

# HFT AND MARKET QUALITY



**BRUNO BIAIS**  
Directeur de  
recherche  
Toulouse  
School of  
Economics  
(CRM/CNRS  
- Chaire FBF/  
IDEI)



**THIERRY  
FOUCAULT\***  
Professor of  
Finance  
HEC, Paris

## I. INTRODUCTION

The rise of high-frequency traders (hereafter HFTs), that is, investors using computers to trade securities over extremely short time intervals, is a major change in securities markets over the last decade. Some analysts view this evolution as beneficial as automation reduces the cost of liquidity provision. Others argue that HFTs essentially exploit their fast access to markets to make profits at the expense of other traders. Fears have also been expressed that HFT could be a source of market instability.

Understanding the effects of HFTs on market quality is important, as the latter will ultimately determine long-term investors' welfare and the cost of capital for firms. The goal of this review is to provide an overview of academic findings on HFT and to examine policy options in light of these findings.

Before stating our main conclusions, the following preliminary remarks must be made:

■ HFTs' strategies are heterogeneous and thereby one expects the effects of HFTs on market quality to depend on the type of their strategy.

■ To analyze trading strategies, it is useful to differentiate orders that take liquidity and trigger immediate trades (market orders or marketable limit orders) from orders that offer liquidity without immediate trade (non-immediately marketable limit orders). For brevity we hereafter simply refer to the former as market orders and to the latter as limit orders. To encourage the supply of liquidity, exchanges often offer rebates or subsidies to limit orders.

■ There exist very few empirical studies on the effects of HFT on market quality, due to the difficulty of obtaining data on HFTs' orders.

With this caveat in mind, our main conclusions are presented below.

Fast access to markets can be used both to (i) reduce intermediation costs and (ii) obtain information in advance of other market participants. A reduction in intermediation costs can benefit all market participants if competition among intermediaries is strong so that the cost reduction is passed to final investors. In contrast, trading on advance information is a source of adverse selection, which hinders the efficiency of risk-sharing in financial markets.

Empirical studies have found that market orders submitted by HFTs contain information, in the sense that they anticipate very short-run price changes. This supports the view that fast access to market data provides HFTs with an informational advantage. As a result, HFTs obtain small, but positive, profits per trade when using market orders. In contrast, HFT's profits on limit orders can be negative before one takes into account the rebates from exchanges.

So far, no significant negative effect of computerized trading on market quality has been evidenced. Yet, the possibility of trading on advance information on market data can generate negative externalities, e.g., induce less market participation by slow traders, overinvestment in trading technologies, and an increase in systemic risk. To mitigate these externalities, we recommend developing trading mechanisms that cater specifically to slow traders. This could require regulatory intervention to overcome exchanges' conflict of interests. We also recommend imposing minimum capital requirements for HFT firms. Moreover we emphasize the need for stress tests to evaluate the robustness of the market to technological problems or high-frequency firms' failure, and for pilot experiments, to assess and fine tune trading rules designed to slow the trading process.

This review is organized as follows. In Section II, we first define HFT and describe the data that researchers have used to analyze the effects of HFTs. This is important

\* Corresponding author: Thierry Foucault  
Dpt of Finance, HEC  
1 rue de la Libération, 78351 Jouy en Josas  
Tel: +33 (0)1 39 67 95 69  
Email: foucault@hec.fr

We thank the Institut Louis Bachelier for financial support. Many thanks to an anonymous referee, for helpful comments.

because researchers still lack account-level data on HFTs (hereafter HFTs). Hence, current results are mainly based on indirect proxies for the activity of HFTs or aggregate data for a subset of HFT firms in the market. These limitations are important to keep in mind when interpreting the empirical findings. We also discuss findings regarding the profitability of HFTs.

In Section III, we consider how, for a given asset, HFT affects (i) “price discovery,” that is, the speed at which new information is impounded into asset prices (ii) liquidity, which is often measured by proxies for the costs borne by investors when they want to buy or sell an asset (e.g., the quoted or effective bid-ask spread) and (iii) short-term volatility (or variability) of the stock, which can be proxied by estimating the standard deviation of returns using high-frequency data. We stress that a major challenge for empiricists is to devise experiments or find instruments to distinguish simple correlations from a true causal effect of HFT on market quality. We also analyze the effect of HFT on market stability and systemic risks.

Finally in Section IV, we review some of the policy options regarding HFT and formulate policy recommendations.

## ■ II. HFT: DEFINITION, DATA, AND PROFITABILITY

In this section, we first define more precisely what is HFT and what HFTs do (Section II.1). We then discuss how researchers have identified HFTs empirically and measured their importance in trading activity (Sections II.2 and II.3). We finally discuss evidence regarding the profitability of HFTs (Section II.4).

### II.1. HFT STRATEGIES ARE HETEROGENEOUS

HFT refers to trading strategies that have two characteristics: (i) they rely on very fast access to trading platforms and market information and (ii) they are highly computerized.

**HFT and algorithmic trading.** HFTs are part of a broader group of traders called algorithmic traders. Algorithmic traders use computer programs to implement investment decisions and trading strategies. However, not all these strategies rely on speed. For instance, brokers often split large orders over time and between trading platforms to achieve small execution costs. They increasingly rely on computers to both determine the optimal splitting strategy for a given order and implement order splitting strategies, yet these order splitting strategies do not necessarily require super fast access to markets. Hence, although HFTs are algorithmic traders, not all algorithmic traders are HFTs.

Speed is critical for HFTs because they specialize in exploiting very short-lived profit opportunities with a “winner takes all” flavor, that is, opportunities whose value is much higher for the first investor who grabs them (see examples below).

To be fast, HFTs strive to minimize so called “latencies:” essentially the time it takes for them to receive “messages” (for instance, a quote update or the status

of their orders) from trading platforms, process this information, and react to it by sending back new orders (market orders, limit orders, cancellations) based on this information.<sup>1</sup> Hasbrouck and Saar (2012) show that some traders react extremely rapidly (sometimes within less than 2 to 3 milliseconds) to market events, such as, e.g., quote improvements.

Such fast reactions are possible only with the help of computers, which can process vast amount of information and act on it much faster than humans.<sup>2</sup> Speed is also achieved by optimizing the access to market information. For instance, HFTs buy “co-location services” that allow them to position their servers in very close proximity to exchanges’ servers. This proximity saves on transmission delays and provides HFTs with quicker access to information on market data than other participants.

As market conditions change frequently, HFT firms can adjust their holdings of a stock very frequently. For instance, they can enter and exit positions or submit orders and then cancel them over extremely short periods of time (e.g., a few milliseconds). For instance, Kirilenko *et al.* (2010) find that HFTs reduce half of their holdings in about two minutes on average.

Because many of their orders are cancelled or modified, HFTs have increased the order-to-trades ratio in securities markets. For instance, Gai, Yao, and Ye (2012) find that a very high fraction of orders are cancelled on Nasdaq stocks (this fraction exceeds 90% for the 118 stocks in their sample) and a large fraction of orders are short-lived (30% of all orders with a life less than one second are cancelled after 5 milliseconds or less).

**Heterogeneity in HFT firms and HFT strategies.** Different types of institutions engage in HFT: proprietary trading firms (e.g., GETCO, Optiver, Tradebot), proprietary trading desks of a broker-dealer firm (e.g., Goldman Sachs or Morgan Stanley) or hedge funds (e.g., Citadel). One can broadly classify HFT trading strategies in five groups: (i) market-making; (ii) arbitrage; (iii) directional trading; (iv) structural and (v) manipulation. None of these strategies are completely new: arbitrage, market-making, informed trading and manipulation have a long history in securities markets. What is new is the use of computing capacity and speed to implement these strategies. Some HFTs firms specialize in some of these trading strategies, very much like hedge funds tend to be specialized. For example, Baron *et al.* (2012) study 31 HFT firms and show that 10 of them complete more than two thirds of their trades with market orders, while 10 other firms complete almost 90% of their trades with limit orders.

**High frequency market-making.** High frequency market-makers primarily submit limit orders that provide liquidity to other traders. For instance, Jovanovic and Menkveld (2010) study one high frequency market-maker in Dutch stocks constituents of the Dutch stock index. They find that this market-maker provides liquidity in about 78% (resp. 74%) of the transactions in which he is involved on Chi-X (resp. Euronext). Brogaard (2011b) also provide evidence that some HFTs do indeed act as liquidity providers. In particular he finds that HFTs in his sample follow a price reversal (or “contrarian”) strategy: that is they buy (sell) stocks whose prices have been declining

(increasing) in the last 10 to 100 seconds. Kirilenko *et al.* (2010) reach a similar conclusion. HFTs in their sample tend to build up their positions (in futures contracts on the S&P500 index) when prices are dropping and exit their positions when prices are increasing.

Speed is important for market-making for several reasons. First, it enables market-makers to quickly react to transient increases in market illiquidity. For instance, the bid-ask spread for a stock may widen after the arrival of a large market order consuming the liquidity available at the best quotes (see Biais, Hillion, and Spatt (1995)). This increase in the bid-ask spread above the competitive level generates a profit opportunity for traders who can intervene by submitting new limit orders within the bid-ask spread, provided they are quick enough to obtain time priority at the best quotes.<sup>3</sup> Investing in speed is therefore a way to capture a larger fraction of these profit opportunities (see Foucault, Kadan, and Kandel (2013) for a theoretical model and Hendershott and Riordan (2012) for empirical evidence).

Second, fast reaction to new information (news releases or quote updates in other related securities) mitigates market-makers' exposure to the risk of being picked off (Copeland and Galai (1983)). Indeed, a dealer's quotes may become stale when news arrive if the dealer does not quickly cancel and resubmit his limit orders to account for the new information. Else, fast traders will hit the dealer's quotes with buy (sell) market orders when his quotes undervalue (overvalue) the security, inflicting a loss to the dealer. Thus, a fast reaction to news for dealers is a way to reduce the likelihood of such a loss and therefore bear lower market-making costs.<sup>4</sup>

Last, speed is a way for market-makers to manage their inventory risk more efficiently. For instance, in today's markets, a stock is often traded in several trading platforms. This possibility implies that there can be a mismatch in "space" between buyers and sellers of the same stock. For instance, a French institutional investor might place a large buy order on Euronext while a U.K. institution places a sell order on Chi-X. High frequency market-makers operating on both platforms can step in in this case by, say, buying the stock on Chi-X at, say, a price of €100 and then reselling it quickly on Euronext at €100.1 or vice versa. By trading fast across trading platforms, the market-maker is able to reduce the time during which his inventory is at risk and thereby his inventory holding costs.

**High-frequency arbitrage.** HFT is also used to exploit arbitrage opportunities, *i.e.*, deviations from parity for the prices of related assets, by taking simultaneous long and short ("hedged") positions in these assets. Speed in this case is important because some arbitrage opportunities are very short-lived and almost riskless to exploit: thus, the first trader detecting the opportunity will exploit it fully, leaving no profit for slow arbitrageurs.

There are several examples of very short-lived arbitrage opportunities. For instance, when a stock is traded on multiple platforms, its ask price on one platform may become temporarily smaller than its bid price on another platform, either because liquidity providers on one plat-

form are slow to adjust their quotes when new information arrives or because one dealer willingly shades his quotes relative to the fundamental value of an asset to quickly unwind his inventory. This "crossed market" situation gives rise to a straightforward arbitrage opportunity whereby one can buy the security at a one price and then resells it immediately at a larger price. Exploiting the opportunity however requires a very fast reaction because (i) computers can detect crossed quotes almost instantaneously and (ii) liquidity providers may adjust their quotes as soon as they realize that quotes are crossed.<sup>5</sup>

Another, related example, are triangular arbitrage opportunities in currency markets. At any point in time, one can trade dollars against euros directly in the euro/dollar market or indirectly by converting euro in pound (or any other currency) and then pounds in dollar (or vice versa). If, for instance, the cost of buying 1 dollar directly with euros is less than the number of euros obtained by converting one dollar in euros indirectly then there is a triangular arbitrage opportunity. Using high frequency data in three currency pairs (euro/dollar, pound/dollar and pound/euro), Foucault, Kozhan and Tham (2012) find that triangular arbitrage opportunities arise frequently (there are 35 to 40 arbitrage opportunities per day on average in their sample) and are very short-lived (they last less than one second).

As high frequency arbitrage opportunities are very short-lived, they require very fast access to the market and are best exploited with market orders. Hence, in contrast to high frequency market-makers, high frequency arbitrageurs mainly use market orders. This feature suggests to use the type of orders (limit vs market) used by HFTs in order to identify the effects of various strategies (*e.g.*, market-making vs arbitrage) that they might use.

That high-frequency arbitrage opportunities are short lived suggests that correcting them via HFT might not be very useful: Without the intervention of HFTs, the arbitrage opportunity would also have been corrected fast, although slightly less rapidly. It is not clear that a decline from, say, 30 seconds to 5 milliseconds is extremely valuable for society.

**Directional strategies.** These strategies consist in taking a directional bet in one asset in anticipation of an impending price change. In contrast to arbitrage strategies, directional strategies do not necessarily involve long/short positions in multiple assets, but they require the acquisition of some signals that help to forecast future price movements. Computers in this case are useful because they can react and process a myriad of signals before humans, even if these signals are already public.

The signals used by HFTs to establish directional bets are very diverse. For instance, Jovanovic and Menkveld (2011) or Zhang (2012) show that HFTs use index futures price to establish positions in underlying stocks. HFTs can also trade on news (unscheduled or scheduled) regarding a stock (see "Computers that trade on the news," *The New-York Times*, May 2012) and data vendors such as Bloomberg, Dow-Jones or Thomson Reuters now provide pre-processed real-time news feed to HFTs.

Brogaard, Hendershott, and Riordan (2012) find that HFTs react to information contained in limit order book updates, market-wide returns, and macro-economic announcements. The value of these signals decays quickly with time because they are made public to all investors. Thus, they can generate profits only if they are exploited very quickly. For instance, Scholtus et al. (2012) show that speed is critical for investors trading on macro-economic announcements. For a sample of 707 announcements, they find that a delay of 300 milliseconds in reacting to the announcement reduces the returns on a strategy exploiting the informational content of the announcement by about 0.44 bps.

Market orders move prices (buy market orders push prices up and sell market orders push prices down). Hence, one way to anticipate future price movements is to anticipate future order flow (the imbalance between the volume buy and sell market orders over a given time interval). Directional strategies that are based on such anticipation are referred to as “anticipation strategies.” A straightforward way to anticipate future order flow is to have direct information on impending trades by one specific investor. Broker-dealers are often well placed to have such direct order flow information because they act as intermediaries between final investors, but they are not allowed to use it for proprietary trading as this would be front running. There are, however, other ways for traders to forecast future order flow or price movements associated with flows. For instance, orders are increasingly split over time which implies that order flow is positively autocorrelated. This positive autocorrelation in turn implies that future order flow can be forecast from past order flow. Faster access to data on trades enables one to better forecast future trades and therefore future price movements.

Using Nasdaq data, Hirschey (2011) finds evidence consistent with the use of order anticipation strategies by some HFTs. Specifically, net order imbalances from HFTs (computed as the difference between the number of shares purchased with marketable orders minus the number of shares sold with marketable orders by HFTs) are positively correlated with lagged, contemporaneous, and future net order imbalances from non HFTs. Moreover, HFTs’ order imbalances are positively correlated with future returns (over time interval of up to 300sds). One possible explanation is that HFTs react to public information faster than non HFTs. However, Hirschey (2011) shows that his results are unchanged when he excludes periods within five minutes of intra-day news. Overall, Hirschey (2011)’s findings are consistent with HFTs forecasting price pressures from other investors and trading on these forecasts.

**Structural strategies.** According to the SEC, “structural strategies” exploit specific features of market structure. The profitability of these strategies is therefore dependent on market organization and could be altered when trading rules change. For instance, high frequency arbitrageurs exploiting crossed quotes exploit the fact that markets are now heavily fragmented. Obviously crossed quotes would not exist if all trading in a stock was centralized in a single market and this source of profits for arbitrageurs would disappear.

Another example pointed to by the SEC is the case in which a high-frequency trader uses his fast access to market data (through, e.g., co-location) to pick off dealers who update their quotes slowly. For instance, the high-frequency trader might observe an increase in bid and ask quotes on platforms A, B and C with quotes on platform D left unchanged. HFTs will then be likely to pick off those slow limit orders. This strategy is “structural” insofar as it would be less profitable if all traders had an equal speed of access to market data.<sup>6</sup>

Yet another example is described in McNish and Upson (2011). In the U.S, the order protection rule (or “no trade through” rule) requires market orders to be routed to the trading platform posting the best consolidated quotes (the “NBBO”) at the time the order is received. For practical reasons, according to the “benchmark quotes exception rule”, the SEC considers that there is a violation of this rule if a trade happens on a platform at a quote inferior to the national best bid and offer prices prevailing over the 1 second time before the trade (the so called “benchmark quotes”). As a result, some trades may not comply with the order protection rule and happen at quotes anywhere between the NBBO at the time of the trade and the benchmark quotes. McNish and Upson (2011) find that about 8% of trades fall in this category. They argue that these trades happen because slow traders see quotes with a delay and therefore route their orders to the “wrong” platform. McNish and Upson (2011) argue that these routing “errors” can be exploited by strategic liquidity suppliers who observe quotes updates in real time faster than liquidity demanders. When an improvement in the NBBO happens on one platform, fast liquidity suppliers deliberately choose to maintain non competitive quotes (outside the NBBO but within benchmark quotes) on other platforms instead of matching or improving upon the NBBO. They may then execute marketable orders at non competitive prices from slow traders who mistakenly miss the opportunity of trading at the best consolidated quotes.

**Manipulation.** One concern is that some HFTs use their fast market access to engage in market manipulation.

For instance, the SEC has pointed to “momentum ignition strategies” that consist in submitting market buy (sell) orders to spark an upward (downward) price movement in the hope that other traders will wrongly jump in the bandwagon and amplify the movement. The high-frequency trader igniting the price movement can then quickly unwind his position at a profit by either selling at artificially inflated prices or buying at discounted prices. Such momentum ignition strategies erode the predictive power of past order flow to predict future order flow, discussed above in our analysis of directional strategies. HFTs are better able to filter this out than slow traders, however. Either, because they are themselves generating the noise in the signal, or because their fast access to the data enables them to react swiftly to the unwinding of others’ momentum ignition strategies.

HFTs have also been accused of engaging in “smoking.” This strategy involves posting alluring limit orders to attract slow traders, then rapidly revising these orders

onto less generous terms, hoping to execute profitably against the incoming flow of slow traders' market orders.

Yet another strategy has been nicknamed "spoofing." Suppose the high-frequency trader's true intention is to buy. Paradoxically, he or she will initially place limit orders to sell in the order book. These orders are not intended to be executed. Therefore they are placed above the best ask. And, since the high frequency trader is faster than the other market participants, he or she can rest assured he or she will have time to cancel the sell orders before they are executed if good news reach the market. With this assurance in mind, the high frequency trader places a sequence of limit sell orders above the best ask, potentially for very large amounts. The hope is to scare the market and induce some naïve participant to sell ... against the limit order to buy the high frequency trader will have discretely placed meanwhile.

"Quote stuffing" is another strategy that has raised concerns. For some strategies, relative speed rather than absolute speed is important. For instance, for a trader taking advantage of stale quotes or crossed quotes, it is important to be fast relatively to the other traders. One way to do so is to slow down other traders by sending a very high number of messages (orders that are subsequently cancelled) just to reduce the speed at which exchanges can inform other traders or process their messages.

Gai, Yao, and Ye (2012) provide interesting empirical evidence consistent with strategic "quote stuffing." They exploit the fact that there are 6 different channels of communication for stocks traded on Nasdaq and stocks are randomly affected among each channel. For instance, stocks with ticker symbols from A to B are affected to channel 1, C to D channel 2 etc. Gai, Yao, and Ye (2012) show that messages in a stock in a given channel covary more strongly with messages for other stocks in the same channel than with stocks in other channels. Moreover, this covariation falls down significantly when a stock is reallocated to another channel because its ticker symbol changes. These findings are surprising since stocks are randomly assigned across channels. Gai, Yao, and Ye (2012) argue that they reflect the fact that HFTs deliberately overflow stocks in one channel with messages when they attempt to exploit a profit opportunity in one stock in this channel.

## II.2. DATA ON HFTS ARE YET LIMITED

To analyze the effects of HFT, one needs to measure it. It is important to bear in mind the limitations of the data, when interpreting empirical results on HFT. In this section, we discuss how HFT activity was measured by the various studies surveyed in this review (a synoptic view is given in Table 1).

The first approach is to build a proxy for the activity of HFTs using data on submitted orders and the speed at which these orders are submitted. For instance, Hendershott *et al.* (2011) and Boehmer *et al.* (2012) note that the ratio of executions to order submissions is lower for algorithmic traders. Thus, Hendershott *et al.* (2011) and Boehmer *et al.* (2012) propose to measure algorithmic trading activity using the number of messages (orders)

normalized by trading volume over some time period (e.g., each month). Hasbrouck and Saar (2012) use the fact that the cancellation of a limit order by a trader followed by the resubmission of another order by the same trader (a "linked message") in less than one second is likely to come from HFTs. Using this intuition, Hasbrouck and Saar (2012) use the number of linked messages ("strategic runs") per 10 minutes interval to build a measure of algorithmic trading.

One drawback of this indirect approach is that the proxies used for HFT may also capture the activity of algorithmic traders operating at lower frequencies (like for instance brokers using algorithms to execute orders at low costs for their clients). In fact, studies using this approach interpret their findings as describing the effects of algorithmic trading, rather than HFT strictly speaking.

In the second approach, researchers use data on the trades of a subset of HFT firms. Several studies (Brogaard (2011a), Brogaard (2011b), Brogaard, Hendershott and Riordan (2011), and Zhang (2012) use data provided by the Nasdaq to academics. These data report aggregated trades for 26 firms identified as HFTs by Nasdaq in 120 randomly selected stocks listed on Nasdaq and the NYSE (over the period 2008 and 2009). Henceforth, we refer to this dataset as the "Nasdaq sample." Nasdaq categorized a firm as an HFT if it engages in proprietary trading only, its net position is often zero, and its limit orders tend to be short-lived. By construction, the Nasdaq sample excludes HFT desks from broker-dealers such as Goldman Sachs, Morgan Stanley, or Merrill Lynch.

The exclusion of some important HFT desks from the Nasdaq sample limits the inferences that can be made from this sample. Moreover, the trades and orders of the 26 trading firms represented in the Nasdaq sample are aggregated.<sup>7</sup> Thus, researchers using these data cannot categorize HFTs in different groups according to, for instance, the type of orders they mainly use (limit or market), the size of their inventories, the frequency with which they post competitive quotes etc. As a result, inferences about the effects of HFTs' trades or orders on market quality are likely to be driven by the strategy that predominates among the HFTs present in the Nasdaq sample, even though HFTs in these data are heterogeneous.

The last approach is to use account-level data for trades made by HFT firms (e.g., Menkveld (2010), Kirilenko, Kyle, Samadi and Tuzun (2010) and Baron, Brogaard and Kirilenko (2012), Malinova, Parks and Riordan (2012), Hagströmer and Norden (2012), Hirschey (2011)). For instance, Kirilenko, Kyle, Samadi and Tuzun (2010) and Baron, Brogaard and Kirilenko (2012) use data on trades by market participants in the CME e-mini S&P500 futures contracts (over two different sample periods, see Table 1). A unique feature of these data is that they provide account-level identification of the traders in each transaction. The authors use these data to identify a subgroup of accounts as being held by HFTs. Specifically, in Kirilenko, Kyle, Samadi and Tuzun (2010) and Baron, Brogaard and Kirilenko (2012), a trader-account is categorized as that of a high-frequency trader if the trader (i) accounts for a sufficiently large fraction of total number of trades, (ii) has a low inventory at

the end of the day (e.g., less than 2% of the total number of futures contracts traded during the day) and (iii) experiences relatively low variations in his inventory position (relative to the number of contracts traded).

By construction, HFTs in Kirilenko *et al.* (2010) or Baron *et al.* (2012) have relatively small inventories and have fast mean reverting inventories. One problem with this approach is that it may select HFTs with a specific trading style (in particular, HFTs paying much attention to inventory risk) while excluding others (e.g., those taking large inventory positions across markets for arbitrage purposes).

Hagströmer and Nordén (2012) use member-level data from the Stockholm Stock Exchange (SSE) for 30 stocks listed on this market (in August 2011 and February 2012). With the help of Nasdaq OMX (the owner of the SSE), they classify members of the SSE in three groups: (i) HFTs (29 members), (ii) non HFTs (49 members), and (iii) hybrid firms (firms that have both agency and proprietary trading activities). The validity of this classification is supported by the fact that 98.2% of all messages initiated by members in the first group are flagged as coming from algorithms. Interestingly, Hagströmer and Nordén (2012) further decompose their group of HFTs in two subgroups: (a) “market-makers” (about 4 to 5 members) and (b) “opportunistic traders.” Firms in the first group provide liquidity to other participants, while opportunistic HFTs use more directional or arbitrage strategies.

Hagströmer and Nordén (2012) show that these two groups of HFTs exhibit very different behaviors. For instance, opportunistic HFTs have a significantly lower message-to-trade ratios than market-makers and the average size of their inventory is significantly higher. Hence, these findings suggest that using these variables to define HFTs is not innocuous. For instance, classifying traders as HFTs if they have a high message-to-trade ratio and small inventories increases the chance of selecting HFTs engaged in market-making. Hagströmer and Nordén (2012) also find that market-makers account for the lion’s share (62% to 80%) of the trading volume by HFTs in their sample.

Finally, it is worth stressing another limitation of existing datasets used in empirical studies on HFT. They typically use data on trades by HFTs in only one asset or one market while in reality HFTs are likely to take positions in multiple markets at the same time. The lack of cross-market data for HFTs can affect inferences, as illustrated by Menkveld (2010). This study uses data on one high-frequency trader active both on Euronext and Chi-X (his sample contains 14 Dutch stocks). Menkveld (2010) shows that this high-frequency trader behaves very much like a market-maker. In particular, the vast majority of his trades (78.1% on Euronext and 78% on Chi-X) occur when his bid or ask quotes are hit on Chi-X or Euronext. Interestingly, Menkveld (2010) shows that the high frequency market-maker’s aggregate inventory position across markets (Chi-X and Euronext) mean reverts (as one would expect for a market-maker) while its position in each individual market does not. This observation underscores the importance of having data

on HFTs’ trades across markets to make inferences about their behavior.

### II.3. IMPORTANCE OF HFT

HFTs can affect market quality if they account for a large fraction of trading activity. Preliminary evidence from empirical analyses using data on HFTs indicates that this is the case.

For instance, Brogaard (2011a) finds that the 26 HFTs in the Nasdaq sample participate in 68.5% of the dollar volume traded on average and account for a larger fraction of the trading volume in large capitalization stocks than in small capitalization stocks. Hirschey (2011) also find that HFTs in his sample are more active in large than in small stocks (41% vs. 15%). HFTs in Kirilenko *et al.* (2010)’s sample account for 34.22% of the daily trading volume in the S&P500 index (for 4 days in May 2010). The high frequency market-maker studied by Menkveld (2010) participates in about 16% of all trades in large stocks and 7% in small stocks. Finally, Hagströmer and Nordén (2012) find that HFTs in their sample account for about 26% to 52% of all trades for the 30 stocks in their sample.

Overall, empirical studies suggest that HFTs account for a significant fraction of trading volume (ranging from 1/3 to 2/3) and that HFTs’ activity might systematically vary according to securities’ characteristics (e.g., small vs. large stocks).

### II.4. PROFITABILITY OF HFTS

It is interesting to study the profitability of HFTs for several reasons. First, one would like to know whether this activity is competitive, that is, whether profits of HFT desks have been declining as the number of HFTs increased. Second, understanding the distribution of trading gains among HFTs and non HFTs is helpful to predict and interpret the effects of HFTs on market quality. For instance, if HFTs placing market orders obtain positive expected profits then they inflict a trading loss on liquidity providers. If this is the case, liquidity providers may then worsen the terms at which they provide liquidity. Third, the size of HFTs’ profits can determine the extent to which a tax on HFTs can affect their activity. Finally, the historical distribution of HFTs’ trading profits is informative about their exposure to tail-risks (risks of large but unfrequent losses).

Hendershott *et al.* (2011) find that the daily average profit per \$10,000 traded for the HFTs in their sample is \$1.45 before fees and \$1.14 after fees.<sup>8</sup> Interestingly, HFTs make significantly larger profit with market orders than with limit orders. Specifically, before fees, their daily average profit per \$10,000 traded is \$2.27 on market orders and -\$0.29 on limit orders. On Nasdaq (as in many other exchanges), limit orders receive a rebate (pay negative fees) funded by a charge on market orders.<sup>9</sup> After accounting for these fees, HFTs’ daily average profit per \$10,000 traded is \$0.5 on market orders and \$1.4 on limit orders.

Baron *et al.* (2012) provide a more detailed analysis of HFTs’ profits using their data on trades in the CME e.mini S&P500 futures in August 2010. They also find that HFTs’ profits

per trade are small, of the order of \$0.77 per contract on average. Furthermore HFTs who predominantly use market orders obtain significantly higher profits on average than HFTs who predominantly use limit orders (\$0.93 vs. \$0.33 per contract). The risk borne by the two types of HFTs is different however: the standard deviation of the per-contract profit of the former HFTs is higher than that of the latter (5.29 vs. 1.62) and as a result the annualized Sharpe ratio of the former is a bit smaller (8.33 vs. 9.94).

Interestingly, Baron *et al.* (2012) can analyze the distribution of trading gains between various types of traders since they have exhaustive trader-level information for all participants in the CME e.mini S&P500 futures. Thus, they compute the average profits of each type of participants conditional on trading with another type. They find that HFTs who predominantly use market orders earn profits. In contrast HFTs who predominantly use limit orders do not systematically earn profits. For instance, they are losing money when trading against HFTs relying on market orders, non-HFT market-makers or small traders (retail investors). Hence, HFTs using market-making strategies do not appear to use superior information on future price changes in contrast to those using market orders.

Menkveld (2010) finds an average profit per trade of €0.88 for the high-frequency market-maker in his sample. He shows that this market-maker earns negative positioning profit (-€0.68). That is, on average, the high frequency market-maker considered by Menkveld (2010) is exposed to adverse selection: he tends to buy the asset before price declines and to sell it before price increases. As predicted by theory (e.g., Glosten and Milgrom (1985)), he makes up for these losses by charging a sufficiently large spread. The spread revenue of the market-maker on limit orders is much higher on Chi-X than on Euronext (€2.52 vs €1.11 per trade) because Chi-X (like Nasdaq) offer rebates to traders submitting limit orders in case of execution. Overall, the high frequency market-maker achieves a Sharpe ratio of 9.32.

HFT market-makers are found to earn razor-blade profits, and sometimes lose money to slower participants. HFT using mainly market orders, which are likely to follow directional or arbitrage strategies, earn larger profits, and do so at the expense of other market participants. In all cases, however, profits per trade are rather small. Yet, HFTs' Sharpe ratios are very large. As their dollar profits per trade are small, such high ratios reflect the relatively small dispersion of their returns, which could be due to efficient risk-management techniques. On the other hand, one can't rule out the hypothesis that HFT is exposed to the risk of extreme losses in very unfrequent states of nature. One should also note that these profits don't take into account the investment costs associated with the setup of a HFT desk.

### ■ III. EFFECTS OF HFT ON MARKET QUALITY

#### III.1. THEORY

Since HFTs' strategies are not new, one can use standard economic tools to predict their potential effects on market

quality. On the one hand, automation and fast reaction to information and market events should allow market-makers to provide liquidity at lower costs. In this case, high frequency market-making should improve market liquidity and ultimately traders' welfare by lowering intermediation costs. On the other hand, if HFTs obtain information faster than other market participants, then slow traders are at an information disadvantage relative to fast traders. In this case, HFT induces adverse selection: slow traders are likely to systematically lose money to HFTs.

Biais, Foucault, and Moinas (2012) analyze the welfare consequences of adverse selection due to unequal speed of access to information. In their model, HFT brings both benefits and costs in terms of welfare. On the one hand, it increases the likelihood that investors can find a mutually profitable trading opportunity. For instance, HFTs can easily trade across multiple trading platforms, easing the transfer of assets from investors with low valuations for the asset to investors with higher valuations when these investors are located in different markets. These trades are welfare enhancing. On the other hand, it generates a negative externality by increasing the risk of adverse selection, which raises the cost of trading (price impacts, i.e., market illiquidity) for all traders. For this reason, some traders (especially the slow ones) decide to trade less frequently when more investors are HFTs and the net effect of an increase in the level of HFT on trading volume is ambiguous: more HFT raises the frequency with which traders find a trading opportunity but traders abstain from exploiting trading opportunities more frequently because trading costs are higher.

Biais *et al.* (2011) show that investment in HFT technologies is in general excessive relative to the socially optimal level.<sup>10</sup> Indeed, in making their investment decision in fast trading technologies, investors do not internalize the negative externality they exert on other investors (they just compare the private benefit of becoming fast with their private cost, rather than accounting for the social cost). Moreover, very much like in an arm's race, investors can decide to invest in HFT technologies simply to avoid being sidelined if they remain slow, even though collectively all investors would be better off being slow. These two effects lead in general to an overinvestment in high trading technologies, relative to what would be socially optimal.

Thus, the net effect of HFT on the liquidity and allocational efficiency of financial markets is ambiguous. Now, turn to its effect on price discovery. By exploiting mispricings very quickly, high frequency arbitrageurs are likely to enhance price discovery. More generally, if HFTs trade on information faster than other traders then they should contribute to price discovery by accelerating the speed at which new information is impounded into prices. This logic is standard in models of informed trading. However, in the case of HFT, it raises several problems.

First, HFT on new short-lived information (e.g., impending news) may induce traders to trade less aggressively on their long-lived information (see Foucault, Hombert, and Rosu (2012)). Eventually, this substitution effect might impair price discovery.

Second, HFTs do not strictly speaking produce or discover new information. In fact, they often trade on market

data: order flow, prices, volume etc. Ultimately, market data are informative because they reflect more primitive signals acquired by other investors (portfolio managers, hedge funds etc.) and these signals would probably find their way into prices even in the absence of HFTs, albeit at a lower rate. In a sense HFTs free ride on the acquisition of information by slower investors, which may reduce incentives for information acquisition in the first place (see Grossman and Stiglitz (1980)). If this is the case the overall informativeness of securities markets may have decreased with the proliferation of HFTs.

Third, there is a trade-off between accuracy and speed. News is sometimes inaccurate or wrong. Hence, investors who react too fast to news could inject noise in the price system rather than making it more efficient, at least at the very high frequency (Dugast and Foucault (2012)).

Hence, there exist several ways in which HFT could push prices away from fundamentals. In line with this possibility, market participants have noticed an increase in so-called “mini flash crashes,” that is, sharp movements in prices in one direction, which are then quickly reverted. For instance, a May article of USA today writes:

*“Mini flash crashes still occur routinely with individual stocks [...] Despite efforts to prevent another flash crash, the infamous day on May 6, 2010, when the Dow Jones industrials fell roughly 900 points, only to quickly recover, regulators and markets have moved to implement safeguards. Yet, traders and market observers are still seeing individual stocks and ETFs suffer flash-crash-like events, when stocks fall suddenly for no reason and quickly rebound, suggesting many of the underlying problems haven’t been solved. [...] The most recent data available are for the first month and three days of 2011. In that period, stocks showed perplexing moves in 139 cases, rising or falling about 1% or more in less than a second, only to recover, says Nanex. There were 1,818 such occurrences in 2010 and 2,715 in 2009, Nanex says.”* (source: “Mini flash crashes worry traders”, USA today, May 17, 2011).<sup>11</sup>

HFTs could play a role in these mini flash crashes for several reasons. First, they may all react at the same time to erroneous signals by sending buy or sell market orders consuming market liquidity, triggering sharp price movements as limit orders get hit by market orders from HFTs. Alternatively, HFTs with limit orders in a stock may all cancel their buy (sell) limit orders after the arrival of a large sell (buy) market order in a stock, for safety reasons and resubmit new limit orders quickly if the large market order does not appear to be informationally motivated. In any case, waves of cancellations or market orders submissions by HFTs reacting to the same event may exacerbate the volatility of the asset they trade.

### III.2. ENDOGENEITY ISSUES

Establishing causal links between HFT and market quality is challenging. It is very likely that HFT and market quality are determined by common factors (e.g., the precision of public news), some of which are difficult to control for or cannot be observed empirically. Moreover, market quality variables and HFT strategies are jointly endogenous. Both directly reflect the optimizing behavior or market participants, their reaction to market conditions, and their response to the strategies of oth-

ers. For instance, while HFT might affect the volatility of the asset they trade, it is also affected by this volatility. Theoretically, Biais, Foucault and Moinas (2013) show that an increase in the uncertainty about the value of an asset increases the incentives to invest in HFT technology. Empirically, Brogaard (2011a) finds that news release affect measures of activity of HFTs on Nasdaq. As news is a source of volatility, this suggests volatility has an effect on the activity of HFTs.

**Endogeneity and instruments.** Due to the joint endogeneity of liquidity, volatility and HFT, correlations between these variables should not be interpreted in terms of causal impact of HFT on market quality. For instance, if volatility induces HFTs to trade more, one would find a positive correlation between volatility and HFT, even if HFT has no effect on volatility. One way to overcome these endogeneity problems is to use instruments, i.e., variables affecting HFT without directly affecting the other variables of interest. Empiricists have argued that some technological changes in market design, such as, e.g., reductions in latency, or the implementation of co-location, should affect HFT without directly affecting market quality. These events have thus been used as instruments to evaluate the effects of HFTs on market quality.

### III.3. HFT AND THE INFORMATIVENESS OF PRICES

Brogaard, Hendershott and Riordan (2011) find that HFTs tend to place buy (sell) market orders just before an increase (a decrease) in the market valuation of assets. This is consistent with the hypothesis that HFTs using market orders possess information. In contrast, for limit orders, Brogaard, Hendershott and Riordan (2011) observe the opposite pattern: HFTs’ buy (sell) limit orders tend to execute when this value is falling (increasing), while those using limit orders are picked off when new information arrives. Correspondingly, Brogaard et al. (2011) find negative trading profits on HFTs’ executed limit orders but positive profits on HFTs’ market orders. These results are consistent with those obtained by Hendershott and Riordan (2012) for the Deutsche Börse. They find that algorithmic traders’ market orders have greater permanent impact than human traders’ market orders.

Boehmer, Fong and Wu (2012) consider data from 39 different countries. Their sample contains about 12,800 stocks per year. They find a negative relationship between a proxy for algorithmic trading (the number of messages normalized by volume in each stock) and the autocorrelation of stock returns (in absolute value) at the 30 minutes horizon, which they interpret as a measure of price inefficiency. Of course, as explained above, due to endogeneity issues, such correlation cannot be interpreted in causal terms. In order to identify causal relations, Boehmer, Fong and Wu (2012) use the introduction of co-location in the countries in their sample as an instrument for HFT. They argue that co-location, in itself, should not have any direct impact on the short-term autocorrelation of stock returns. Yet, they find that an increase their proxy for HFT, instrumented on co-location, leads to a decline in the autocorrelation of stock returns. Even then, however, it



is difficult to reach a conclusion about the consequences of HFT, because the proxy could reflect, more generally, algorithmic trading, which includes, as discussed above, other forms of trading than HFT.

Chaboud *et al.* (2009) focus on algorithmic trading in the foreign exchange market. Using a Vector Autoregressive Approach, they estimate the contributions of algorithmic and human trades to the variance of returns over a 30 minutes horizon. These contributions are interpreted in terms of contribution to price discovery. Overall they find that human trades contribute more to price discovery than algorithmic trades.

### III.4. HFT AND LIQUIDITY

As explained above, it is possible that automation of the trading process by HFT firms reduces the cost of liquidity provision. If this is the case, and if the market is competitive, the costs of trading for investors should decrease.

Hendershott, Jones, and Menkveld (2011) study the effect of algorithmic trading on liquidity, using the number of electronic messages normalized by trading volume as a proxy for algorithmic traders' activity. In order to address endogeneity issues, they use a technological change in the organization of the NYSE in 2003, namely the implementation of the "autoquote" functionality. Until Autoquote, specialist clerks had to update manually best bid and offer prices for stocks listed on the NYSE. This manual procedure was slowing down the speed at which algorithmic traders could receive information about market conditions on NYSE stocks. The implementation of Autoquote considerably increased this speed and therefore made algorithmic trading easier. For this reason, as shown by Hendershott *et al.* (2011), Autoquote is associated with a significant increase in their proxy for algorithmic trading activity. To the extent Autoquote does not directly affect market liquidity, it can be used as an instrument to study the effect of algorithmic trading on liquidity.

Hendershott *et al.* (2011) find that standard measures of market liquidity (the quoted bid-ask spread and the effective bid-ask spread) improve after the introduction of Autoquote for large capitalization stocks.<sup>12</sup> In contrast, they do not find any significant effect of algorithmic trading on market liquidity for stocks with small capitalizations. Moreover, they show that the reduction in trading costs documented by their study is driven by a reduction in the adverse selection component of the bid-ask spread, which more than offsets a simultaneous increase in realized spreads (a measure of the average profit per trade for liquidity suppliers). Hence, the reduction in the cost of liquidity provision is not entirely passed to liquidity demanders. This suggests competition among fast liquidity providers was imperfect in 2003.<sup>13</sup>

Using the same proxy for algorithmic trading as in Hendershott *et al.* (2011), Boehmer, Fong and Wu (2012) provide similar findings for a larger sample of stocks and countries. For all countries in their sample, they find a negative association between algorithmic trading activity and bid-ask spreads. Moreover, using co-location as an instrument for algorithmic trading, they find that liquidity improves when co-location is introduced.

Interestingly, Boehmert, Fong and Wu (2012) find cross-sectional variations in the effect of HFT on liquidity. The beneficial effect of algorithmic trading on liquidity accrues mainly to large, high-priced, and low volatility stocks. In contrast, algorithmic trading appears to have a negative effect on liquidity for small capitalization stocks and no effects for stocks with low prices or high volatility.

These studies provide very interesting findings regarding the effects of automated trading on liquidity. However, one must be careful in interpreting them for two reasons. First, proxies for HFT based on the normalized number of messages are likely to reflect the activity of both HFTs and other algorithmic traders operating at lower frequencies (such as brokers using computers to split the orders of their clients). Thus, Hendershott *et al.* (2011) or Boehmer *et al.* (2012)'s findings may reflect the effect of algorithmic traders operating at relatively slow speed rather than the effect of HFTs. In fact, prior to 2006-2007, the latency for trades on the NYSE was about 4 to 10 seconds. This is too high for HFT. Hence, the sample period used by Hendershott *et al.* (2011) most likely predates the years in which HFT started in NYSE stocks.

Second, Hendershott *et al.* (2011) or Boehmer *et al.* (2012) use measures of liquidity that are relevant for investors submitting market orders (e.g., bid-ask spreads). In limit order markets, however, investors can choose to submit both market and limit orders. It could be that more intense competition from HFTs reduces the likelihood of profitable execution for slow investors submitting limit orders, forcing them to switch to, costlier, market orders or to stop providing liquidity (a crowding out effect). This negative effect of HFT on investors' welfare cannot be captured by standard liquidity measures.

Very recent empirical studies address the first issue by focusing more specifically on trades by HFTs. On April 1st, 2012, the Investment Regulatory Organization of Canada began charging fees based on the number of electronic messages used by traders. This can be viewed as an exogenous increase in cost for HFTs. Correspondingly, the share of HFTs in total message traffic fell from 81% to 5.65%. In line with the findings in Hendershott, Jones, and Menkveld (2011), Malinova, Parks, and Rjordán (2012) find that this reduction in message traffic (HFT activity) is associated with an increase in bid-ask spreads.<sup>14</sup> Moreover, they find that the drop in HFT worsens execution quality for retail investors.

Brogaard, Hendershott, Hunt, Latza, Pedace and Ysusu (2012) study trades by institutional investors and HFTs for stocks in the FTSE250 index.<sup>15</sup> They use reductions in latency on the London Stock Exchange to measure the causal effect of HFT on the price impacts of trades for institutional investors in their sample. They do not find any effect, positive or negative. Furthermore, they show that institutional investors' impacts have been steadily declining from 2003 to 2011 (except during the subprime crisis), in contrast to investors' perception.

### III.5. HFT AND VOLATILITY

From a theoretical standpoint, it is not clear whether HFT should increase or decrease volatility. To the extent that i)

short-term volatility reflects the transient impact of liquidity demand and that ii) HFT improve the supply of liquidity, HFT should reduce short-term volatility. But, if HFT reduces the supply of liquidity (e.g., because it worsens adverse selection for slow liquidity suppliers) then the opposite would obtain. Also, if HFT contributed to the occurrence of transient market disruptions (such as the “flash crash”), then it would increase volatility. In line with the ambiguity of theoretical predictions, empirical evidence is mixed.

Chaboud *et al.* (2009) study the effect of algorithmic trading on volatility in three currency pairs (dollar/yen, dollar/euro and euro/yen) using data on trades taking place on EBS (one of the two main electronic trading platforms in currency markets), over the period 2006-2007. For each order placed on EBS, they know whether the order is submitted by a human trader or generated by an algorithm. Although algorithmic trading and HFT are not equivalent, they are likely to be strongly correlated. For each day, Chaboud *et al.* (2009) measure the algorithmic trading activity on EBS by the fraction of total trading volume accounted by algorithmic traders. In simple OLS regressions, they find a positive relationship, at the daily frequency, between the volatility of the currency pairs in their sample and their measure of algorithmic trading. This positive correlation however may only reflect the fact that algorithmic trading desks choose to be more active on days with high volatility.

In order to identify the causal effect of algorithmic trading on volatility, they use the monthly number of trading desks equipped for algorithmic trading on EBS as an instrument. Algorithmic traders use a special interface to interact with EBS and therefore EBS knows the number of users of this interface at each point in time. The number of users is unlikely to be affected by daily variations in volatility since setting up a trading desk for algorithmic trading takes time (more than one day). Hence, monthly variations in the number of algorithmic trading desks on EBS can be used to identify the causal effect of algorithmic trading on daily volatility because it affects the volume of algorithmic trading on EBS without being directly affected by volatility.

Using this approach, Chaboud *et al.* (2009) find a (weak) negative effect of algorithmic trading on volatility, in contrast to the result they obtain with the simple OLS analysis. This finding underscores the importance of controlling for endogeneity issues in analyses of the effects of HFT.

Other empirical studies reach a similar conclusion with different methods and for different markets. Hasbrouck and Saar (2012) address the endogeneity problem by estimating a system of equations in which volatility can influence algorithmic trading and vice versa. They use the number of “linked messages” (see Section 2.2) over these intervals as a proxy for algorithmic trading and find a negative effect of algorithmic trading on volatility.

Brogaard (2011a) uses the ban on short-sales that affected financial stocks in the U.S. for about three weeks in September and October 2008. This ban applied to thirteen stocks in the Nasdaq sample used by Brogaard and affected HFTs who are not registered market-makers. Brogaard (2011a) finds that HFTs’ fraction of daily trading volume fell sharply for the stocks affected by the ban relative to unaffected stocks. Hence, the short sale ban is indeed a

negative shock to the activity of HFTs. Moreover, Brogaard (2011a) shows that stocks in which short-sales by HFTs are the most affected (relative to other stocks) experience a relatively bigger increase in volatility (measured over different time intervals). This finding again is consistent with a negative effect of HFT on volatility.

In contrast, Boehmer, Fong and Wu (2012) obtain different conclusions for their international sample of stocks. Using various proxies for volatility (e.g., the daily realized volatility or the standardized intraday price range), they find a positive association between their measure of algorithmic trading and volatility. The same finding holds when they use co-location as an instrument for algorithmic trading.

### III.6. CROWDING OUT AND RESILIENCE TO DEMAND SHOCKS.

High frequency market-makers are low cost competitors for more traditional market-makers. Accordingly, they are often able to position their limit orders ahead of the queue of limit orders: Brogaard (2011b) find that Nasdaq HFTs post quotes at least equal to the best quotes 50% of the time and stand alone at the best quotes 19% of the time. Because of their speed advantage, HFTs can supply liquidity on better terms than slower traders, cancel their quotes rapidly when the market moves against them, and hit stale limit orders resting in limit order books. This creates an adverse selection problem for slow liquidity suppliers, who can be crowded out of the market by HFTs.<sup>16</sup> In normal times, when order imbalances between buy and sell orders oscillate around zero, HFTs can offer a good substitute for slow market makers. When there are larger and sustained imbalances, however, HFTs may be unable to provide sufficient liquidity, if their risk-bearing capacity is too small. In such circumstances, the absence of slow market makers could impair the resilience of the market.

### III.7. CORRELATION AMONG HFT STRATEGIES

HFTs extract informational signals from market data, and then automatically feed these signals into trades. When there is only one high-frequency trader, or a few small ones, such behavior is not likely to affect the market. In contrast, when HFTs amount for a large fraction of the volume, their behavior is likely to impact the market. Now, it is quite possible that many HFTs would focus on the same market signals, and therefore trade in the same direction. To the extent that these trading algorithms would not be programmed to take such correlation into account, HFT could trigger cascades and spirals, amplifying market shocks. This is in line with what happened during the August 2007 mini-crash. At that time, many quants were using similar strategies. Thus, they were simultaneously hit by a shock, and reacted similarly, which generated a downward spiral in the market.

While, in 2007, HFT was likely to be much lower than it is nowadays, more recent empirical studies offer evidence consistent with the case discussed above. Chaboud *et al.* (2009) find that algorithmic trading strategies are correlated

and not as diverse as those used by non-algorithmic traders. Brogaard (2011b) finds that Nasdaq HFTs tend to place orders (market or limit) in the same direction over various time intervals (10 seconds, thirty seconds, 2 minutes and 15 minutes). Egginton *et al.* (2012) focus on periods of very high activity by HFTs for Nasdaq stocks in 2010. They identify these periods as one minute periods in which quoting activity (cancellations and new quotes arrival) exceeds by 20 standard deviations the mean number of quotes per minute over the past twenty days. They find about 125 events of this type per day in their sample and 74% of the stocks in this sample experience at least one event of very high quoting activity. These bursts in traffic are consistent with clustering of HFTs' decisions. Egginton *et al.* (2012) show that they are associated with greater volatility and larger effective bid-ask spreads.

There is strong evidence that market orders from HFTs have superior short-term information, and weaker evidence that algorithmic trading may improve the informational efficiency of prices. Empirical studies so far have not found significant negative effects of HFT on liquidity, while the findings regarding the effects of HFT on volatility are rather mixed. Empirical findings also suggest that HFT's trades are clustered and correlated with one another.

### III.8. DOMINO EFFECTS

While HFT amounts to a very large fraction of trading volume, HFTs most often hold risky positions for only brief periods of time. Correspondingly, HFT firms usually have relatively limited risk-bearing capacity. They are not subject to prudential regulation and in practice often operate with very little capital. Again, in normal times, this is quite all right, but if HFT firms were to be hit by large market shocks, the lack of capital for HFTs could generate failures.

As HFT firms often take similar positions, there could be a wave of such failures. These defaults could prove very difficult to handle. One would need to determine the net position of each market participant towards each defaulting HFT firm. That would be made difficult by the discrepancy between the low frequency at which clearing and settlement systems operate (e.g., daily) and the vibrant pace of HFT. These failures could then propagate to other market participants with open positions with HFT firms.

## IV. POLICY

In this section we first outline the market failures HFT could generate because of negative externalities. Then we discuss several possible policy responses to these market failures.

### IV.1. MARKET FAILURES

Regulatory intervention is justified when the activity of one group of agents exerts negative externalities on other agents. The discussion above suggests that HFTs can generate several types of negative externalities: congestion externalities, adverse selection, crowding out fundamental liquidity suppliers, systemic risk due to default and contagion.

First, HFTs can generate congestion externalities, relative to the access to exchanges' trading platforms or to market information. Optimal allocation of trading and information dissemination require that such congestion externalities be priced.

Second, HFTs can generate negative externalities for other traders, under the form of adverse selection. In the theoretical model of Biais, Foucault, and Moinas (2012) the ability of HFTs to react faster to information generates adverse selection costs for slow traders. This, per se, is a source of welfare loss as it can lead to a lower participation rate of some traders (and in turn to lower sharing of gains from trade). Moreover, investors can be trapped in a technological race just to reduce the risk of trading with faster and therefore better informed investors. Biais *et al.* (2012) show that this race results in excessive investment in HFT technologies relative to the social optimum. In this context, taxing investment in HFT technology can improve efficiency.

Third, HFTs might crowd out slow liquidity providers, who trade on long term fundamental information but are exposed to the risk of being picked off in the short term. Now, these slow liquidity providers have greater long-term risk-bearing capacity than HFTs. Hence the latter exert negative externality on other market participants by depriving them from liquidity supply at the time of significant shock that only slow traders could accommodate.

Fourth, to the extent that i) lightly capitalized HFT firms are exposed to the risk of default waves in case of large market shocks, and ii) such default waves can destabilize other institutions, HFT exerts a negative externality on other market participants by raising systemic risk.

### IV.2. POLICY RESPONSES

**Taxes and pricing.** One approach to curb excessive investments in HFT technologies is to tax them. In August 2012, France adopted a tax on HFTs (defined as traders who use algorithms to place orders and modify their orders in less than half a second). Specifically, traders can cancel and modify up to 80% of their orders free of charge. Above this threshold, traders pay a tax of 1bps of the value of cancelled or modified orders. This tax scheme is similar in spirit to the pricing scheme introduced on NYSE Euronext, which charges a fee of €0.1 on each order when the order-to-execution ratio of a trader exceeds 100.

Hagströmer and Nordén (2012) find that high frequency market-makers tend to have relatively high order-to-execution ratios relative to other HFTs. This is because liquidity suppliers need to frequently cancel and revise their orders, to avoid adverse execution. Thus, taxes based on order-to-execution ratios may curtail high-frequency market-making rather than HFT strategies that are a source of adverse selection for slower market participants, e.g., those that specialize in picking off stale quotes in case of information arrival. To discourage the latter rather than the former, one way is to charge higher fees or taxes on marketable orders (orders that are for immediate execution) since empirical evidence suggest that these orders are used by HFTs to exploit their advance information.

**Market mechanisms.** Another approach would be to offer market mechanisms mitigating the adverse selection problem induced by fast traders and reduce their incentives to overinvest in fast trading technologies.

One could wonder why policy intervention would be needed to implement such beneficial trading mechanisms. Shouldn't they emerge spontaneously? Unfortunately, there are several reasons why market forces could fail to bring about optimal mechanisms.

First, exchanges' incentives may not be fully aligned with those of slow traders. Indeed, exchanges recover a fraction of the trading profits earned by HFT firms, by charging trading fees, co-location fees, or fees for the sale of information. This possibility may distort their decisions regarding market structure or the pricing of their product in a way that protects the interest of HFT firms.<sup>17</sup> Cespa and Foucault (2012) provide a theoretical analysis of such a distortion. They show that an exchange has an incentive to curb investors' access to real-time price information (e.g., by charging high fees for co-location and their real time datafeed) to enhance the trading profits of sell-side traders because it recovers a fraction of these profits through trading fees.

Second there might be barriers to entry for slow-traders friendly exchanges. Investors benefit from trading where other participants trade. Hence, expectations that a new market will attract few trades can be self-fulfilling, even if many market participants would benefit from collectively joining the new platform. This coordination problem could explain why slow platforms do not arise even though many market participants would collectively benefit from using them.

Third, the current market structure does not necessarily facilitate the adoption of new trading mechanisms by exchanges. Consider the two following examples. In the U.S., market orders must be routed to the market posting the best quotes at any point in time according to the so called order protection rule. This rule therefore links markets together making it more difficult for slow markets to insulate themselves from fast markets. In the U.S. and in Europe, equity trading is heavily fragmented as trading for a stock can occur on multiple platforms. The implementation of periodic batch auctions would therefore require coordination on the times at which competing platforms will run these auctions. Indeed, in an environment in which batch auctions are very frequent (say, every second) mismatches in auction times across competing platforms (say, Chi-X and NYSE-Euronext) would again give an advantage to traders who can observe the outcome of one auction before other traders know this outcome. Coordination would also be required between cash markets and derivatives markets as some traders often wish to establish simultaneously offsetting positions in these instruments for hedging.

**Prudential regulation.** The European Commission has included the analysis of HFTs in its review of the Market In Financial Instruments Directive. It considers the possibility to subject HFT firms to regulatory oversight and capital requirements. This would help prevent systemic risk creation by HFT firms.

**Pilot experiments.** It can be very difficult to identify causal links between the presence of HFTs and market outcomes because both are endogenous (see Section 3.2). In this context, it is difficult to predict the effects of specific regulatory proposals regarding HFTs. In light of this uncertainty, we recommend conducting pilot experiments to evaluate effects of these proposals on a limited but representative sample of stocks before implementing them at a larger scale.

For instance, a high fraction of HFTs' orders are cancelled, sometimes very quickly after submission. This high cancellation-to-trade ratio is a source of concerns for various reasons: it might be a source of congestion slowing down the entire trading process; it makes it more difficult for slow market participants to figure out exact terms of trade; cancellations might be used for manipulative purposes. There are many ways to cope with this problem. One can for instance impose cancellation fees or minimum resting times on limit orders, whereby a limit order cannot be cancelled before a certain amount of time (see SEC (2010)). The effects of these measures however are a priori unclear. For instance, rather than improving market quality, they might raise the exposure of traders submitting limit orders to the risk of being picked off and thereby result in larger bid-ask spreads. Thus, the net effect of minimum resting times (or fees on cancellations-to-trade ratios) on market quality is difficult to predict. The same is true for many of the measures considered to curb HFT.

## ■ V. CONCLUSION

Levying taxes on institutions with high message traffic or high cancellation rate is not likely to be an optimal policy. The problem with such a policy is that it might primarily affect market-making, as opposed to other, potentially less useful, high-frequency strategies.

Changes in market structure might offer a more effective response. One possibility would be to move from continuous trading to periodic call auctions, organized, say, every 100 milliseconds. It is hard to think of any reasonable situation where trading only every 100 milliseconds would make the market less useful for society. Another possibility would be to offer slow traders the possibility to place orders that could be executed only against slow orders or to create platforms that could be accessed only by slow traders. If slow traders adopted these mechanisms, they would be (at least partly) protected from predatory behavior by HFTs. This would reduce predatory profits for HFTs and, in turn, curb excessive investment in HFT technology. An alternative to directly regulating HFTs could then be to regulate exchanges, to ensure that the trading mechanisms and pricing schemes they offer are likely to lead to efficient outcomes.

It would also be prudent to ensure that HFT does not create systemic risk. For example, capital buffers would reduce the likelihood that HFT firms would be destabilized by liquidity shocks and would in turn destabilize their counterparties. Also, capital requirements could increase the "skin in the game" of the manager owners

of HFT firms, and reduce the moral hazard problem associated with limited liability. Stress tests should be conducted, at the level of each trading firm, and also at the market level, to evaluate the likely effects of market shocks, and the ability of the market to cope with them without and avoid systemic crises.

Finally, in light of the uncertainty about the effects of regulations, it would be useful to conduct pilot experiments before implementing policy measures. Such pilot experiments have been conducted in the U.S to evaluate the costs and benefits of imposing greater post trade transparency in bond markets (see Goldstein, Hotchkis and Sirri (2007)). The same type of approach would be very useful to design the optimal regulation of HFT. ■

- 1 Trading platforms define latency as the communication time between a trader's server and the platform (e.g., the time for the platform to acknowledge reception of an order submitted by the trader). This delay is just one component of the relevant latency for traders, who are also concerned by the speed at which they can process and react to information received from trading platforms.
- 2 Human reaction times to events are of the order of 200 milliseconds. See Kosinski (2010). The amount of financial data that can be used to make trading decisions at any point in time is extremely large. Hendershott (2011) provides an interesting calculation. He notes that there are about 2.5 billion orders per day over 23,400 seconds of trading, which implies more than 100,000 messages per second for U.S. equity markets alone. Only computers can process such a rich, near continuous, flow of signals.
- 3 For instance, in Sandas (2001), limit orders at the top of the queue of limit orders at a given price (that is with time priority) earn a higher expected profit than traders at the back of the queue. Hence, acquiring time priority has value.
- 4 Foucault, Röell and Sandas (2003) consider a model in which dealers incur monitoring costs. Their speed of reaction to events is determined by the size of these costs. They show that lower monitoring costs induce dealers to react faster to events and reduce their exposure to the risk of being picked off. The use of computers to automatically update quotes is a way to reduce monitoring costs.
- 5 Storckenmaier and Wagener (2011) report the duration of crossed quotes for stocks constituents of the FTSE 100 and traded on the LSE, Chi-X, BATS and Turquoise. The find that this duration is 16 minutes in April/May 2009 and only 19.8 seconds in April/May 2010. This dramatic decline is most likely due to an intensification of automated arbitrage between these markets.
- 6 This strategy is not new. For instance, in the 90s, day traders (so called SOES bandits) took advantage of the automation of quote execution on Nasdaq to pick off dealers who were slow to update their quotes relative to other dealers (see Foucault, Röell, and Sandas (2003)).
- 7 Hirschey (2011) also uses data from Nasdaq. These data enable him to identify trading by individual HFTs in 96 Nasdaq and NYSE stocks for 2009. Thus, his data are more precise than those in the Nasdaq sample. However, in contrast to the Nasdaq sample, they are unavailable to other researchers.
- 8 Hendershott *et al.* (2011) treat all the HFTs in their sample as a single HFT. They calculate the daily profit of this trader as the cumulative cash received on sell orders minus the cumulative cash paid on buy orders plus the value of the trader's inventory at the end of the day, marked-to-market at the closing price.
- 9 See Foucault, Kadan, and Kandel (2013) and Colliard and Foucault (2012) for theoretical analyses of liquidity rebates, their effects on bid-ask spreads, and on incentives for liquidity provision.
- 10 Jovanovic and Menkveld (2011) also study the effect of high-frequency trading on welfare but take as exogenous the level of high-frequency trading. They point out that HFTs can quickly integrate new information into their quotes. In this way, they help to mitigate informational asymmetries among final buyers and sellers in assets where these asymmetries are high. On the other hand, as in Biais *et al.* (2011), by reacting fast on information, HFTs can be a new source of asymmetric information. These two effects have opposite effects on trading volume and welfare, so that the net effect of high-frequency trading on welfare is ambiguous in Jovanovic and Menkveld (2011).
- 11 For graphical illustrations of mini flash crashes, see <http://www.nanex.net/FlashCrash/OngoingResearch.html> (the "strange days" column).
- 12 One exception is quoted depth (*i.e.*, the number of shares offered at the best quotes) which has decreased for these stocks. However, this decline seems too small to offset the decline in bid-ask spreads and Hendershott *et al.* (2011) argue that the net effect of algorithmic trading on trading costs is negative for large capitalization stocks.
- 13 Riordan and Storckenmaier (2012) study the effect of a reduction in latency on Xetra (the Deutsche Boerse trading system) in 2007. Their findings are very similar to those in Hendershott, Jones and Menkveld (2011). In particular, the reduction in latency is associated with a drop in effective spreads (from 7.72bps to 7.04bps), a reduction in price impacts (from 6.87bps to 2.65 bps) and an increase in realized spreads (from 0.97bps to 4.45bps).
- 14 As Kirilenko *et al.* (2010), Malinova, Parks, and Riordan (2012) use trader-level data to identify HFTs on the Toronto Stock Exchange. They classify a trader as a high-frequency trader if it has a very high message-to-trade ratio and if its absolute number of messages (market orders, limit orders, cancellations, fill-or-kill order) is very high. Using these criteria, they identify 107 and 88 HFTs in March and April, 2012, respectively.
- 15 Data on institutional investors' trades are provided by Ancerno, a provider of data for trading costs analysis. The HFT data are provided by the Financial Services Authority. It contains data on trades by HFTs who have to report their trades to the FSA or who trade through a broker.
- 16 Foucault, Kandel, and Kadan (2013) show that a market-maker who reacts faster than his competitors gets a larger market share. See also Cartea and Penalva (2012) for a model in which slow market-makers are crowded out by faster market-makers.
- 17 On this point, see "*For Superfast Stock Traders, a Way to Jump Ahead in Line,*" Wall Street Journal, September 19, 2012. This article explains how some U.S. exchanges are suspected of giving the possibility to HFTs to use special orders giving them priority of execution over slower traders.

## Appendix

**Table 1:** This table provides a synoptic view of the samples used in the empirical studies discussed in our paper. The sample periods are not always precise because some papers use different sample periods when conducting different tests.

Author(s)	Asset Class	Sample	Sample Period
Chaboud, Chiquoine, Hjalmarsson and Vega (2009)	Currencies	Euro-Dollar, Dollar-Yen, Euro-Yen	01/01/2006-31/12/2007
Brogaard (2011a)	U.S. stocks	120 stocks listed on Nasdaq and NYSE	2008-2010
Brogaard (2011b)	U.S. stocks	120 stocks listed on Nasdaq and the NYSE	2008-2010
Hasbrouck and Saar (2012)	U.S Stocks	345 and 394 stocks listed on Nasdaq	10/2007 and 06/2008
Hendershott, Jones and Menkveld (2011)	U.S Stocks	NYSE stocks	2001-2005
Hendershott and Riordan (2009)	Dax stocks	30 stocks listed on Deutsche Börse	01/1/2008-18/01/2008.
Brogaard, Hendershott, and Riordan (2012)	U.S. Stocks	120 stocks listed on Nasdaq and NYSE	2008-2009
Brogaard, Hendershott, Hunt, Latza, Pedace and Ysusi (2012)	FTSE 250 stocks	250 stocks	2007-2011
Boehmer, Fong, and Wu (2012)	Stocks	39 different exchanges	2001-2009
Jovanovic and Menkveld (2011)	Dutch stocks	14 stocks constituents of of the AEX index	01/1/2008-04/23/2008
Hirschey (2011)	U.S stocks	96 Nasdaq and NYSE stocks	01/01/2009-12/31/2009
Kirilenko, Mehrdad, Kyle, and Tugkan (2010)	Futures	E-mini futures contracts on the S&P500	05/03/2010-05/08/2010
Baron, Brogaard, and Kirilenko (2012)	Futures	E-mini futures contracts on the S&P500	August 2010
Menkveld (2011)	Dutch stocks	14 stocks constituents of of the AEX index	01/01/2007-17/06/2008
Riordan and Stockenmaier	German stocks	110 stocks from the HDAX index	02/22/2007-06/19/2007
Gai, Yao and Ye (2012)	U.S. Stocks	120 Nasdaq and NYSE stocks	2010-2011
Malinova, Parks and Riordan (2012)	Canadian Stocks	248 stocks listed on the Toronto Stock Exchange	March-April 2012.
Hagströmer and Nordén (2012)	Swedish stocks	30 stocks part of the OMXS index	August 2011 and February 2012

## References

- BARON M., BROGAARD J. AND KIRILENKO A. (2012), The Trading Profits of High-Frequency Traders, Working Paper, University of Washington.
- BIAIS B. AND WOOLLEY P. (2012), HFT, Working Paper, IDEI Toulouse.
- BIAIS B., FOUCAULT T. AND MOINAS S. (2011), Equilibrium Algorithmic Trading, Working Paper, Université de Toulouse.
- BIAIS B., HOMBERT J. AND WEILL P.O. (2008), Trading and Liquidity with Limited Cognition, Working Paper, Toulouse University, IDEI.
- BIAIS B., HILLION P. AND SPATT C. (1995), An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse, *Journal of Finance* 50, p. 1655-1689.
- BOEHMER E., FONG K. AND WU J. (2012), International Evidence on Algorithmic Trading, Working Paper, EDHEC.
- BROGAARD J. (2011A), HFT and Volatility, Working Paper, Washington University.
- BROGAARD J. (2011B), The Activity of HFTs, Working Paper, Washington University.
- BROGAARD J., HENDERSHOTT T. AND RIORDAN R. (2011), HFT and Price Discovery, Working Paper, UC Berkeley.
- BROGAARD J., HENDERSHOTT T., HUNT S., LATZA T., PEDACE L. AND YSUSI C. (2012), HFT and the Execution Costs of Institutional Investors, *Foresight Driver Review*.
- CARTEA A. AND PENALVA J. (2012), Where is the Value in HFT?, *Quarterly Journal of Finance* 2, p. 1-46.
- CHABOUDA., CHIQUOINE B., HJALMARSSON E. AND VEGA C. (2009), Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, Working Paper, Federal Reserve Board.

## References (continued)

- CESPA G. AND FOUCAULT T. (2012), Sale of Price Information by Exchanges: Does it Promote Price Discovery?, *Management Science*, forthcoming.
- COLLIARD J-E. AND FOUCAULT T. (2012), Trading Fees and Efficiency in Limit Order Markets, *Review of Financial Studies*, 25, p. 3389-3421.
- COPELAND T. AND GALAI D. (1983), Information Effects on the Bid-Ask Spread, *Journal of Finance* 38, p. 1457-1469.
- DUGAST J. AND FOUCAULT T. (2012), HFT, Market Efficiency, and Mini Flash Crashes, *Work in progress*, HEC Paris.
- EGGINTON J., BONNIE F., VAN NESS F. AND VAN NESS R. (2012), Quote stuffing, *Working Paper*, University of Mississippi.
- FOUCAULT T., RÖELL A. AND SANDAS P. (2003), Market Making with Costly Monitoring: an Analysis of SOES Trading, *Review of Financial Studies* 16, p. 345-384.
- FOUCAULT T., KADAN O. AND KANDEL E. (2013), Liquidity Cycles, and Make/Take Fees in Electronic Markets, *Journal of Finance* 68, p. 299-339.
- FOUCAULT T., HOMBERT J. AND ROSU I. (2012), News Trading and Speed, *Working Paper*, HEC Paris.
- FOUCAULT T., KOZHAN R. AND THAM W. (2012), Toxic Arbitrage, *Work in progress*, HEC Paris.
- GAI J., YAO C. AND YE M. (2012), The Externalities of HFT, *Working Paper*, University of Illinois.
- GLOSTEN L. AND MILGROM P. (1985), Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders, *Journal of Financial Economics* 13, p. 71-100.
- GOLDSTEIN M., HOTCHKISS E. AND SIRRI E. (2007), Transparency and Liquidity: A Controlled Experiment on Corporate Bonds, *Review of Financial Studies* 20, p. 235-273.
- GROSSMAN S. AND STIGLITZ J. (1980), On the Impossibility of Informationally Efficient Markets, *American Economic Review* 70, p. 393-408.
- HAGSTRÖMER B. AND NORDEN L. (2012), The Diversity of HFTs, *Working Paper*, Stockholm university school of business. Hasbrouck J. and Saar G. (2012), Low-latency Trading, *Working Paper*, New-York University.
- HENDERSHOTT T. AND RIORDAN R. (2009), Algorithmic Trading and Information, *Working Paper*, University of Berkeley.
- HENDERSHOTT T. AND RIORDAN R. (2012), Algorithmic Trading and the Market for Liquidity, *Journal of Financial and Quantitative Analysis*, forthcoming.
- HENDERSHOTT T., JONES C. AND MENKVELD A. (2011), Does Algorithmic Trading Improve Liquidity?, *The Journal of Finance* 66, p. 1-33.
- HENDERSHOTT T. (2011), HFT and Price Efficiency, *Foresight Driver Review* (DR12).
- HIRSCHEY N. (2011), Do HFTs Anticipate Buying and Selling Pressures?, *Working Paper*, London Business School.
- JOVANOVIĆ B. AND MENKVELD A. (2010), Middlemen in Limit Order Markets, *Working Paper*, VU University Amsterdam.
- KIRILENKO A., KYLE A.S., SAMADI M. AND TUZUN T. (2010), The Flash Crash: the Impact of HFT on an Electronic Market, *Working Paper*, University of Maryland.
- KOSINSKI R.J. (2010), A Literature Review on Reaction Time, *Working Paper*, Clemson University.
- MALINOVA K., PARKS A. AND RIORDAN R. (2012), Do Retail Traders Suffer from HFTs?, *Working Paper*, University of Toronto.
- MCLNISH T. AND UPSON J. (2011), Strategic Liquidity Supply in a Market with Fast and Slow Traders, *Working Paper*, University of Memphis.
- MENKVELD A. (2011), HFT and the New-Market Makers, *Working Paper*, VU University of Amsterdam.
- RIORDAN R. AND STORKENMAIER A. (2012), Latency, Liquidity and Price Discovery, *Journal of Financial Markets*, p. 416-437.
- SANDAS P. (2001), Adverse Selection and Competitive Market Making: Empirical Evidence from a Limit Order Market, *Review of Financial Studies* 14, p. 795-734.
- SEC (2010), *Concept Release on Equity Market Structure*, Release No. 34-61358; File No. S7-02-10.
- SCHOLTUS M., VAN DIJK D. AND FRIJNS B. (2012), Speed, Algorithmic Trading and Market Quality around Macroeconomic News Announcements, *Working Paper*, Erasmus University, Rotterdam.
- STORKENMAIER A. AND WAGENER M. (2011), Do we need a European "National Market System"? Competition, Arbitrage, and Suboptimal Executions, *Working Paper*, Karlsruhe Institute of Technology.
- ZHANG S. (2012), Need for Speed: an Empirical analysis of Hard and Soft Information in a High Frequency World, *Working Paper*, Karlsruhe Institute of Technology.