Oxera

Agenda

Advancing economics in business

High-frequency trading: towards capital market efficiency, or a step too far?

The growth in high-frequency trading has been a significant development in equity markets, with regulators and governments under pressure to 'do something' about it, notwithstanding the diversity of views on whether this growth is good or bad for investors. What is the likely impact of some of the proposals that have been put forward to control HFT, and how will these interventions actually affect the market place, if at all?

The past five years have seen a sharp growth in high-frequency trading (HFT) on the world's major stock exchanges and other trading venues. Although there is no agreed, detailed definition of what HFT is, there are some characteristics that distinguish this development from what went before. In particular:

- traders have emerged who tend to trade a complete cycle (in its simplest form, buying and then selling the same security) over very short time periods—seconds, or even parts of seconds, rather than minutes, hours or days;
- the speed with which these traders can react to detailed changes in market conditions is important to making a success of this type of trading. Both absolute speed and speed relative to others attempting to use the same strategies are important, and, as a result, very small changes in speed become critical;
- the reaction to changing market conditions leads to traders sending very high levels of orders to trading venues, and then cancelling a fairly high proportion of these orders before they actually execute. (For some of the techniques used, only one in 500 or so orders sent to the venue will execute, notwithstanding the fact that, in order for the strategy to be profitable, some trading must actually take place);
- these traders are usually deciding, on the basis of detailed analysis of past market behaviour, precisely when to trade, using detailed monitoring of current market conditions to create very short-term (within seconds or minutes) predictions of whether they can execute the full sequence of trades that returns them to a neutral position within that timeframe. They may also be trying to identify how prices will change; or find fleeting anomalies in the price of the same securities trading in different locations, or anomalies in the price of different securities that are linked in some way (eg, the price of a derivative and the price of the underlying security).

The techniques used by high-frequency traders to make money from trading are, however, generally the same as those used before—ie:

- simultaneously offering to buy a security at one price and offering to sell it at a higher price (market making);
- predicting that the price of a security will change, and buying if the prediction is a price rise, then selling the security if and when the predicted price change occurs (and vice versa);
- arbitrage between different venues or between different, but linked, securities.

Although these techniques are the same, what has changed is the speed with which a high-frequency market-maker can complete the round trip (for example, the buy and sell of the same security); the successful trading of small (and therefore frequent) price changes; and the speed with which price anomalies can be spotted and used to trade profitably.

This article is based on a published Oxera report commissioned by the Foresight Programme to assess proposed European Commission rules aimed at regulating HFT. Oxera (2012), 'What is the Economic Impact of the MiFID Rules aimed at Regulating High-Frequency Trading?', available at http://www.bis.gov.uk/assets/foresight/docs/computer-trading/12-1080-eia21-economic-impact-mifid-rules-high-frequency-trading.pdf. The Foresight Programme advises the UK government on how to ensure that today's policy decisions are robust to future uncertainties. For details, see http://www.bis.gov.uk/foresight.

Rather surprisingly, however, given the name, the actual average frequency of trading has not increased that much, and nor has the value of trading. What has increased, by orders of magnitude, is the volume of orders and subsequent cancellations.

What impact has this had?

Depending on which reports you read, these developments have made it harder for other market participants to successfully make their trades; have exacerbated the fundamental instability of capital markets; have increased liquidity (making it easier for others to trade); have reduced spreads (making it cheaper for others to trade); or have increased *and* decreased the depth of the order book—and the list of effects goes on.

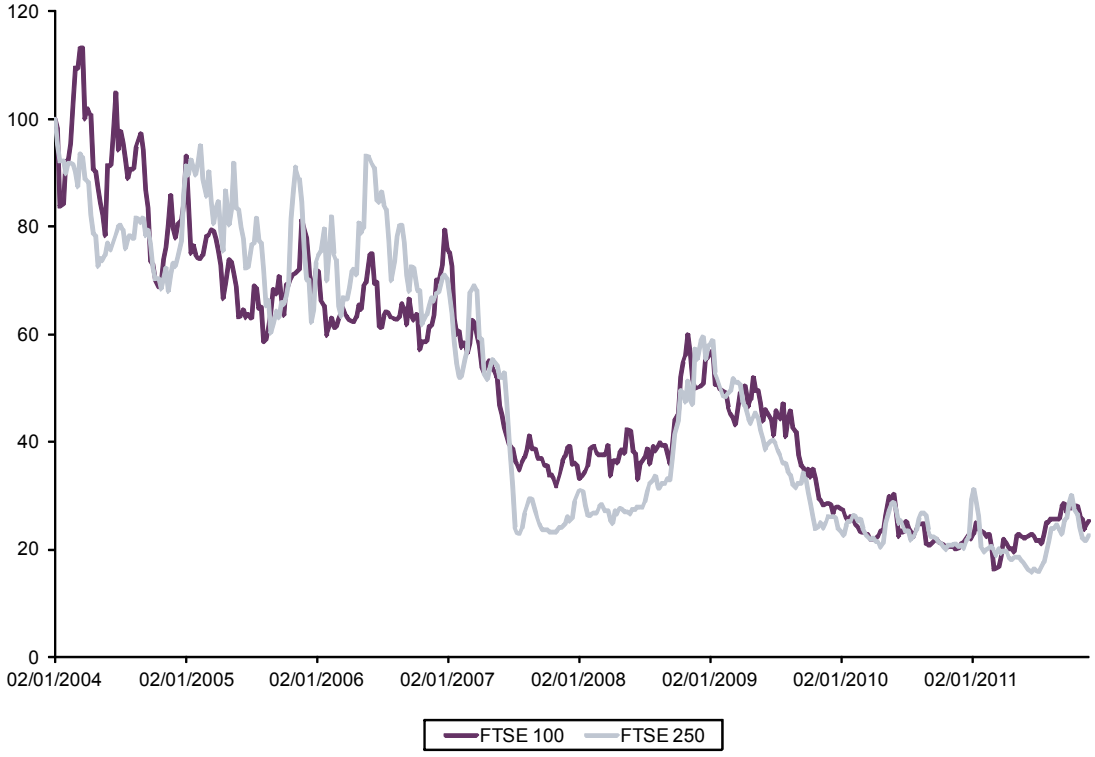
If HFT were having serious effects on the trading ability of others, it would be expected that there would be some significant difference in the performance of markets with and without the prevalence of HFT, as well as an impact on the ability of end-investors to trade.

To take a European example, the average spread on the FTSE 100 and FTSE 250 shares has been falling, at least until recently. Figure 1 below shows that the average closing spread (indexed to 100 at January 2004) has been falling both since the financial crisis and relative to its value before the crisis, for both indices.

HFT occurs mostly in highly liquid securities in the FTSE 100, and not (in general) in FTSE 250 securities. The patterns of changes in the spreads over time are remarkably similar, notwithstanding this difference. In addition, there are measures relating to how successful, or otherwise, end-investors are at executing trades. Figure 2 overleaf shows the evolution of one of these measures[1] for two sizes of US securities, and all UK securities. On this measure, there has been little change in the implicit costs of trading over the period during which HFT has been growing—and, again, the pattern in markets where HFT is thought to be more prevalent (US large cap) is similar to that in the markets where HFT is thought to be less prevalent (US small cap).

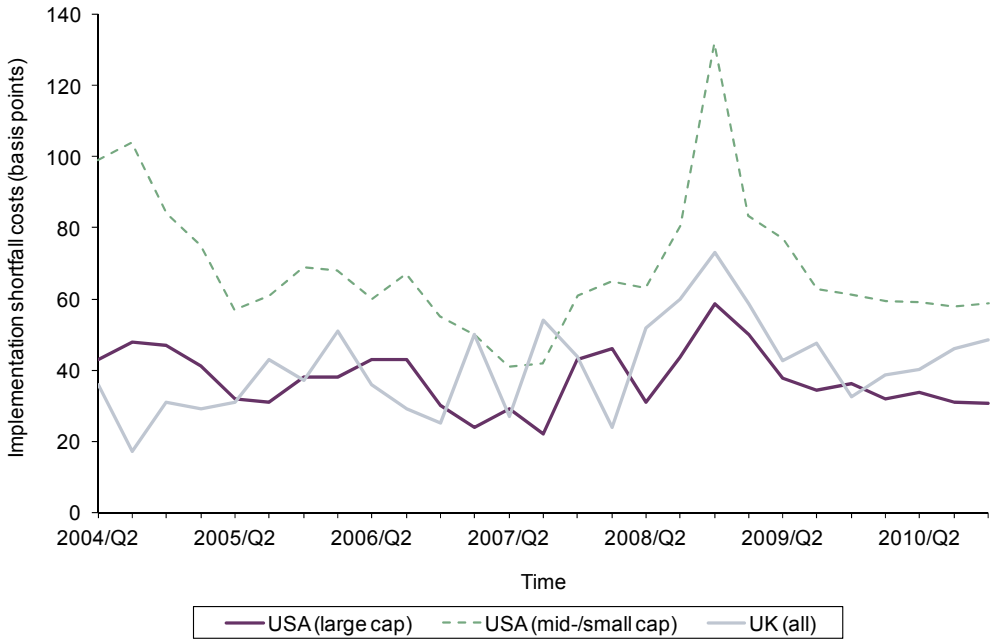
It may just be that HFT is having a bad (good) effect but that, where this happens, other participants take actions which then counteract this effect. A much more detailed analysis of the market would be required to conclude on this point.

Figure 1 Average closing spread (market cap-weighted), one-month rolling average, indexed to January 2004 (=100) for the FTSE 100 and FTSE 250



Note: Data relating to the period around Christmas has been removed. Source: Bloomberg, and Oxera calculations.

Figure 2 Implementation shortfall in US and UK markets



Note: Implementation shortfall: 'the difference, or slippage, between the arrival price and the average execution price for a trade'.
Source: Investment Technology Group (2011 and 2008, various dates), 'ITG's Global Cost Review', available at http://itg.com/news_events/papers/ITGGlobalCostReview_2010Q3_Final.pdf.

Intervention to stop the 'bad' effects of HFT (which may also risk stopping the 'good' effects)?

Notwithstanding the lack of agreement on the impact of HFT on end-investors, various financial authorities have applied, or are proposing to apply, new rules and/or taxes to activities that are classified as HFT and/or stakeholders classified as high-frequency traders. Analysing the impact of these proposals would appear to be in line with the principles of evidence-based policy-making. However, there is currently little in the way of robust analysis, partly as a result of confusion about the potential positive and negative effects of HFT (or even exactly what it is).

The proposed rules for Europe include:[2]

- imposing a minimum time between an order being sent to a trading venue and when it can be cancelled (minimum resting periods);
- imposing a maximum ratio between orders that are cancelled and those that execute (maximum cancellation ratios).

These two rules could be seen as targeting the behaviour of high-frequency traders (and which might be seen as 'bad'):

- they make decisions on trading, and react to events, very quickly, and more quickly than those with smaller computers or those located further away from the venue in question (which has been seen as being unfair to other market participants, including investors);
- they cancel most (or even all) of their orders, implying that most orders must somehow not be real (which has been seen as being designed just to confuse other market participants).

By looking at how high-frequency traders actually trade within markets, it may be possible to obtain some idea of how these rules would affect HFT.

Minimum resting periods

In most exchanges, when an order is sent to an exchange it can either interact with an opposite order that is already there—in which case it will execute immediately—or, if there is no counterparty immediately available, it will rest until either a new order arrives from a counterparty that will execute, or the order is cancelled. The minimum resting time would specify how quickly this cancellation could take place. (The time periods under discussion are in the order of 200 or 500 milliseconds.) A rule of this sort will have an effect only when an order that did not execute on arrival (ie, becomes a resting order) would have executed within the specified time period, but the person who sent the order now does not want it to execute (ie, they have 'changed their mind' within the 200 or 500 milliseconds).

There is little available analysis, however, on the time that resting orders that do execute have actually rested for. The evidence available for ten of the most traded securities on NASDAQ in 2009 (ie, after HFT became fairly widespread) shows that the average was around 75 seconds, although a significant proportion do execute within 200 or 500 milliseconds (around 10% and around 20% respectively)—see Figure 3 below.

To execute, the order that rests is placed at the back of the queue within the tick (the 'tick' is the pre-set price level that orders can be placed at—for example, an offer to buy at £10.00 or £10.05, but offering to buy at £10.03 is not allowed). Any order at the same price that has been sent earlier will be in front, and will have to execute (or be cancelled) before the order that has just been placed. So, an order placed at the back of a queue with, say, four orders in front of it will not execute until those four have executed (or been cancelled). Over a normal trading day for FTSE 100 shares, a resting buy or sell order for a specific security will execute every 16 seconds (in 2011, traded on the London Stock Exchange).[3] For heavily traded securities (eg, shares in Vodafone), the average period is shorter, but still only once every two or three seconds.

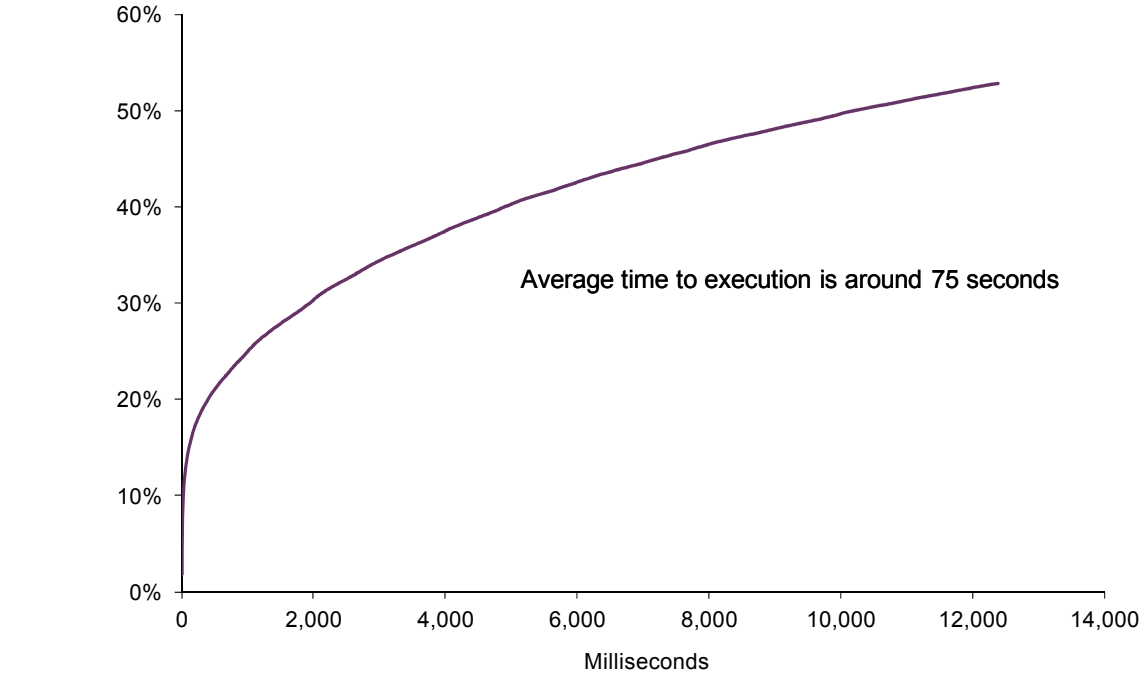
An order placed at the back of the tick at the price at which the security is currently trading ('at the touch', or at the best bid or offer) with four or five orders ahead of it might not, therefore, have a high probability of executing within even 500 milliseconds; hence, in a large proportion of cases the proposed rule would have no impact.

There are other orders which, although they rest, will execute much faster. These are orders placed either into a tick at the touch which is almost empty, or into an empty tick ahead of the current touch. Analysis of the same NASDAQ stocks indicates that orders sent to ticks inside the touch and which rest, tend to rest for an average of only 400 milliseconds. However, any competent high-frequency trader will have a good idea of how many orders will be in front of them in the tick at the touch, and whether they are sending an order to rest in the empty tick ahead of the touch. Since complex prediction is central to HFT, predicting (on a probabilistic basis) the minimum likely time to execution would be expected to be incorporated into a high-frequency trader's analysis, with their orders and cancellations modified as a result. Given the timings set out above, apart from orders sent to (empty) ticks within the touch, the modification to high-frequency traders' trading patterns as a result of the proposed rule may be quite small.

Maximum cancellation ratios

Some, but not all, HFT strategies have high cancellation ratios. This will generally be because the trader does not want a resting order to execute at the

Figure 3 Cumulative frequency distribution of resting time to execution (ten securities on NASDAQ 2009)



Source: Oxera, based on Hautsch, N. and Huang, R. (2011), 'Limit Order Flow, Market Impact and Optimal Order Sizes: Evidence from NASDAQ TotalView-ITCH data'.

wrong time.[4] When the order is placed, there will generally need to be some probability that it will execute (although this probability may be low), since executing is a necessary part of a high-frequency trader having an economic business. As such, applying a maximum cancellation ratio will reduce the number of orders submitted that have a low probability of executing at the right time. As probabilities are core to HFT, a rule of this sort could be expected to drive changes in the placing, and cancellation, of orders.

That said, as long as the trader has some idea of the probability of individual orders actually being allowed to execute, eliminating (by not posting) the lowest-probability orders can be expected to drive up the execution ratio but without significantly changing the orders that actually do execute. The same effect can be expected on the cancellation of existing orders: orders with a high probability of executing at the wrong time still get cancelled, but orders with a low probability of executing at the wrong time get left in the order book. The impact of what actually happens in the market may, therefore, be rather slight.

In addition, if that does not work, a high-frequency trader with a high cancellation ratio could team up with a market participant with a much lower cancellation ratio (say, a broker acting for an investor), and together their combined ratio would fall to meet the regulatory requirement, but without changing their respective trading behaviours.

Conclusions

A careful analysis of how the proposed rules designed to regulate HFT would affect what high-frequency traders actually do, and what trades actually take place, suggests that the outcome may not be as is generally expected. To understand the actual implications, policy-makers will need to undertake extensive testing of any proposed rules on the messaging and trading patterns actually taking place. Any proposed intervention will also need to state clearly what problem the rules are designed to address, and demonstrate how the rule would have the desired impact. Without this analysis, there is a risk that the types of rule being proposed either do not have the desired effect, or could have undesirable effects on those stakeholders whom the rules are designed to benefit.

[1] The measure used is the implementation shortfall, which is the difference between the price actually achieved for an investor's trade and the last price obtained in that security prior to the investor starting to buy (or sell) it. The measure combines the impact of the prevailing spread and the impact on the price while the order is being executed, and thus takes account of the liquidity in the market.
[2] Other proposed rules include imposing a minimum tick size, and coordinated 'circuit breakers' (which automatically suspend trading when a certain speed of price changes is reached) on exchanges; a continuous market-making requirement on high-frequency traders; registration and/or approval of algorithms by regulators before they can be used; and requiring trading venues to charge for messaging or cancellations of orders, rather than charging when a transaction is actually undertaken.
[3] Oxera calculations, using London Stock Exchange trading data.
[4] There are examples of market manipulation strategies where an order is designed never to execute, but to create a false impression of where the market is heading and then to execute different trades to exploit this false market. However, such behaviour is generally a breach of the existing rules and so does not require special HFT rules.