Government Office for Science

Foresight

Minimum resting times and transaction-to-order ratios

Review of Amendment 2.3.f and Question 20

Economic Impact Assessment EIA2

Foresight, Government Office for Science

Contents

Minimum resting times and transaction-to-order ratios: review of Amendment 2.3.f and Question 20

European Commission Public Consultation

Review of the Markets in Financial Instruments Directive[1]

J. Doyne Farmer

Santa Fe Institute

jdf@santafe.edu

Spyros Skouras

Athens University of Economics and Business

skouras@aueb.gr

This review has been commissioned as part of the UK Government's Foresight Project, The Future of Computer Trading in Financial Markets. The views expressed do not represent the policy of any Government or organisation.

[1] We would like to thank three referees for many extremely useful comments on a previous version of this regulatory impact appraisal which led to very substantial improvements.

Proposed amendments focused on in this document

Amendment 2.3. f: Automated trading and related issues:

"Market operators would be required to ensure that orders would rest on an order book for a minimum period before being cancelled. Alternatively they would be required to ensure that the ratio of orders to transactions executed by any given participant would not exceed a specified level. In either case, further specification would be needed on the specific period or level."

Question 20:

"What is your opinion about requiring orders to rest on the order book for a minimum period of time? How should the minimum period be prescribed? What is your opinion of the alternative, namely of introducing requirements to limit the ratio of orders to transactions executed by any given participant? What should be the impact on market efficiency of such a requirement?"

1. Objective

The proposed measure of minimum resting times or transaction to order ratios is intended to decrease the speed at which markets operate and the cost in monitoring the information they generate.

In what follows, our analysis suggests market speed has gone far beyond the point where it is beneficial to the point where it is harmful. This is for the simple reason that the private benefits of speed come from *relative* speed (i.e. being faster than others) while market quality is determined by *absolute* speed levels. Beyond some level of absolute speed which has long been exceeded markets are harmed by speed, even though the private incentives for relative speed remain unchanged *regardless of the overall level of speed* in the market.

The proposed measure is one way of dealing with market speed. While the intent of the measure is in the right direction, we argue the measure itself is ineffective and may cause more harm than good. We develop an alternative proposal which we believe achieves the same goals much more effectively.

2. Background: review of relevant research

A rigorous regulatory impact assessment of the proposed rules faces two major obstacles, largely common to assessing the impact of any change in the details of market microstructure. First, generally accepted scientific knowledge about market microstructure is usually not powerful enough to answer such questions based on theoretical reasoning alone – data analysis is essential. Second, the empirical analyses required to answer such questions, while sometimes feasible, are complex, expensive and time-consuming. We believe appropriate empirical analysis could be conducted to predict the likely outcome of rule changes, and that such analysis should be done before such changes are enacted. Lacking such an analysis our evaluation will be based purely on reasoning from first principles and we acknowledge this is an unfortunate handicap which leaves plenty of scope for disagreement.

Unfortunately existing literature has limitations as a basis for answering the issues of interest. We will discuss this literature wherever relevant, but we note at the outset that many papers discussed involve studies in non-European markets, may be outdated, or may be based on experimental or theoretical idealizations. Few researchers have been given the access to data that would be required to begin distinguishing the effects of different kinds of algorithms at the level of granularity where very clear empirical answers could be given to the questions posed. Many of the key issues relate to the behavior of high frequency trading (HFT) algorithms but it remains extremely difficult to measure HFT or distinguish among its different varieties ,

Both proposed regulations have the direct effect of reducing the number of cancellations in relation to the number of transactions. Indirectly they also both reduce the number of passive order submissions and the total number of market events, as many submissions will not occur if it is harder to cancel them.

However, these two rules are likely to be different in terms of their effect on the overall functioning of the market. In particular we will argue here that the minimum resting time rule will give a substantial advantage to aggressive orders, and could result in significant increases in the bid-ask spread and possibly also in volatility, without solving the real problems caused by market speed. The negative effects of a limit on passive order submissions relative to transactions are not likely to be severe, but we also see very little benefit. Our reasons for this prognosis are developed below.

Our perspective highlights the fact that markets and trading algorithms form a rich ecology co-habited by a plethora of diverse strategies, some of which may be harmful while others may be beneficial. Even once we narrow attention to fast strategies, there may be both beneficial and harmful algorithms and their ecology may depend in delicate and subtle ways on strategies the role of which may not be obvious. In an ideal world, policy might target individual strategies or niches of the ecology based on detailed ecology analyses like the ones routinely used to assess the impact of large scale environmental policy initiatives (see e.g. Farmer and Skouras, 2011). We are probably some distance from narrowly targeted and precise policy both because of limitations of the current state of knowledge about how markets operate and because of the crudeness of policy instruments that are currently feasible. We should acknowledge that to the extent that innovation in the financial services industry is overall beneficial, crude regulatory changes will punish innovators both good and bad and the regulatory uncertainty it creates may discourage future innovation.

2.1. Review of the policy case for curbing market speed

We want to make it clear that we strongly support the intent of the proposed regulations to slow markets down. We thus think it is useful to include some discussion of why the overall goal of the proposed regulations is desirable (even though we argue that the particular implementations that are offered are insufficient). Here we discuss specifically the policy case for curbing market speed drawing from Skouras and Farmer (2011).

Currently, cutting edge HFT systems and the markets they operate in are designed around latency considerations at the level of single digit microseconds (millionths of a second) and there is widespread agreement that this will go down to nanoseconds (billionths of a second) in the next few years. We will argue that this is “too much speed”. Markets need to allow frequent trading which regulates the dissemination of private information, provides incentives for participants to provide liquidity and can therefore promote market quality and the objectives of MIFID II. There are trade-offs involved in more versus less speed, but we believe we are far beyond the optimal speed level and need to develop mechanisms to control market speed without disrupting markets too heavily. We are strongly in favour of empirical research to more accurately study this trade-off and in the meantime sketch some of the relevant considerations below. Inevitably decisions need to be made before a full understanding of the effects of market speed are developed, so we need to take a stance based on the available evidence. In our view imperfect understanding is little basis for sticking with the status quo. Furthermore, we will make a proposal that allows fine tuning of market speed as a deeper understanding of market speed is developed.

2.1.1 Understanding the causes of market speed

Market speed is largely the outcome of the investment of private traders in low latency technologies. This quest for speed is driven by a clear private incentive: In an electronic order book market, speed is worth a lot of money to traders. This money derives from *three and only* sources:

1. Aggressively exploiting or "picking off" stale passive orders.
2. Keeping passive orders from becoming stale or from being picked off by aggressive orders.
3. Obtaining a better position in order book queues than competitors with similar strategies and information.

It should be noted that all these private benefits come from *relative* speed, so that the incentives for relative speed do not decrease with the absolute speed level of the market. Of course to the extent that increasing speed eventually harms markets, this will also be felt by private traders but this overall effect is privately negligible to each of them (it is a large overall effect divided across millions or billions of investors). So private incentives will cause market speed and therefore social costs and this could lead to indefinite increases in market speed as the driver is *relative* speed which will always be improvable with further investments.

Based on these observations, it is important to try to quantify the incentives for private speed as this can help understand the enormous magnitude of the divide between social and private incentives for market speed.

Most studies so far have focused on the profit potential involved in (1). HFT strategies of this type necessarily involve aggressive trading, in which high frequency traders pick off stale quotes or exploit momentary profit opportunities using market orders. The estimates for the profit potential of aggressive trading are surprisingly small, ranging from $1.5 billion to $21 billion per year worldwide.[2] We should note however we believe this number is a huge

---

[2] According to one early report (Martin, 2007), a 1-millisecond speed advantage can be worth $100 million a year to a major brokerage firm, though this estimate is not substantiated in any way. Kearns, Kulesza and Nevmyaka (2011), further develop the approach of Aldridge (2009) to make detailed calculations for an upper bound on the profitability of *aggressive* HFT on US equity markets. They arrive at estimates in the region of the single digit billions of dollars, which seem fairly low considering the extremely optimistic assumptions they make for the HF trader's forecasting ability, response times and transaction costs. They also report estimates for the profitability of HFT of several other authors which reach up to $21 billion per year with estimates to aggressive HFT of up to $3 billion per year. Brogaard (2010) finds that HFT on the US equity market earns roughly $3 billion per year.

underestimate because it does not take into account the fleeting near-arbitrage opportunities generated from trading the same or similar instruments across different exchanges, which is known to be the bread-and-butter of many HFT firms. Unfortunately it is impossible for an academic or a regulator to estimate the profits available from fleeting near-arbitrage opportunities as this would require data synchronized at the location of trading servers and time-stamped within microsecond accuracy, which is available only to HFT firms. Even exchanges and regulators do not have access to this kind of data, let alone academics.

The profit potential for (2) is by construction the opposite of (1) – this can be thought of as an avoided loss rather than a profit. The losses to be avoided by keeping passive orders from becoming stale are by definition equal to the profit potential from picking off stale orders. This highlights the fact that the private value of speed is not determined by the *level* of speed but rather by *relative* speed. In this sense, much HFT is the outcome of a needless arms race, in which aggressive "predators" continually attempt to prey on passive prey, and both of them evolve to be faster and faster without any clear social purpose.

Taking a different direction to the extant literature, we have recently considered the profit-making potential for the third source of profits, i.e. obtaining a better position in order book queues, and found that it is enormous (Skouras and Farmer, 2011). A rough estimate is that speed in limit order placement is worth at least $500 billion per year worldwide. This is at least one and possibly two orders of magnitude higher than the profits due to aggressive orders. The enormous size of these profits makes it clear why the evolutionary pressure driving speed is so intense.

Our argument hinges on analysis of the reasons that queue priority is advantageous. This is because (1) limit orders with queue priority trade more often and (2) orders with queue priority more often trade against small market orders, which have lower market impact, i.e. they move the market less in the opposite direction of passive orders. The latter effect is the more important of the two: Large market impact can undo the profits from passive quoting.

The estimate for the value of speed rests on assuming linear market impact of individual orders as a function of size, a power law distribution of these incoming aggressive orders and symmetric information across market participants, all of which are widely used reasonable idealizations though nevertheless imperfect idealizations. Under these assumptions the profitability of a passive order is a decreasing function of its order priority because it orders with lower priority get executed against larger orders on average and therefore incur larger market impact. Orders at the end of a queue are just breaking even otherwise they either wouldn't be there (if they were loss making) or there would be other orders with lower priority (if they were making a profit) and we can use this zero-profit condition to derive the profits of all other orders as a function of their priority. Calibrating against real data for the distribution of queue sizes, trade sizes and frequencies delivers our estimate for the value of speed.

Certainly refinements are possible but we believe this calculation is a powerful way to show that the value of *relative* speed is extremely high in electronic limit books and that this may be largely due to the popularity of the time priority rule, which is not a law of nature – barring political barriers due to vested interests, this is easy to change. We note that two top HFT firms, Getco and Quantlab Financial each spent around $300,000 on lobbying and $100,000 on political contributions in 2010 (Bowley, 2011). We mention this in passing as an indication that vested interests are present and are no doubt attempting to influence policy outcomes, which may create pressure in favour of the status quo. The point of our estimates is not to deliver very precise valuations of speed, but to illustrate the mechanisms and incentives at work and to explain why the value is certainly large, even if its exact value remains elusive.

From the point of view of society, these arguments suggest that a limit on the duration of passive orders will have an effect on overall market speed that is at least partially offset by increases in the value of relative speed for aggressive orders since more opportunities for picking off stale passive orders will become available. And that to a good approximation, the quest for speed is an arms race. Since it is this arms race that determines the level of absolute speed, society will overinvest relative to what would be an ideal optimum unless this race creates some kind of beneficial externality.

Defendants of market speed argue that the arms race for speed creates a beneficial externality for society because faster markets mean faster allocation of resources and more beneficial trading. We discuss this possibility as well as its opposite, i.e. that speed causes several *negative* externalities in the next section 3. We list the potential sources of positive and negative externalities so the reader can form his own view of their relative importance. Our conclusion is that the negative externalities are likely to be much stronger so we believe HFT is most likely a *negative sum arms race*.

2.1.2 Empirical evidence on the effect of market speed

As an empirical matter the effect of market speed has been analyzed in the context of HFT and the evidence is somewhat mixed. On the negative side, BMO Capital Markets (2009) argues that the effect of HFT on the Canadian Market has been detrimental and Zhang (2010) finds that HFT is correlated with volatility; while he interprets this as evidence of the potential harmfulness of HFT, he only evidences correlation stopping short of proving causation. Zhang and Powell (2011) suggest that HFT was responsible for the flash crash of 2010 and caused an adverse impact on market confidence that they believe has been large. Jarrow and Protter (2011) and Cartea and Penalva (2010) develop models in which HFT is generally harmful.

On the positive side, Jovanovic and Menkveld (2010) find that a HFT market maker in Dutch stocks leads to a reduction of spreads by 29% while Brogaard (2010) finds that in the US equity markets HFT contributes to price discovery and is generally beneficial.

In our view this mixed evidence is due to the fact that HFT is a very rich subset of strategies, some of which are beneficial for markets while others are not, with these effects being potentially dependent on other market conditions. It is misleading to lump all of HFT together and a detailed ecological approach is necessary to empirically evaluate the effect of HFT. This is impossible without data that identifies participants and a new body of research making use of this data. Farmer and Skouras (2011) develop this perspective and propose a research agenda in this direction.

One thing that is clear about HFT is that it has distributional consequences for the profits available to traders: Tradeworx (2010) describes one situation in which HFT can – and in their experience does – exploit market microstructure imperfections to the detriment of other traders. A QSG study (2009) reports that HFT strategies have been designed to exploit features in execution algos. One way to interpret these reports is that there is a predator-prey relationship between fast and slower traders.

Acknowledging the limitations current empirical evidence has for evaluating the effects of HFT, we turn to a discussion of the effect of market speed which is based on more theoretical reasoning from first principles.

3. Risk assessment: the potential risks from no intervention to contain market speed

Several market observers have questioned whether market speed is harmful or should be contained. In this view, view speed is the outcome of an inexorable process of competition and technological innovation that in many other domains has been beneficial to markets (Jain and Johnson, 2008) and to society at large. Since HFT technology is available to anyone who is willing to make the investment in it, it seems subject to the usual competitive pressures and so markets should function properly to deliver reasonable outcomes. Furthermore, as discussed in the introduction, financial markets have always conferred an advantage to speed and considerable resources and ingenuity have always been spent to reduce latencies so it is possible to argue that HFT is not a qualitatively new phenomenon, even if it is more important now in terms of scale than it ever has been.

But it is difficult to find theoretical reasons why current speed levels might confer significant benefits to society. One argument might be that if prices reflect information sooner rather than later, then more of the information that influences the prospects of an investment is available at the time of a decision and therefore its risk is smaller. Taking as an example an investment in the shares of a biotech company developing a new technology, this effect would mean the company would be able to raise capital at a slightly lower cost. Society at large would benefit because of reduced uncertainty about the future. The problem with this argument is that any benefits of uncertainty reduction from receiving information milliseconds or microsecond earlier are surely negligible yet these are the speed levels at which markets operate.

A second argument might be based on the idea that one of the main benefits of markets come from the information they generate about what is the appropriate price of securities. If faster markets mean markets perform this role more effectively then social benefits could arise because of this improved transmission of information. However there are no obvious reasons why this would be the case.

Finally, a third argument in favour of market speed stems from the idea that markets create “gains from trade”. Trade is on average beneficial to both parties, for example because it allows risks to be shared. Ergo obtaining these gains sooner must be beneficial. But at plausible discount rates, obtaining these benefits sooner by a few milliseconds surely has negligible consequences. Similarly, the opportunity to share risks of negative events that might occur within milliseconds seems an unlikely source of private or social benefits. Indeed, we doubt even the existence of HFT algorithms serving this purpose.

On the other hand, there are several reasons why HFT might be harmful and why it will be increasingly harmful if nothing is done to contain it. Farmer and Skouras (2011) discuss in some detail the reasons for which algorithmic trading (more broadly than HFT) can cause market instabilities and why it requires a carefully designed competition policy, so here we will focus only on reasons which are specific to HFT beyond the general problems with algorithmic trading not specific to speed.

Speed as a perverse systemic outcome: Although speed has always had value in markets, it is not in any clear sense "natural" or inevitable that it should have as much value as it does presently. We believe its value now is at historical highs based on calculations that place it in the order of hundreds of billions of Euros per year. Our calculations reveal that this is so high because of a combination of the increase in the volume of trading and a reduction in frictions such as costs which previously cancel any small edge that a small differential information speed could offer. Indeed, we believe the importance of speed emerged as a byproduct of the symbiotic relationship of exchanges and HFT which itself is caused by the fact that exchanges have volume based fees (e.g. they charge per transaction) and HFT is by its nature high volume since it involves exploiting very many small profit opportunities. Small profits per transaction are made but because volumes are very large the profits become significant. In this context, exchanges competing for the business of HFR firms have an incentive to adopt market designs that are attractive to HFT, such as time priority electronic order books. This allows HFT to extract profits from slower traders, part of which is extracted by the exchange in the form of fees. This explains why markets use volume rebates in their pricing models to attract high volume HFT and why the Deutsche Börse recently went at great lengths to get permission from its regulator to allow algorithmic trading to have lower prices than other types of trading. This is consistent with the evidence presented by Menkveld (2011), which suggests a symbiotic relationship between Chi-X and a single large HFT market maker; it also indicates there are strong path dependencies in how the industry has evolved and why HFT has achieved its current levels of importance.

The profits from queue priority do not exist at all in "pro-rata" limit order book markets in which there is no time priority but instead orders at the same price execute against incoming aggressive orders in proportion to their relative size (we discuss a version of the pro rata approach in detail in sections 4.3 and 5.3).[3] The fact that new market designs are emerging such as "dark pools" which are designed to exclude HFT suggests that the value of speed may be an artifact of coincidental systemic characteristics of today's markets. De facto weak monitoring and regulation of HFT is another indication that institutions are set-up to favor HFT in ways that would not be considered acceptable for more mature types of trading.

Speed creates tradeoffs that increase risk: In order to increase speed, there is no doubt that algorithmic traders take short-cuts when it comes to implementing risk controls or in handling unusual situations. If speed is a priority, this necessarily limits the processing time that is needed for other tasks, such as risk control or "intelligent" information processing. Obtaining the sophisticated understanding needed to evaluate the meaning of information requires large programs that consume many clock cycles. Without such processing it is impossible to make intelligent decisions. Even though processing power gets continually faster, speed requirements also get faster, and the sophistication of information processing never improves – the algorithms just get faster."Naked access" is a type of extremely fast market access that became popular (until it was eventually banned in the US) because it allowed HFT to bypass risk controls on the broker side. While adding safety for the client, broker and market, and having been universally required for many years, HFT firms managed to convince brokers to give them naked access to reduce the latency of their orders. The drawbacks of this are obvious.

---

[3] This pro-rata approach is common in futures contracts, especially for interest rates.

It may be objected that HFT algos do not typically trade huge volumes in timescales over which risk checks are a limiting factor so the risk checks can be performed *after* each trade to allow the algorithm to continue trading. However, poor risk checks may actually cause a large volume of trading in a very small amount of time even if that is not the intent (e.g. in the event of a bug or hardware malfunction triggering a large number of large orders simultaneously). Perhaps more importantly, if many separate but correlated algos neglect the same risk-check this can aggregate to a huge systemic impact even if it is a sensible design decision for each individual algo.

Controlling market speed is important in the same way that it is in NASCAR racing. Drivers have such strong incentives to go fast that they are almost forced to use technology that causes risks that will eventually kill them unless appropriate rules are imposed by a regulator. Imagine the outcome for example if racing cars had no constraints on how they could be designed or powered while drivers were allowed to dope themselves as they saw fit in order to optimize their performance.

While it is possible that speed can be used for hedging or risk-mitigation sooner rather than later we doubt it is of much consequence whether this happens in nanoseconds rather than seconds.

Speed as a barrier to entry: Current HFT institutions have a huge incumbent advantage, with costly barriers to entry and an oligopolistic ownership structure. For example, Arnuk and Saluzzi (2009) cite a TABB report according to which expenditures on colocation and facilities for fast access amount to $1.8 *billion* per year (we assume this number refers only to expenses in US equity markets). In addition, markets are also spending huge sums and the NYSE alone is investing in facilities at a cost of $500 million. According to Price (2009), the cost of in-house solutions for competitive data feeds is of the order of $260,000 *per month* and a start-up cost of $270,000 per data center. Just to begin recording the kind of data that is necessary as a first step in designing a HFT strategy, a potential entrant would need several million per year. Furthermore, brokers typically require very large amounts of capital under management in order to provide competitive commission schedules (of the order of several tens of millions). The industry is very opaque in terms of information regarding fees and latencies of competing brokers and connectivity providers and it is a very difficult task to collect and compare such information across the large menu of available options.

It may at first sight seem that market pressures would limit the scope for this type of barrier. After all, a broker or other large market participant could share its HFT infrastructure with smaller entities charging a pay-as-you-go fee structure. However, commercially available HFT systems lag far behind custom solutions in term of speed. The reason for this is that commercial products sacrifice performance in order to meet commercial product standards and to provide functionality of interest to a wide customer base. Given the rapid evolution of markets, connectivity and hardware, systems need to be rewritten faster than the typical product lifecycle in order to remain at the cutting edge and furthermore they need to be designed to implement a very specific algorithm.

One reason for the oligopolistic nature of speed is that a low frequency trader does not have access to the historical data with accurate, high resolution timestamps that are necessary to develop and backtest many types of HFT strategies. Furthermore, in our experience HFT firms demand ownership of any strategies that are developed using their data or infrastructure, making it very difficult for new start-ups to enter the HFT space.

While it is true that competition for speed has always been present in financial markets, it is also true that regulators have always been involved in trying to level the playing field as much as they could. This has become increasingly difficult because of speed itself which makes it much harder for regulators to even observe what is going on in markets. The fact that there are other barriers to entry such as the availability of expensive research or information is no reason to not try to level the playing field where this is feasible, if there is strong evidence that eliminating such barriers is beneficial.

Indeed, there *is* evidence that policy to eliminate barriers to entry or to level the playing field for traders can have *very* positive effects. For example, Easley, Hendershott and Ramadorai (2009) study a natural experiment in which market access was changed to significantly equalize speed access of on and off-floor traders on NYSE and found that it led to an increase in prices of 3%, suggesting that this kind of issue can have a huge effect on the behavior of markets. Easley, O'Hara and Yang (2011) use a theoretical model to argue against practices which allow differential access to information (including speed advantages to HFT, an issue which seems related to the Net Neutrality debate relating to whether homogenous access to the Internet should be enforced by regulation – see e.g. Economides, 2008). That *equal* access to markets is viewed as a very important consideration is underlined by the fact that co-location providers make every effort to ensure that an identical service is provided to all co-location customers (e.g. the length of wires from market servers to all customers' co-located servers is the same regardless of where the servers happen to be located in the market data centre).

In sum, we acknowledge that barriers to entry are inevitable in markets, that speed is only one of them and there is not necessarily any foul play in how they emerge. However we believe that regulators should nevertheless try to contain these barriers to entry where practical as this is beneficial for markets. It may also be beneficial for market participants including those that build the barriers who may be subject to winners' curse (i.e. they spend so much to be the fastest that this offsets their advantage).

Speed can cause systemic instability and collapse: HFT has dominated the market ecology but may disappear in abnormal conditions, which can cause a disruption in the ecology through a sudden reduction in diversity. Research in market ecology suggests this can be very dangerous for stability (see Farmer and Skouras, 2011). HFT can pull out of the market very easily as HFT firms usually have small inventories.

There is one line that is clearly the fastest connecting the futures markets in Chicago with the Equity / ETF markets in New York. The highest frequency traders all use this same line and are responsible for a large fraction of the volume on both exchanges. If this line fails there would be a huge disruption in both markets that could easily spread globally and indeed, such a vulnerability could be attractive to terrorists. The broader point is that this vulnerability is caused by high frequency trading, while if relative speed did not confer such a huge advantage, network speed would not be a bottleneck.

It is widely acknowledged by practitioners that the strategies used for HFT are highly correlated so HFT can lead to systemic risk (evidence that HFT algos are less diverse than other types of trading is presented in Brogaard 2010). The heterodox explanation for the Flash-crash given by Nanex (2010) suggests it occurred because highly correlated automated market making algos caused a surge in activity, which caused an overwhelming data processing task, which triggered the algos to attempt to close positions more or less simultaneously in an extremely aggressive sell-off. This illustrates a problem with HFT whether or not the explanation is correct. Less pressure for speed would allow the algorithms to be more complex and more diverse, and therefore (likely) more stable.

Speed causes a wasteful arms race: The analysis of section 3 leaves no doubt that speed competition has become an arms race with "Red Queen" characteristics, i.e. "It takes all the running you can do, to keep in the same place". Effectively this results in a transfer of wealth from investors and market participants to technology participants who are the "arms dealers" in this situation. Further improvements in latency technology are unlikely to improve the behavior of the markets but will perpetuate this transfer of wealth indefinitely if the arms race is not contained.

A further disturbing feature of this arms race is that it may have a winner-takes-all outcome, i.e. if an institution invents a way to be faster than everybody else they can take *all* the opportunities known to be available to the very fastest algos (e.g. pure arbitrage opportunities). This will eventually cause competitors to exit the arms race, decreasing competitive pressure on the winner. Indeed, market lore suggests that some latency sensitive near-arbitrage opportunities are already entirely dominated by single institutions that have tailored their infrastructure especially to exploit the specific opportunity.

Speed causes instability in market *location*: The increasing demand for market speed means there are huge benefits to markets locating in physical proximity to each other. This is already disruptive to the historical pattern of the existence of national exchanges and may well lead to all markets being concentrated in a single location. While this is not in itself harmful on a global level, it will be damaging for most current financial centers which – if trends continue – will eventually cease to exist. This may be relevant for the future of the UK as a market centre.

Speed makes effective regulation much harder: Speed itself causes regulators to have a hard time forming effective policies. The technological arms race between HFT firms means that their data analysis resources are far superior to those of regulators. Both regulators and academic researchers have incomparably smaller personnel and incentives to collect and analyze market data than HFT firms. As a result the capability to analyze many issues of concern lags reality because markets change faster than the needed studies can be done. It is indicative that while the Flash Crash lasted half an hour, it took more than six months for the SEC team of experts to prepare its report. Slowing down this arms race would make it easier for regulators to understand markets and perform their functions. It is an extremely complex task for regulators to evaluate all the market innovations that speed leads to and whether they are harmful or beneficial, including flash orders, naked access and dark pools. When they do conclude they are harmful this is always with a lag during which harm has been done; this lag is the time period over which HFT innovators can expect to make profits from new trading technologies.

Of course this problem can be solved by equipping the regulators with technology and personnel of a calibre similar to that of HFT firms. However, the cost of doing this is likely prohibitive.

In sum there are many reasons to believe that speed is likely to have harmful effects at its current level. Furthermore, studies of the effect of market speed-ups in the range of seconds do not seem to find a significant improvement in market quality (e.g. Webb et al, 2007) so we believe that we have long reached the point where diminished returns to speed have kicked in. At the same time, we fully acknowledge the need for rigorous empirical analysis of each of the points we have made, a task which would no doubt involve tens of man-years of research. Unfortunately research is not as fast as markets and policy makers will need to make judgments based on the available qualitative evidence.

4. Policy options

4.1. Minimum resting times

The first proposed policy option measure we consider is that of minimum resting times. This would be a rule that makes it illegal or technically impossible to cancel or modify an order until some amount of time elapsed from its submission (or its most recent modification).

The direct effect of this rule would be to eliminate "fleeting orders" or orders that appear and then disappear within a short period of time. However, there would also be a broader effect because order submissions would need to be evaluated over longer horizons. Without the rule, submitted orders can be re-evaluated continuously and therefore the decision to submit carries only as much commitment as the latency involved in this re-evaluation. By contrast, with the rule a submission becomes a commitment which may be very significant depending on the parametrization of the resting time minimum.

No doubt this rule can be parametrized to achieve significant reductions in the number of unexecuted submissions and cancellations and hence the overall number of events occurring in the market.

4.2. Minimum transaction to order ratio rule

The second proposed policy option is to impose a minimum transactions to order ratio. This means that any trader would need to participate in a trade before they could submit a new order once they hit the rule limit. Currently many markets operate with an average ratio of around 1% so any ratio above that is likely to significantly impact market dynamics.

Introducing a minimum transaction to order ratio has in common with the minimum resting time rule the property of reducing the number of limit orders and the number of cancellations. However, in other respects the effect of this rule is likely to be quite different. In particular, the key difference is that because the passive orders are never frozen, they are not as vulnerable to being picked off – the situation is just as it is in a market without restrictions. A participant placing a passive order can move it at will, until the minimum transaction to order ratio is hit. At this point this participant will be forced to place an aggressive order but in reality all participants will become more hesitant to modify or cancel their orders, so that they will do so only subject to stronger (and therefore rarer) incentives than would be the case without the regulation.

4.3. Our proposal: frequent pro rata randomized call auctions

Building on the detailed considerations regarding the risks of market speed discussed in section 3 we propose a more subtle proposal. Our idea is to contain the private benefits to speed in markets which will thereby cause a reduction in speed without having a direct impact on trading strategies that are not predicated on speed. This achieves the goal of reducing the value of speed while otherwise affecting markets as little as possible.

Our proposal is to require the following features for automated trading mechanisms:

1. **Pro rata:** Electronic order books should replace time priority rules with pro-rata rules. In modern electronic order book markets, orders can be sent at prices such that there is no immediate interest against which they might execute. Such orders are "passive limit orders" and there will often be several orders at the same price. Markets have rules regarding how such passive orders will execute in the event that there are incoming aggressive orders which can fill some but not all of them. In the usual "time priority" markets, passive orders at the same price are allocated against incoming orders on a first-come first-serve basis. By contrast, in pro rata markets, all passive orders get partially executed, each in proportion to its size. This will drastically reduce the advantage to processing information and sending orders before competitors. A cap on the size of orders will be required so that institutions with large capital do not overwhelm the queue, but this cap can be quite large and should not inhibit the proper functioning of markets.

2. **Frequent sealed bid auctions:** All market venues will be allowed to process events and disseminate information only once every second, within a few microseconds of a globally common clock reference. According to the terminology of Schwartz and Francioni (2006) these would be 'sealed bid auctions'. As discussed below, this time interval is short enough to allow for many opportunities for intraday price discovery but long enough so that algorithmic trading does not become a technological arms race. Economides and Schwartz (1995) were very influential in the adoption of call auctions side-by-side to continuous trading mechanisms, especially at the open and close and the empirical literature examining the impact of these auctions has found overwhelmingly positive effects. However, this does not mean that frequent sealed bid auctions are necessarily a good replacement for continuous trading mechanisms and indeed the reasons we suggest for this mechanism have not been addressed in the literature as far as we are aware. The mechanism has been used (at up to 10 second intervals and without the other elements we propose) on TAIFEX for Taiwan Stock Index Futures Markets quite successfully at least relative to open outcry continuous trading according to several studies (see e.g. Huang 2004). Evidence from Webb et al (2007) suggests market behavior is not very sensitive to the frequency of trades, and indeed in TAIFEX the frequency gradually increased from once every two minutes to once every 10 seconds without a clear impact on market quality. Cheng, Kang and Fu (2008) find that the transition from this mechanism to continuous trading had mild consequences that depend on the liquidity of the underlying instrument.

3. **Randomization:** The 1 second duration between event processing should in fact be random, with the probability that they will occur at any given instance being unpredictable, so that it is the average duration that is 1 second. The random duration (common across all markets) would also eliminate the possibility that exogenous events such as macroeconomic / earnings announcements could create predictable opportunities for institutions that know they are faster (for example if the announcement takes places 1 millisecond before the time at which it is known with certainty that a market event will happen, fast traders will know they can exploit slower traders who will not be able to react in time). By making the timing of the market event uncertain, the profits available to fast traders will also go down as they will not be guaranteed a "last-mover advantage". Randomization is commonly used in opening and closing auctions especially in European exchanges but we are not aware of any studies specifically of the effect of randomization.

This shift from continuous-double auction electronic order book markets with time priority to markets with very frequent pro-rata call auctions will level the playing field so that it is accessible to anyone with an infrastructure equivalent to a Bloomberg terminal (a cost of around $30k per year). We mention this just as an example, since a Bloomberg terminal coupled with a high level programming environment can support the development of trading algorithms with around one second latency. Arguably, the chunking may be even more effective if it is less frequent, e.g. once every 10 seconds to allow humans to compete with algos in many activities where this is now impossible, but we believe it would be preferable to begin with a less disruptive intervention.

This is a drastic change given that as we have seen, cutting edge players now spend tens of millions per year in infrastructure. Our proposal would simultaneously end the arms race for speed and reduce the barriers to entry in the algorithmic trading space. It has the advantage that it allows markets to remain located at any distance from each other across the globe. Meanwhile, at one event per second markets can still generate almost 30,000 trades per (eight hour) day to facilitate price discovery through the trading process. There is no clear evidence that this approach or the pro-rata rule would have an adverse effect on liquidity which a critic might want to counterpoise to the obvious beneficial effects of our proposal.

Any policy should ideally be global, i.e. implemented at the G-20 level to avoid regulatory arbitrage. Care must be taken to ensure that internal crossing networks or over the counter markets are not used to bypass regulations. However, we think our proposal below may work even if imposed unilaterally. Our reason for this is that it offers a trading environment that most market participants would welcome, and gives no selective advantage to either aggressive orders or passive orders. It is an environment that would make participants feel safe. Some evidence in this direction is that increasing volume in the last few years is being traded in ‘dark pools’ and other exchange mechanisms designed to exclude high frequency traders. However, we emphasize that more work needs to be done in order to evaluate our proposal broadly and work out many important details including whether a unilateral implementation would be successful.

5. Costs and benefits of each option

5.1. Minimum resting time

5.1.1 Potential benefits of a minimum resting time

We expect the following beneficial consequences:

- Almost by definition, the imposition of a minimum resting time makes the limit order book more stable, in the sense that each order that enters the book stays there for at least the minimum resting time, so changes in the book are going to be less frequent.

- This will reduce the volume of passive order submissions and cancellations.

- Both of these effects make it easier for traders to monitor the book, update their decisions and make the effect of sending an order more predictable. This would incur the following benefits:

  i) *Easier market monitoring*. With a high volume of market events, an accurate reading of the current state of the market requires sophisticated software and hardware. Under the proposed change (and with a sufficiently long resting time), the order book will change more slowly, and thus the required infrastructure would be cheaper. This will save firms from expensive technology investments generating little or no social value.

ii) *Making more thoughtful decisions*. With fewer market events and a slower-changing order book, traders will have more time to analyze the state of the market before making trading decisions, and thus can make such decisions more intelligently.

iii) *Having a predictable environment*. In a market where an order can be cancelled at any instant, there is a high probability of submitting an order based on one set of information, and finding that by the time the order arrives the state of the orderbook has changed. This can result in unintended consequences. A minimum resting time ensures that with large probability the order book will not have changed between the time when a trader generates a decision and the time when the associated order actually arrives on the market. It thus ensures that traders who want to make transactions are more likely to get what they intended to get. In particular, if a trader who wants to make a transaction sees an order arrive in the book, he or she will know that it will persist at least for the minimum resting time, and therefore can send an order to trade against it.

- Some types of manipulative behaviors that generalize “layering and spoofing” will become much harder. Of course manipulative behaviors are already illegal but the reality is that the definition of what is manipulative is sufficiently ambiguous that such rules are difficult to enforce. Layering and spoofing refer to practices where passive orders are sent in the opposite direction to the one the trader actually wants to execute a trade and are cancelled without being executed (e.g. buy passive orders are sent and then cancelled before an aggressive sell order is sent or hidden sell orders are sent together with visible buy orders). Such orders can influence short-term order book dynamics in a way favorable to the price of the intended trade (e.g. the sell trade price will be higher thanks to the spoofing) *because* they are misleading to other market participants. Certainly a clear audit trail and the resources to analyze it might be another extremely useful tool in weeding out such behaviors.

5.1.2 Potential costs of a minimum resting time

The proposed rule handicaps passive trading without placing any corresponding handicap on aggressive trading. Thus one expects that it will reduce the profits for passive trading while increasing the profits for predatory aggressive trading. This will in turn reduce the volume of limit orders being placed, which reduces the number of orders sitting in the book. Thus the main consequence of this rule will be to reduce liquidity, which is bad for liquidity takers in general. This will almost certainly manifest itself in terms of increased spreads and increased market impact; it may also manifest itself in terms of increased volatility, though this is more controversial. We now explain our reasoning for this in more detail.

- *Picking off stale orders*. Preventing passive orders from being cancelled means that such orders are in constant danger of becoming *stale.* To see what we mean, suppose the minimum residence time for an order is one second. Suppose that in between the time when a passive sell order is placed and when it can be cancelled, there is a positive news event that makes most market participants believe the price should be higher. The passive order is unable to respond to this event, thus it becomes a target to be picked off by an aggressive market order that buys the shares in the passive order and then immediately sells them at a higher price. The aggressive market order thus profits at the expense of the passive order. Such an effect is called *adverse selection.* This rule change will artificially increase adverse selection effects for passive orders and will therefore cause an increase in the bid-ask spread as passive orders will need to be compensated for this additional adverse selection risk.

- *Raising transaction costs.* Increasing profits for aggressive orders at the expense of passive orders creates an incentive for participants to place more aggressive orders and less passive orders. Aggressive orders remove passive orders from the book, which means less liquidity. Similarly, fewer passive order placements means less liquidity. The net result is to increase the bid-ask spread, which increases the cost to make a round-trip transaction. In general, aggressive orders have more market impact, i.e. they cause larger changes in price. Thus transaction costs will likely go up across the board, for both large and small orders, and for both passive and aggressive orders.

- *This rule change will hurt retail investors.* Slowing down passive orders relative to aggressive orders is particularly perverse as it will cause an increase in bid-ask spreads and therefore increased costs to retail investors not involved in the high frequency trading zero-sum game. Wider bid-ask spreads will also mean that less trades happen as they impose a larger cost to trading that needs to be exceeded by the benefits of each voluntary trade. Therefore the gains-from-trade will be reduced.

- *Increased volatility?* Less liquidity in the book may result in increased volatility. There are two different ways of thinking about this, that lead to opposite conclusions. One view is that prices are strictly about information, so that except for small effects such as rounding errors or increases in the bid-ask spread, prices are unaffected by market microstructure. An alternative view is that price changes are noisy. Under this view, decreased liquidity leads to larger price changes when an aggressive order of a given size hits the market, thus causing more volatility.

- *It is unfair t*o impose a regulation that will so clearly benefit one type of HFT technology (aggressive) at the expense of another (passive). Firms specialize and this will lead to a shuffling in the market ecology, with possibly unpredictable consequences. Worse still, many market participants feel that it is fast predatory market orders that cause “unstable” or volatile price moves, which may be an additional reason to expect increases in volatility.

- *This change might drive away activity to other markets*, unless policy was co-ordinated internationally, e.g. on a G-20 level.

- As is usual in dealing with externalities, it seems more efficient to *tax* events than impose artificial limits the consequences of which are intractable. Such taxes, fees or penalties could be targeted specifically at orders with duration below a minimum resting time if such behaviors are deemed harmful.

By giving more profits to aggressive order placement and reducing profits to passive order placement, this rule change will introduce a substantial change in the market ecology – there will be more participants placing aggressive orders. A lower limit on the size of this change can be crudely estimated by calculating the profits for aggressive orders picking off stale orders. This calculation follows a logic similar to that used by Copeland and Galai (1983), who introduced the notion that a limit order can be treated as an option.

Consider a passive order to sell, and let the minimum residence time be T. Assume that the logarithm of the "efficient price" *p* (reflecting the fair price for the financial instrument) moves according to a continuous Brownian motion with diffusion rate σ. Assume also that the volatility σ is unaffected by the rule change (the optimistic scenario about the impact of this rule on volatility as discussed above). In units where T and σ are measured consistently, the distribution of logarithmic price movements *Δp* after time T has a normal distribution $N_{\sigma\sqrt{T}}(\Delta \mathrm{p})$ with zero mean and standard deviation $\sigma\sqrt{T}$. If the bid-ask spread is *s*, an aggressive order can make a round-trip profit whenever the increase in the efficient price exceeds *s/2* in time less than T because then the aggressive order can take advantage of the staleness of the passive sell order (which is constrained by the minimum residence time) and buy at a price lower than what the efficient price is.

However, this is not the only situation in which an aggressive order makes a profit. For a trading strategy that already wishes to buy for reasons unrelated to moves in the midprice, it is advantageous to buy even if it is not possible to make a round-trip profit by immediately selling. As long as there is any increase in the efficient price at all, executing against the stale order (rather than an order that is properly placed relative to the current efficient price) correspondingly increases profits. Under the assumption that an aggressive order picks off the passive order at time T (and not sooner), the expected profit to the aggressive order is the expected price movement multiplied by the size of the trade, which is

$$\pi = v\int_0^\infty \Delta p N_{\sigma\sqrt{T}}(\Delta p) d\Delta p = v\sigma\sqrt{\frac{T}{2\pi}}.$$

The expected profit for aggressive orders is therefore proportional to the volatility σ, the value *v* of a typical trade and the square root of the minimum residence time T. In a market where there are M trades per year, the average profits is roughly $\pi = Mv\sigma\sqrt{T/2\pi}$. For a minimum residence time of T = 1 second, if there are 8 hours in a typical trading day and the average daily volatility is 2% then then σ=0.0001 (1basis point per second). Considering that across Europe in one year M= 1 billion and that a typical transaction size is Euros 20,000 this gives a total profit of around 940 million Euro.

We do not believe 940 million is a definitive estimate, rather that this calculation helps illustrate how to think about the source of predatory profits and their ballpark range. We believe plausible alternative calculations may well deliver numbers in the region of Euros 100 million-5 billion. In particular, our estimate contains several approximations that should be acknowledged:

- We have assumed that the passive order is not picked off until the end of the minimum residence time. In reality competition will force aggressive orders to pick off stale orders sooner. This may substantially lower the estimate given above.

- We have assumed a normal distribution. In reality short time scale price movements have a distribution with power law tails, which should increase the size of the profit opportunities (and could raise the estimate).
- We have assumed that σ is unaltered by the rule change. It is possible that σ will increase due to the rule change, which would raise the estimate.
- We have neglected the opportunity cost of not being able to move a stale sell order if the price goes down. That is, if the price moves down other passive orders may be placed at prices below that of the frozen order we are considering, and may capture the transactions that the frozen order would have liked to make. This effect is difficult to estimate, as it depends on the cost to the frozen order of not making these transactions. We believe a plausible guesstimate is that this effect would raise the estimate by as much as a factor of two.

5.1.3 Overall assessment for minimum resting time proposal

This proposal has some benefits in slowing down the order book and making a more predictable environment for liquidity takers. However, because it handicaps limit orders and does not place a corresponding handicap on market orders, we predict that the negative effects will far outweigh the benefits. In particular, we predict that if one uses minimum resting times sufficiently large to slow the order book down to a comfortable rate of change, predatory aggressive orders will make huge profits at the expense of passive orders. This will result in a dry up of liquidity in the book and might even cause an increase in volatility. To be honest, we think this proposal is a terrible idea unless it is complemented with several other interventions in market rules which we discuss in Section 4.3 (a pro rata frequent call auction mechanism). By itself we do not feel it serves the goals of MiFID II of increasing liquidity, efficiency and investor confidence.

5.2.Minimum transaction to order ratio

5.2.1 Potential benefits of a minimum transaction to order ratio rule

The benefits are similar to those for the minimum resting time rule, in that this rule will cause fewer cancellations and therefore will make the limit order book more stable, and make execution more predictable. However, we predict that the benefit will not be as great, for the simple reason that a limit order can still be cancelled at any time, so the order book is not as predictable.

5.2.2 Potential costs of a minimum transaction to order ratio rule

The costs of implementing a rule of this type are correspondingly lower than for the minimum resting time proposal. The key difference is that passive orders are not forced into a situation where they can be easily picked off, so there are not windfall profits to predatory aggressive orders as there are for the other proposal.

However, there will necessarily be less passive order placement, and because this rule puts pressure on passive orders to make transactions, passive orders will have to be more aggressive which means a reduction in liquidity. However, in contrast to the other proposal, we don’t think the effect will be large unless the ratio is set to a very small number, by which we mean a number that would drastically reduce the current market-wide average. Typical ratios of orders (new orders + modifications + cancellations) to transactions are around 20:1 to 40:1 across liquid markets dominated by HFT. Thus a restriction that capped this ratio at, say, 20:1, would probably not have a strong harmful effect though of course a cap at 2:1 would have a huge effect. Implementation would involve many important design details beyond the ratio itself such as the time interval over which it is calculated, the segment of the market and so on.

It is also worth noting that the monitoring costs and the costs of practically implementing the restriction would also be significant.

5.2.3 Overall assessment of a minimum transaction to order ratio rule

We anticipate that if this rule were implemented to contain individual market member's ratios to around the market average, there would not be a dramatic effect. However, if the ratio were set to a small number there would be a substantial negative effect due to the loss of liquidity in the book, which would result in increased spreads and increased transaction cost. We think this rule is a bad idea, though for a somewhat different reason: This rule forces market participants into doing something that they otherwise would not want to do, for no clear public welfare benefit. Again we do not think this proposal is in line with the goals of MiFID II.

Unlike the minimum residence time proposal, there are several examples of markets where minimum transaction-to-order ratio rules have been used as part of exchange fee schedules. For example, in several European equity markets (e.g. Euronext) when some member institution sends too many passive orders relative to transactions the price it pays for each passive order is increased. A careful empirical study of the effect of the introduction of such pricing schemes would be a very effective way of determining the impact of a regulation imposing a minimum transaction to order ratio so we consider such a study essential before any such regulation is pursued. However, because existing schemes impose costs at transaction to order parameters very far from the ones we believe a regulator might want to impose even such studies may not be very informative. The only feasible way to definitively evaluate this proposal is to try it, optimize the parametrization based on observed impact and then do an ex post empirical evaluation. We acknowledge that reasonable people can disagree on the merits of this kind of experimental policymaking, but it not our preferred flavour.

5.3. Our proposal for complementary regulatory interventions

5.3.1 Potential benefits

The main benefits of our proposal are the following:

- The private value of speed can be controlled by parameters such as the frequency of pro rata auctions so that it is equalized to the social value of speed. We believe our 1Hz proposal is in the right ballpark to ensure effective intraday price discovery while eliminating the incentives pushing algorithmic trading towards a technological arms race. Under our proposal, each market can still generate almost 30,000 trades per (eight hour) day to facilitate price discovery through the trading process. At the same time, our proposal levels the playing field so that it is accessible to anyone with an infrastructure equivalent to a Bloomberg terminal (a cost of around $30k per year). This is a drastic change given that as we have seen, cutting edge players now spend tens of millions per year in infrastructure.

- balanced incentives for passive and aggressive orders which may be important for better price discovery.

- Our proposal will eliminate all the risks from increasing speed discussed in Section 3: it will decrease the risks assumed by algorithmic systems as clock cycles could be allocated to risk control and calculations associated with intelligent and stable price formation; it would reduce flash-crash type systemic risks stemming from a crowded algo trading ecology; it will reduce waste and barriers to entry; it will stabilize the location of market centers; and it will make markets more transparent to regulators looking to control abuse, manipulation or fraud.

- Most importantly, this proposal can be implemented without disrupting other kinds of algorithmic trading that are not speed intensive and which we view as beneficial to market efficiency and price discovery. Unlike the other policy proposals it would not disrupt the incentives to submit passive orders. Furthermore, by reducing the barriers to entry in the algorithmic trading space it may even benefit algorithmic trading as a whole.

5.3.2 Potential costs

No market structure is free of shortcomings and no regulatory change will be beneficial to all market participants. Our proposal is no exception, however we believe the overall costs are fairly contained.

Our proposal will increase uncertainty about whether an order will get executed. It may also increase uncertainty about the price at which it will get executed if it gets executed. This may discourage some types of trading algorithms and it could be argued that this might be harmful for markets. Indeed, there is some evidence that pro-rata markets are less liquid than continuous markets (Lepone and Yang 2012). However, it is difficult to draw general conclusions from specific studies as the details are likely to matter and our proposal combines several elements. For example, it is quite plausible that continuous trading is preferable to daily call auctions (as reported e.g. by Amihud, Mendelson and Lauterbach 1997) even though it may be very similar to call auctions every ten seconds (e.g. as in Cheng, Kang and Fu, 2008). A more targeted study of our proposal would be welcome but we do not see a mechanism through which it could do significant harm to market quality.

Any restrictive market regulation will increase the incentives for regulatory arbitrage. A unilateral adoption of our proposal by one regulator may push market activity to the jurisdiction of another regulator. This suggests it is important for regulators to co-ordinate policy, e.g. at the G-20 level, but even in the absence of such co-ordination we believe that a market in which the private and social value of speed are aligned may in fact prove attractive to market participants which would limit the incentives for regulatory arbitrage. If our analysis is correct, regulated markets will be systemically safer and in many respects better so they should retain a significant share of customers.

An interesting question is how currently available alternative market mechanisms such as dark pools would be affected by our proposed changes. From discussions with market participants we understand that an important reason for which they use these mechanism is to avoid being gamed by high frequency traders. To the extent this is true, if our proposal were to be successful it would undermine the business model of dark pools and related markets. This would seem of relatively little import other than to the stakeholders in the companies that own these market mechanisms.

5.3.3 Overall assessment

Based on our analysis, we are confident that the beneficial effects of adopting our proposal are far more important than the potential costs. Other proposals in the same direction achieve the intended regulatory goals only very imperfectly and with significant side-costs.

Unfortunately we believe mechanisms such as those we propose are unlikely to evolve without regulatory intervention because competition among markets has essentially become a competition to attract the handful of institutions that provide liquidity and these institutions use high frequency technology. It is therefore very difficult for a “slow” market to threaten incumbent fast markets, even though this might be a more efficient market structure and could have been the observed outcome were it not for path dependency and various historical accidents including previous regulations. For the same reason, it seems unlikely that markets will choose to slow themselves down voluntarily.

6. Future: how might costs and benefits evolve

If there is no regulatory intervention to control market speed, we expect the technological arms race for speed will lead to ever faster markets. The private value of more relative speed would remain large even though the social value of more speed is negative.

No doubt the market’s acceleration rate will at some point decline, but it will likely remain positive for decades to come. The gap between the private and social value of speed will widen, which will make all the problems we have discussed in section 4 worse. The longer it takes for regulators to deal with the issue, the more likely it becomes that a catastrophic systemic failure might occur, which could undermine market confidence or the appetite of retail investors to allocate their savings to stocks and related financial instruments.

One way or another the future will be shaped by regulators as well as technology and market forces. It may be shaped by regulation that is instructed by recent market acceleration or by legacy regulation which has not kept up with recent market developments.

7. Summary

We have argued that the private value of increases in *relative* speed is much larger than has previously been appreciated and that this is the main reason we are seeing such an explosion in high frequency trading as technology has improved and transaction costs have gone down. We also argue that while difficult to measure, the social value of increases in *absolute* speed is very negative – in particular many individual investors are harmed by market speed, each by a relatively small amount which however is much larger than the benefits from speed which are extracted by just a few institutions.

The proposed rules *by themselves* are an inefficient way of redistribute the benefits of speed from passive to aggressive orders without having a clear impact on the overall levels of market speed. However, if they are complemented with some *additional* rules, market speed could be contained without a major impact on other aspects of markets which are overall extremely beneficial to society. An effective improvement in market design could be built around the idea of a minimum resting time for passive orders but would also need the following additional elements:

- Markets would publish information and execute trades once per second (1 Hertz). This would effectively impose a minimum resting time for passive orders without giving a corresponding advantage to predatory aggressive orders -- both passive and aggressive orders would be slowed down, so the playing field would be leveled.

- Passive orders would execute pro rata instead of according to time priority. (*Pro rata* means that trades are allocated to passive orders at any given price level in proportion to the volume of shares of the orders rather than by time priority). This would eliminate the benefits to speed that go to passive orders when there is time priority in order books.

- The benefits of this scheme would be particularly clear if markets could be synchronized globally which would reduce the benefits of low latency predatory inter-market arbitrage. However, unlike the previous proposal, we believe such a proposal can be successful even if implemented unilaterally. More study is needed on this point to identify possible inter-market arbitrage.

- The 1 Hertz rate of markets would be an *average* rate, with some intentional randomness to further reduce the gains from low latency trading.

Such changes would slow trading down and ensure sufficient time between trades for deliberate, intelligent trading. Our analysis suggests these changes would drastically reduce the incentives market participants have to invest in HFT infrastructure as the overall benefits available to speed (which have no general welfare value) would be reduced by as much as $500 billion per year. Therefore the deleterious effects of speed on market quality would be contained with modifications that should be innocuous to the main purpose of markets, which is to provide benefits from trade and pricing information and through this cheaper financing for society’s investments for the future.

References

Amihud, Yakov and Haim Mendelson and Beni Lauterbach, 1996, "Market microstructure and securities values: Evidence from the Tel Aviv Stock Exchange", Journal of Financial Economics 17, 223-249.

Arnuk S. and J. Saluzzi, 2009, "Latency arbitrage: The real power behind predatory high frequency trading", Themis Trading LLC White Paper

Bowley, G., 2011, "Fast traders, in spotlight, battle rules", New York Times, July 17, 2011.

Brogaard, J.A., 2010, High frequency trading and its impact on market quality, working paper

BMO Capital Markets, 2009, The Impact of High Frequency Trading on the Canadian Market, July 22, 2009

Cartea A, and J. Penalva, 2010, "Where is the value in high frequency trading", working paper

Chaboud, A., Hjalmarsson E., Vega, C. and B. Chiquoine, 2009, "Rise of the Machines: Algorithmic trading in the foreign exchange market", FRB International Finance Discussion Paper No. 980

Cheng, M.S., Kang, H.H. and Y-F Fu, 2008, "Trading activity and the effects of trading mechanism", working paper.

Copeland T.E. and D. Galai, 1983, "Information effects on the bid-ask spread", Jounal of Finance, XXXVIII(5).

Easley, D., Hendershott, T. and T. Ramadorai, 2009, Levelling the trading field, working paper

Easley, D., M. O'Hara and L. Yang, 2011, "Differential access to price information in financial markets, working paper

Economides, N. and R.A. Schwartz, 1995, *Journal of Portfolio Management*, 21(3), 10-18.Farmer J.D. and S. Skouras, 2011, "An ecological perspective on the future of computer trading", Foresight Driver Review DR6.

Farmer J.D. and S. Skouras, 2011, "An ecological perspective on the future of computer trading", Foresight Driver Review DR6.

Huang, Y.C., 2004, "The market micrtostructure and relative performance of Taiwan stock index futures: a comparison of the Singapore exchange and the Taiwan futures exchange", Journal of Financial Markets, 7, 335-350.

Jain P.K. & W.F. Johnson, 2008, "Trading Technology and Stock Market Liquidity: A Global Perspective", in, Stock Market Liquidity, F-S. Lhabitant and G.N. Gregoriou, John Wiley and Sons Inc.

Jarrow R.A. and P. Protter, 2011, “A dysfunctional role of high frequency trading in electronic markets”, Johnson School Research Paper Series No. 08-2011

Jovanovic, B. and A.J. Menkveld, 2011, “Middlemen in Limit-order markets”, working paper.

Lepone, Andrew and Jin Young Yang, 2012, “The impact of a pro-rata algorithm on liquidity: Evidence from the NYSE LIFFE”, Hournal of Futures Markets, 32(7): 660-682.

Menkveld, A, 2011, “High frequency trading and the New-Market makers”, working paper

Nanex, 2010, “May 6th 2010 Flash Crash Analysis: Final Conclusion”, October 14, 2010

Price, T. 2009, Market Data Management: In-house vs outsourced solutions, Towergroup, February.Schwartz, R.A. and R. Francioni, 2006, “Call auction trading”, *Encyclopedia of Finance*, Lee C.F. and A.C. Lee (Eds), Springer.

QSG, 2009, “QSG study proves higher trading costs incurred for VWAP algorithms vs Arrival Price algorithms, High Frequency Trading Contributing Factor”, November 24, 2009

Skouras S and J.D. Farmer, 2011, The value of queue priority in limit order books, working paper.

Tradeworx, 2010, “Public commentary on SEC market structure concept release”, April 21.

Webb, R, Muthuswamy, J. and R. Segara, 2007, Journal of Futures Markets, 27)12), 1219-1243

Zhang, F, 2010, “High frequency trading, stock volatility and price discovery”, Yale university working paper

Zhang, F, and S.B. Powell, 2011, “The impact of high-frequency trading on markets”, CFA magazine, March-April.

Foresight
1 Victoria Street
London SW1H 0ET
www.foresight.gov.uk
URN: 12/1064