Equilibrium High Frequency Trading

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

October 20, 2011

1This 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.
Abstract

High Frequency Trading (HFT) enables investors 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

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. While some trading algorithms can operate at low frequency, High Frequency Trading (hereafter HFT) now accounts for a large fraction of total activity in financial markets. 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.

Is the growth of HFT socially valuable? Does it improve the workings of markets? Or does it benefit some investors at the expense of others? Is there excessive investment in this business line? Is policy intervention called for? The goal of this paper is to offer a theoretical framework to shed some light on these issues.

The next section presents our model. Its three key ingredients are the following:

- Computers can mitigate the imperfection of traders’ cognition, helping them to locate and seize trading opportunities faster and more effectively. Thus, HFT can generate gains from trade.¹

- High Frequency traders can trade upon new information faster than slow traders. Empirically, Hendershott and Riordan (2010), Brogaard (2010), Kirilenko et al (2010), and Hendershott and Riordan (2011) find that high frequency traders’ orders are better informed than other orders. That high frequency traders have access to advance information generates adverse selection costs for the other traders.²

- HFT involves significant fixed investments. One must acquire the hardware, obtain ultra-rapid connections to the exchanges’ trading systems (via, e.g., colocation, whereby the High Frequency trader locates its computer just next to that of the exchange), develop and maintain codes, hire highly qualified personnel,³ and subscribe to expensive real time data

¹See e.g. the theoretical analyses of Biais, Hombert and Weill (2007), Foucault, Kadan and Kandel (2010) and the empirical findings of Hendershott, Jones and Menkveld (2010) and Hendershott and Riordan (2010).

²In 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.

³For example, in its East Setauket office, one of the leading HFT funds, Renaissance, employs around a hundred Ph.Ds in mathematics, physics, computer science, statistics, etc...
feed. Market participants with large trading volume are in a better position than small investors to recoup the cost of these investments.

In Section 3, we analyze equilibrium transactions for a given fraction of high-frequency traders. An increase in HFT has two effects. On one hand it increases the fraction of investors that find a trading counterparty. This can raise trading volume and gains from trade. On the other hand, it increases information asymmetries in the market, and, in particular, raises adverse selection costs for slow traders. This negative externality can reduce overall volume and gains from trade. When there is a lot of HFT, slow traders are crowded out of the market. We also find that, for a given level of HFT, there can be multiple equilibria. 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.)

In Section 4, we turn to the equilibrium determination of the aggregate level of HFT. We consider a population of financial institutions, differing in terms of size. We show there is a critical size above which financial institutions choose to incur the fixed investment cost necessary to engage in HFT. Thus, in equilibrium there is a non-level playing field: a few actively trading fast institutions coexist with smaller, slower and less active institutions incurring adverse selection costs. We also show that the decisions to invest in the HFT technology can be strategic complements. As more and more traders invest in HFT, this raises the cost of remaining slow, and thus increases the incentive to also opt for HFT. In contrast, if many institutions remain slow this reduces the incentives to become fast. This is in line with the externality first identified by Admati and Pfleiderer (1984) who observe that the value of information for one trader depends on the amount and quality of information other agents possess. Admati and Pfleiderer (1984) show this generates an externality. In our framework it gives rise to multiplicity in the equilibrium level of investment in the HFT technology. And some of these equilibria are characterized by individually rational but socially inefficient investment waves in that technology.

Section 5 complements this analysis by investigating welfare issues. Because HFT generates

---

4 On July 4, 2007, an article in the International Herald Tribune, entitled "Citigroup to expand electronic trading capabilities by buying Automated Trading Desk," wrote "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."

5 This is in line with the comments quoted in the IOSCO consultation report of July 2011, entitled “Regulatory Issues Raised by the Impact of Technological Changes on Market Integrity and Efficiency.” On page 10 of this report one can read: "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."

6 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.

7 This is in line with the empirical observation that 2% of the 20,000 trading firms operating in the US markets account for 73% of trading volume (Iati, 2009, TAAB Group.)

8 Thus investment in HFT can be compared an arm’s race, in line with the analysis of financial expertise as an arms race by Glode, Green and Lowery (2011).
negative externalities, equilibrium investment in this business line typically exceeds its utilitarian–welfare maximizing level. Thus there is excessive investment in HFT and this suboptimal allocation of resources leads to suboptimal workings of markets.

Section 6 concludes and discusses some policy implications of our analysis. Proofs not given in the text are in the appendix.

2 Model

Consider a unit mass continuum of risk–neutral, profit maximizing financial institutions. Until Section 4 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.\textsuperscript{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 $C$.\textsuperscript{10} We refer to these players as fast institutions and the denote the fraction of institutions that are fast by $\alpha$. The remaining fraction is referred to as slow institutions. HFT technology helps fast institutions in two ways.

• 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.

\textsuperscript{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).

\textsuperscript{10}This decision is endogenized in Section 4.
• 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: quotes for the same security are posted in various trading venues.\textsuperscript{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.\textsuperscript{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. 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)).

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

• At $\tau = 0$, institutions decide whether to pay $C$ and become fast or not.

• At $\tau = 1$,

1. Each institution observes its private valuation $\delta$ 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 $\omega$ at price $E(v|\omega)$.

• At $\tau = 2$, $v$ is realized.

\textsuperscript{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.

\textsuperscript{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.
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 ($L$).

3 Price formation and trading with fast and slow investors

This section analyzes equilibrium transaction prices and trading volume, for a given level of $\alpha$. 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 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.

3.1 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 $\mu$. 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, $\rho$.

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 $\beta^F_j$ the probability that a fast institution $j \in \{GH, GL, BH, BL\}$ buys, and by $\beta^S_j$ 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., $\beta^F_{GH} = 1$. Symmetrically, fast institutions with bad news and low private valuation never buy, i.e., $\beta^F_{BL} = 0$. Hence, by Bayes law, buy orders execute at price

$$a = E(v | \text{buy}) = \mu + \frac{\alpha(1 + \beta^F_{GL} - \beta^F_{BH})}{2((1 - \alpha)\rho(\beta^S_H + \beta^S_L) + \frac{\alpha}{2}(1 + \beta^F_{GL} + \beta^F_{BH}))}\epsilon. \quad (1)$$

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 $\mu$. Since $a \geq \mu$, institutions with a low private valuation never buy, even if they are fast and receive a good signal, i.e., $\beta^F_{GL} = \beta^S_L = 0$. 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 $\mu$, 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 in the benchmark case 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 our 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 ($\frac{\epsilon}{2} < \delta$). 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

\footnote{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.}
private value. Equation (3) implies there are 5 different possible equilibria, corresponding to increasingly high ask prices. In all equilibria, $\beta_{GH}^F = 1$ and $\beta_{BL}^F = \beta_L^S = 0$, as mentioned above. Furthermore, $\beta_{BH}^F = 0$ 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, $\beta_{GL}^F = 1$ and $\beta_H^S = 1$.

- **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., $\beta_H^S = 1$. Fast institutions with good news but low private value are indifferent between buying and not trading. They play mixed strategies, buying with probability $0 \leq \beta_{GL}^F \leq 1$.

- **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., $\beta_{GL}^F = 0$. Slow institutions with high private value buy, i.e., $\beta_H^S = 1$.

- **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., $\beta_{GL}^F = 0$. Slow institutions with high private value are indifferent between buying or not trading. They play a mixed strategy, buying with probability $\beta_H^S \in [0, 1]$.

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

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 our assumptions (2) and $\delta < \epsilon$, 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:

$$
\alpha_{P1} = \frac{\rho(\epsilon - \delta)}{\rho(\epsilon - \delta) + \delta}, \quad \alpha_{P2} = \frac{\rho(\epsilon - \delta)}{\rho(\epsilon - \delta) + \frac{\delta}{2}}, \quad \alpha_{P3} = \frac{2\rho\delta}{2\rho\delta + \epsilon - \delta}.
$$
Relying on these notations, and noting that $\alpha P_1 < \alpha P_2 < \alpha P_3$, we state our next proposition.

**Proposition 3**

1. If $0 < \alpha \leq \alpha P_3$ there exists an equilibrium of type M2, in which $\beta^S_{II} = \frac{\alpha}{2(1-\alpha)\rho}(\epsilon - \delta)$.

2. If $\alpha < \alpha P_1$, there exists an equilibrium of type P1, in which $a = \mu + \frac{\alpha}{\alpha + (1-\alpha)\rho} \epsilon$.

3. If $\alpha P_1 \leq \alpha \leq \alpha P_2$, there exists an equilibrium of type M1, in which $\beta^E_{GL} = \frac{2(1-\alpha)\rho}{\alpha}(\epsilon - \delta) - 1$.

4. If $\alpha P_2 < \alpha < \alpha P_3$, 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 $0 < \alpha < \alpha P_3$ 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 1** For each value of $\alpha$, there is a unique Pareto dominant equilibrium: P1 when $0 < \alpha < \alpha P_1$, M1 when $\alpha P_1 \leq \alpha \leq \alpha P_2$, P2 when $\alpha P_2 < \alpha < \alpha P_3$, M2 when $\alpha = \alpha P_3$, and P3 when $\alpha > \alpha P_3$.

Hereafter, for each value of $\alpha$, 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 $\alpha$. In our framework, this is a measure of the informativeness of trades. It weakly increases in $\alpha$ 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 $\alpha > \alpha P_3$).\(^{14}\) Let $\psi(\alpha)$ and $\phi(\alpha)$ be the expected gain for slow and fast institutions respectively. Using Proposition 3, we obtain\(^{15}\)

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

(4)

and

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

(5)

which yields the following corollary.

\(^{14}\)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 High Frequency Trading easier on the NYSE. Note however that the change considered in Hendershott et al.(2011) may also have helped slow traders to better find trading opportunities (an increase in $\rho$). Our model predicts that for a fixed value of $\alpha$, the informational impact of trades declines in $\rho$.

\(^{15}\)A derivation of $\phi(\alpha)$ and $\psi(\alpha)$ is given in the proof of Lemma 1. The case where $\alpha = \alpha P_3$ and where the Pareto-dominant equilibrium is M2 is actually a limit case of $\alpha P_2 < \alpha < \alpha P_3$. 

8
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 $\alpha$. 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, 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 these externalities 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 5.

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

Corollary 2 Equilibrium trading volume is:

$$\text{Vol}(\alpha) = \begin{cases} 
\alpha + (1 - \alpha)\rho & \text{for } 0 \leq \alpha < \alpha_P1, \\
\frac{\alpha - \alpha P1}{\rho} & \text{for } \alpha_P1 \leq \alpha \leq \alpha_P2, \\
\frac{\alpha - \alpha P2}{\rho} & \text{for } \alpha_P2 < \alpha \leq \alpha_P3, \\
\frac{\alpha - \alpha P3}{\rho} & \text{for } \alpha > \alpha_P3.
\end{cases}$$

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 $\alpha$, as illustrated in Figure 4. When $\alpha$ is very low, adverse selection is limited and the main effect of an increase in $\alpha$ is to increase the probability that an institution finds a counterparty. Furthermore, when $\alpha$ is very large, most institutions participating in trading are fast, and an increase in $\alpha$ increases total trading volume. Therefore, when there is either little
HFT or a lot of it, trading volume is increasing in $\alpha$. In contrast, for intermediate values of $\alpha$, trading volume can decrease in the level of high frequency trading. 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).\footnote{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 $\alpha_{P1} \leq \alpha \leq \alpha_{P2}$.} There is a discrete drop in trading when $\alpha$ increases beyond $\alpha_{P3}$, due to the fact that at this point slow institutions stop trading. More precisely, a small increase in $\alpha$ at $\alpha = \alpha_{P3}$ implies that trading volume drops from $\frac{\alpha_{P2}}{2} + (1 - \alpha_{P3})\rho$ 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.\footnote{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.}

4 Scope of High Frequency Trading

In the previous section, the “scope of High Frequency Trading”, $\alpha$, was exogenous. We now study its equilibrium determination.

4.1 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 $f(t)$ such that:

$$f(t) = \frac{N}{n(t)}, \quad (7)$$
$f(t)$ is decreasing, which 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, $n(t)$.

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: 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 ($n(t_1) = 1$). 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 $t$. In this context, there is exactly one trader of each type in each market (to build intuition, Figure 5 illustrates this setting in a simple discrete case). Consequently, as mentioned above, that one trader is active in a market does not convey any information about its size.

### 4.2 Investing in High Frequency Trading

For a given level of HFT, $\alpha$, 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:

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

where $\phi$ and $\psi$ 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

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

---

18 For an application of Zipf’s law in finance and economics see Gabaix and Landier (2003).

19 Also, in keeping with Zipf’s law, the most frequent type of trader ($t$) “occurs” twice as often as the second most frequent type ($t_2$), and three times as often as the third most frequent type ($t_1$) and six times most often as the less frequent type ($t$).
or, as \( n(.) \) is increasing,

\[
t \geq n^{-1}\left(\frac{C}{\phi(\alpha) - \psi(\alpha)}\right).
\]

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

\[
t^*(\alpha) = n^{-1}\left(\frac{C}{\phi(\alpha) - \psi(\alpha)}\right),
\]

we obtain the following result.

**Lemma 2** 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, \( \alpha \). As this level depends on other institutions’ choices, institutions’ technological choices are interdependent. The logic is the following: An increase in \( \alpha \) 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 2 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 \( \alpha \), then fast investors lose more than slow ones when \( \alpha \) goes up. Hence \( t^*(\alpha) \) increases in \( \alpha \), 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 \( \alpha \), that is, slow investors are hurt more than fast investors by an increase in \( \alpha \). Consequently, \( t^*(\alpha) \) is decreasing in \( \alpha \), 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 \( \alpha \) 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 \( \alpha \). When \( \alpha_{P2} < \alpha < \alpha_{P3} \), \( \phi(\alpha) - \psi(\alpha) \) is increasing in \( \alpha \) 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 $P1$ and $P2$. When $0 \leq \alpha < \alpha_{P1}$ and $P1$ prevails, then slow traders buy if and only if they have located a trading opportunity and their private valuation is high so that their expected profit is:

$$\psi(\alpha) = \Pr(+\delta)\rho(\mu + \delta - a(\alpha)),$$

where $a$ is written as a function of $\alpha$ to emphasize that the price impact of trades depends on the fraction of fast investors. In this equilibrium, fast traders buy if and only if they have observed good news about $v$ and their expected profit is:

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

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

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

This is positive if and only if $\rho > \frac{1}{2}$. In that case, as $\rho$ 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 $\alpha$, and show that when the decisions to invest in HFT are strategic complements equilibrium multiplicity can arise.

### 4.3 Corner equilibria

Denote by $\alpha^*$ 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:

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

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

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

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

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

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

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

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

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

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

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

**Proposition 7** If $\rho > \rho^*$, then $C_{\text{max}} < C_{\text{min}}$ and if $C \in (C_{\text{max}}, C_{\text{min}})$ 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 Lemma ??, when $\rho > \frac{1}{2}$, investments in HFT are strategic complement, and, hen $\rho > \rho^* > \frac{1}{2}$, 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.
4.4 Interior equilibria

So far we derived conditions for corner equilibria, we now consider interior equilibria. From Lemma 2, when market participants expect that a fraction $\alpha$ 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

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

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

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

(14) there exists an interior equilibrium. (14) states that, when institutions expect a fraction $\alpha^*$ to invest in High Frequency Trading, then $\alpha^*$ 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 algorithmic 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.

For tractability, we hereafter assume that the density of $t$ is linear, and, more specifically, that

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

(15)

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 $N$. Let $\kappa = \frac{\Delta t}{2(1 - \Delta t)}$. Denoting

$$C = \frac{N \Delta t}{2(1 - \Delta t)(\alpha P_3 + \kappa)} \delta,$$

15
and

$$\mathcal{C} = \frac{N \Delta t}{2(1 - \Delta t)(\alpha_{P3} + \kappa)} \varepsilon.$$

we obtain the following proposition.

**Proposition 8** If $\rho \leq 1/2$, then $C_{\text{max}} > \mathcal{C} > \mathcal{C} > C_{\text{min}}$ and

1. If $C \in (C_{\text{min}}, C_{\text{max}}) \setminus (\mathcal{C}, \mathcal{C})$, there exists a unique equilibrium fraction of fast institutions $0 < \alpha^* < 1$, solving (14). In equilibrium, the level of high frequency trading, $\alpha^*$, decreases with $C$.

2. If $C \in (\mathcal{C}, \mathcal{C})$, there is no solution to (14).

3. If $C \leq C_{\text{min}}$ then $\alpha^* = 1$ and if $C \geq C_{\text{max}}$ then $\alpha^* = 0$.\(^{20}\)

When $\rho > 1/2$, if $C \geq C_{\text{max}}$ then $\alpha^* = 0$, while if $C < C_{\text{max}}$, there can be multiple equilibria, but $\alpha^* > 0$ in any equilibrium.

First consider the case $\rho \leq 1/2$, in which $C_{\text{min}} < C_{\text{max}}$. Consistent with intuition, if the cost of the HFT technology is very large then all institutions choose to remain slow, while if this cost is very low all institutions choose to become fast.

- For intermediary values of $C$ in $(C_{\text{min}}, C_{\text{max}}) \setminus (\mathcal{C}, \mathcal{C})$, there exists an interior equilibrium $\alpha^* \in (0, 1)$. In line with intuition, this equilibrium fraction of fast institutions increases as the the cost of HFT declines from $C_{\text{max}}$ to $\mathcal{C}$. Improvements in internet and communication technology, are likely to have induced a decline in $C$, leading to an increase in HFT.

- At $C = \mathcal{C}$, the equilibrium fraction of fast institutions is $\alpha^* = \alpha_{P3}$. Beyond that point, the benefits of being fast experience a discontinuous downward jump, as slow institutions exit the market and liquidity drops (See Figure 3). This discontinuity precludes existence of a solution to the fixed–point problem (14) as long as $C \in [\mathcal{C}, \mathcal{C})$. Inexistence is driven by the simultaneity of investment decisions. Suppose instead institutions reach decisions one after another, starting with the highest type ($t$) and ending with the lowest one ($t$). When making that choice, each institution observes previous decisions and rationally anticipates the decisions that will be taken afterwards. In that case, for $C \in [\mathcal{C}, \mathcal{C})$, the only subgame perfect equilibrium is such that all institutions with $t \geq t^*(\alpha_{P3})$ invest in HFT whereas smaller institutions don’t.\(^{21}\) To see this, suppose that $C \in (\mathcal{C}, \mathcal{C})$. In this case, institutions

\(^{20}\)Cases where $C = C_{\text{min}}$ and $C = C_{\text{max}}$ are derived as limit cases of Proposition 8 Part 1.

\(^{21}\)For other values of $C$, the unique subgame perfect equilibrium of this sequential investment game is identical to that obtained when institutions make their investment decision simultaneously.
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. 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. Hence institutions smaller than \( t^*(\alpha_{P3}) \) choose not to invest in HFT.

Second turn to the case \( \rho > 1/2 \). Similarly to the case where \( \rho \leq 1/2 \), when \( C \) is very large (above \( C_{\text{max}} \)) then no institution invests in the HFT technology, while, if \( C \) is lower, some institutions do. In contrast with the case \( \rho \leq 1/2 \), there can be multiple equilibria. The intuition for such multiplicity is similar to that of Proposition 7. As stated in Lemma ??, since \( \rho > 1/2 \) the decisions to invest in HFT are strategic complements. This implies that, for the same parameter values one can have both an equilibrium with little HFT and an equilibrium with lots of HFT.\(^\text{22}\) 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.\(^\text{23}\)

5 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

\[
\alpha = \frac{\tilde{t} - t^*}{\tilde{t} - \underline{t}}.
\]

Utilitarian welfare is:

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

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

\[
W(\alpha) = N\left[ \int_{\underline{t}}^{t^*} \psi(\alpha)dt + \int_{t^*}^{\tilde{t}} \phi(\alpha)dt \right] - C(1 - F(t^*)).
\]

That is:

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

\(^\text{22}\)For instance, suppose that \( \epsilon = 1, \delta = 0.9, \Delta t = 0.9, C = 4.145 \) and \( N = 10 \). If \( \rho = 0.6 \) then we have \( C_{\text{max}} = 4.6 \). Using equation (14), one can check that there are two possible HFT equilibrium levels: \( \alpha^* \simeq 6.5\% \) and \( \alpha^* \simeq 85\% \). In contrast if \( C = 3.8 \), the equilibrium is unique and equal to \( \alpha^* = 1 \).

\(^\text{23}\)Consider the same parameter values as in the previous footnote and suppose that the cost of HFT decreases from \( C = 4.145 \) to \( C = 4.1 \). There are again two possible equilibria: \( \alpha^* \simeq 10\% \) and \( \alpha^* \simeq 50\% \) when \( C = 4.1 \). If \( \alpha^* \) was 6.5\% when \( C = 4.145 \) then the level of algorithmic increases when \( C \) drops to 4.1, as one would expect. However if \( \alpha^* \) was 85\% when \( C = 4.145 \) then this conclusion does not hold.
Since, by (16), \( t^* = \bar{t} - \alpha(\bar{t} - t) \), (17) rewrites as:

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

Maximizing \( W(\alpha) \) and comparing the solution to the equilibrium fraction of fast investors, \( \alpha^* \) 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., \( \frac{\partial W}{\partial \alpha} \big|_{\alpha=\alpha^*} \leq 0 \), with a strict inequality for some parameter values.

Consider an interior equilibrium \( \alpha^* \in (0, 1) \).\(^{24}\) 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, for \( \rho > 1/2 \) one obtains a stronger, global, result, as stated in the next proposition.

**Proposition 10** If \( \rho > 1/2 \), the level of investment in High Frequency Trading maximizing utilitarian welfare is \( \alpha = 0 \).

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 \( C < C_{\text{max}} \). Hence, for \( C < C_{\text{max}} \) and \( \rho > 1/2 \), equilibrium HFT is always excessive. In contrast, when \( \rho \leq 1/2 \), some investment in HFT can be socially optimal.\(^{25}\)

---

\(^{24}\)Proposition 9 does not cover the case \( \alpha^* = 1 \). When \( \rho < 1/2 \), \( \alpha^* = 1 \) if and only if \( C < C_{\text{min}} \). In general, the socially optimal level of High Frequency Trading will be smaller than 1 unless \( C \) is sufficiently small relative to \( C_{\text{min}} \).

\(^{25}\)For instance, consider the following numerical example: \( \epsilon = 1, \delta = 0.9, \Delta t = 0.9, \rho = 0.3 \) and \( N = 10 \). In this case \( C_{\text{min}} = 3.68 \). 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 \( \alpha_{\text{max}} = 0.562 \).
6 Conclusion

While HFT can increase gains from trade, it can also generate adverse selection. Because of the negative externality it thus generates, HFT will attract an equilibrium amount of investment exceeding the utilitarian optimum. This suggests that Pigovian taxes, such as, for example, taxes on colocation, could improve utilitarian welfare.

Other costs of HFT, related to operational and systemic risk, are outside the scope of this paper. Yet, these risks might be quite significant, as suggested by the recent report of the SEC and CFTC on the flash-crash of May 6, 2010. They underscore the need to further study HFT, both theoretically and empirically.
Appendix: Proofs

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

\[
\begin{align*}
\Pi_{GH}^F &= (\mu + \epsilon + \delta - a), \\
\Pi_{BH}^F &= (\mu - \epsilon + \delta - a), \\
\Pi_{GL}^F &= (\mu + \epsilon - \delta - a), \\
\Pi_{BL}^F &= (\mu - \epsilon - \delta - a).
\end{align*}
\]

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

\[
\begin{align*}
\Pi_L^S &= (\mu - \delta - a) \\
\Pi_H^S &= (\mu + \delta - a)
\end{align*}
\]

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 $\mu$. 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, $\beta_L^S = 0$, $\beta_{GH}^F = 1$ and $\beta_{BL}^F = 0$. 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.\[ \blacksquare \]

Proof of Proposition 1. Provided that $\mu \leq a \leq \mu + \epsilon$, $\beta_{BL}^F = 0$ and $\beta_{GH}^F = 1$ 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 $\beta_{BH}^F = \beta_H^S = 1$ and $\beta_{GL}^F = \beta_L^S = 0$. If institutions behave in this way, using equation (1), we deduce that $a = \mu$.\[ \blacksquare \]

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 \( \beta_L^S = 0, \beta_H^S = \beta_{GL}^F = \beta_{BH}^F = 0 \). 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 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: \( \beta_H^S = \frac{\alpha}{2(1-\alpha)\rho \delta}(\epsilon - \delta) \). Hence \( \beta_L^S = 0, \beta_{GL}^F = \beta_{BH}^F = 0, \) and \( \beta_H^S = \frac{\alpha}{2(1-\alpha)\rho \delta}(\epsilon - \delta) \). Then using equation (1), we deduce that \( a = \mu + \delta \). We define as \( \alpha_{P_3} \) the threshold such that \( \beta_H^S \leq 1 \) for \( \alpha \leq \alpha_{P_3} \).

**Part 2:** \( \alpha < \alpha_{P_1} \). Suppose that that institutions expect buy orders to execute at \( a = \mu + \frac{\alpha}{\alpha+(1-\alpha)\rho} \epsilon \). We define as \( \alpha_{P_1} \) the threshold such that \( a \geq \mu + \epsilon + \delta \) for \( \alpha \geq \alpha_{P_1} \). Thus, as \( \alpha < \alpha_{P_1}, a < \mu + \epsilon + \delta : \) 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:** \( \alpha_{P_1} \leq \alpha \leq \alpha_{P_2} \). 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 \( \alpha_{P_2} \) the threshold such that \( \beta_{GL}^F \geq 0 \) for \( \alpha \leq \alpha_{P_2} \), and we check that indeed \( \beta_{GL}^F \leq 1 \) for \( \alpha \geq \alpha_{P_1} \).

**Part 4:** \( \alpha_{P_2} < \alpha < \alpha_{P_3} \). Suppose that institutions expect buy orders to execute at price \( a = \frac{\alpha}{\alpha+2(1-\alpha)\rho} \epsilon \). As \( \alpha_{P_3} < \alpha, a < \mu + \delta \). 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 1.** 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 \( H(\alpha) \) be this expected profit. A necessary condition for fast institutions to buy the asset is that they receive good news. Hence given that \( \beta_{GH}^F = 1 \)

\[
H(\alpha) = \frac{1}{2}(\mu + \epsilon + \delta - a) + \frac{1}{2}(\mu + \epsilon - \delta - a) \times \beta_{GL}^F.
\]

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 \( \beta_{GL}^F \) in the various types of equilibria, we obtain:

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

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 \( 0 \leq \alpha < \alpha_{P1}, \frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho} \epsilon > \epsilon/2 \).
- In the region where a type \( M1 \) equilibrium exists, that is, for \( \alpha_{P1} \leq \alpha \leq \alpha_{P2}, \delta > \epsilon/2 \).
- In the region where a type \( P2 \) equilibrium exists, that is, for \( \alpha_{P2} \leq \alpha < \alpha_{P3}, \frac{1}{2}(\delta + \frac{(1-\alpha)\rho}{\alpha/2+(1-\alpha)\rho} \epsilon) > \epsilon/2 \).

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 \( \alpha = \alpha_{P3} \), fast institutions’ expected profit in types \( M2 \) equilibrium is higher than in an equilibrium of type \( P3 \).

Now let:

\[
\phi(\alpha) = \begin{cases} 
\frac{(1-\alpha)\rho}{\alpha+(1-\alpha)\rho} \epsilon & \text{for } 0 \leq \alpha < \alpha_{P1}, \\
\delta & \text{for } \alpha_{P1} \leq \alpha \leq \alpha_{P2}, \\
\frac{1}{2}(\delta + \frac{(1-\alpha)\rho}{\alpha/2+(1-\alpha)\rho} \epsilon) & \text{for } \alpha_{P2} < \alpha < \alpha_{P3}, \\
\frac{\epsilon}{\delta/2} & \text{for } \alpha > \alpha_{P3}.
\end{cases}
\]
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 \( \lim_{\eta \to 0} \phi(\alpha P_3 - \eta) = \xi = \phi(\alpha P_3) \). Thus, \( \phi(.) \) is continuous over \( [0, \alpha P_3] \). It is then immediate that \( \phi(.) \) decreases over \( [0, \alpha P_3] \).

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 \( \alpha P_1 \leq \alpha < \alpha P_2 \), 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, \( H(\alpha) = \phi(\alpha) \) for \( 0 \leq \alpha < \alpha P_1 \); in a type M1 equilibrium, \( H(\alpha) = \phi(\alpha) \) for \( \alpha P_1 \leq \alpha < \alpha P_2 \); in a type P2 equilibrium, \( H(\alpha) = \phi(\alpha) \) for \( \alpha P_2 < \alpha < \alpha P_3 \); and in a type M2 equilibrium, \( H(\alpha) = \phi(\alpha) \) 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 valuation. 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 \( \rho \), (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 it sells the asset. We focus on \( \alpha < \alpha P_3 \), since for \( \alpha \geq \alpha P_3 \), the unique equilibrium is a type P3 equilibrium. Using the expression for the equilibrium value of \( \alpha \) and \( \beta_H^S \) in the various types of equilibria, we obtain:

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

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 \( \alpha \).

Last, using the expressions for \( H(\alpha) \) 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

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

\[\square\]

**Proof of Corollary 1.** We have already proved in the proof of Lemma 1 that \(\phi(\alpha)\) decreases in \(\alpha\) and is discontinuous at \(\alpha = \alpha_{P_3}\). Now consider the slow institutions. Obviously, \(\psi(\alpha)\) decreases in \(\alpha\) over the following intervals \(\alpha \in [0, \alpha_{P_1})\) and \(\alpha \in (\alpha_{P_2}, \alpha_{P_3})\). Otherwise its is constant. Moreover calculations yield

\[
\psi(\alpha_{P_1}) = (2\delta - \epsilon) \rho.
\]

\[
\lim_{\eta \to 0} \psi(\alpha_{P_1} + \eta) = (2\delta - \epsilon) \rho
\]

\[
\lim_{\eta \to 0} \psi(\alpha_{P_1} - \eta) = 0
\]

\[
\lim_{\eta \to 0} \psi(\alpha_{P_1} + \eta) = 0
\]

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

**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 likely in equilibrium). Thus

\[Vol(\alpha) = 2\left(\frac{\alpha}{2}(1 + \beta_{GL}^E) + (1 - \alpha)\rho \beta_{H}^S\right).
\]

We can then obtain the expressions for the trading volume in the Pareto dominant equilibrium by replacing \(\beta_{GL}^E\) and \(\beta_{H}^S\) by their expression for each possible value of \(\alpha\). For instance, if \(\alpha_{P_1} \leq \alpha \leq \alpha_{P_2}, \beta_{H}^S = 1\) and \(\beta_{GL}^E = \frac{2(1-\alpha)\rho}{\alpha\delta}(\epsilon - \delta) - 1\). Hence:

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

\[\square\]

**Proof of Lemma 2.** Immediate from the arguments in the text.\[\square\]

**Proof of Proposition 5.** Immediate from the arguments in the text.\[\square\]

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

**Proof of Proposition 7.** Immediate from the arguments in the text.\[\square\]
Preliminary remarks for the Proposition 8. As explained in the text, if there is an interior equilibrium with fast and slow institutions then the fraction of fast institutions, \( \alpha^* \), solves:

\[
\alpha^* = \frac{\bar{t} - n^{-1} \left( \frac{C}{\phi(\alpha^*) - \psi(\alpha^*)} \right)}{\bar{t} - \bar{t}},
\]

or \( G(\alpha^*, C) = 0 \) where

\[
G(\alpha, C) = \frac{\bar{t} - n^{-1} \left( \frac{C}{\phi(\alpha) - \psi(\alpha)} \right)}{\bar{t} - \bar{t}}.
\]

Note that if \( G(\alpha, C) > 0 \), the fraction of fast institutions in the market is higher than the fraction of fast institutions for which the gain of High Frequency Trading exceeds the cost. Thus, in this case, the fraction of fast institutions is too high relative to the equilibrium fraction of fast institutions. Similarly, if \( G(\alpha, C) < 0 \), the fraction of fast institutions is too small relative to the equilibrium fraction. As \( n(t) = \frac{N}{1 + b(t - \bar{t})} \) with \( b = \frac{2(1 - \Delta t)}{(\Delta t)^2} \), we obtain:

\[
G(\alpha, C) = \alpha - \frac{N}{C} \times (\phi(\alpha) - \psi(\alpha)) + \kappa.
\]

where \( \bar{C} = bC\Delta t = 2(1 - \Delta t)C/\Delta t = C/\kappa. \)

Using Equations (5) and (4), we deduce, after some algebra, that:

\[
G(\alpha, C) = \begin{cases} 
\alpha - \frac{N}{C} \times (\rho(\alpha) + \delta) + \kappa & \text{if } \alpha \in [0, \alpha_1] \\
\alpha - \frac{N}{C} \times (\delta - (2\delta - \epsilon) \times \rho) + \kappa & \text{if } \alpha \in [\alpha_1, \alpha_2] \\
\alpha - \frac{N}{C} \times (\delta(\frac{1}{2} - \rho) + \frac{\rho\epsilon}{\alpha(1-2\rho)+2\rho}) + \kappa & \text{if } \alpha \in (\alpha_2, \alpha_3] \\
\alpha - \frac{N}{C} \times (\frac{\delta}{2}) + \kappa & \text{if } \alpha \in (\alpha_3, 1]
\end{cases}
\]

It is easily checked that the piecewise function \( G(\alpha, C) \) is continuous, except at \( \alpha = \alpha_3 \), where it experiences a positive jump. Indeed, \(^{26}\)

\[
G(\alpha_3, C) = \alpha_3 - \frac{N}{C} \frac{\epsilon}{2} + \kappa,
\]

while

\[
\lim_{\eta \rightarrow 0} G(\alpha_3 + \eta, C) = \alpha_3 - \frac{N}{C} \frac{\delta}{2} + \kappa.
\]

Hence, \( G(\alpha, C) < G(\alpha_3, C) \) for \( \alpha \) higher than but sufficiently close to \( \alpha_3 \) since \( \delta < \epsilon \).

Moreover for all values of the parameters, \( G(\alpha, C) \) is increasing in \( C \). The following lemma is useful for the proof of Proposition 8.

Lemma 3
1. If \( \rho \leq 1/2 \), \( G(\alpha, C) \) increases in \( \alpha \).
2. If \( \rho > 1/2 \) and \( \alpha \notin (\alpha_2, \alpha_3] \), \( G(\alpha, C) \) increases in \( \alpha \).

\(^{26}\)To obtain this expression, observe that \( \lim_{\eta \rightarrow 0} \phi(\alpha_3 + \eta) = \frac{\delta}{2} \) and \( \lim_{\eta \rightarrow 0} \psi(\alpha_3 + \eta) = 0 \) and use equation (20).
3. If \( \rho > 1/2 \) and \( \alpha \in (\alpha_{P_2}, \alpha_{P_3}] \), \( G(\alpha, C) \) increases in \( \alpha \) or decreases in \( \alpha \). Thus, over this interval, it reaches its minimum either at \( \alpha = \alpha_{P_2} + \eta \) with \( \eta \to 0 \) or at \( \alpha = \alpha_{P_3} \).

The proof of lemma 3 is immediate as computations yield:

\[
\frac{\partial G(\alpha, C)}{\partial \alpha} = \begin{cases} 
1 + \frac{N \rho}{\alpha(1-\rho)} \frac{1}{2} & \text{if } \alpha \in [0, \alpha_{P_1}) \\
1 & \text{if } \alpha \in [\alpha_{P_1}, \alpha_{P_2}] \\
1 + \frac{N \rho}{\alpha(1-\rho)+\rho_2} & \text{if } \alpha \in (\alpha_{P_2}, \alpha_{P_3}] \\
1 & \text{if } \alpha \in (\alpha_{P_3}, 1]
\end{cases}
\] (22)

\[\square\]

**Proof of Proposition 8.** Let define \( C_1 \) as the value of \( C \) such that \( G(\alpha_{P_1}, C_1) = 0 \). Similarly, we define \( C_2 \), and \( \overline{C} \) such that \( G(\alpha_{P_2}, C_2) = 0 \), \( G(\alpha_{P_3}, \overline{C}) = 0 \). Using Equation (21), we obtain

\[
C_1 = \frac{N(\rho + \delta(1-2\rho))\Delta t}{2(1-\Delta t)(\alpha_{P_1} + \kappa)},
\]

\[
C_2 = \frac{N(\rho + \delta(1-2\rho))\Delta t}{2(1-\Delta t)(\alpha_{P_2} + \kappa)},
\]

\[
\overline{C} = \frac{N\Delta t}{2(1-\Delta t)(\alpha_{P_3} + \kappa) \delta}.
\]

Last, let

\[
\underline{C} = \frac{N\Delta t}{2(1-\Delta t)(\alpha_{P_3} + \kappa) \delta}.
\]

Note that \( \underline{C} \) is such that \( \lim_{\eta \to 0} G(\alpha_{P_3} + \eta, \underline{C}) = 0 \).

Notice that for all values of \( \rho \), we have \( \overline{C} > \underline{C} > C_{\min} \) and \( C_{\max} > C_1 > C_2 \).

**Part 1:** If \( \rho \leq 1/2 \), then \( C_2 \geq \overline{C} \), so we have

\[
C_{\max} > C_1 > C_2 \geq \overline{C} > \underline{C} > C_{\min}.
\]

1. If \( C > C_{\max} \), then by definition of \( C_{\max}, G(0, C) > 0 \). As \( G(., C) \) is monotonically increasing in \( \alpha \), there is no interior equilibrium, and the corner equilibrium \( \alpha^* = 0 \) given in Proposition 5 is the unique equilibrium.

Conversely, if \( C < C_{\min} \), then by definition of \( C_{\min}, G(1, C) < 0 \). As \( G(., C) \) is monotonically increasing in \( \alpha \), there is no interior equilibrium, and the corner equilibrium \( \alpha^* = 1 \) given in Proposition 6 is the unique equilibrium.

2. If \( C \in [\overline{C}, C_{\max}] \) then we have \( G(0, C) \leq 0 \) as \( C \leq C_{\max} \) and \( G(\alpha_{P_3}, C) \geq 0 \) as \( C \geq \overline{C} \). Since \( G(\alpha, C) \) is monotonically increasing in \( \alpha \), there is a unique value of \( \alpha^* \in [0, \alpha_{P_3}] \) such that \( G(\alpha^*, C) = 0 \). Hence the equilibrium fraction of fast institutions is unique. More precisely,

(a) The case where \( C = C_{\max} \) is a limit case, as \( G(0, C) = 0 \) implies that the unique value of \( \alpha^* \in [0, \alpha_{P_3}] \) such that \( G(\alpha^*, C) = 0 \) is precisely \( \alpha^* = 0 \).
Part 2: If \( C \in (C_1, C_{\text{max}}) \) then \( 0 < \alpha^* < \alpha_{P_1} \) since \( \lim_{\eta \to 0} G(\alpha_{P_1} - \eta, C) > 0 \) and \( G(0, C) < 0 \).

The equilibrium is of type P1.

(c) If \( C \in [C_2, C_1] \) then \( \alpha_{P_1} \leq \alpha^* \leq \alpha_{P_2} \) since \( G(\alpha_{P_1}, C) \leq 0 \) and \( G(\alpha_{P_2}, C) \geq 0 \).

Moreover \( \alpha^* = \frac{N}{C}(\delta - \rho(2\delta - \epsilon)) - \kappa \) and the equilibrium is of type M1.

(d) If \( C \in (C, C_2) \) then \( \alpha_{P_2} < \alpha^* < \alpha_{P_3} \) since \( G(\alpha_{P_2}, C) < 0 \) and \( G(\alpha_{P_3}, C) > 0 \). The equilibrium is of type P2.

(e) If \( C = \overline{C} \) then \( \alpha^* = \alpha_{P_3} \) since \( G(\alpha_{P_3}, C) = 0 \). The equilibrium is of type M2.

Conversely, if \( C \in [C_{\text{min}}, C] \), then we have \( \lim_{\eta \to 0} G(\alpha_{P_3} + \eta, \overline{C}) < 0 \) as \( C < \overline{C} \) and \( G(1, C) \geq 0 \) as \( C \geq C_{\text{min}} \). Since \( G(\alpha, C) \) is monotonically increasing in \( \alpha \), there is a unique value of \( \alpha^* \in (\alpha_{P_3}, 1] \) such that \( G(\alpha^*, C) = 0 \). Hence the equilibrium fraction of fast institutions is unique. More precisely,

(a) The case where \( C = C_{\text{min}} \) is a limit case, as \( G(1, C) = 0 \) implies that the unique value of \( \alpha^* \in (\alpha_{P_3}, 1] \) such that \( G(\alpha^*, C) = 0 \) is precisely \( \alpha^* = 1 \).

(b) If \( C \in (C_{\text{min}}, C) \) then \( \alpha^* < 1 \). The equilibrium is of type P3.

3. Finally if \( C \in [\overline{C}, \overline{C}] \), there is no equilibrium. Indeed, for \( \alpha \leq \alpha_{P_3}, G(\alpha, C) < G(\alpha_{P_3}, C) < 0 \) since \( C < \overline{C} \). But for \( \alpha > \alpha_{P_3} \), we have \( G(\alpha, C) > \lim_{\eta \to 0} G(\alpha_{P_3} + \eta, \overline{C}) \geq 0 \) since \( C \geq \overline{C} \). Hence there is no equilibrium fraction of fast institutions in the market.

Notice that in equilibrium, \( \alpha^* \) solves \( G(\alpha^*, C) = 0 \). As \( G \) increases in \( \alpha \) and in \( C \), we deduce that \( \frac{\partial \alpha^*}{\partial C} < 0 \).

Part 2: If \( \rho > \rho^* \). We know that when \( \rho > \rho^* \), \( C_{\text{min}} > C_{\text{max}} \). Thus, we have:

\[
\overline{C} > C > C_{\text{min}} > C_{\text{max}} > C_1 > C_2
\]

Observe first that if \( \overline{C} \geq C > C_2 \) and \( G(.) \) decreases over \( (\alpha_{P_2}, \alpha_{P_3}) \), then there is always an equilibrium with \( \alpha^* \in (\alpha_{P_2}, \alpha_{P_3}) \). Indeed, in this case, we have \( G(\alpha_{P_2}, C) > G(\alpha_{P_2}, C_2) = 0 \) and \( G(\alpha_{P_3}, C) \leq G(\alpha_{P_3}, \overline{C}) = 0 \). As \( G(.) \) is continuous over the interval \( (\alpha_{P_2}, \alpha_{P_3}) \), we deduce that there is one \( \alpha^* \in (\alpha_{P_2}, \alpha_{P_3}) \) such that \( G(\alpha^*, C) = 0 \). Let \( \alpha_0^* \) be this equilibrium. We now explore the other possible equilibria that can emerge for the different values of \( C \).

1. If \( C \geq \overline{C} \), \( \alpha^* = 0 \) is always a possible equilibrium as explained in Section 4.3 since \( C > C_{\text{max}} \).

This is the unique equilibrium. Indeed for \( C \geq \overline{C} \) and \( \alpha > \alpha_{P_3} \), \( G(\alpha, C) > G(\alpha, \overline{C}) > G(\alpha_{P_3}, \overline{C}) = 0 \). Thus there is no equilibrium with \( \alpha > \alpha_{P_3} \). Moreover, for \( \alpha \leq \alpha_{P_2} \), \( G(\alpha, C) > G(\alpha, C_{\text{max}}) > G(0, C_{\text{max}}) = 0 \). Thus there is no equilibrium with \( \alpha \leq \alpha_{P_2} \). Last, for \( \alpha_{P_3} < \alpha \leq \alpha_{P_3} \), Lemma 3 implies that either \( G(\alpha, C) > G(\alpha_{P_2}, C) > 0 \) or \( G(\alpha, C) > G(\alpha_{P_3}, C) > G(\alpha_{P_3}, \overline{C}) = 0 \). Thus, there is no equilibrium with \( \alpha_{P_2} < \alpha \leq \alpha_{P_3} \).
2. If \( \overline{C} > C \geq C \), \( \alpha^* = 0 \) is always a possible equilibrium as explained in Section 4.3 since \( C > C_{\max} \). Moreover if \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \) then \( \alpha^* = \alpha^*_0 \) is another equilibrium.

3. If \( C > C \geq C_{\min} \), \( \alpha^* = 0 \) is always a possible equilibrium. Moreover if \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \), there are two other equilibria: one for which \( \alpha^* \in (\alpha_{p2}, \alpha_{p3}] \) and one for which \( \alpha^* \in (\alpha_{p3}, 1] \). The second equilibrium exists since we have in this case \( \lim_{\eta \to 0} G(\alpha_{p3} + \eta, C) < 0 \) (since \( \overline{C} > C \)) but \( G(1, C) \geq 0 \) since \( C \geq C_{\min} \). Hence, there is one value of \( \alpha \in (\alpha_{p3}, 1] \) such that \( G(\alpha^*, C) = 0 \). Notice that when \( C > C_{\min}, \alpha^* < 1 \) since \( G(1, C) < 0 \).

4. If \( C_{\min} > C > C_{\max} \), there are always two possible equilibria: \( \alpha^* = 0 \) and \( \alpha^* = 1 \) as explained in Section 4.3. Moreover if \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \), then \( \alpha^* = \alpha^*_0 \) is another equilibrium.

5. If \( C_{\max} \geq C > C_1 \), then \( \alpha^* = 1 \) is always a possible equilibrium as explained in Section 4.3 since \( C < C_{\min} \). Moreover there is an equilibrium with \( \alpha^* \in [0, \alpha_{p1}] \) (the argument is as in Case 1 when \( \rho \leq 1/2 \)) and if \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \), then \( \alpha^* = \alpha^*_0 \) is another equilibrium.

6. If \( C_1 \geq C \geq C_2 \) then \( \alpha^* = 1 \) is always a possible equilibrium as explained in Section 4.3 since \( C < C_{\min} \). Moreover there is an equilibrium with \( \alpha^* \in [\alpha_{p1}, \alpha_{p2}] \) (the argument is as in Case 1 when \( \rho \leq 1/2 \)) and if \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \), then \( \alpha^* = \alpha^*_0 \) is another equilibrium.

7. If \( C_2 > C \) then \( \alpha^* = 1 \) is always a possible equilibrium as explained in Section 4.3. This is the unique equilibrium. Indeed if \( G(.) \) increases over \( (\alpha_{p2}, \alpha_{p3}] \), then the maximum over \( (\alpha_{p2}, 1] \) is either \( \alpha = \alpha_{p3} \) or \( \alpha = 1 \). But since \( C < C_{\min} \), we have \( G(1, C) < 0 \), and since \( C < \overline{C}, G(\alpha_{p3}, C) < 0 \). Hence, in this case \( G(\alpha, C) < 0 \) for all values of \( \alpha \). If instead \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \) then \( G(\alpha, C) < G(\alpha_{p2}, C) < G(\alpha_{p2}, C_2) = 0 \). Hence, again, \( G(\alpha, C) < 0 \) for all values of \( \alpha \). Thus, \( \alpha^* = 1 \) is the unique equilibrium.

**Part 3:** If \( \frac{1}{2} < \rho \leq \rho^* \), then \( C_{\max} > C_{\min} \). If \( C_{\max} > C_1 > C_2 > \overline{C} > C > C_{\min} \), the analysis of this case is identical to Case 1 when \( \rho \leq \frac{1}{2} \). Otherwise for some parameter values \( \overline{C} \) and/or \( \overline{C} \) may exceed \( C_2 \) or \( C_1 \) and even \( C_{\max} \). Similarly, \( C_1 \) and/or \( C_2 \) can be smaller than \( C_{\min} \). All the possible cases for the value of \( C \) can then be analyzed as when \( \rho > \rho^* \). For instance, suppose that \( \overline{C} > \overline{C} > C_{\max} > C_1 > C_2 > C_{\min} \). In this case if \( \overline{C} > C > C_2 \), then there is always an equilibrium with \( \alpha^* \in (\alpha_{p2}, \alpha_{p3}] \) if \( G(.) \) decreases over \( (\alpha_{p2}, \alpha_{p3}] \). Indeed, in this case, we have \( G(\alpha_{p2}, C) > G(\alpha_{p2}, C_2) = 0 \) and \( G(\alpha_{p3}, C) < G(\alpha_{p3}, \overline{C}) = 0 \). As \( G(.) \) is continuous over the interval \( (\alpha_{p2}, \alpha_{p3}] \), we deduce that there is one \( \alpha^* \in (\alpha_{p2}, \alpha_{p3}] \) such that \( G(\alpha^*, C) = 0 \).
Proof of Proposition 9. Using equation (18), 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

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

With our specification for \(f\),

\[
f(t) = 1 + 2 \left(\frac{\bar{t} - t}{t - \underline{t}}\right) \left(\frac{1}{t - \underline{t}} - 1\right)
\]

Consequently:

\[
f(\bar{t} - \alpha(\bar{t} - \underline{t})) = 1 + 2\alpha\left(1 - \frac{\bar{t} - \underline{t}}{t - \underline{t}}\right)
\]

Therefore:

\[
\frac{dW(\alpha)}{d\alpha} \times \frac{1}{N(\bar{t} - \underline{t})} = (1 - \alpha)\psi'(\alpha) - \psi(\alpha) + \alpha\phi'(\alpha) + \phi(\alpha) - \frac{C}{N}(1 + 2\alpha\left(1 - \frac{\bar{t} - \underline{t}}{t - \underline{t}}\right)).
\]

With our notations:

\[
\kappa = \frac{\Delta t}{2(1 - \Delta t)},
\]

and

\[
\widehat{C} = \frac{C}{\kappa},
\]

we have:

\[
\frac{dW(\alpha)}{d\alpha} = N \times (\bar{t} - \underline{t}) \times [(1 - \alpha)\psi'(\alpha) + \alpha\phi'(\alpha) - \frac{\widehat{C}}{N}(\alpha - \frac{N}{\varepsilon}(\phi(\alpha) - \psi(\alpha)) + \kappa)].
\]

Finally:

\[
W'(\alpha) = N \times (\bar{t} - \underline{t}) \times [(1 - \alpha)\psi'(\alpha) + \alpha\phi'(\alpha) - \frac{\widehat{C}}{N}G(\alpha, C)],
\]

where \(G(\alpha, C)\) is defined in equation (21). Remember that the equilibrium level of High Frequency Trading, \(\alpha^*\), is such that \(G(\alpha^*, C) = 0\) 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 \(\phi'(\alpha) < 0\). Indeed for \(0 \leq \alpha^* < \alpha_{P1}\):

\[
\phi'(\alpha^*) = \frac{\rho\varepsilon}{(\alpha(1 - \rho) + \rho)^2};
\]

\[
\psi'(\alpha^*) = \frac{\rho^2\varepsilon}{(\alpha(1 - \rho) + \rho)^2}.
\]
and for \( \alpha^* \in (\alpha_{P2}, \alpha_{P3}) \):

\[
\phi'(\alpha^*) = -\frac{\rho e}{(\alpha(1 - 2\rho) + 2\rho)^2},
\]

\[
\psi'(\alpha^*) = -\frac{2\rho^2 e}{(\alpha(1 - 2\rho) + 2\rho)^2}.
\]

Consequently, as \( G(\alpha^*) = 0 \), equation (24) implies that \( W'(\alpha^*) < 0 \) for \( \alpha^* \in (\alpha_{P2}, \alpha_{P3}) \) or \( 0 < \alpha^* < \alpha_{P1} \). In this case a reduction in \( \alpha^* \) raises social welfare.

Last observe that the set of parameters for which \( W'(\alpha^*) < 0 \) is not empty. For instance, when \( \rho \leq 1/2 \), the equilibrium is unique and we know that \( 0 < \alpha^* < \alpha_{P1} \) for \( \max > C > C_1 \) and \( \alpha^* \in (\alpha_{P2}, \alpha_{P3}) \) for \( C \in [C_1, C_2] \), where the thresholds \( \max, C_1, C_2, \overline{C} \) and \( \overline{C} \) are defined in the proof of Proposition 8.■

**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 (18) and the expressions for \( \phi(\alpha) \) and \( \psi(\alpha) \), we obtain that:

\[
W(\alpha) = \begin{cases} 
N\Delta t(1 - \alpha)\delta\rho - \Gamma(\alpha) & \text{if } \alpha \in [0, \alpha_{P1}) \\
N\Delta t((1 - \alpha)(2\delta - \epsilon)\rho + \alpha\delta) - \Gamma(\alpha) & \text{if } \alpha \in [\alpha_{P1}, \alpha_{P2}) \\
N\Delta t((1 - \alpha)\delta\rho + \alpha\delta) - \Gamma(\alpha) & \text{if } \alpha \in (\alpha_{P2}, \alpha_{P3}] \\
N\Delta t(\alpha\delta) - \Gamma(\alpha) & \text{if } \alpha \in (\alpha_{P3}, 1]
\end{cases}
\]

with \( \Gamma(\alpha) = C(1 - F(\bar{t} - \alpha(\bar{t} - \bar{\xi}))) > 0 \). Observe that \( \Gamma(\alpha) \) increases with \( \alpha \) and \( \Gamma(0) = 0 \). First, observe that \( W(\alpha) \) decreases in \( \alpha \) for \( \alpha \in [0, \alpha_{P1}) \). Thus:

\[
W(0) \geq W(\alpha) \text{ for } \alpha \in [0, \alpha_{P1}).
\]

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

\[
W(0) \geq W(\alpha) \text{ for } \alpha \in (\alpha_{P3}, 1].
\]

Moreover:

\[
W(0) - W(\alpha) = N\Delta t\alpha\delta(\rho - \frac{1}{2}) + \Gamma(\alpha) > 0 \text{ for } \alpha \in (\alpha_{P2}, \alpha_{P3}].
\]

Last

\[
W(0) - W(\alpha) = N\Delta t ((1 - \alpha)e\delta - \delta(\rho + (1 - 2\rho)\alpha)) + \Gamma(\alpha) \text{ for } \alpha \in [\alpha_{P1}, \alpha_{P2}].
\]

As \( \rho > \frac{1}{2} \) and \( \epsilon > \delta \), the first term decreases in \( \alpha \) and \( \Gamma(.) \) is always positive. Thus, a sufficient condition for \( W(0) - W(\alpha) \) to be positive for \( \alpha \in [\alpha_{P3}, \alpha_{P2}] \) is:

\[
(1 - \alpha_{P2})e\delta - \delta(\rho + (1 - 2\rho)\alpha_{P2}) > 0
\]

which is always true.■
Bibliography


