# Is Trading Fast Dangerous?

Thierry Foucault[1] and Sophie Moinas[2]

January 18, 2018

**Abstract**

The speed of trading has considerably increased in recent years, due to progress in information technologies and automation of the trading process. This evolution raises many questions about the effects of trading speed. In this chapter we discuss the findings of the growing theoretical and empirical literature on trading speed in financial markets. We argue that an increase in trading speed raises adverse selection costs but increases competition among liquidity providers and the rate at which gains from trade are realized. Thus, the effect of an increase in trading speed on market quality and welfare is inherently ambiguous. This observation is important for assessing empirical findings regarding the effects of trading speed and policy making.

[1] HEC, Paris, GREGHEC, and CEPR. 1 rue de la Liberation, 78351 Jouy en Josas, France. Tel: (33) 1 39 67 95 69; E-mail: foucault@hec.fr

[2] Toulouse School of Economics, and CEPR (Toulouse 1 Capitole University, TSM). Manufacture des Tabacs, 21 allées de Brienne, 31000 Toulouse, France. Tel: (33) 5 61 12 85 73; E-mail: sophie.moinas@ut-capitole.fr

# 1 Introduction

Progress in information and trading technologies have considerably increase the "speed" at which trading takes place in financial markets and trading speed is a defining characteristic of high frequency trading firms (hereafter HFTs), such as Citadel, Virtu, Flow Traders or Jump Trading.[3]

These firms account for a significant fraction of the trading activity in electronic markets. For instance, according to a recent report for the Congressional Research Service (Miller and Shorter, 2016), HFTs account for roughly 55% of trading volume in U.S. equity markets (40% in European equity markets), 80% of trading volume in currency futures and about 66% of trading volume in treasury markets. For 2013 alone, the Tabb Group estimates the investment in fast trading technologies at $1.5 billion, twice the amount invested in 2012. This predominance of high frequency traders and the size of their investment in speed raise concerns about whether high frequency traders use speed at the expense of slower traders (as argued by Michael Lewis in his best seller, "Flash boys: a Wall Street revolt") and whether their activity make markets less stable. For instance, analysts have questioned the role played by high frequency traders in several disrupting market events such as the May 6, 2010 flash crash in U.S. stock markets, the October 15, 2014 U.S. treasury flash crash, the January 1, 2016 Shanghai flash crash, or the British pound flash crash of October 7, 2016.[4]

These concerns have led to several proposals to slow down markets and calls for regulatory interventions. For instance, some trading platforms (e.g., IEX or Alpha) now use so called "speed bumps" that delay the time elapsed between the moment at which an order arrives to a trading platform and the moment in which the order enters into the trading platform's matching engine. Alternatively, Budish, Cramton and Shim (2015) proposed to switch from continuous electronic order books (the current market organization in many electronic markets) to frequent batch auctions, in which markets clearing would take place only at discrete points in time (say, every seconds).

[3] The SEC defines high frequency traders as proprietary traders who use "*extraordinary high speed and sophisticated computer programs for generating, routing, and executing orders.*"

[4] For an analysis of high frequency traders' behavior during the May 6, 2010 Flash crash, see Kirilenko et al. (2017).

What is the economic rationale for such regulatory interventions? What are the market failures that they intend to fix? Why should we a priori expect the speed of trading to be excessive? Economic analysis of these questions for high frequency trading is nascent. To address them, one must first understand why trading speed is important for high frequency traders and through which channels it can affect liquidity and pricing efficiency. We analyze these issues in Sections 2 and 3, respectively. Then, in Section 4, we discuss possible reasons for which market forces might not generate the socially optimal level of speed and why therefore regulatory intervention might be needed in this area.

## 2 Matching and Informational Speed

One way to assess speed in financial markets is to measure how fast traders (or their algorithms) and trading platforms communicate with each other, i.e., send "messages" to each other and process these messages. Messages are orders/instructions and reports of various types. For instance, a "buy market order" is an instruction to buy a given number of shares sent by a trader to a trading platform. After receiving this message, the trading platform "processes" it (i.e., matches the market order against standing sell limit orders) and sends back a message to the sender confirming receipt and processing of his message. Taking this perspective, one can measure trading speed (e.g., for a given trader and for a given message) by the time elapsed between the moment a message (e.g., the market order to buy shares in our example) is sent to a trading platform and the moment the sender receives feedback from the trading platform (e.g., on the status of his order). This time is often refered to as "latency".

Latency has become extremely short in electronic securities markets, close to physical limits. For instance, Baldauf and Moller (2017) report that the average exchange-to-trader latency (the average time elapsed between the moment a message is sent by an exchange to the moment it is received by the firm) is 31 microseconds on average for messages pertaining to the SPDR S&P500 ETF (so called SPY, traded on the New York Stock Exchange). High frequency traders make investments to minimize the time required to process messages and send messages to exchanges, so presumably their overall latency is also extremely small. For instance, they have invested in fast communication lines between markets, either using fiber optics or microwave transmission (which is 30% faster than fiber optic lines), provided by firms such as Hiberna or McKay brothers.[5] Using microwave transmission, latency between trading platforms based in the Chicago area and those in the New-York area is 8.02 milliseconds (source: McKay brothers website), close to the theoretical lower bound (about 7.9 milliseconds).

Speed of trading is important for high frequency traders traders because it is key for the profitability of their strategies, in particular high frequency market making, high frequency arbitrage, and directional trading on very short lived signals. The role of speed in these strategies is best explained by considering a few examples.

**Speed and Trading Strategies.** First, consider a trading algorithm exploiting arbitrage opportunities between an ETF on an index (say, the SPY ETF traded on the NYSE) and a futures on the same index, e.g., the E.mini futures on the S&P500, traded on the Chicago Mercantile Exchange (CME). When the price of the ETF is high relative to the futures price, the algorithm will send buy marketable orders in the futures market and sell marketable orders in the ETF market. It takes time for these orders to travel between the computer running the algorithm and the trading platforms on which the ETF and the futures are traded (New York and Chicago respectively). Moreover, once these orders have reached these platforms, it takes time for the platforms to execute them against posted quotes, confirm their status (whether they have been fully executed or not), and report the prices at which they have been executed.

The high frequency arbitrageur's profitability is higher if the time elapsed between the moment at which his algorithm sends marketable orders and the moment it gets confirmation of their execution is short, i.e., if trading is fast, for at least two reasons. First, the high frequency arbitrageur faces competition from other arbitrageurs. If competitors are faster, they will take advantage of the arbitrage opportunity before the high frequency trader and the latter will end up trading at bad prices (see Kozhan and Wah (2012) for

[5] The first microwave network linking Chicago and New-York started operating in 2010 (see Shkilko and Sokolov, 2016) and is used by traders to obtain fast information on quotes of futures on indexes traded on the Chicago Mercantile Exchange (CME), and on that of Exchange Traded Funds (ETFs) on these indexes traded on equity markets located in the New-York area (e.g., trading plaforms such as the New York Stock Exchange (NYSE), BATS, or DirectEdge).

evidence). Second, quick feedback from trading platforms is key for risk management. For instance, the buy marketable order placed by the arbitrageur might execute only partially if the arbitrage opportunity vanishes before this order is processed by the CME. In this scenario, the arbitrageur takes the risk of taking a non fully hedged position, which in turn requires new quick actions. Chaboud et al. (2014), Budish et al. (2015), and Foucault et al. (2017) provide evidence about high frequency arbitrage strategies.

Consider another example, namely a high frequency trader using price changes in the futures market to forecast price changes in the index constituents. For instance, when the price of the futures increases, the trader buys the index constituents that are the most correlated with the index (e.g., which are more heavily weighted in the index).[6] This is a directional trading strategy based on very short lived signals, i.e., price changes in the futures market. The profitability of this strategy crucially depend on the high frequency trader being able to observe and react to these signals *before* the price of the index constituents adjust to reflect the information contained in futures price changes. This requires very fast access to the futures market datafeed and very fast access to trading platforms on which the index constituents are traded. Brogaard et al. (2014) provide evidence of such directional strategies on short lived signals (e.g., macro economic announcements or changes in limit order books).

Finally, consider a market maker in one stock traded on two platforms, e.g., BATS Europe (now Cboe) and the London Stock Exchange (LSE). Suppose that a buy market order just get executed on BATS, consuming all the available liquidity at the best ask price in this market. The bid-ask spread on BATS is now wider (say, that it increases from 1 cent to 5 cents). Quick information on this event is useful for the market maker for several reasons. First, the arrival of a buy order on BATS is in itself a source of information, which should lead the market maker to revise his quotes upward. The market maker might then optimally decide to update his quotes on the LSE.[7] Second, the new bid-ask spread of five cents on BATS might now be too large relative to the competitive level (say, one cent). There is therefore an opportunity for submitting new, more competitive, sell

[6] See Hendershott and Riordan (2013) or Zhang (2017) for evidence that HFTs use future prices as a source of information.

[7] This scenario is predicted by models of informed trading, such as Glosten (1994); see van Kervel (2015) for evidence.

limit orders on BATS. Again the market maker must be fast to take advantage of this opportunity. Speed matters because if the market maker moves first, he will acquire time priority at the competitive quotes leaving smaller or no profit to other market makers. Last, suppose that the market maker is the first to set a new competitive bid-ask spread of one cent on BATS and that his offer is lift by another buy market order. The market maker has now a short position. To reduce his position risk, he can place a buy market order on the LSE. The faster this order is filled, the smaller is the risk faced by the market maker. Lescourret and Moinas (2017) and Menkveld (2013) provide evidence of multi-venue inventory management by high frequency market makers.

High frequency traders do not necessarily specialize in using one trading strategy (market making, directional trading, and arbitrage) and might opportunistically use each of them (although empirical evidence suggests that there is some specialization, see Boehmer, Li, and Saar, 2017). Moreover, there are concerns that speed might also be used by some traders for manipulating market prices and deceive other investors. For instance, the CFTC has charged a U.K. trader (Navider Singh Sarao) of using a "layering algorithm" to manipulate prices (see "Complaint in U.S. *vs.* Navinder Singh Sarao").[8]

Layering is a manipulative scheme that consists in sending limit orders that are not intended for execution and are therefore quickly cancelled after submission. The goal of this scheme is to distort other traders' beliefs regarding the value of the asset and profit from such distorsions. According to the CFTC, N. Sarao's layering algorithm repeatedly submitted large sell limit orders away from the best quotes in the E.mini S&P500 futures limit order book, without intending these orders to be executed (in fact 99% of these were cancelled). The CFTC claims that these orders artificially depressed the E-mini S&P500 price and increased its volatility, enabling N. Sarao to make profits by buying the E-mini S&P500 at depressed prices. Similarly, the French market regulator (AMF) charged Madison Tyler Europe (MTE) (now part of Virtu) or layering in December 2015. However, defining price or market manipulation is difficult, both in legal and economic terms (see Fishel and Ross, 1991 and Kyle and Viswnanathan, 2008). In particular, as Kyle and Viswanathan (2008) point out, it is difficult to distinguish between trading strategies that

[8] See http://www.cftc.gov/idc/groups/public/@lrenforcementactions/documents/legalpleading/enfsaraocomplaint041715.pdf.

undermine both price informativeness and liquidity (which they view as manipulative) from trading strategies that just rationally exploit market power and private information.

**Matching speed and Informational speed.** The previous examples show that two types of speeds matter for traders in financial markets (e.g., arbitrageurs, directional traders, or market makers): (i) "informational speed", i.e., the speed at which traders receive information (about their own trades, changes in quotes, others' transactions etc.) from trading platforms and (ii) "matching speed", i.e., the speed at which orders are processed and filled by the trading platform. Conceptually, matching and informational speeds are different, e.g., a trading platform which is fast in executing incoming market orders could be slow in providing information to traders. In practice, however, they are bundled and difficult to disentangle.

For instance, consider co-location, i.e., the possibility for traders to locate their algorithms in close proximity to a trading platform's server. Co-location is a way to reduce latency by minimizing the physical distance between traders' algorithms and the trading platform. It increases both the speed at which the orders sent by an algorithm are received and processed by a trading platform (matching speed) and the speed at which the algorithm receives information from the trading platform (informational speed).[9] For instance, Nasdaq OMX offers three different types of co-location: basic, premium, and 10G (10 Gigabyts). The premium colocation reduces the latency for order entry (matching speed) and order book information retrieval (informational speed) while the 10G colocation reduces the time from order submission to order confirmation for traders' algorithms (see Brogaard et al., 2015).

The dramatic increase in informational speeds in recent years is just the terminal point of a long historical process. Astute speculators and exchanges have always sought to speed up the transmission of information on trades and quotes using new information technologies (such as carrier pigeons, the telegraph, the telephone, the ticker etc.). The first undersea

[9] Some market participants might exploit only some aspects of co-location that suit best their trading or investment strategies. For instance, some active funds may choose to be co-located when they rebalance their portfolio in order to execute their trades in different stocks simultaneously and minimize their price impact.

transatlantic cable between Europe and the U.S. dates back to 1866 and was quickly used to arbitrage price differentials between Wall Street and the City (see Garbade and Silber, 1978 and Hoag, 2006). Stock tickers were introduced in 1867 and the Boston Stock Exchange and the NYSE were linked by the telegraph in 1889 to speed up the transmission of information on trades and prices between these two markets. Similarly, speculation on advanced access to news (public information) is in no way novel. For instance, around 1835, the reporter Daniel Craig started selling advanced access to news from Europe arriving in the U.S. by steamships. Steamships stopped in Halifax, allowing Craig's agents to read the news and deliver them in Boston through carrier pigeons *before* the steamships arrived in America. Craig was charging $500 for each hour of advance access to the news.[10]

In sum, the value of informational speed has long be recognized by traders and the various strategies exploiting informational speed (cross-market arbitrage, market making and directional trading) are not new. More novel is the scale at which high frequency traders can deploy these strategies. To leverage their investments in informational speed, high frequency traders typically trade in a large number of stocks and process at any given point in time a very large amount of data. Thus, they must automate information processing and the order submission process ("algorithmic trading"). In turn the automation of order submission requires order processing by exchanges to be automated as well. Thus, informational and matching speeds are complementary and they have been feeding on each other (the automation of matching engines by stock exchanges began in the 1970s; see Jain, 2005). This complementarity is important for regulatory interventions, as one can affect the profitability of high frequency trading by reducing informational or matching speeds or both.

# 3 Effects of informational and matching speeds: theory and evidence

There are four channels through which a change in traders' informational and matching speed can affect costs of liquidity provision and gains from trade: (i) search, (ii) adverse

[10] See "Wall Street, 1889: The telegraph ramps up trading speed," Wall Street Journal, July 7, 2014.

selection, (iii) inventory risk, and (iv) competition among liquidity providers.

**Search.** Speed enables traders to find a counterparty faster. In Pagnotta and Philippon (2016), an exchange's matching speed allows buyers and sellers to be matched and realize gains from trade more quickly. Biais, Foucault, and Moinas (2015) argue that, in fragmented markets, informational speed allows traders to detect trading platforms with the best quotes, and thereby to complete their desired trades, faster. For instance, investors can buy fast access to exchange data feed to receive accurate information on posted quotes, or use smart routers that instantaneously compare quotes across trading venues to split their orders across platforms optimally. Similarly, in Foucault, Kadan and Kandel (2014) or Bongaerts, Kong, and Van Achter (2016), a trader's likelihood of trading increases in the speed with which he observes posted quotes. In all these models, trading faster shortens the time required to realize gains from trade. We are not aware of any direct empirical evaluation of this (potentially first order) benefit of trading speed.

**Inventory Risk management.** Speed helps market makers to better manage their inventory risk. Ait-Sahalia and Sağlam (2017) analyze theoretically the optimal pricing behavior of a high frequency market maker. The market maker receives buy or sell market orders at random points in time and bears inventory holding costs. In choosing his quotes, the market maker optimally balances bid-ask spread revenues with his inventory holding cost (as, e.g., in Ho and Stoll, 1981). For instance, after accumulating a long position, the market maker optimally stop quoting on the bid side (or shades his bid price) to reduce his inventory. Ait-Sahalia and Sağlam (2017) characterize the market maker's optimal pricing policy and show that his bid-ask spread is smaller on average when he is faster. Intuitively, trading faster enables the market maker to revise his quotes more frequently and thereby to better manager his inventory position (e.g., to shade his bid more quickly after accumulating a long position).

Brogaard et al. (2015) find evidence for the inventory risk channel. Indeed, they show that traders who buy the fastest co-location services on Nasdaq OMX are predominantly market makers. Moreover, these market makers hold their inventory for a longer period of time after buying the fastest co-location service, which suggests that their inventory holding cost is smaller when they are faster. Consistent with this conjecture, Brogaard et al. (2015) also find that fast market makers' quotes are less sensitive to an increase in their inventory positions than slow ones.

**Adverse selection.** Speed allows traders to get information and use it faster, possibly at the expense of slower traders. For instance, consider the arrival of positive news for the S&P500 index. Dealers in the E.mini futures mark up their quotes while liquidity providers in ETFs on the S&P500 are slow in doing so. In this scenario, a high frequency trader who observes the quote update in the E.mini futures can make a profit by buying the ETFs at stale quotes. This profit is obtained at the expense of dealers in the ETFs' market.

This source of adverse selection (known as "picking off", "sniping" or "free option" risk) is due to differential in speeds of reaction to news. It is by no way specific to modern markets. Speculators buying advanced information on news from Daniel Craig in the $19^{th}$ century (see Section 2) were already exploiting a faster access to news (see Copeland and Galai,1983 or Foucault, Röell and Sandas, 2003 for other examples and models of the picking off risk that predate the development of high frequency trading). However, automation of the trading process and the considerable increase in the speed at which traders can react to news has made picking off risk for liquidity providers much more acute. Accordingly several theoretical papers have analyzed how the acceleration of access and reaction to news (broadly defined, i.e., including quote updates for instance) affects adverse selection costs in financial markets (see, for instance, Hoffman, 2014, Biais, Foucault, and Moinas, 2015, Foucault, Hombert, and Rosu, 2015, Budish, Cramton, and Shim, 2015, Jovanovic and Menkveld, 2016, Foucault, Kozhan, and Wah, 2017, or Menkveld and Zoican, 2017).

Liquidity providers can reduce their exposure to picking off risk by better monitoring the flow of news and cancelling their quotes when they become stale, before they are lift by other traders ("quote snipers"). News arrival is therefore followed by a race between traders attempting to lift stale quotes before they are cancelled and liquidity providers attempting to cancel their quotes before they are picked off. This race is one reason why cancellations-to-trade ratios have considerably increased in recent years and has been one driver of the massive investment in speed observed in the industry (see Foucault, Röell and Sandas, 2003, Biais, Foucault, and Moinas, 2015, Cramton, Budish, and Shim, 2015, or Foucault, Kozhan and Tham, 2017 for models of investments in speed).

Consequently the net effect of an increase in trading speed on adverse selection costs depends on whether it makes liquidity providers relatively faster than quote snipers (directional traders or arbitrageurs) or vice versa. For instance, in Foucault, Kozhan, and Wah (2017), an increase in arbitrageurs' speed *relative* to dealers raises adverse selection costs while a decrease in their relative speed decreases adverse selection costs.

Ultimately, which effect dominates is an empirical question. Several empirical papers find support for the prediction that an increase in trading speed can raise adverse selection costs. Chakrabarty et al. (2015) consider the SEC ban on unfiltered market access (so called "naked access"). This ban is a negative shock on high frequency traders' speed of access to the market since unfiltered access enabled them to directly connect to exchange servers, thereby bypassing brokerage controls. Moreover, it did not affect traders registered as broker-dealers. Hence, it most likely affected high frequency traders that are not specialized in market making. Consistent with the view that fast trading by HFT can raise adverse selection costs, Chakrabarty et al. (2015) find that the unfiltered access ban is associated with a significant drop in adverse selection costs and accordingly a drop in trading costs. Foucault, Kozhan, and Wah (2017) consider a technological change that increases arbitrageurs' relative speed in the foreign exchange market and find that this shock results in an increase in adverse selection costs. In a similar vein, Shkilko and Sokolov (2016) show that rain precipitations, which slow down the microwaves transmission of information between Chicago and New-York, result in a decrease in adverse selection costs in ETFs traded in New York. Again, this finding suggests that slowing down arbitrageurs or directional traders exploiting lagged adjustments of prices across markets reduces adverse selection. Brogaard, Hendershott and Riordan (2016) show that high frequency traders' short selling activity has a negative effect on liquidity because high frequency traders raise adverse selection for other market participants.

There is also evidence that improving the speed at which liquidity providers can cancel their quotes is associated with improvements in liquidity and a reduction in adverse selection costs. For instance, Hendershott et al. (2011) find that the introduction of Autoquote on the NYSE (which allows liquidity suppliers to become more quickly informed about changes in the limit order book) is associated with an improvement in liquidity. Brogaard et al. (2015) show that market makers become significantly better at avoiding adverse selection after upgrading their co-location with Nasdaq OMX and Biais, Declerck and Moinas (2015) also find that fast traders provide liquidity without making losses.

**Competition.** Speed intensifies competition among liquidity providers. Indeed, fast feedback on the state of the limit order book allows market makers to quickly identify profitable trading opportunities for two reasons. First, time priority has value in limit order book markets. Indeed, price discreteness (a non zero tick) implies that limit orders placed ahead of the queue at a given quote in the book are more profitable than limit orders placed further in the queue, as implied by models such as Glosten (1994), Parlour and Seppi (2003), or Foucault and Menkveld (2008). Thus, after a transient increase in the bid-ask spread due to the arrival of one market order, liquidity suppliers have an incentive to race to be first to supply liquidity at price points at which liquidity has vanished. Faster traders are more likely to win this race and pocket the rents associated with time priority (see, for instance, Foucault, Kadan, and Kandel, 2014, Section V.B). The value of time priority is higher in stocks with larger tick sizes (as predicted by models such as Glosten, 1994, Parlour and Seppi, 2003, Foucault and Menkveld, 2008, or Foucault, Kadan, and Kandel, 2014). Thus, high frequency market makers should be particularly active in stocks with large tick size. Consistent with this prediction, Yao and Ye (2017) find that HFTs provide a larger fraction of liquidity in stocks with relatively large tick sizes.

Second, and relatedly, fast reaction to a change in the limit order book enables market makers to undercut their competitors' quotes more quickly. Thus, when market makers receive quicker feedback on the state of the book (e.g., because they can observe it more frequently), quotes become competitive more quickly (see Cordella and Foucault, 1999 and Bongaerts and Van Achter, 2016).

In sum, an increase in the speed at which liquidity providers react to the limit order book should intensify competition among liquidity providers and thereby improve liquidity. There is yet little empirical evidence on this channel. Brogaard et al. (2016) find that constraints on short selling activity by high frequency traders *increases* competition among liquidity providers, which goes opposite to the idea that their presence enhances competition. More empirical research is needed on this question.

**Net effects on market liquidity and efficiency.** Overall, these channels suggest that the effects of speed on the costs of liquidity provision are complex. An increase in speed can increase adverse selection costs for liquidity suppliers while increasing competition among liquidity providers and reducing their inventory holding costs. Empirical studies should therefore carefully analyze the effects of changes in trading speeds (due for instance to technological upgrades) on *each* component of bid-ask spreads separately.[11] Theory predicts that the effect on adverse selection costs should be positive while the effect on inventory holding costs and order processing costs (which also include rents for liquidity providers) should be negative. The net effect on transaction costs is theoretically ambiguous and might vary across samples, which explains why conclusions of empirical studies regarding the effect of high frequency trading on liquidity are not clear cut.

Speed can also affect the pricing efficiency of securities markets. Intuitively, accelerating the flow of information on trades and prices help to make markets better integrated, i.e., it brings prices more quickly in line with no arbitrage relationships. Garbade and Silber (1978) show that the introduction of the telegraph in the U.S. had exactly this effect. In a similar vein, Chaboud et al. (2014) show that the introduction of algorithmic trading on foreign exchange trading platforms has made triangular arbitrage opportunities in currencies markets more short lived. Brogaard et al. (2014) also find that high frequency traders' orders tend to reduce noise in prices, making the latter closer to a random walk.

The social benefit of the gains in pricing efficiency brought up by super fast communication lines between markets is not clear, however. Price differentials for identical assets across markets suggest that gains from trade are not exploited. For instance, if a trader is willing to buy an asset in market A at price higher than the price at which the asset is offered by another trader in market B, both traders should trade together. By buying the asset in market B and selling it in market A, arbitrageurs correct the price differential and help traders located in different markets to realize gains from trade (see Foucault,

[11] Bid-ask spreads in financial markets can be decomposed in three components: (i) adverse selection costs, (ii) inventory costs, and (iii) order processing costs. see Foucault, Pagano, and Röell (2013), Chapter 5 for empirical techniques used to estimate each cost.

Kozhan and Tham, 2017). Doing so at the "speed of light" is valuable only if traders highly discount the time required for realizing gains from trade. More evidence is needed on this question.

Moreover, although high frequency trading makes prices more efficient, they do not necessarily make prices more informative. In fact, high frequency traders' ability to quickly extract signals from order flows (trades and price changes), could increase the rate at which informed investors' informational advantage ("alpha") decays. If this is the case then high frequency trading might reduce the profitability of producing fundamental information and thereby make asset prices less informative about firms' future cash-flows in the long run (see Dugast and Foucault, 2017 and Draus, 2017 for formal analyses of this possibility in the broader context of automated information processing). Consistent with this prediction, Weller (2017) finds a negative association between algorithmic trading activity in a stock and the informativeness of the stock price about future earnings.

# 4 Are financial markets too fast or too slow?

Concerns about the speed at which trading takes place in financial markets has triggered several proposals to slow down trading in financial markets (see the introduction). Implicit in these proposals is the view that market forces alone cannot set the right level of trading speed and that trading speed is excessive for social welfare. Hence, regulatory intervention is needed. What are the market failures that justify such intervention?

This is a difficult question. As explained previously, investment in speeds by some traders generates gains for these traders, possibly at the expense of other traders. For instance, traders who get faster access to information make profits at the expense of liquidity providers who are slow to update their quotes when new information arrives. Market makers who are fast in obtaining time priority capture a larger fraction of the rents associated with time priority at the expense of slower market makers. In these examples, investment in speeds redistributes gains from trade from slow to fast traders. However, per se, this transfer is not sufficient to conclude that fast traders should be slowed down, unless one takes the view that slow traders' welfare is more important than that of fast traders.

A stronger case for regulating speed in financial markets requires asking why (i) market forces will not lead to the optimal level of speed from a social welfare viewpoint and (ii) whether markets are too fast or too slow relative to this social optimum. These are complex questions. To address them, one must develop economic models endogenizing both the demand for speed by investors and the supply of speed by trading platforms, information sellers (such as Bloomberg or Thomson-Reuters) or infrastructure providers (e.g., providers of antenna for microwaves transmission). Some recent models (discussed below) have made progress in this direction but more research is needed in this area.

Pagnotta and Philippon (2016) is a good example of such models. They analyze how trading platforms compete in speed. In their model, the speed of trading on a platform determines the rate at which buyers and sellers are matched. Thus, Pagnotta and Philippon (2016) focus on the provision of matching speed and on the search benefit of speed. They show that trading platforms relax competition by choosing different speed levels, thereby attracting different clienteles (as in models of vertical differentiation). The equilibrium is such that the levels of speed chosen by competing platforms is in general too *low* compared to the level that would maximize welfare. The reason is that choosing a low speed is a way for a trading platform to differentiate from a faster market and retain market power. Pagnotta and Phillipon (2016)'s model however does not consider the role of informational speed.

Biais, Foucault, and Moinas (2015) and Budish et al. (2015) develop models in which traders can make investments to increase their speed of access to information. In Biais, Foucault, and Moinas (2015), traders differ in their private valuations for a risky asset, so that gains from trade exist. Each trader can choose, before learning his private valuation for the asset, to invest in a trading technology to obtain (i) fast access to information about the payoff of the asset (e.g., incoming news about earnings for a stock) and (ii) fast access to trading opportunities. That is, the trading technology enables traders to acquire both informational and matching speed. In this setting, an increase in the mass of fast traders has an ambiguous effect on aggregate welfare. On the one hand, speed has a search benefit: it enables fast traders to find a counterparty more quickly (as in Pagnotta and Philippon, 2016). This is valuable because traders are impatient, so that they value more gains from trades realized earlier than later. On the other hand, as the mass of fast traders increases, adverse selection increases. As a result, bid-ask spreads charged by liquidity providers become larger and, for this reason, some traders optimally abstain from trading (to avoid trading costs). This is a social loss that adds to the real costs of the trading technology. The socially optimal level of investment in speed balances the search benefit of speed with the adverse selection cost of speed and the technological cost. This level is in general strictly positive (i.e., allowing some traders to trade fast is socially optimal).

However, the equilibrium level of investment in speed exceeds the socially optimal level. Indeed, in making their decision to invest in speed, traders balance the speculative and search benefits they can obtain by becoming fast with the cost of being fast (e.g., co-location fees, technological investments etc.) but they do not internalize the negative externality of their decision (their effect on adverse selection cost) on other traders. This negative externality generates a welfare loss since higher adverse selection costs ultimately lead more investors to abstain from trading.

In sum, Biais et al. (2015) suggest two market failures that can justify the regulation of trading speed in securies markets. First, asymmetric information raises adverse selection costs and thereby prevents traders from maximizing gains from trade. Second, investors do not internalize the negative externality that their investment speed imposes on other market participants. So there is excessive investment in speed from a social standpoint. Budish et al. (2015) reach a similar conclusion in a model in which speed is only a way to get quicker access to information. In this framework, speed has no social value and any investment in speed is excessive.

As regulatory intervention, Biais et al. (2015) propose a Pigovian tax on investment in speed as a way to align private decisions with the social optimum. Instead, Budish et al. (2015) propose to change the organization of electronic markets, moving from continuous limit order books to batch auctions (uniform double sided price auctions) ran at frequent points in time (e.g., every second). They show that such a shift reduces incentives to invest in speed for the sole purpose of picking off ("sniping") stale quotes after news arrival. As a result, liquidity is improved and investment in speed is closer to the social optimum (zero in their model).

# 5 Conclusion

Theory and evidence suggest that informational speed (fast access to information) can be harmful for market quality and welfare: it raises adverse selection costs for slow traders, leading to less trading, and possible wasteful investments to be fast. However, theory also suggests that trading speed has benefits: it intensifies competition among liquidity providers, lowers their inventory holding costs, and expedites the search for trading counterparties.

Evidence on the magnitude of these benefits is scarce so far. It is often claimed that high frequency traders have reduced trading costs. However, direct empirical evidence that this is the case is limited and, in any case, the exact channels (inventory holding costs, monitoring, competition, etc.) through which this happened are still not well understood.[12]

One reason is due to specific data challenges. Consider inventory holding costs first and the conjecture that they are smaller for high frequency traders, other things equal (i.e., holding inventory size constant). Ideally, one would like to estimate inventory holding costs for high frequency and non high frequency market makers and compare them (again, other things equal). Estimating inventory holding costs however require data on dealers inventories (for estimations of inventory holding costs for non high frequency market makers, see Hendershott and Menkveld (2014) and references therein). Unfortunately, to our knowledge, there is yet no long time series of individual inventories for a cross-section of high frequency and non high frequency market market makers. Indeed, one would need to observe all the trades made by each market maker in all the different platforms to be able to compute the consolidated inventory. It is not clear that market makers (whether slow or fast) would be willing to provide such data to researchers. Now, consider competition among liquidity providers. Ideally, one would like to analyze how entry by high frequency market makers intensifies competition among liquidity providers and reduce bid-ask spreads. However, identifying the causal impact of such entry is difficult because high frequency traders endogenously choose to be active or not in a given market or under

[12] Moreover, some channels can simultaneously have positive and negative effects on trading costs. For instance, entry by high frequency traders can both increase competition between liquidity suppliers (which reduces trading costs) and raise adverse selection costs for slower traders (which increases adverse selection costs). Measuring separately each effect is therefore very important to have a good understanding of high frequency trading.

specific market conditions. Empiricists must therefore rely on "natural experiments," i.e., events that trigger "exogenous" entries or exits of high frequency traders (i.e., entries or exits unrelated to unobserved variables that could also affect liquidity or price efficiency). Good experiments of this type are rare (for examples, see Brogaard and Gariott (2017) and Breckenfelder (2017)). In either case (inventory costs or competition), researchers should also be able to identify high frequency traders and non high frequency traders. Some existing datasets allow such identification to some extent but for some limited periods of time.

A second, maybe more serious, problem is that some of the benefits of high frequency trading, in particular the utility benefit of reduction in search time, are intrinsically difficult to measure in electronic limit order markets. In contrast to liquidity or price efficiency, there are no easily observable proxies for improvements in gains from trade due to reduction in search frictions in electronic markets. Moreover, it is not even clear that standard measures of liquidity such as bid-ask spreads and market depth constitute good proxies for gains from trade (see Dugast (2017)). To overcome this problem, researchers could rely on structural estimation of models of trading in limit order markets (as for instance Hollifield et al.(2006)). This approach could potentially allow empiricists to quantify the welfare gains or costs of increasing trading speed in securities markets and regulatory actions to control it.

# Bibliography

Ait-Sahalia, Yacine, and Sağlam, Mehmet (2017) "High Frequency Market Making: Optimal Quoting," Working paper, Princeton University.

Baldauf, Markus, and Joshua Mollner (2017), "High-Frequency Trading and Market Performance," Working paper, available at https://ssrn.com/abstract=2674767.

Biais, Bruno, Thierry Foucault, and Sophie Moinas (2015), "Equilibrium Fast Trading," *Journal of Financial Economics*, 116 (2015), 292-313.

Biais, Bruno, Fany Declerck, and Sophie Moinas (2015), "Who supplies liquidity, how and when?, " Working paper, BIS.

Boehmer, Ekkehart, Dan Li, and Gideon Saar (2017), "The Competitive Landscape of High-Frequency Trading Firms ," Working paper, available at SSRN: https://ssrn.com/abstract=2745367.

Bongaerts, Dion, and Mark Van Achter (2016), "High-Frequency Trading and Market Stability," Working paper, available at https://ssrn.com/abstract=2698702.

Bongaerts, Dion, Lingtian Kong, and Mark Van Achter (2016), "Trading speed competition: Can the arms race go too far?," Working paper, available at https://ssrn.com/abstract=2779904.

Breckenfelder, Johannes (2017), "Competition among high frequency traders, and market quality," Working paper, ECB.

Brogaard, Jonathan, Björn Hagströmer, Lars Nordèn, and Ryan Riordan (2015), "Trading fast and slow: Colocation and liquidity," *Review of Financial Studies*, 28, 3407-3443.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2014), "High frequency trading and price discovery," *Review of Financial Studies*, 27, 2267-2306.

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan (2016), "High frequency trading and the 2008 Short Sale Ban," *Journal of Financial Economics*, forthcoming.

Brogaard, Jonathan and Correy Garriott (2017), "High frequency trading competition," Working paper available at https://dx.doi.org/10.2139/ssrn.2435999.

Budish, Eric, Peter Cramton, and John Shim (2015), "The high-frequency trading arms race: Frequent batch auctions as a market design response," *Quarterly Journal of Economics*, 30, 1547-1621.

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega (2014), "Rise of the machines: Algorithmic trading in the foreign exchange market," *The Journal of Finance*, 69, 2045-2084.

Chakrabarty, Bidisha, Pankaj K. Jain, Andriy Shkilko, Konstantin Sokolov (2015) "Speed of market access and market quality: Evidence from the SEC naked access ban, " Working Paper, Wilfrid Laurier University.

Chen, Haoming, Sean Foley, Michael Goldstein and Thomas Ruf (2017) "The value of a millisecond: Harnessing information in fast, fragmented markets," Working Paper, University of New South Wales.

Cordella, Tito, and Thierry Foucault (1999), "Minimum Price Variations, Time Priority, and Quote Dynamics," *Journal of Financial Intermediation*, 8, 141-173.

Draus, Sarah (2017), "High Frequency Trading and Fundamental Trading," Working Paper, University Paris Dauphine.

Dugast, Jérôme, and Thierry Foucault (2017), "Data abundance and Asset Price Informativeness", forthcoming *Journal of Financial Economics*.

Dugast, Jérôme (2017), "Inefficient market depth", Working paper, available at https://ssrn.com/abstra

Fishel Daniel, and David Ross (1991), "Should the lawprohibit manipulation?," *Harvard Law Review*, 503.

Foucault, Thierry, Johan Hombert, and Ioanid Rosu (2016), "News Trading and Speed," *Journal of Finance*, 71, 335-382.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel (2014), "Liquidity cycles and make/take fees in electronic markets," *The Journal of Finance*, 68, 299-341.

Foucault, Thierry, Roman Kozhan, and Wing Wah Tham (2017), "Toxic Arbitrage," *Review of Financial Studies*, 30, 1053?1094.

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

Foucault Thierry, Ailsa Röell, and Patrik Sandås (2003), "Market making with costly monitoring: An analysis of the SOES controversy," *Review of Financial Studies*, 16, 345-384.

Foucault Thierry, Marco Pagano, and Ailsa Röell (2013), "Market Liquidity: Theory, Evidence and Policy," *Oxford University Press.*

Garbade, Kenneth, and William Silber (1978), "Technology, communication and the performance of financial markets: 1840-1975," *Journal of Finance*, 819-832.

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

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

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

Hendershott, Terrence and Albert Menkveld (2014), "Price Pressures", *Journal of Financial Economics*, 114, 405-423.

Hollifield, Burton, R. A. Miller, Patrick Sandas, and J. Slive (2006), "Estimating the gains from trade in limit order markets", *Journal of Finance* 61, 2753-2804.

Ho, Thomas, and Hans Stoll (1983),"The dynamics of dealer markets under competition," *Journal of Finance*, 38, 1053-1074.

Hoag, Christopher (2006), "The Atlantic Telegraph Cable and Capital Market Information Flows," *The Journal of Economic History*, 66, 342-353.

Hoffman, Peter (2014), "A dynamic limit order market with fast and slow traders," *Journal of Financial Economics*, 113, 298-302.

Jain, Pankaj (2005), "Financial market design and the equity market premium: electronic vs. floor trading," *Journal of Finance*, 2956-2985.

Jovanovic, Boyan, and Albert Menkveld, 2016, "Middlemen in Limit-Order Markets," Working Paper, available at SSRN: https://ssrn.com/abstract=1624329 .

van Kervel, Vincent (2015), "Competition for Order flow with fast and slow traders," *Review of Financial Studies*, 23, 2094-2127.

Kirilenko, Andrei, Albert S. Kyle, Mehrdad Samadi, and Tugkan Tuzun (2017), "The Flash Crash: High-Frequency Trading in an Electronic Market," *Journal of Finance*, 72, 967-998.

Kozhan, Roman, and Winh Wah Tham (2012), "Execution Risk in High-Frequency Arbitrage, " *Management Science*, 58, 2131-2149.

Kyle, Albert S., and S. Viswanathan (2008). "How to define illegal price manipulation?," *American Economic Review*, 98, 274-279.

Lescourret, Laurence, and Sophie Moinas (2017), "Fragmentation and Strategic Market-Making," Working Paper, available at SSRN: https://ssrn.com/abstract=2498277.

Michael Lewis (2014), "Flash Boys: A Wall Street Revolt," *W. W. Norton & Company.*

Menkveld, Albert (2013),"High frequency trading and the new market makers," *Journal of Financial Markets*, 16, 712-740.

Menkveld, Albert, and Marius Zoican (2017), "Need for Speed? Exchange Latency and Liquidity," *Review of Financial Studies*, 30, 1188-1128.

Miller, Rena S., and Gary Shorter (2016), "High Frequency Trading: Overview of Recent Developments," *Congressional Research Service.*

Parlour, Christine, and Duane Seppi (2003), "Liquidity-Based Competition for Order Flow," *Review of Financial Studies*, 16, 301-343.

Pagnotta, Emiliano, and Thomas Philippon (2016), "Competing on Speed," NBER Working Paper 17652.

Shkilko, Andriy, and Konstantin Sokolov (2016), "Every cloud has a silver lining: Fast trading, microwave connectivity and trading costs," Working Paper, available at SSRN: https://ssrn.com/abstract=2848562.

Weller, Brian M., (2017) "Does Algorithmic Trading Deter Information Acquisition?" Available at SSRN: https://ssrn.com/abstract=2662254.

Yao, Chen, and Mao Ye (2017), "Why Trading Speed Matters: A Tale of Queue Rationing under Price Controls," forthcoming in the *Review of Financial Studies.*

Zhang, Sarah (2017), "Need for speed: Hard information processing in a high-frequency world," *Journal of Futures Markets*, 2017, 1-19.