

In whose best interest? An agent-based model of High Frequency Trading

Paolo Pellizzari

Abstract We study a stylized model of High Frequency Trading in which traders equipped with private values and costs operate in a Continuous Double Auction. They can revise their orders with different frequencies and, hence, (only) some agents can repeatedly revise and resubmit orders in the same session, mimicking the behaviour of high frequency traders. All agents attempt to maximize profits, learning which bid and ask is to be posted in a given configuration of the book. We analyze the efficiency of the resulting market and the way the surplus from trading is apportioned among agents as a function of the number and type of high frequency traders. We find that the presence of a small proportion of high frequency traders increases the overall efficiency of the market; secondly, the ones who have the chance to frequently revise the offers learn to extract a disproportionate fraction of the profits that ordinarily would belong to slow traders.

1 Introduction

High Frequency Trading (HFT) broadly refers to trading strategies involving fast submission, cancellation and revision of orders in a Continuous Double Auction (CDA). Often, a massive number of orders is submitted to be left on the book for very short times and it is estimated that more than 50% of daily stocks are exchanged by HFT financial firms. There is anecdotal evidence that such firms are willing to spend hundreds of millions to increase the speed of their operations by a few milliseconds through the installation of fiber cables in the ocean or purchasing other related IT infrastructure, [Philips, 2012].

The remarkable surge of HFT in recent years was investigated to assess whether and to what extent such practices can provide liquidity to slow and more tradi-

Paolo Pellizzari
Dept. of Economics, Ca' Foscari University, Venice
e-mail: paolop@unive.it

tional financial operators, alter the volatility of the market, [Biais et al., 2014] or contribute to the destabilization of the regular functioning of exchanges in some specific instances, [Easley et al., 2011].

Clearly, there are issues worth investigating for the policy maker: HFT may improve the liquidity in the market, as the order book is flooded with thousands of bids or asks. In particular, while market orders consume liquidity, plenty of limit orders are typically interpreted as a way to provide liquidity or a way to manage trading akin to what an ordinary market-maker would do. However, it was insightfully pointed out by [Hasbrouck and Saar, 2009, Hasbrouck and Saar, 2013] that the picture is more blurred: a limit order certainly offers an option to trade to others but, if this opportunity is (only) available for a fraction of a second, it may be difficult for the “ordinary” trader to use it and, indeed, the triggered order may reach the market too late or generate other unexpected and potentially negative effects. Moreover, HFT may generate intermittent volume and suddenly reduce the number or orders that are posted, thus enlarging the spread and failing to be a reliable source of liquidity.

In this paper we describe a stylized model of HFT in which a set of agents want to buy or sell a single stock in a trading session. This can be done submitting limit orders, with no certain execution, or marketable orders that instead immediately hit the opposite side of the book and are executed at less advantageous prices. Every trader maximizes his expected profits and must decide the limit price for a single unit of the stocks. We assume that the limit price is a simple linear combination of the outstanding best asks and bid at the time of submission, as in [Ladley and Pellizzari, 2014].

HFT is included in the model allowing some traders to cancel and resubmit multiple times during a trading session: in contrast to the ordinary agent, who can send an order only once, the fast trader will have f additional chances to cancel the previous order (if it was lying in the book unexecuted) and revise the submission posting a fresh quote or a market order. Clearly, this permits to exploit orders that appeared after the first submission or, say, to change the degree of aggressiveness of the order with the conditions of the market. The value of f spans the speed of the traders and while ordinary, slow traders have $f = 0$ (no further chance beyond the unique one granted to everyone in the session), larger values of f are meant to model various intensities of the high frequency attribute.

We are mainly interested in quantifying the efficiency, if any, that the HF traders would bring with respect to the same market with slow agents. As a second target we aim at describing the way profits from trading are distributed among traders, when a fraction of the agents are engaged in HFT and with a special concern for the (differential) effects on the remaining slow traders.

In the next Section, we present the model describing the agents, the order book and how strategies are learned. Section 3 uses simulations to show results related to the efficiency of the market with/without HFT and how profits from trading are shared. Finally, we close with some additional remarks.

2 The model

2.1 The traders

We consider a standard Continuous Double Auction (CDA) where at each time step a single trader enters the market. An equal number N of buyers and sellers are endowed with a positive reservation value $v_i, i = 1, \dots, N$ or positive cost $c_i, i = N + 1, \dots, 2N$. In every trading session they can buy or sell a single unit of stock and, if a transaction occurs at a price p , the profit π_i awarded to the agent is $v_i - p$ or $p - c_i$, respectively. As no profit is made in the lack of a trade, i.e., $\pi = 0$, traders have the obvious incentive to exchange their unit, bidding (asking) less (more) than the value (cost) they are endowed with. Scholars have discussed other mechanisms to model trading: in [Rosu, 2009], traders are equipped with a discount rate that penalizes delayed or missed trading differently, allowing to define sets of patient or impatient traders. We prefer to use private values and costs because an equilibrium price is easily computed and still impatience is immediately and intuitively related to the magnitude of values and costs. In fact, buyers with high values or sellers with low costs suffer large losses if they do not trade and can be thought as impatient traders; conversely, buyers/sellers with values and costs close to the equilibrium price can generally obtain small gains and are nearly indifferent between trading and non-trading, hence resembling patient traders.

We deal with cancellation in a simplified and non-strategic way: at the end of every time step each order stored in the book is cancelled¹ with (a small) exogenous probability $P_c > 0$ that is independent of time, state of the book and of the specific agent acting in that period. This is the only source of cancellation for traders whose $f_i = 0$. Fast traders with $f_i > 0$ can have their order exogenously canceled as just described or intentionally cancel (and resubmit) their order up to f times after the first submission, if they have not traded yet.

2.2 The book

At any time t the book is a double sequence of outstanding unit orders

$$S_t = \{0 \leq \dots \leq b_{3t} \leq b_{2t} \leq b_{1t} < a_{1t} \leq a_{2t} \leq a_{3t} \leq \dots\},$$

where b_{1t}, b_{2t}, \dots and a_{1t}, a_{2t}, \dots are the lists of buy and sell orders in the books. We often omit the time index for simplicity. The highest bid b_1 and lowest ask a_1 are referred as best bid and best ask, respectively.

Traders submit a single order when they enter the market in a random period, according to the rules mentioned in the previous subsection. As the quantity is fixed

¹ We never cancel the order in the time step in which it is submitted.

at one unit, the trader must decide the limit price that is computed using a function of the state of the book and his valuation/cost.

The submission of an (unit) order with limit price l changes the book and results in an immediate trade, a marketable order, if the bid (ask) is greater (smaller) than or equal to the best ask (bid). In this case, the two agents involved in the transaction get the associated profits and the book is then updated, changing one between the best bid or ask. If instead the new order is not marketable, it is inserted in the book, maintaining its ordering, to be possibly used in future trades. Any profit occurring after the time of submission t is accrued in the same way to the parties involved whenever a transaction takes place.

2.3 The strategy

In real markets, traders can decide the time at which to submit and the limit price l they want to use (in our model the quantity is fixed at one). The time of first submission is random and drawn independently from any other variable of the model. We further assume, as in [Ladley and Pellizzari, 2014, Pellizzari, 2011], that each agent develops his own trading strategy picking the limit price using the relation

$$l_{it} = \alpha_i a_{1t} + \beta_i b_{1t} + \gamma_i, \quad (1)$$

where α_i, β_i and γ_i are individual constant to be found maximizing expected profits and a_{1t}, b_{1t} are the best ask and bid at the time t in which the order is issued. Agents for which $f_i > 0$ have multiple chances to cancel the orders: in more detail, in f random periods after the first submission, they can cancel their unexecuted order and resubmit at different limit prices (that are computed using (1) with the outstanding best bid and ask at the time of resubmission).

In other words, the model assumes that HFT really is only a matter of how often one can cancel the order, reevaluate his/her limit price and resubmit the order. We are aware that this simple approach has limitations and real HF traders may be more sophisticated in several ways. However, it is clear that having the chance to cancel and resubmit at high frequency should give an edge to some traders who can, say, wait-and-see for better market conditions or have a much higher chance to selectively pick the most beneficial quotes left on the book for some time.

The constants α_i, β_i and γ_i in (1) are individually tuned using a Evolution Strategies (ES) numerical scheme to maximize expected profit (hence, many sessions are averaged and as a consequence we also smooth the effect of the random position in the submission queue). The ES method belongs to the family of evolutionary algorithms and superficially resembles the well known Genetic Algorithms but requires no binary-coding or discretization of $\alpha_i, \beta_i, \gamma_i$ and has self-adapting meta-parameters that can be used to gauge whether convergence has been reached, see [Beyer and Schwefel, 2002] for an overview.

Technically speaking, maximization of profits is jointly performed by all the traders and ES determine a numerical equilibrium in which no agent has further incentive to alter his/her constants. In such a sense, the outcome is (just) one of the multiple trading equilibria that can appear given the values of v_i, c_i, f_i . We do not investigate here the number and properties of different equilibria and are content with the observation that, whatever they are, the selected triples $\alpha_i, \beta_i, \gamma_i$ form a stable configuration that is learnable as proved by the fact that it was indeed built by the ES optimizing procedure, that can be thought as a learning device in which traders (and their strategies) compete as in the marketplace.

3 Results

3.1 Implementation and parameters

We simulated the model described in the previous section using $N = 120$, effectively allowing 240 traders to participate in every trading session. To assess the variability inherent in any simulation, we have run 20 independent simulations in which 5000 trading sessions take place. Expected profits are computed averaging 100 trading sessions, after which ES performs selection, recombination and the so-called s-mutation (for meta-parameters) and y-mutation (for trading strategies).

All the results are based on the average of 20 simulations, in which we keep the last trading session (out of the 5000 that were used to “train” the traders towards optimal extraction of profits). Values and costs are symmetric around 10, which is therefore the theoretical equilibrium price and cancellation probability P_c is set at 0.01.

We performed 5 experiments to investigate how HFT affects efficiency and the way profits are apportioned among traders. Table 1 describes our design in which, starting from a benchmark case of slow traders alone ($f = 0$), we progressively add sets of HF traders ($f = 3$) with different values and costs. It is convenient in the following to label agents according to their values and costs: in particular, we nickname marginal agents whose value or cost is 10 as calm, referring to the fact that they have nothing to lose if they fail to trade at the equilibrium price; using the same intuition, buyers or sellers with $v = 14$ or $c = 6$ are nervous, as they feel more the need to trade; finally, when values reach 18 and costs as are low as 2, traders are named as hysteric, to capture the fear for sizeable losses in case of no trade. We remark that the terms calm, nervous and hysteric are related to personal values and costs and have nothing to do with the high frequency feature that depends on f .

Figure 1 shows the allocative efficiency reached in the experiments. Observe that when no HFT is present, agents nevertheless learn to extract 95% of the trading surplus (median value over 20 simulations). As more HF traders are added, efficiency increases reaching a median value of 100% in Experiment 3, in which some hysteric traders use HFT. The effect is smaller if calm or nervous HFT is included in

| | <i>HFT</i> | Buyers' values v_i | | | | | | Sellers' costs c_i | | | | | |
|------------|------------|----------------------|----|----|------------|------------|------------|----------------------|---|----|-----------|-----------|------------|
| Benchmark | 0 | 18 | 14 | 10 | 18 | 14 | 10 | 2 | 6 | 10 | 2 | 6 | 10 |
| Experim. 1 | 40 | 18 | 14 | 10 | 18 | 14 | 10* | 2 | 6 | 10 | 2 | 6 | 10* |
| Experim. 2 | 40 | 18 | 14 | 10 | 18 | 14* | 10 | 2 | 6 | 10 | 2 | 6* | 10 |
| Experim. 3 | 40 | 18 | 14 | 10 | 18* | 14 | 10 | 2 | 6 | 10 | 2* | 6 | 10 |
| Experim. 4 | 120 | 18 | 14 | 10 | 18* | 14* | 10* | 2 | 6 | 10 | 2* | 6* | 10* |

Table 1 Description of the 5 experiments. The first row describes the benchmark case with no HFT and each listed value/cost represents a set of 20 traders. Groups of HF traders are boldfaced and starred and, for instance, in Experiment 3 there are 20 hysteric HF buyers (sellers) with value (cost) equal to 18 (2).

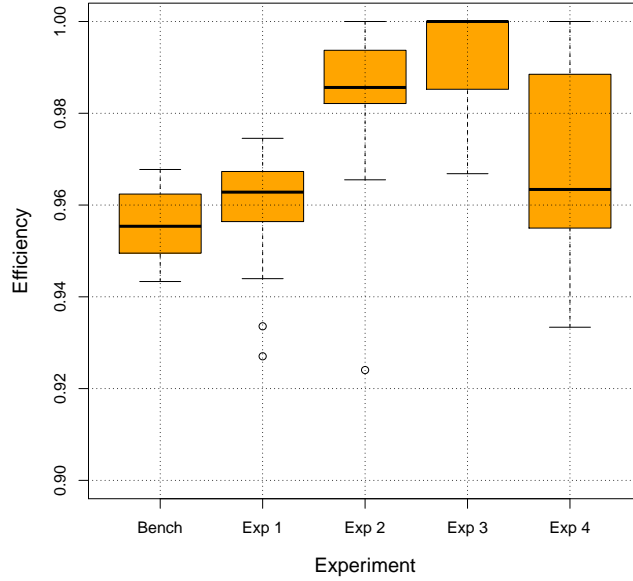


Fig. 1 Efficiency of the markets in the simulation experiments.

the market and the same can be said for Experiment 4 when HF traders belong to all types (calm, nervous and hysteric), thus proving that a large portion of HFT results in a negligible efficiency supplement with respect to the benchmark case.

Efficiency is a global measure of the ability of traders to transact when it is convenient given the individual valuations and the equilibrium price driven by the demand and supply for the asset. However, it says nothing about the precise way in which the available surplus is shared among traders. Table 2 shows how profits are distributed in our experiments. The profits of the first row, relative to the benchmark, are used to normalize the entries of the other rows and cases where HF traders are in action are boldfaced.

In Experiment 1, we see the effect the introduction of calm HF traders. With respect to the benchmark case, the average profit of all calm traders increase by 38% mainly at the expenses of the nervous agents who see their profits reduced by 6%.

When HFT is unleashed for some nervous traders (Experiment 2), they see their own profit increase by 8%, slashing by 35% the share of the calm agents.

| | Calm | Nervous | Hysteric |
|------------|-------------|-------------|-------------|
| Benchmark | 1.00 | 1.00 | 1.00 |
| Experim. 1 | 1.38 | 0.94 | 0.99 |
| Experim. 2 | 0.65 | 1.08 | 1.01 |
| Experim. 3 | 0.76 | 0.96 | 1.10 |
| Experim. 4 | 0.18 | 0.98 | 1.02 |

Table 2 Fractions of profit gained by calm, nervous and hysteric traders when HFT enters the market. All figures are normalized with respect to the benchmark case with no HFT.

The most pronounced effect can be seen in Experiment 3 when hysteric HFT enters the market. The profit of all hysteric traders, slow and fast, increase overall by 10%, with reductions ranging from 4 to 24% for all other market's participants.

At least two things are worth noticing: firstly, adoption of HFT increases the profits of the agents that use this trading approach. Effects are relatively large for calm traders (+38%), even though the absolute magnitude of the variation is small (being traders' valuations equal to the equilibrium price), and are smaller but still sizeable in absolute terms for hysteric traders. Secondly and quite remarkably, if HFT is used by one subgroup alone, negative effects on profits are detectable for all other traders.

In a nutshell, HFT benefits the ones who use it but damages the other agents that are worse off, despite the increment of the efficiency that was shown in Figure 1. If the model realistically depicts the market, traders may increasingly be pushed to employ HFT, producing the outcomes shown in "Experiment 4". In this case, in which agents equally split between slow and HFT, the profits of the calm traders are nearly wiped out (-82%), whereas nervous agents experience some losses (-2%) and hysteric traders keep some gains (+2%).

4 Discussion

In this paper, we present a simple model of HFT where some traders have in every trading session $f > 0$ additional chances to cancel and resubmit not yet unexecuted orders. The higher f , the more frequently traders can adjust their orders, if needed, and exploit transitory opportunities in the book. In contrast, traditional (i.e., slow) traders have a single chance to deposit an order per trading session, with no possibility to alter it.

Despite the simplicity of the model, we obtain two main insights: the introduction of HFT marginally increases the market efficiency and, hence, the surplus extracted is larger than in the presence of slow traders alone. In this sense, the pie gets slightly bigger as efficiency steps up from 95 in the benchmark case to 98% or more when few HF traders are introduced. We also find that HFT benefits the investors who are using it at the expenses of other traditional investors that have their profits eroded significantly or slashed.

In [Bernales, 2014] it is claimed that algorithmic traders with a trading speed advantage only reduce global welfare and this statement is to some extent at odd with our first finding. We instead strongly corroborate, in our agent-based setup, Bernales' result that HFT *de facto* predates non-HFT agents, seizing a portion of their theoretical gains. One novel insight of our model follows the observation that, under the dynamics that we have just described, many traders would be tempted to invest in IT infrastructure and actually jump in the HFT arena. If this be the case, however, efficiency would drop almost to the levels of the benchmark, as shown in the last row of Table 2. Hence, when too many HF traders operate in the market, their profit would shrink again as if few were present. In a situation resembling a "prisoner dilemma", policymakers may attempt to regulate access to and activities of HFT firms, to curb arms races potentially able to waste costly private investments and hamper public welfare.

References

- [Bernales, 2014] Bernales, A. (2014). Algorithmic and high frequency trading in dynamic limit order markets. Technical report, SSRN.
- [Beyer and Schwefel, 2002] Beyer, H.-G. and Schwefel, H.-P. (2002). Evolution strategies: a comprehensive introduction. *Natural Computing*, 1:3–52.
- [Biais et al., 2014] Biais, B., Foucault, T., and Moinas, S. (2014). Equilibrium fast trading. Technical report, SSRN.
- [Easley et al., 2011] Easley, D., Lopez de Prado, M., and O'Hara, M. (2011). The microstructure of the flash crash: Flow toxicity, liquidity crashes and the probability of informed trading. *The Journal of Portfolio Management*, 37(2):118–128.
- [Hasbrouck and Saar, 2009] Hasbrouck, J. and Saar, G. (2009). Technology and liquidity provision: The blurring of traditional definitions. *Journal of Financial Markets*, 12(2):143 – 172.
- [Hasbrouck and Saar, 2013] Hasbrouck, J. and Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, 16(4):646 – 679. High-Frequency Trading.
- [Ladley and Pellizzari, 2014] Ladley, D. and Pellizzari, P. (2014). The simplicity of optimal trading in order book markets. In Dieci, R., He, X.-Z., and Hommes, C., editors, *Nonlinear Economic Dynamics and Financial Modelling*, pages 183–199. Springer International Publishing.
- [Pellizzari, 2011] Pellizzari, P. (2011). Optimal trading in a limit order book using linear strategies. Working Papers 16, Department of Economics, University of Venice "Ca' Foscari".
- [Philips, 2012] Philips, M. (2012). High-speed trading: My laser is faster than your laser. *BloomberBusinessweek*. <http://www.businessweek.com/articles/2012-04-23/high-speed-trading-my-laser-is-faster-than-your-laser>.
- [Rosu, 2009] Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies*, 22(11):4601–4641.