Nikolaus Hautsch
Professor
University of Vienna

High-frequency trading: risks and benefits

1 Introduction

Nowadays a substantial part of trading activity in equity, derivative and currency markets is due to algorithmic high-frequency trading (HFT). The role and effect of HFT on financial markets is controversially discussed and in the center of attention of market operators, regulators, and market participants. This article briefly introduces to the concepts of HFT, reviews its developments through the last decade and summarizes the current state of discussion. It moreover gives an overview of current empirical evidence on the effects of HFT and provides an outlook on its future in light of upcoming regulation.

HFT is characterized as automated trading that employs (i) algorithms for order execution and automatic order routing, i.e., the distribution of (large) orders through time and across different market places, (ii) low-latency technology and co-location services, and (iii) high message rates. HFT is typically carried out by proprietary firms, hedge funds or broker-dealer proprietary desks. High-frequency traders (HFTs) use short holding periods and do no take significant over-night positions. They neither take highly leveraged positions, but face rather low margins per trade, while making profits by executing many (small) trades through a day. Accordingly, they typically focus on highly liquid assets.

A central aspect of HFT is to exploit speed advantage. A central requirement is that the server of the HFT firm is co-located, i.e., it is placed in near distance to the server of the exchange. Exchanges offer this as a paid service, and promise certain latencies. Likewise, HFTs pay for high-speed connections between different market places. While fiber-optic cable connections have been used in the early days of HFT (around 2005), microwave connections and laser links are the current state of the art and push the latency, i.e., the time it takes for a signal to travel from point A to point B, close to natural limits induced by the speed of light.

HFTs perform various kinds of strategies where speed advantages, low reaction times and the ability to post (and cancel) a large amount of orders within very short time periods, are beneficial. One major strategy is market making, i.e., providing liquidity on both sides of the market. Accordingly, HFTs post limit orders on the best ask and best price level and earn the bid-ask spread, similarly to designated market makers in classical floor trading, see, e.g. Demsetz (1968). In some markets, liquidity providers additionally earn a liquidity rebate offered by the exchange to reward market participants for providing market making service. This incentivizes HFTs to serve as passive liquidity suppliers in possibly many transactions.

Other examples are order detection strategies. In many markets, posted limit orders are partly or entirely hidden. The motivation for hiding an order is to get protected from front-running and to avoid price impact, i.e., unfavorable market movements as a reaction to a posted limit order, see, e.g., Cebiroglu, Hautsch and Horst (2014). For HFTs it is beneficial to identify hidden orders placed in

Nikolaus Hautsch

the spread as they induce lower transaction costs. A common way to identify hidden liquidity is to post so-called immediate-or-cancel (IOC) orders that are automatically canceled if they do not get executed. An obvious downside of such strategies is that they create substantial message traffic with high order-to-trade and cancelation ratios. For instance, on Nasdaq, up to 90%–95% of all posted limit orders are canceled shortly after submission and thus get never executed.

Another important strategy is statistical arbitrage. Traders try to make profits by exploiting temporary inconsistencies in prices between different exchanges or assets. Due to a high market fragmentation, such strategies are particularly pronounced in the U.S.A. Other dominant strategies are latency arbitrage, exploiting direct market access and the possibility to receive market data a few milliseconds earlier than other market participants. With such a speed advantage it is possible to react faster on corresponding trading signals or to anticipate order flow that is automatically routed through smart order routers.

Momentum ignition strategies involve posting and cancelling a large number of trades and orders in a particular direction in order to trigger a price movement and to cause other algorithms to react on it. HFT firms have established a position earlier on and can benefit by leveraging the subsequent price movement. "Spoofing" is a more extreme form of it, where orders are posted with the intent to cancel them before they get filled. This is a manipulative strategy, which is illegal. A further example for an illegal strategy is "quote stuffing" with the aim to increase the message traffic, such that the bandwidth and thus the access of other market participants is slowed down.

2 Discussion and history of HFT

The role of HFT is controversially discussed. The public discussion and media coverage is dominated by the view that HFT degrades the function of the market, discriminates non-HFTs, makes markets less stable and wastes resources by an unreasonable technological arms race. The most famous critique comes from Michael Lewis in his book Flash Boys (Lewis, 2014). According to Lewis, "speed traders prey on retail investors and rig the stock market". Likewise, Stiglitz (2014) argues that HFT steals information rents and that markets are ultimately too active and too volatile. He asserts that there is no social value as HFT degrades the market function.

Public perception, however, often tends to regard HFT as an isolated phenomenon disconnected from general developments in the trading landscape. In fact, HFT is a consequence of technological progress and regulatory changes during the last decade. The starting point was the change from classical floor trading to electronic trading and the introduction of ECNs in the 1990s. In 1998, the U.S. Securities and Exchange Commission (SEC) passed the Regulation Alternative Trading Systems (so-called Reg. ATS) to restrict the monopoly enjoyed by NYSE and NASDAQ in the U.S.A. This was the starting point for an increase of market fragmentation. In 2001, U.S. stock exchanges began quoting prices in decimals instead of fractions, bringing down the minimum spread. In 2005, the SEC passed the Regulation National Market System (Reg. NMS) requiring trade orders to be posted nationally and not on individual exchanges ("trade-through rule"). Simultaneously, in Europe, the Markets in Financial Instruments Directive (MiFID) introduced a principles-based best execution regime compared to the rules based U.S. approach. This openend the door for Smart Order Routing. In 2007, new market access models have been introduced. Some market participants obtained "direct market access" without sufficient control mechanisms on the validity of orders. Finally, regulation allowed exchanges to introduce co-location and proximity services. Hence, unequal market access has been systematically established.

Regulation thus established a level playing field for HFT. Accordingly, the extent of HFT rapidly increased since 2005. In 2010, HFT accounted for approximately 56% by volume of the entire equity turnover in the U.S., see, Agarwal (2012). In Europe, this percentage amounts to approximately 38% in 2010. Since 2009/10, the extent of HFT in U.S. and Europe equity trading declines and ranges between 40% and 50% in 2014. Similar developments and quantities are observed in U.S. futures trading and FX trading. In Asia, the extent of HFT is generally lower, while in China HFT basically does not exist.

In public perception, HFT is often associated with male-functioning algorithms getting out of control as, e.g., in case of the Knight Capital Group on August 1, 2012, or with (flash) crashes, such as the flash crash on May 2010, where leading U.S. indices dropped by nearly 10% within a few minutes. While incidences based on male-functioning algorithms seem to be an existing (though low) operational risk which typically mostly harms the HFT firms themselves, there is no convincing evidence for HFT causing flash crashes, see, e.g., Kirilenko et al. (2017). In fact, recent empirical research predominantly shows that HFT improves liquidity and market efficiency, thus showing a positive effect of HFT. Hendershott et al. (2011) find that algorithmic trading enhances the informativeness of quotes. Hasbrouck & Saar (2013) show that increased low-latency activity decreases spreads, increases displayed depth and lowers short-term volatility. Menkveld (2013) stresses the role of HFTs as high-frequency market makers. Brogaard et al. (2014) show that HFTs facilitate price efficiency. Hence, the criticism of HFT degrading the market function is empirically not necessarily confirmed.

Nevertheless, some evidence supports more critical views. While Kirilenko et al. (2017) find that HFTs did not trigger the May 2010 flash crash, it is shown that HFTs are nonetheless not helpful in stabilizing markets in such a situation. The authors find evidence for latency arbitrage and inventory changes of HFTs being positively related to contemporaneous price changes. Thus HFTs tend to trade in the direction of the market, which is contrary to "classical" market making. Budish et al. (2015) argue that the high-frequency trading arms race is a symptom of flawed market design and that the re-introduction of (high-frequency) batch auctions would be a favorable alternative to continuous trading.

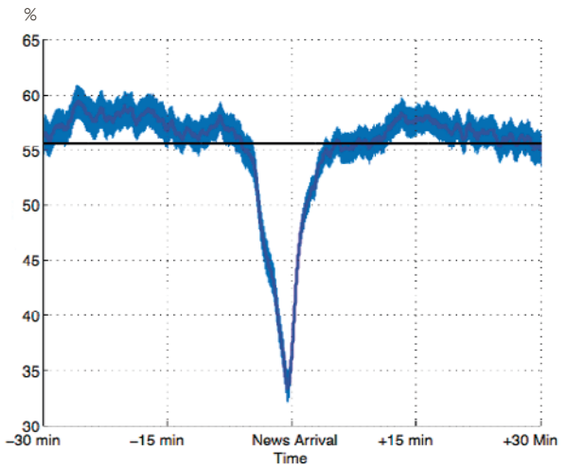
3 Empirical evidence

Hautsch, Noé and Zhang (2017) (henceforth HNZ) provide evidence on the role of HFTs as market makers in Bund Futures trading at the derivatives exchange Eurex. HNZ exploit access to proprietary order-level message data with member ID and trader ID allowing for an institutional HFT identification. In addition, they

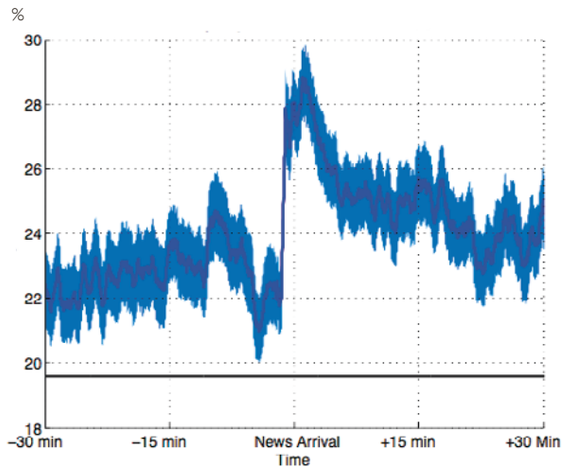
Chart 1

HFT Liquidity in the Bund Futures market

HFT Liquidity supply participation rate



HFT Liquidity demand participation rate



Source: Chart reproduced from Hautsch, Noé and Zhang (2017).

Note: HFT liquidity supply participation rate (left) and demand participation rate (right) in traded contracts through 60 minute windows around scheduled macroeconomic announcements with extreme price movements in Eurex Bund Futures trading, 2014–2015. Averages across announcements with shaded areas indicating 95% confidence intervals. The solid line is the overall mean across all trading days excluding the one hour window around the release.

employ a statistical identification of HFT activity by considering trading desks with a given number of order submissions per day, low end-of-day positions and very short order life times. This allows for a quite precise identification of activity stemming from HFTs and non-HFTs. HNZ particularly focus on turbulent market periods during periods around scheduled macroeconomic news announcements.

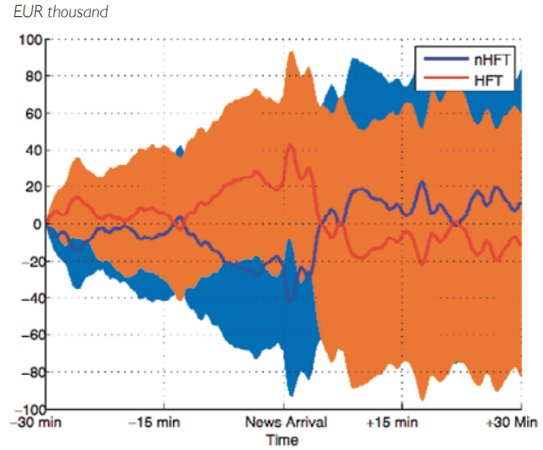
Chart 1 shows the development of liquidity supply and liquidity demand participation ratios around scheduled news announcements creating large price changes. The liquidity supply and demand participation ratio corresponds to the percentage of trading volume where liquidity is supplied and demanded, respectively, by HFTs. In general up to 60% of all liquidity supply in the market stems from HFTs. In contrast, less than 25% of all trades are *initiated* by HFTs. Thus, HFTs are rather passive and tend to do more market making (i.e. liquidity provision) than aggressive trading (i.e. liquidity demand). The only exception is shortly before the news arrival. In the last minute before the news is released, HFT liquidity supply drops from roughly 50% to less than 35%. At the same time, their trading strategies become more aggressive and they significantly increase their liquidity demand. Hence, in periods where uncertainty becomes very high and the risk of a limit order becoming mispriced peaks, HFTs considerably reduce their inventory and at the same time increase their trading activities, presumably trying to exploit latency arbitrage.

As documented by HNZ, this is also reflected in the bid-ask spreads as a measure for trading costs in the market. They show that bid-ask spreads, where HFTs make the market on both sides are generally lower than bid-ask spreads originating from order submissions by nHFTs. Shortly before the news arrival, however, HFT-implied bid-ask spreads widen by approximately 25%. While this behavior is widely in line with the behavior of a "classical" (designated) market maker, such a drop in liquidity provision can happen very rapidly.

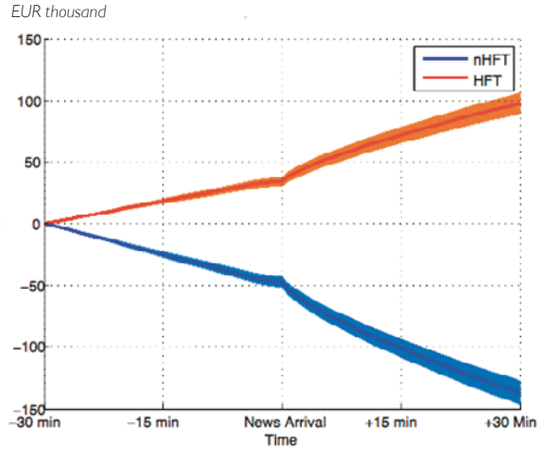
Chart 2

Trading profits of high- frequency trades in the Bund Futures market

Positioning profit



Net spread



Source: Chart reproduced from Hautsch, Noé and Zhang (2017).

Note: Profits through trade positions (left) and through market making during 60 minute windows around scheduled macroeconomic announcements with extreme price movements in Eurex Bund Futures trading, 2014–2015. Averages across announcements with shaded areas indicating 95% confidence intervals.

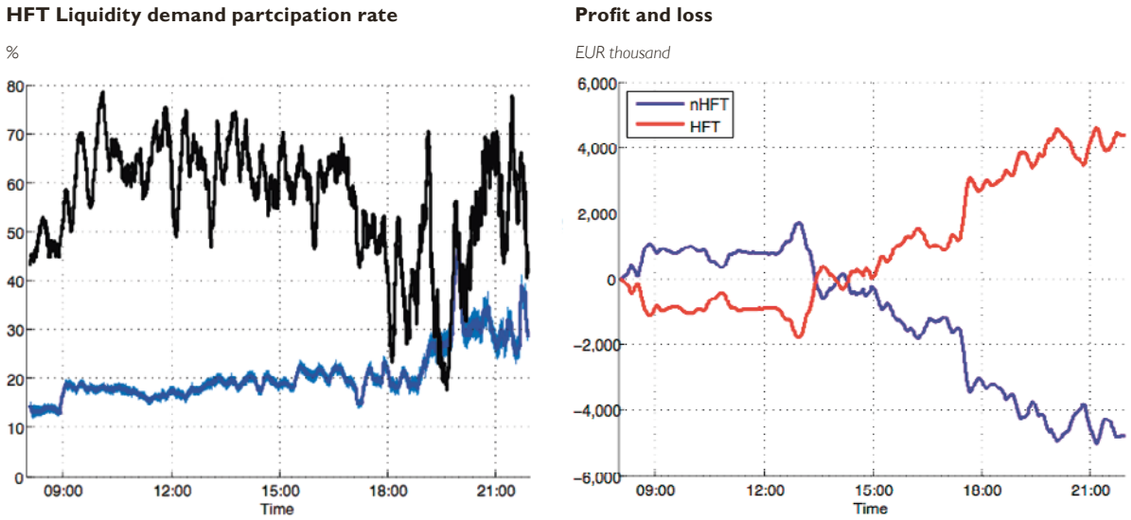
Chart 2 gives the average profits and losses of HFTs and non-HFTs during the 60 minute period around an announcement under the assumption of an initial inventory of zero 30 minutes before the news release. The red curve in the left picture gives the average profit made by the entire HFT activity in periods around news announcements. The blue curve indicates the average profits and losses of all non-HFTs. The two pictures are based on a decomposition of the overall (average) profits into the profits from pure market making, i.e., earning the bid-ask spread (chart 2, right picture), and the profits by active directional trading through this period (chart 2, left picture). Correspondingly, during the hour around a news announcement, HFTs in the Bund future market earn on average close to EUR 100,000. Conversely, non-HFTs loose up to EUR 130,000.[1] Chart 2, however, shows that basically all the profits made by HFTs during this period result from liquidity provision. Likewise, non-HFTs repeatedly pay the spread as they predominantly act as liquidity demanders. Hence, these results seem to suggest that HFTs serve as high-frequency market makers and just take over the function of designated market makers, standing ready to supply liquidity whenever needed and earning the bid-ask spread. HNZ, however, show that this conclusion can be misleading.

Chart 3 shows HFT activity from 9:00 a.m.to 21:00 p.m. on the day after Brexit (June 24, 2016). The black line is the proportion of all trades initiated by HFTs, whereas the blue curve is the corresponding proportion on "normal" days without any particular news events. We observe that on this day, the order aggressiveness of HFTs is significantly higher than during periods around scheduled news announcements. Hence, on such a day of high market turbulence, HFTs do *not* serve as passive market makers but are heavily involved in directional trading. This behavior is in stark contrast to the behavior of a "classical" (designated) market maker.

[1] *The fact that the losses of non-HFTs are higher than the gains of HFTs is due to the existence of trading fees. As non-HFTs tend to initiate trades much more often than HFTs, non-HFTs face significantly higher transaction costs.*

Chart 3

HFT liquidity demand and trading profits in the Bund Futures market after the Brexit decision



Source: Chart reproduced from Hautsch, Noé and Zhang (2017).

Note: Left picture: HFT participation rate in liquidity demand on June 24, 2016 (black line). The blue line presents the average across normal trading days. Right picture: Total profits and losses by HFTs and non-HFTs on June 24, 2016 in Eurex Bund Futures trading.

The right picture in chart 3 displays the profits and losses by HFTs and nHFTs. Hence, HFTs earned approximately EUR 4 million on this day. As shown by HNZ, these profits, however, predominantly result from active (directional) trading but *not* from market making. Further evidence provided by HNZ shows that HFTs obviously do not replace "classical" designated market makers but differ in an important respect: HFTs (rapidly) change their strategy according to the market situation. In a situation where trading opportunities through latency arbitrage come up, they become aggressive and exploit their speed advantage. Then, market making functionality can be severely limited.

Currently, it is still very unclear what the effects on volatility and market stability in such extreme situations might be. It is still widely unknown whether HFT increase the risk of tail events and increase volatility on turbulent days, as the day after the Brexit. More research is clearly needed to gain a better understanding of the possibly dangerous sides of HFT.

4 Future of HFT – regulatory perspectives

Current evidence shows that the extent of HFT will decline due to increasing costs of infrastructure, increasing competition among HFTs and the introduction of alternative trading systems which partly rule out HFT, e.g. via dark pools, or so-called "speed bumps", where market access is artificially and randomly delayed, such that millisecond speed advantages disappear. According to estimates by Kaya (2016), the overall HFT revenues in the U.S. decline from approximately USD 7 billion in 2009 to less than USD 2 billion in 2014. Though the very glory times of HFT in 2009 seem to be over, it is not expected that HFT will disappear. As long as trading designs are not systematically changed, HFT will remain an integral part of electronic trading and should be understood as a consequence of market evolution and past regulation. Correspondingly, the future of HFT will strongly depend on upcoming regulation.

In fact, currently there exists severe regulatory uncertainty. In the U.S., as a response to the May 2010 flash crash, the SEC introduced trading pause regulation preventing further flash crashes. In the same year, the Dodd-Frank legislation restricted the so-called proprietary trading of banks and the SEC (Security Exchange Commission) issued a ban on "naked" (unfiltered) market access. Current developments go into the direction of more monitoring, recording and registration of HFT activity. Since 2015, the SEC forces certain HFT broker-dealers to register with the Financial Industry Regulatory Authority (FINRA) strengthening the SEC's oversight of proprietary firms. In 2015, the CFTC proposed rules for a regulation of automated trading (Reg AT), governing certain HFT practices.

In Europe, MiFID II will become in force in 2018. MiFID will require HFT firms to provide details on the nature of algorithmic trading strategies, trading parameters, and risk controls. There will be specific obligations for trading venues in terms of monitoring, circuit breakers, capacity requirements, pre-trade and post-trade transparency and manual "kill functionality". HFT firms will also have to test the conformance of their trading systems and algorithms. There will be algorithm-tagging rules in order to identify market manipulation. Finally, HFTs have to fulfill certain obligations if they want to pursue market making strategies.

There are, however, potential regulatory pitfalls. First, there might be too much focus on monitoring, registration and (massive) data collection. Though it opens up the possibility to investigate potential market manipulation and to detect fraud, it will require substantial resources to process and analyze this massive data. Currently, it is unclear whether this will be efficient and effective. Second, in the MiFID II regulation, it is unclear how to implement all the planned details on risk control and the testing of algorithms. Third and most importantly, current regulation plans are too rigid for market making. Empirical evidence, as discussed above, demonstrates that HFT market making is beneficial for market quality and other market participants. Regulation thus should try to strengthen these strategies and thus to preserve the benefits from HFT while simultaneously mitigating risks. Hence, a negative scenario could be that of a too rigid and misguided regulation, which will reduce HFT, but will also reduce market quality in terms of lower liquidity, higher transaction costs and higher volatility. HFT and liquidity will flee into other – potentially non-regulated – markets while we are confronted with high (maybe too high) regulation costs.

Ideally, technological innovation should go hand in hand with smart regulation. For instance, the idea of limiting latency differences is a good way to stop the arms race for speed and to reduce predatory trading and a major amount of harmful HFT strategies. Such an artificial delay of trading, a so-called "speed bump", is the major concept of the exchange IEX, which has been officially approved by the SEC in June 2016 and is a growing exchange that is even partly supported by HFT firms themselves. With such a speed bump (on IEX it is 350ms), HFT market making could be still a beneficial strategy, while the downsides of HFT, such as predatory trading and the arms race for millisecond speed advantages could be limited. In combination with a well-balanced use of "circuit breakers" and general safeguards, we could accept HFT as a "normal" integral part of modern trading, which would settle down to a moderate level and will predominantly concentrate on strategies which are favorable for market quality, such as liquidity provision.

References

Agarwal, A. 2012. High Frequency Trading: Evolution and the Future. www.capgemini.com/capitalmarkets.

Brogaard, J., Hendershott, T. and R. Riordan. 2014. High Frequency Trading and Price Discovery. Review of Financial Studies 27(8). 2267–2306.

Budish, E., P. Cramton and J. Shim. 2015. The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. In: Quarterly Journal of Economics 130(4). 1547–1621.

Cebiroglu, G., N. Hautsch and U. Horst. 2014. Does Hidden Liquidity Harm Price Efficiency? Equilibrium Exposure under Latent Demand. CFS Working Paper 468.

Demsetz, H. 1968. The Cost of Transacting. In: Quarterly Journal of Economics 82. 33–53.

Hautsch, N., M. Noé and S. S. Zhang. 2017. The Ambivalent Role of High-Frequency Trading in Turbulent Market Periods. Working Paper.

Hasbrouck, J. and G. Saar. 2013. Low-latency trading. In: Journal of Financial Markets 16(4). 646–679.

Hendershott, T., C. M. Jones and A. J. Menkveld. 2011. Does Algorithmic Trading Improve Liquidity? In: Journal of Finance 66(1). 1–33.

Kaya, O., 2016. High-frequency trading – Reaching the limits. Research Briefing Global Financial Markets. Deutsche Bank Research. May 24.

Kirilenko, A., A. S. Kyle, M. Samadi and T. Tuzun. 2017. The Flash Crash: High-Frequency Trading in an Electronic Market. In: Journal of Finance 72. 967–998.

Lewis, M. 2014. Flash Boys. W. W. Norten & Company.

Menkveld, A. J. 2013. High Frequency Trading and The New-Market Makers. In: Journal of Financial Markets 16. 712–740.

Stiglitz, J. E. 2014. Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy? Working Paper presented at the Federal Reserve Bank of Atlanta 2014. Financial Markets Conference "Tuning Financial Regulation for Stability and Efficiency".