on the other hand find optimal to behave as described in Propositions 1, 2 and 3.■

Proof of Proposition 1. Provided that $\mu \leq a \leq \mu+\epsilon$, $\beta^F_{BL} = 0$ and $\beta^F_{GH} = 1$ as explained in the text. Now suppose that institutions expect to be able to buy at $a = \mu$. In this case and when $\delta > \epsilon$, using the expressions for the expected profit of a slow institution and a fast institution, it is immediate that it is optimal for institutions with a high private valuation to buy the asset whereas it cannot be optimal for institutions with a low private valuation to buy it (since this results in a negative expected profit). This yields $\beta^F_{BH} = \beta^S_H = 1$ and $\beta^F_{GL} = \beta^S_L = 0$. If institutions behave in this way, using equation (1), we deduce that $a = \mu$.■

Proof of Proposition 2. Suppose that institutions expect buy orders to execute at $a = \mu + \epsilon$. In this case, it is immediate that only fast institutions with good news and a high private valuation find optimal to buy. Now suppose that institutions behave in this way. This implies $\beta^S_L = 0$, $\beta^S_H = \beta^F_{GL} =^F_{BH}= 0$. Then using equation (1), we deduce that $a = \mu + \epsilon$.■

Proof of Proposition 3.

Part 1: Suppose that institutions expect buy orders to execute at $a = \mu + \delta$. In this case, fast institutions with good news and a high private valuation find optimal to buy. Slow institutions with a high private private valuation are just indifferent between buying or not. Hence, playing a mixed strategy is optimal for these institutions.

Now suppose that institutions behave as described in the text for a M2 equilibrium, with slow institutions with a high private valuation buying with the following probability when they find a counterparty: $\beta^S_H = \frac{\alpha}{2(1-\alpha)\rho\delta}(\epsilon - \delta)$. Hence $\beta^S_L = 0$, $\beta^F_{GL} = \beta^F_{BH} = 0$, and $\beta^S_H = \frac{\alpha}{2(1-\alpha)\rho\delta}(\epsilon - \delta)$. Then using equation (1), we deduce that $a = \mu + \delta$. We define as α_{P_3} the threshold such that $\beta^S_H \leq 1$ for $\alpha \leq \alpha_{P_3}$.

Part 2: $\alpha < \alpha_{P1}$. Suppose that that institutions expect buy orders to execute at $a = \mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon$. We define as α_{P_1} the threshold such that $a \geq \mu + \epsilon - \delta$ for $\alpha \geq \alpha_{P_1}$. Thus, as $\alpha < \alpha_{P1}$, $a < \mu + \epsilon - \delta$: the expected profit of a fast institution with good news is strictly positive if it buys whatever its private valuation and we have observed that such an institution never finds optimal to sell (see preliminary remarks). Hence, a fast institution with good news always buys the asset. The expected profit of a slow institution with a high private valuation is positive and we have observed that such an institution never finds optimal to sell the asset. Hence, a slow institution with a high private valuation optimally buys the asset.

Now suppose that institutions behave as described in the proposition. This implies $\beta_L^S = 0$, $\beta_H^S = 1$, $\beta_{GL}^F = 1$, and $\beta_{BH}^F = 0$. We then deduce using equation (1) that $a = \mu + \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon$.

Part 3: $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$. Suppose that institutions expect buy orders to execute at $a = \mu + \epsilon - \delta$. The expected profit of a slow institution with a high private valuation is positive and we have observed that such an institution never finds optimal to sell. Thus, it optimally submits a buy market order. A fast institution with good news and a low private valuation is just indifferent between buying and doing nothing since if it trades, it gets an expected profit equal to zero. Hence purchasing the security with probability $\beta_{GL}^F = \frac{2(1-\alpha)\rho}{\alpha\delta}(\epsilon - \delta) - 1$ is optimal.

Now suppose that institutions behave as described in the proposition. $\beta_L^S = 0$, $\beta_H^S = 1$, $\beta_{GL}^F = \frac{2(1-\alpha)\rho}{\alpha\delta}(\epsilon-\delta)-1$,and $\beta_{BH}^F = 0$. We then deduce using equation (1) that $a = \mu+\epsilon-\delta$. We define as α_{P_2} the threshold such that $\beta_F^{GL} \geq 0$ for $\alpha \leq \alpha_{P_2}$, and we check that indeed $\beta_F^{GL} \leq 1$ for $\alpha \geq \alpha_{P_1}$.

Part 4: $\alpha_{P2} < \alpha < \alpha_{P3}$. Suppose that institutions expect buy orders to execute at price $a = \frac{\alpha}{\alpha+2(1-\alpha)\rho}\epsilon$. As $\alpha_{P3} < \alpha$, $a < \mu + \delta$. The expected profit of a slow institution with a high private valuation is positive and we have observed that such an institution never finds optimal to sell. Thus, it optimally submits a buy market order. But as $\alpha > \alpha_{P2}$, $a > \mu + \epsilon - \delta$. A fast institution with good news and a low private valuation makes a negative expected profit if it buys the asset and we have observed that such an institution never finds optimal to sell. Thus, a fast institution with a low private valuation and good news does not trade.

Now suppose that institutions behave as described in the proposition. This implies $\beta_L^S = 0$, $\beta_H^S = 1$, $\beta_{GL}^F = \beta_{BH}^F = 0$. We deduce using equation (1) that $a = \mu + \frac{\alpha}{\alpha+2(1-\alpha)\rho}\epsilon$.■

Proof of Lemma 2. When $\alpha > \alpha_{P3}$, the unique equilibrium is a type $P3$ equilibrium. Now consider the case in which $\alpha \leq \alpha_{P3}$. First we write the expected profit of fast institutions conditional on buying the asset. Let $K(\alpha)$ be this expected profit. A necessary condition for fast institutions to buy the asset is that they receive good news. Hence given that $\beta^F_{GH} = 1$

$$K(\alpha) = \frac{1}{2}(\mu + \epsilon + \delta - a) + \frac{1}{2}(\mu + \epsilon - \delta - a) \times \beta^F_{GL}.$$

The total gains from trade for fast institutions is just $H(\alpha)$ since the sell side is symmetric and fast institutions receive good and bad news with equal probabilities. Using the expression for the equilibrium value of a and β^F_{GL} in the various types of equilibria, we obtain:

$$K(\alpha) = \begin{cases} \frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho}\epsilon & \text{in a type P1 equilibrium,} \\ \delta & \text{in a type M1 equilibrium,} \\ \frac{1}{2}(\delta + \frac{(1-\alpha)\rho}{\alpha/2+(1-\alpha)\rho}\epsilon) & \text{in a type P2 equilibrium} \\ \delta/2 & \text{in a type P3 equilibrium} \\ \epsilon/2 & \text{in a type M2 equilibrium} \end{cases}$$

As $\epsilon > \delta$, fast institutions are better off in a type M2 equilibrium than in a type P3 equilibrium. Besides, comparing expected profits yields:

- In the region where a type P1 equilibrium exists, that is, for $0 \leq \alpha < \alpha_{P_1}$, $\frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho}\epsilon > \epsilon/2$.
- In the region where a type M1 equilibrium exists, that is, for $\alpha_{P_1} \leq \alpha \leq \alpha_{P_2}$, $\delta > \epsilon/2$.
- In the region where a type P2 equilibrium exists, that is, for $\alpha_{P_2} \leq \alpha < \alpha_{P_3}$, $\frac{1}{2}(\delta + \frac{(1-\alpha)\rho}{\alpha/2+(1-\alpha)\rho}\epsilon) > \epsilon/2$.

We deduce that fast institutions' expected profit in types P1, M1 and P2 equilibria, when they exist, is higher than in an equilibrium of type M2 (this than an equilibrium of type P3). At $\alpha = \alpha_{P_3}$, fast institutions' expected profit in types M2 equilibrium is higher than in an equilibrium of type P3.

Now let:

$$\phi(\alpha) = \begin{cases} \frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho}\epsilon & \text{for } 0 \leq \alpha < \alpha_{P_1}, \\ \delta & \text{for } \alpha_{P_1} \leq \alpha \leq \alpha_{P_2}, \\ \frac{1}{2}(\delta + \frac{(1-\alpha)\rho}{\alpha/2+(1-\alpha)\rho}\epsilon) & \text{for } \alpha_{P_2} < \alpha \leq \alpha_{P_3}, \\ \delta/2 & \text{for } \alpha > \alpha_{P_3}. \end{cases}$$

Observe that $\lim_{\eta \to 0} \phi(\alpha_{P_1} - \eta) = \delta = \phi(\alpha_{P_1})$, $\lim_{\eta \to 0} \phi(\alpha_{P_2} + \eta) = \delta = \phi(\alpha_{P_2})$, and $\lim_{\eta \to 0} \phi(\alpha_{P_3} - \eta) = \frac{\epsilon}{2} = \phi(\alpha_{P_3})$. Thus, $\phi(.)$ is continuous over $[0, \alpha_{P3}]$. It is then immediate that $\phi(.)$ decreases over $[0, \alpha_{P3}]$.

Now recall that a type P1 equilibrium is obtained and provides fast institutions with higher profits than other equilibria iff $0 \leq \alpha < \alpha_{P_1}$, similarly for a type $M1$ equilibrium iff $\alpha_{P_1} \leq \alpha \leq \alpha_{P_2}$, a type P2 equilibrium iff $\alpha_{P_2} < \alpha < \alpha_{P_3}$, and a type $M2$ equilibrium iff $\alpha = \alpha_{P_3}$. Thus, in a type P1 equilibrium, $K(\alpha) = \phi(\alpha)$ for $0 \leq \alpha < \alpha_{P_1}$; in a type M1 equilibrium, $K(\alpha) = \phi(\alpha)$ for $\alpha_{P_1} \leq \alpha \leq \alpha_{P_2}$; in a type P2 equilibrium, $K(\alpha) = \phi(\alpha)$ for $\alpha_{P_2} < \alpha < \alpha_{P_3}$; and in a type M2 equilibrium, $K(\alpha) = \phi(\alpha)$ for $\alpha = \alpha_{P_3}$.

Now we write the expected profit of slow institutions conditional on buying the asset. Let $S(\alpha)$ be this expected profit. A necessary condition for slow institutions to buy the asset is that they have a high private valuations. Hence

$$S(\alpha) = (\mu + \delta - a) \times \beta_H^S.$$

The expected gains from trade for slow institutions is just $\rho S(\alpha)$ since (i) a slow institution finds a trading opportunity with probability ρ, (ii) has a high or a low private valuation with equal probabilities, and (iii) the expected payoff of a slow institution with a high valuation when it buys the asset is identical to the payoff of a slow institution with a low private valuation when when it sells the asset. We focus on $\alpha < \alpha_{P3}$, since for $\alpha \geq \alpha_{P3}$, the unique equilibrium is a type $P3$ equilibrium. Using the expression for the equilibrium value of a and β_H in the various types of equilibria, we obtain:

$$\rho S(\alpha) = \begin{cases} (\delta - \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon)\rho & \text{in a type P1 equilibrium} \\ (2\delta - \epsilon)\rho & \text{in a type M1 equilibrium} \\ (\delta - \frac{\alpha/2}{\alpha/2+(1-\alpha)\rho}\epsilon)\rho & \text{in a type P2 equilibrium} \\ 0 & \text{in a type P3 equilibrium} \\ 0 & \text{in a type M2 equilibrium} \end{cases}$$

Clearly, the expected gains from trade of slow institutions is strictly higher in equilibria of types P1, M1 or P2 than in equilibria of types P3 and M2.

We conclude that when $\alpha < \alpha_{P_3}$, equilibria of types P1, M1 and P2 Pareto dominate equilibria of types P3 and M2. Equilibria P1, M1 and P2 cannot be obtained simultaneously. When $\alpha = \alpha_{P_3}$, equilibrium of type M2 Pareto dominates equilibrium of type P3 as fast institutions get a higher profit, while for $\alpha > \alpha_{P_3}$ the type P3 equilibrium is unique. Therefore, we deduce that there is a unique Pareto dominant equilibrium for each value of α.

Last, using the expressions for $K(\alpha)$ and $\rho S(\alpha)$ and the ranges of values for which the various equilibria are obtained, we deduce that the expected profit of fast institutions in the Pareto dominant equilibrium is $\phi(\alpha)$ and the expected profit of the slow institutions in the Pareto dominant equilibrium is

$$\psi(\alpha) = \begin{cases} (\delta - \frac{\alpha}{\alpha+(1-\alpha)\rho}\epsilon)\rho & \text{for } 0 \leq \alpha < \alpha_{P_1}, \\ (2\delta - \epsilon)\rho & \text{for } \alpha_{P_1} \leq \alpha \leq \alpha_{P_2}, \\ (\delta - \frac{\alpha/2}{\alpha/2+(1-\alpha)\rho}\epsilon)\rho & \text{for } \alpha_{P_2} < \alpha \leq \alpha_{P3}, \\ 0 & \text{for } \alpha > \alpha_{P_3}. \end{cases}$$

■

Proof of Corollary 1. We have already proved in the proof of Lemma 2 that $\phi(\alpha)$ decreases in α and is discontinuous at $\alpha = \alpha_{P3}$. Now consider the slow institutions. Obviously, $\psi(\alpha)$ decreases in α over the following intervals $\alpha \in [0, \alpha_{P1})$ and $\alpha \in (\alpha_{P2}, \alpha_{P3})$. Otherwise its is constant. Moreover calculations yield

$$\begin{aligned} \psi(\alpha_{P_1}) &= (2\delta - \epsilon)\,\rho. \\ \lim_{\eta \to 0} \psi(\alpha_{P_1} + \eta) &= (2\delta - \epsilon)\,\rho \\ \lim_{\eta \to 0} \psi(\alpha_{P_3} - \eta) &= 0 \\ \lim_{\eta \to 0} \psi(\alpha_{P_3} + \eta) &= 0 \end{aligned}$$

Hence $\psi(\cdot)$ is decreasing and continuous over $[0, 1]$.■

Proof of Corollary 2. In a given equilibrium, the likelihood of trade by a given institution is simply twice the likelihood that an institution buys the asset (since, by symmetry, buy and sell orders are equally likelyb in equilibrium). Thus

$$Vol(\alpha) = 2\left(\frac{\alpha}{2}(1 + \beta^F_{GL}) + (1 - \alpha)\rho\beta^S_H\right).$$

We can then obtain the expressions for the trading volume in the Pareto dominant equilibrium by replacing β^F_{GL} and β^S_H by their expression for each possible value of α. For instance, if $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$, $\beta^S_H = 1$ and $\beta^F_{GL} = \frac{2(1-\alpha)\rho}{\alpha\delta}(\epsilon - \delta) - 1$. Hence:

$$Vol(\alpha) = \frac{(1 - \alpha)\rho\epsilon}{\delta}.$$

■

Proof of Lemma 3. Immediate from the arguments in the text.■

Proof of Proposition 4. Using equations (4) and (5), we obtain that $\psi'(\alpha) = \phi'(\alpha) = 0$ if $\alpha \in [\alpha_{P1}, \alpha_{P2}]$ or $\alpha > \alpha_{P3}$. Thus, $\phi(\alpha) - \psi(\alpha)$ is constant when $\alpha \in [\alpha_{P1}, \alpha_{P2}]$ or $\alpha > \alpha_{P3}$. Moreover, for $0 \leq \alpha < \alpha_{P1}$:

$$\begin{aligned} \phi'(\alpha) &= -\frac{\rho\epsilon}{(\alpha(1 - \rho) + \rho)^2}, && (20) \\ \psi'(\alpha) &= -\frac{\rho^2\epsilon}{(\alpha(1 - \rho) + \rho)^2}. && (21) \end{aligned}$$

Thus, $\phi'(\alpha) - \psi'(\alpha) = \phi'(\alpha)(1 - \rho) < 0$ for $0 \leq \alpha < \alpha_{P1}$. Hence $\phi(\alpha) - \psi(\alpha)$ decreases over this interval. Last, for $\alpha \in (\alpha_{P2}, \alpha_{P3}]$:

$$\begin{aligned} \phi'(\alpha) &= -\frac{\rho\epsilon}{(\alpha(1 - 2\rho) + 2\rho)^2}, && (22) \\ \psi'(\alpha) &= -\frac{2\rho^2\epsilon}{(\alpha(1 - 2\rho) + 2\rho)^2}. && (23) \end{aligned}$$

Thus, $\phi^{'}(\alpha)-\psi'(\alpha) = \phi'(\alpha)(1-2\rho)$ for $\alpha \in (\alpha_{P2}, \alpha_{P3}]$. Hence, over this interval, $\phi(\alpha)-\psi(\alpha)$ decreases if $\rho \leq 1/2$ and increases otherwise.■

Proof of Proposition 5. Immediate from the arguments in the text.■

Proof of Proposition 6. The condition $t^*(1) < \underline{t}$ is equivalent to $\frac{C}{\phi(1)-\psi(0)} < n(\underline{t})$. The result is then immediate using the fact that $\phi(1) = \frac{\delta}{2}$ and $\psi(1) = 0$.■

Proof of Proposition 7. Immediate from the arguments in the text.■

Proof of Proposition 8. As explained in the text, for a fixed C, the equilibrium levels of high frequency trading solve:

$$H(\alpha^*, C) = \alpha^*,$$

where $H(\alpha, C) = Min\{Max\{\frac{\bar{t}-t^*(\alpha)}{\bar{t}-\underline{t}}, 0\}, 1\}$. Thus, an equilibrium exists if this equation has at least one solution and the number of equilibria is equal to the number of solutions to this equations, i.e., the number of times $H(\cdot)$ crosses the 45° line. We first prove the following lemma.

Lemma 4 *The function $H(\cdot)$ has the following properties:*

1. *It weakly decreases in α for $\alpha \leq \alpha_{P_1}$, is constant for $\alpha \in (\alpha_{P_1}, \alpha_{P_2}]$, and is constant for $\alpha > \alpha_{P3}$. For $\alpha \in (\alpha_{P_2}, \alpha_{P_3}]$, it increases in α if $\rho > \frac{1}{2}$ and decreases in α if $\rho \leq \frac{1}{2}$.*

2. *It is continuous in α, except maybe at $\alpha = \alpha_{P3}$ where it experiences a downward jump iff $lim_{\alpha \to \alpha_{P3}}\ t^*(\alpha) < \bar{t}$.*

3. *It weakly decreases in C and is continuous in C.*

Proof: Remember that $t^*(\alpha) = n^{-1}(\frac{C}{\phi(\alpha)-\psi(\alpha)})$. Hence, as $n^{-1}(.)$ is decreasing, Proposition 4 implies that $t^*(\alpha)$ increases in α for $\alpha \leq \alpha_{P_1}$, is constant for $\alpha \in (\alpha_{P_1}, \alpha_{P_2}]$, and is constant for $\alpha > \alpha_{P3}$. Moreover it decreases in α for $\alpha \in (\alpha_{P_2}, \alpha_{P_3}]$ if $\rho > \frac{1}{2}$ and increases in α for $\rho \leq \frac{1}{2}$. Last, observe that $\phi(\alpha) - \psi(\alpha)$ is continuous, except at $\alpha = \alpha_{P3}$ where it experiences a jump downward (see Figure 3). Thus, $t^*(\alpha)$ is continuous, except at $\alpha = \alpha_{P3}$ where it experiences a a positive jump. The lemma is then immediate using the definition of $H(\alpha, C)$. Last, as as $n^{-1}(.)$ is decreasing, we immediately deduce that $H(.)$ weakly decreases in C and is continuous in C.

Using this result, we can now determine the number of equilibria for various configurations of the investment required for HFT and parameter ρ.

Case 1: $\rho > \rho^*$. In this case $C_{\max} < C_{\min}$.

Case 1.a. Suppose first that $C < C_{\max}$. This implies that $t^*(0) < \bar{t}$ and $t^*(1) < \underline{t}$. Thus, $H(0, C) > 0$ and $H(1, C) = 1$. Thus, $\alpha^* = 1$ is an equilibrium in this case. Moreover if $H(\alpha_{P1}) < \alpha_{P1}$, we deduce from Lemma 4 that $H(., C)$ cuts the 45° lines two times: one for $\alpha < \alpha_{P1}$ and one for $\alpha \in (\alpha_{P_2}, \alpha_{P_3}]$. Thus, there are 3 possible equilibria in this case. If $H(\alpha_{P1}, C) = \alpha_{P1}$, we deduce from Lemma 4 that there are just two equilibria, $\alpha^* = 1$ and $\alpha^* = \alpha_{P1}$. Finally if $H(\alpha_{P1}, C) > \alpha_{P1}$, $\alpha^* = 1$ is the unique equilibrium. In all cases, $\alpha^* > 0$.

Case 1.b. Now suppose that $C_{\max} < C < C_{\min}$. Then $t^*(0) > \bar{t}$ and $t^*(1) < \underline{t}$. Thus, we deduce $H(0) = 0$ and $H(1, C) = 1$. Hence, we deduce from Property 1 that $H(\alpha, C) = 0$ for $\alpha \leq \alpha_{P_2}$ and $H(\alpha, C) = 1$ for $\alpha \geq \alpha_{P3}$. Thus, $\alpha^* = 1$ and $\alpha^* = 0$ are two possible equilibria in this case. Moreover as $H(\alpha)$ increases in α for $\alpha \in (\alpha_{P_2}, \alpha_{P_3}]$ and is such that $H(\alpha_{P3}, C) = 1$ while $H(\alpha_{P2}, C) = 0$, we deduce that there is a third equilibrium for which $\alpha^* \in (\alpha_{P_2}, \alpha_{P_3}]$.

Case 1.c. Finally, suppose that $C > C_{\min}$. Then $t^*(0) > \bar{t}$ and $t^*(1) > \underline{t}$. Thus, we deduce $H(0, C) = 0$ and $H(1, C) < 1$. Thus $\alpha^* = 0$ is always a possible equilibrium in this case. In addition, if $H(\alpha_{P3}, C) > \alpha_{P3}$, proceeding as in the previous case, we deduce that there is another equilibrium for which $\alpha^* \in (\alpha_{P_2}, \alpha_{P_3}]$. Finally if $H(1, C) > \alpha_{P_3}$ then $\alpha^* = H(1, C)$ is a third equilibrium. In contrast, if $H(\alpha_{P3}, C) < \alpha_{P3}$, $\alpha^* = 0$ is the unique equilibrium.

Case 2: $\frac{1}{2} < \rho < \rho^*$. In this case $C_{\max} > C_{\min}$.

Case 2.a: $C < C_{\min}$. This case is identical to Case 1.a.

Case 2.b: $C_{\min} < C < C_{\max}$. Then $t^*(0) < \bar{t}$ and $t^*(1) > \underline{t}$. Thus, $H(0, C) > 0$ and $H(1, C) < 1$.

Case 2.b.i: Suppose first that $H(1, C) > \alpha_{P3}$. If $H(\alpha_{P1}, C) < \alpha_{P1}$, we deduce from Lemma 4 that $H(.)$ cuts the 45° lines two times: one for $\alpha < \alpha_{P1}$ and one for $\alpha \in (\alpha_{P_2}, \alpha_{P_3}]$. Moreover, we have $H(1, C) = H(H(1, C), C)$ since $H(1, C) > \alpha_{P3}$. Thus, there are 3 equilibrium levels of high frequency trading in this case, including $\alpha^* = H(1, C)$. If $H(\alpha_{P1}, C) = \alpha_{P1}$ then $\alpha^* = H(1, C)$ and α_{P1} are the only two possible equilibria.If $H(\alpha_{P1}, C) > \alpha_{P1}$, then $\alpha^* = H(1, C)$ is the unique equilibrium.

*Case 2.b.ii:*Suppose now that $H(1, C) > \alpha_{P3}$. If $H(\alpha_{P1}, C) < \alpha_{P1}$, we deduce from Lemma 4 that $H(., C)$ cuts the 45° lines two times: one for $\alpha < \alpha_{P1}$ and one for $\alpha \in (\alpha_{P_2}, \alpha_{P_3}]$. Moreover, we have $H(1, C) > H(H(1), C)$ since $H(1, C) < \alpha_{P3}$. Thus, in this case, there are only two possible equilibria, one for which $\alpha^* \in (\alpha_{P_2}, \alpha_{P_3}]$ and one for which $\alpha^* \in (0, \alpha_{P_1}]$. If $H(\alpha_{P1}, C) = \alpha_{P1}$ and $H(1, C) < \alpha_{P3}$ then $\alpha = \alpha_{P1}$ is the unique equilibrium and if $H(\alpha_{P1}) > \alpha_{P1}$, there is no equilibrium.

Case 2.c: $C > C_{\max}$. Then $t^*(0) > \bar{t}$ and $t^*(1) > \underline{t}$. This case is identical to case 1.c.

Case 3: $\rho < \frac{1}{2}$. In this case $C_{\max} > C_{\min}$ and $H(.)$ decreases in α. When $C > C_{\max}$, then $t^*(0) > \bar{t}$. Hence, $H(0,C) = 0$. As $H(.,C)$ declines in α, we deduce that $\alpha^* = 0$ is the unique equilibrium. When $C < C_{\min}$ then $t^*(1) > \underline{t}$ and $H(1,C) = 1$. As $H(\alpha, C) \geq 1$ for $\alpha < 1$ then $\alpha^* = 1$ is the unique equilibrium. Now suppose that $C_{\min} < C < C_{\max}$. In this case, $t^*(0) < \bar{t}$ and $t^*(1) < \underline{t}$. Thus, $H(0,C) > 0$ but $H(1,C) < 1$. Now suppose that $H(\alpha_{P3}, C) < \alpha_{P3}$. Then, the function $H(.,C)$ crosses the 45° only once at $\alpha^* < \alpha_{P3}$ since since $H(.,C)$ decreases and is continuous over $[0, \alpha_{P3})$. If $H(\alpha_{P3}, C) > \alpha_{P3}$ and $H(1,C) < \alpha_{P3}$ then the equilibrium does not exist since $H(.,C)$ is constant over $(\alpha_{P3}, 1]$. This possibility arises because $H(.,C)$ is discontinuous at $\alpha = \alpha_{P3}$. Finally, if If $H(\alpha_{P3}, C) > \alpha_{P3}$ and $H(1,C) > \alpha_{P3}$. Then $\alpha^* = H(1,C)$ is the unique equilibrium.

In this case, as $H(\alpha, C)$ decreases in C and α, it is immediate that α^* decreases in C.■

Proof of Proposition 9. Using equation (19), we obtain

$$\frac{dW(\alpha)}{d\alpha} = N(\bar{t} - \underline{t})[(1-\alpha)\psi'(\alpha) - \psi(\alpha) + \alpha\phi'(\alpha) + \phi(\alpha)] - (\bar{t} - \underline{t})Cf(\bar{t} - \alpha(\bar{t} - \underline{t})).$$

Simplifying this yields

$$\frac{dW(\alpha)}{d\alpha} = N \times (\bar{t} - \underline{t}) \times [(1-\alpha)\psi'(\alpha) + \alpha\phi'(\alpha) - \frac{C}{N}G(\alpha, C)], \tag{24}$$

where

$$G(\alpha, C) = \phi(\alpha) - \psi(\alpha) - \frac{C}{N}f(\bar{t} - \alpha(\bar{t} - \underline{t}))$$

The equilibrium level of HFT, α^*, is such that $G(\alpha^*, C) = 0$ when $0 < \alpha^* < 1$. To see this, observe that equation (8) implies that:

$$n(t^*(\alpha)) = \frac{C}{\phi(\alpha) - \psi(\alpha)}$$

Moreover equation (7) implies $N = f(\bar{t} - \alpha(\bar{t} - \underline{t}))n(\bar{t} - \alpha(\bar{t} - \underline{t}))$. Thus,

$$G(\alpha, C) = \frac{C}{n(t^*(\alpha))} - \frac{C}{n(\bar{t} - \alpha(\bar{t} - \underline{t}))}$$

Now if $\alpha = \alpha^*$ and $0 < \alpha^* < 1$, $\bar{t} - \alpha^*(\bar{t} - \underline{t}) = t^*(\alpha)$. Thus, $G(\alpha^*, C) = 0$ when $0 < \alpha^* < 1$.

If $\alpha^* \in [\alpha_{P1}, \alpha_{P2}]$ or $\alpha^* > \alpha_{P3}$ then $\psi'(\alpha^*) = \phi'(\alpha^*) = 0$. Hence in this case, using equation (24), we have:

$$W'(\alpha^*) = 0,$$

and the equilibrium locally maximizes $W(\alpha)$. Otherwise, if $0 < \alpha^* < \alpha_{P1}$ or $\alpha^* \in (\alpha_{P2}, \alpha_{P3}]$, then $\psi'(\alpha) < 0$ and $\phi'(\alpha) < 0$ (see equations (20), (21), (22), and (23)) Consequently, as $G(\alpha^*) = 0$, equation (24) implies that $W'(\alpha^*) < 0$ for $\alpha^* \in (\alpha_{P_2}, \alpha_{P_3}]$ or $0 < \alpha^*$ $< \alpha_{P1}$. In this case a reduction in α^* raises social welfare.■

Proof of Proposition 10

We show that if $\rho > 1/2$, $W(\alpha)$ (defined in equation (25) below reaches its maximum in $\alpha = 0$. Using Equation (19) and the expressions for $\phi(\alpha)$ and $\psi(\alpha)$, we obtain that:

$$W(\alpha) = \begin{cases} N\Delta t(1-\alpha)\delta\rho - \Gamma(\alpha) \text{ if } \alpha \in [0, \alpha_{P_1}) \\ N\Delta t((1-\alpha)(2\delta - \epsilon)\rho + \alpha\delta) - \Gamma(\alpha) \text{ if } \alpha \in [\alpha_{P_1}, \alpha_{P_2}] \\ N\Delta t((1-\alpha)\delta\rho + \alpha\frac{\delta}{2}) - \Gamma(\alpha) \text{ if } \alpha \in (\alpha_{P_2}, \alpha_{P_3}] \\ N\Delta t(\alpha\frac{\delta}{2}) - \Gamma(\alpha) \text{ if } \alpha \in (\alpha_{P_3}, 1] \end{cases} \tag{25}$$

with $\Gamma(\alpha) = C(1 - F(\bar{t} - \alpha(\bar{t} - \underline{t}))) > 0$. Observe that $\Gamma(\alpha)$ increases with α and $\Gamma(0) = 0$. First, observe that $W(\alpha)$ decreases in α for $\alpha \in [0, \alpha_{P_1})$. Thus:

$$W(0) \geq W(\alpha) \text{ for } \alpha \in [0, \alpha_{P_1}).$$

Second, since $\Gamma(1) > \Gamma(0) = 0$ and $\rho \geq \frac{1}{2}$, it is immediate that:

$$W(0) \geq W(\alpha) \text{ for } \alpha \in (\alpha_{P_3}, 1].$$

Moreover:

$$W(0) - W(\alpha) = N\Delta t\alpha\delta(\rho - \frac{1}{2}) + \Gamma(\alpha) > 0 \text{ for } \alpha \in (\alpha_{P_2}, \alpha_{P_3}].$$

Last

$$W(0) - W(\alpha) = N\Delta t\left((1-\alpha)\epsilon\rho - \delta(\rho + (1-2\rho)\alpha)\right) + \Gamma(\alpha) \text{ for } \alpha \in [\alpha_{P_1}, \alpha_{P_2}].$$

As $\rho > \frac{1}{2}$ and $\epsilon > \delta$, the first term decreases in α and $\Gamma(.)$ is always positive. Thus, a sufficient condition for $W(0) - W(\alpha)$ to be positive for $\alpha \in [\alpha_{P_1}, \alpha_{P_2}]$ is:

$$(1 - \alpha_{P_2})\epsilon\rho - \delta(\rho + (1-2\rho)\alpha_{P_2}) > 0$$

which is always true.

Now, if $\rho \leq \frac{1}{2}$, given system of equations (25), we have:

$$W'(\alpha) = \begin{cases} W'_0(\alpha) = -N\Delta t\delta\rho - \Gamma'(\alpha) \text{ if } \alpha \in [0, \alpha_{P_1}) \\ W'_1(\alpha) = N\Delta t(\epsilon\rho + (1-2\rho)\delta - \Gamma'(\alpha) \text{ if } \alpha \in [\alpha_{P_1}, \alpha_{P_2}] \\ W'_2(\alpha) = N\Delta t(\frac{1}{2} - \rho)\delta - \Gamma'(\alpha) \text{ if } \alpha \in (\alpha_{P_2}, \alpha_{P_3}] \\ W'_3(\alpha) = N\Delta t\frac{\delta}{2} - \Gamma'(\alpha) \text{ if } \alpha \in (\alpha_{P_3}, 1] \end{cases}$$

with $\Gamma'(\alpha) = C\Delta t f(\bar{t} - \alpha(\bar{t} - \underline{t})) = \frac{CN\Delta t}{n(\bar{t} - \alpha(\bar{t} - \underline{t}))} > 0$.

Besides, $W(.)$ is continuous in α_{P1}, in α_{P2}, and not in α_{P3} but:

$$W_2(\alpha_{P3}) < \lim_{\alpha \to \alpha_{P3}} W_3(\alpha)$$

Part a On the one hand, for $\rho \leq \frac{1}{2}$, we have:

$$\forall C \geq 0, W_0'(\alpha) < W_2'(\alpha) < W_3'(\alpha) < W_1'(\alpha)$$

Let C_1 be such that $G(\alpha_{P1}, C_1) = 0$.

$$C_1 \equiv (\epsilon\rho + (1-2\rho)\delta)\, n(\bar{t} - \alpha_{P1}\Delta t)$$

By definition of C_1, we have $W_1'(\alpha) < 0$ for $C > C_1$. Therefore, if $C > C_1$ then $W(.)$ is strictly decreasing in α, and $\alpha^{OPT} = 0$.

On the other hand, we know that $\alpha^* = 0$ if $C \geq C_{\max}$, but that $\alpha^* \in (0, \alpha_{P1})$ if $C \in (C_1, C_{\max})$. Consequently, if $C \in (C_1, C_{\max})$ then $\alpha^* > \alpha^{OPT}$.

Part b Let us define:

$$\begin{aligned} C_2 &\quad s.t \quad G(\alpha_{P2}, C_2) = 0 \\ \overline{C} &\quad s.t \quad G(\alpha_{P3}, \overline{C}) = 0 \\ \underline{C} &\quad s.t \quad \lim_{\alpha \to \alpha_{P3}} G(\alpha, \underline{C}) = 0 \end{aligned}$$

For $\rho \leq \frac{1}{2}$, notice that:

$$C_{\min} < \underline{C} < \overline{C} < C_2 < C_1 < C_{\max}$$

Below we compare the equilibrium level to the optimal level of investment in HFT. However, it is non possible to determine α^{OPT} without assumptions on the distribution of institutions' size, $n(.)$. Consequently, instead of focusing on the optimal level α^{OPT}, we consider the set of possible optima, namely A^{OPT}, such that $\alpha^{OPT} \in A^{OPT}$, and we prove that $\forall \alpha \in A^{OPT}, \alpha \leq \alpha^*$.

- If $C_2 \leq C \leq C_1$ then $W(.)$ is decreasing on $[0, \alpha_{P1})$, increasing up to α_1 defined below then decreasing on $[\alpha_{P1}, \alpha_{P2}]$, then is monotonically decreasing on $(\alpha_{P2}, 1]$. Consequently, $A^{OPT} = \{0, \alpha_1\}$, where:

$$\alpha_1 = \frac{\bar{t} - n^{-1}\left(\frac{C}{\epsilon\rho + (1-2\rho)\delta}\right)}{\Delta t}$$

  Besides, by definition of C_1 and C_2, if $C_2 \leq C \leq C_1$ then $\alpha_{P1} \leq \alpha^* \leq \alpha_{P2}$. Notice first that $\alpha^* \geq \alpha_{P1} > 0$. Second, we have shown (proof of Proposition 9) that if $\alpha^* \in [\alpha_{P1}, \alpha_{P2}]$ then $W'(\alpha^*) = 0$. Therefore, $\alpha^* = \alpha_1$. Consequently, $\forall \alpha \in A^{OPT}, \alpha \leq \alpha^*$.

- If $\overline{C} \leq C < C_2$ then $W(.)$ is decreasing on $[0, \alpha_{P1})$, increasing on $[\alpha_{P1}, \alpha_{P3}]$ either up to α_{P2}, or up to α_2 defined below then decreasing, then is monotonically decreasing on $(\alpha_{P3}, 1]$. Consequently, $A^{OPT} = \{0, \alpha_{P2}, \alpha_2\}$, where:

$$\alpha_2 = \frac{\bar{t} - n^{-1}(\frac{C}{(\frac{1}{2}-\rho)\delta})}{\Delta t}$$

  Besides, by definition of C_2 and $\overline{C}$, if $\overline{C} \leq C < C_2$ then $\alpha_{P2} < \alpha^* \leq \alpha_{P3}$. Notice first that $\alpha^* > \alpha_{P2}$ and $\alpha^* > 0$. Second, if $\alpha^* \in (\alpha_{P2}, \alpha_{P3}]$, then given $\phi(\alpha)$ and $\psi(\alpha)$ (given in the proof of Proposition 9), the equilibrium is such that:

$$\alpha^* = \frac{\bar{t} - n^{-1}\left(\frac{C}{(\frac{1}{2}-\rho)\delta + \frac{\rho\epsilon}{(1-2\rho)\alpha^*+\rho}}\right)}{\Delta t}$$

  It is straightforward to see that $\alpha_2 < \alpha^*$ since $n(.)$ is decreasing in t. Consequently, $\forall \alpha \in A^{OPT}, \alpha \leq \alpha^*$.

- If $\underline{C} \leq C < \overline{C}$ then there is no equilibrium.

- If $C_{\min} < C < \underline{C}$ then $W(.)$ is decreasing on $[0, \alpha_{P1})$, increasing either up to α_2 defined above, or up to α_{P3} then decreasing on $[\alpha_{P1}, \alpha_{P3}]$, then is increasing up to α_3 defined below then decreasing on $(\alpha_{P3}, 1]$. Consequently, $A^{OPT} = \{0, \alpha_2, \alpha_{P3}, \alpha_3\}$.

$$\alpha_3 = \frac{\bar{t} - n^{-1}(\frac{C}{\frac{\delta}{2}})}{\Delta t}$$

  Besides, by definition of $\overline{C}$ and $C_{\min}$, if $C_{\min} < C < \underline{C}$ then $\alpha_{P3} < \alpha^* < 1$. Notice first that $\alpha^* > \alpha_{P3}$ and $\alpha^* > 0$. Second, the existence condition for α_2 ensures that $\alpha_2 < \alpha_{P3}$ so $\alpha^* > \alpha_2$. Third, we have shown (proof of Proposition 9) that if $\alpha^* > \alpha_{P3}$,then $W'(\alpha^*) = 0$. Therefore, $\alpha^* = \alpha_3$. Consequently, $\forall \alpha \in A^{OPT}, \alpha \leq \alpha^*$.

- If $C_{\min} \geq C$ then $\alpha^* = 1$, so in any case $\forall \alpha \in A^{OPT}, \alpha \leq 1$.

■

Bibliography

Admati, Anat, and Paul Pfleiderer, 1984, The Value of Information in Speculative Trading, Research paper 782, Graduate School of Business, Stanford University.

Admati, Anat and Paul Pfleiderer, 1986, A monopolistic market for information. *Journal of Economic Theory 39* , 400–438.

Admati, Anat, and Paul Pfleiderer, 1988, A Theory of Intraday Patterns: Volume and Price Variability, *Review of Financial Studies*, 1, 2–40.

Biais, Bruno, Johan Hombert and Pierre Olivier Weill, 2010, Trading and liquidity with imperfect cognition, Working paper, Toulouse University.

Brigagliano, James, 2009, "Testimony Concerning Dark Pools, Flash Orders, High Frequency Trading, and Other Market Structure Issues," Before the Senate Banking Subcommittee on Securities, Insurance, and Investment. October 28, 2009.

Brogaard, Jonathan, 2010, High frequency trading and its impact on market quality, ssrn working paper.

Carteas, Alvaro and Penalva, José, 2010, Where is the value in high frequency trading, Working paper, Carlos III University.

Chaboud, Alain., Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega, 2010, Rise of the machines: Algorithmic trading in the foreign exchange market, International Finance Discussion Papers, Board of Governors of the Federal Reserve System.

Chamley, C. (2007). "Complementarities in information acquisition with short-term trades," *Theoretical Economics*, 441-467.

Chen, Zhizhua, Aziz Lookman, Schürhoff, Normann and Duane Seppi, 2011, Why Ratings Matter: Evidence from the Lehman Brothers Index Rating Redefinition, Working paper, HEC Lausanne.

Dow, James, 2004, "Is liquidity self-fulfilling," *Journal of Business*, 77, 895-908.

Dow, James and Gary Gorton (1994), "Arbitrage chains." *Journal of Finance*, 49, 819–849.

Foucault, Thierry, Ailsa Röell, and Patrik Sandas, 2003, Market making with costly monitoring: an analysis of the SOES controversy, *Review of Financial Studies* 16, 345–384.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2010, Liquidity Cycles and Make/Take Fees in Electronic Markets, Working Paper, HEC.

Foucault, Thierry, 1999, Order Flow Composition and Trading Cost in a Dynamic Limit Order Market, *Journal of Financial Markets*.

Gabaix, Xavier and Augustin Landier, 2008, "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123, 49-100.

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

Garvey, Ryan, and Fei Wu, 2010, Speed, Distance, And Electronic Trading: New Evidence On Why Location Matters, Journal of Financial Markets 13, 367-396.

Glode, Vincent , Richard Green and Richard Lowery, 2011, "Financial Expertise as an Arms Race", forthcoming *Journal of Finance*.

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.

Goettler, R., Christine Parlour and Uday Rajan, 2005, Equilibrium in a Dynamic Limit Order Market, *Journal of Finance*, 2149–2192.

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

Hendershott, Terrence, Charles Jones and Albert Menkveld, 2010, "Does Algorithmic trading improve liquidity," *Journal of Finance* forthcoming.

Hendershott, Terrence and Ryan Riordan, 2010, "Algorithmic trading and information," NET Institute WP 09-08.

Hendershott, Terrence and Ryan Riordan, 2011, "High Frequency Trading and Price Discovery," Working Paper, UC Berkeley.

Hoffman, Peter, "Algorithmic trading in a dynamic limirt order market," mimeo, work in progress.

Iati, Robert (2009), "The real story of software trading espionage," Advanced Trading, available at http://http://www.advancedtrading.com/algorithms/218401501.

IOSCO (2011), "Consultation Report on Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency" (July 2011).

Jovanovic, Boyan, and Albert Menkveld, 2010, "Middlemen in limit order markets," Working paper, New York University.

Kirilenko, Andrei, Pete Kyle, M. Samadi & T. Tuzun, 2010, "The impact of high-frequency trading on an electronic market," Working paper, University of Maryland.

Martinez Victor and Rosu Ioanid (2011), "High frequency traders, news and volatility," mimeo, HEC Paris.

Parlour, Christine, 1998, Price Dynamics in Limit Order Markets, *Review of Financial Studies*, 789–816.

Saraiya, Nigam and Hitesh Mittal (2009), "Understanding and avoiding adverse selection in dark pools," Investment Technology Group, November 2009.

SEC (2010), Concept Release on equity market structure, available at www.sec.gov/rules/concept/2010/34-61358.pdf.

Figure 1

Possible equilibria for different values of α

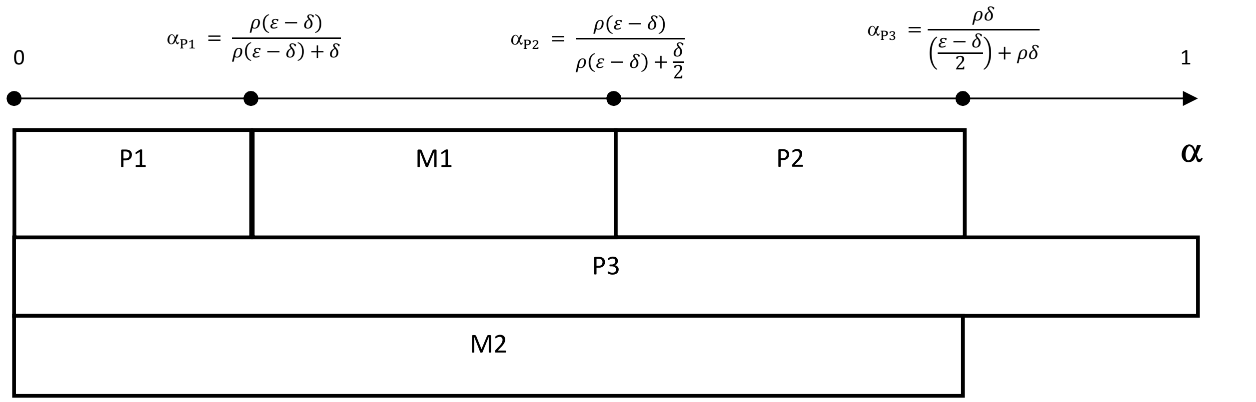


Figure 2

Informational impact of trades (a-μ) in the Pareto dominant equilibrium

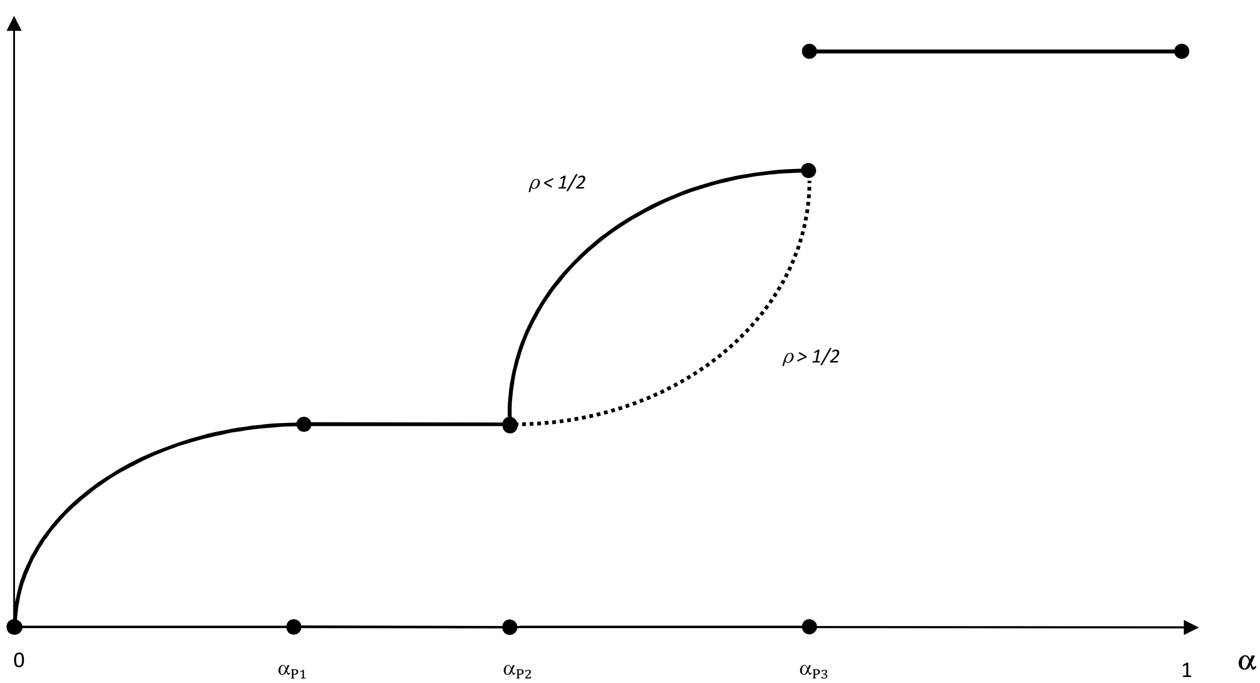


Figure 3: Expected profits of slow and fast institutions

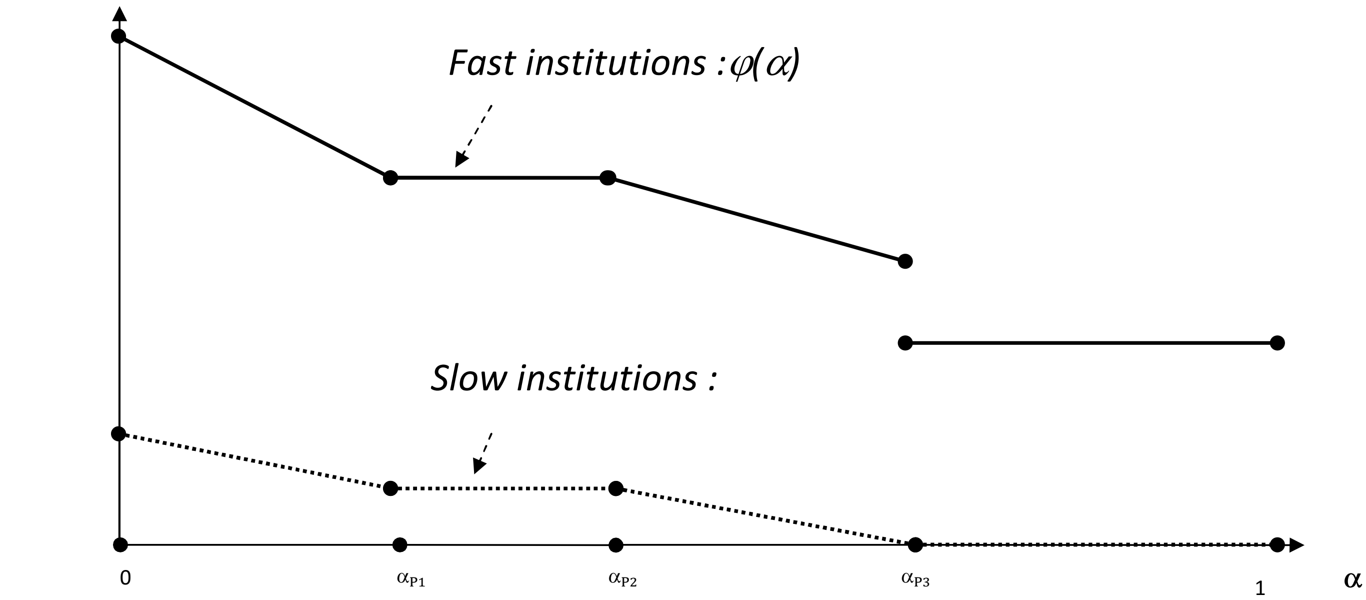


Figure 4

Volume of trade in the Pareto dominant equilibrium

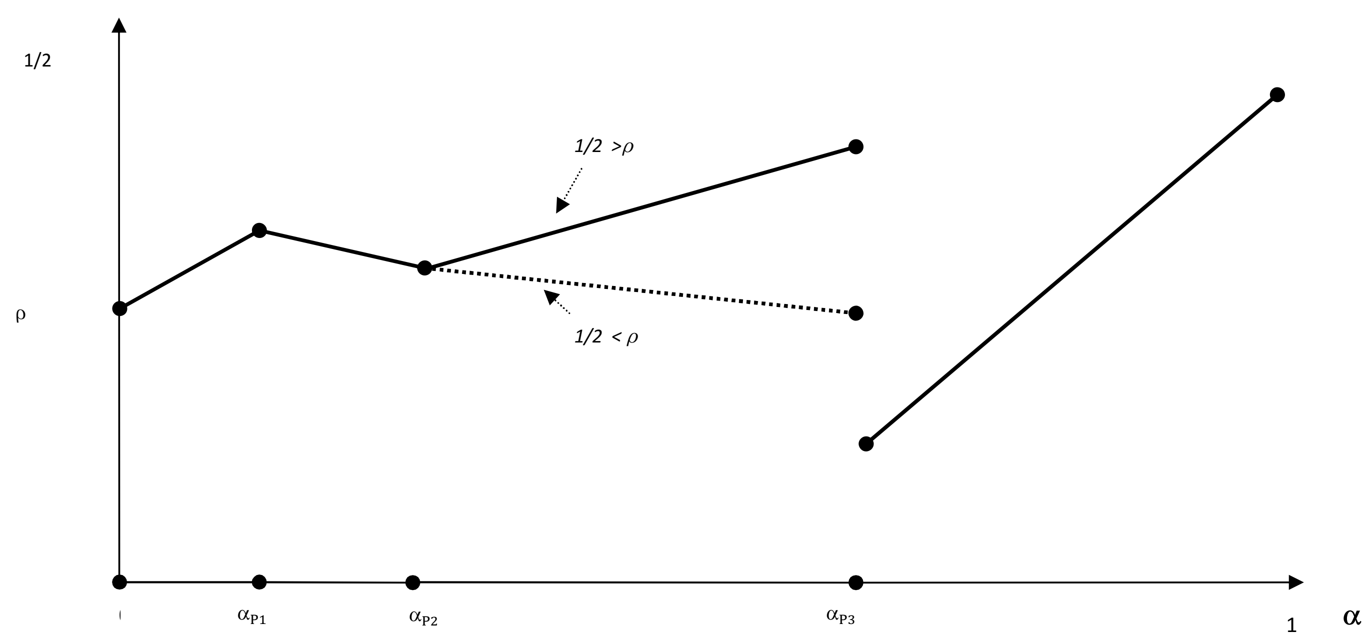


Figure 5

Illustration of our assumptions on the distribution of traders in a discrete case

We assume that investors have potential access to a N=6 markets. The scale of an institution is defined by the number of markets to which it can participate. Namely, an institution of type t can participate to $n(t) \leq N$ markets (or has n(t) trading opportunities) and n(t) increases in t, i.e., a higher value of t corresponds to a bigger institution. Thus, we will refer to t as the size of an institution. Institutions' sizes are distributed over $[t_{..}, t^{..}]$ with a frequency f(t) such that: $f(t) = N/n(t)$.
Thus, the total number of trading opportunities faced by all type t institutions is exactly equal to the number of markets. We assume that these opportunities are uniformly distributed across markets so that in each market there is an equal proportion of each type of trader. In other words, within each market, investors' types have a uniform distribution.

N=6 markets

Institutions which size is $t^{..}$ such that $n(t^{..})=6$. Frequency: $f(t^{..})=6/6=1$.

Institutions which size is t_1 such that $n(t_1)=3$. Frequency : $f(t_1)=6/3=2$.

Institutions which size is t_2 such that $n(t_2)=2$. Frequency : $f(t_2)=6/2=3$.

Institutions which size is $t_{..}$ such that $n(t_{..})=1$. Frequency : $f(t)=6/1=6$.

In each market, there is an equal number of 4 institutions.

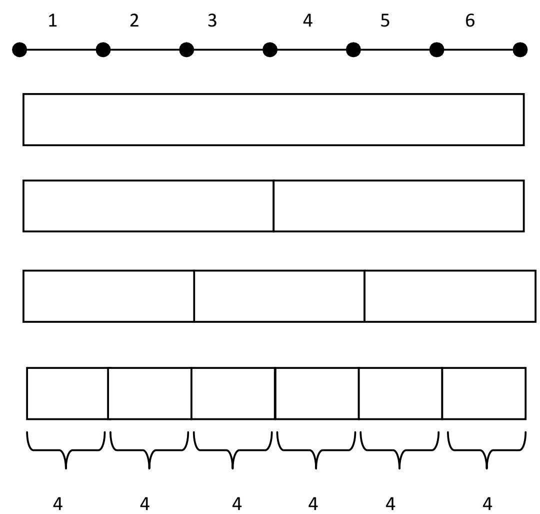
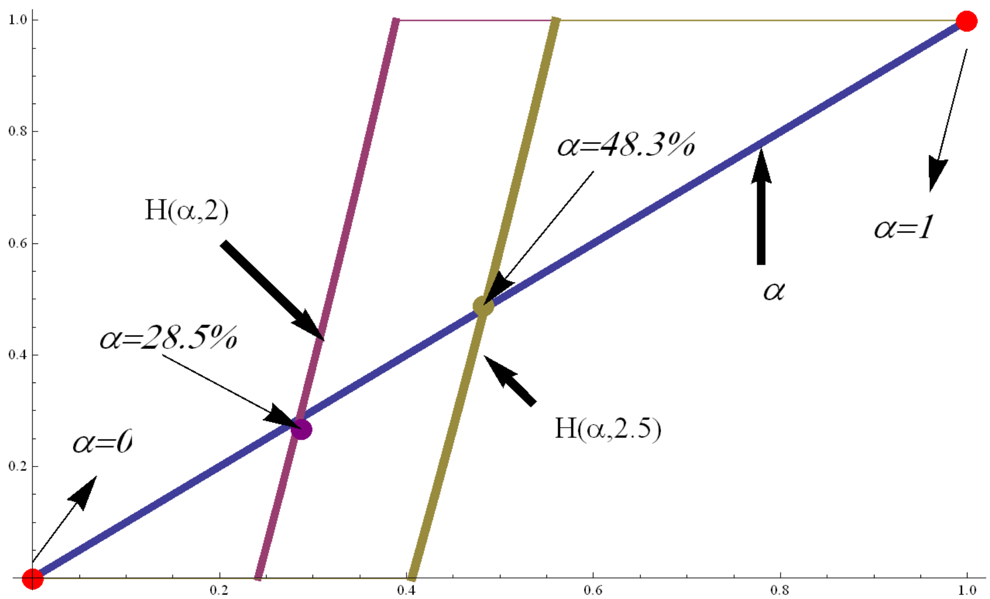


Figure 6

Panel A



Panel B