Government Office for Science

Foresight

High-frequency trading and the execution costs of institutional investors

Driver Review

Foresight, Government Office for Science

Contents

High-frequency trading and the execution costs of institutional investors

Jonathan Brogaard, Terrence Hendershott, Stefan Hunt, Torben Latza, Lucas Pedace, Carla Ysusi[1]

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

[1] Jonathan Brogaard, Foster School of Business, University of Washington; Terrence Hendershott, Haas School of Business, University of California Berkeley; Stefan Hunt, Financial Services Authority; Torben Latza, University of Bristol; Lucas Pedace, Her Majesty's Treasury; and Carla Ysusi, Financial Services Authority (author for correspondence, carla.ysusi@fsa.gov.uk). This paper uses data from various trading venues and market participants that was provided to the FSA. The data was not provided specifically with this research in mind. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Financial Services Authority or any of the trading venues or market participants. They do not. All errors or omissions are the authors' sole responsibility. The London Stock Exchange is not party to this research or responsible for the views expressed in it. References in this report to data from the London Stock Exchange refers to information provided by the London Stock Exchange to the UK Financial Services Authority, which regulates the London Stock Exchange, at the FSA's request. The data provided by the London Stock Exchange to the FSA relates to historic trading data from 2010 and such data was made anonymous, aggregated and did not identify member firms or their activity.

Abstract

This paper studies whether high-frequency trading (HFT) increases the execution costs of institutional investors. We use technology upgrades that lower the latency of the London Stock Exchange to obtain variation in the level of HFT over time. Following upgrades, the level of HFT increases. Around these shocks to HFT, as far as can be measured, institutional traders' costs remain unchanged. We find no clear evidence that these increases in HFT activity impacted institutional execution costs.

> Institutional investors also have expressed serious reservations about the current equity market structure. … [I]nstitutional investors questioned whether our market structure meets their need to trade efficiently and fairly, in large size.
>
> ~ Mary Schapiro, SEC Chairman, September 7, 2010 speech.

1. Introduction

Transaction costs matter.[2] Financial markets exist so investors can efficiently transfer assets and their associated payoffs and risks; the cheaper it is to transfer an asset, the more likely the most-suited investor will end up holding the asset. In addition, with lower transaction costs, investors with private information can more readily buy and sell, aiding price discovery.

High-frequency trading (HFT) accounts for an increasingly large fraction of financial market trading, potentially affecting transaction costs. HFT is a subset of computer-based trading, defined by the use of sophisticated trading algorithms and the ability to trade rapidly to make proprietary returns. Until recently, human intermediaries, such as NYSE Specialists and registered market makers, facilitated the smooth transfer of assets. Now, many human market makers have been substituted by computers. Some high frequency trading firms (HFTs) engage in active trading such as arbitrage, structural, and directional strategies (Hendershott, 2011).

While the rise of machines has raised concern, most academic evidence suggests it has improved measures of market quality such as volatility, price discovery, and liquidity (e.g. see Linton and O'Hara, 2011). Even though raw measures of market liquidity may improve, including the spread, this does not necessarily imply that institutional investors are better off.[3] Some claim that execution costs, a component of transaction costs, could be increasing because of HFT. Possible reasons include faster reaction to public information by HFTs, which pick off orders from slower market participants, or trading in front of institutional investors through the detection of autocorrelation in order flow caused by institutional investors entering large trades. Important gaps exist in the literature on the impact HFT has on the different components of the transaction costs of institutional investors including execution costs.

This paper aims to address one of these gaps. We construct measures of HFT activity and institutional investor execution costs. We show that HFT activity increases following improvements in exchange speed. From 2007 to 2011, the London Stock Exchange (LSE) implemented a variety of improvements to its technology that dramatically increased exchange speed. These changes in exchange speed are used to study the role of HFT in institutional execution costs. We find no relationship between these shocks to the activity of HFTs and institutional execution costs.[4]

We study trading in the UK and show that institutional trading costs in the FTSE's 250 largest stocks have been decreasing since 2003, albeit with an interruption during the financial crisis,

---

[2] We use the term 'transaction costs' to mean all costs incurred in financial trading, including execution cost, commissions and rebates, IT costs and other costs. We use the term 'execution costs' – synonymously, trading costs – to mean execution shortfall, the volume-weighted percentage difference between the price available in the market when brokers receive institutional orders and the price at which the order is executed. Our definition follows Anand et al (2010a). See Section III for further details.

[3] We use institutional investors, as with other papers in the literature, to refer to buy-side institutions such as pension plans and money managers. Our data come from ANcerno, a well-known consulting firm that works with pension plan sponsors and money managers to monitor their equity trading costs.

[4] As noted, this is only one component of their transaction costs. We do not study, for example, whether HFTs have increased commission costs by increasing the number of trades to fill an order.

using a dataset from Ancerno. This trend is consistent with work done by others on trading costs in the US (Anand, Irvine, Puckett, and Venkataraman, 2010a). Like other papers in the academic literature, the measures used in this paper capture only the execution component of trading costs and do not include other trading costs such as commissions or technology costs.

The remainder of the paper focuses on November 2007 to August 2011. During this period we can identify HFT activity using the Financial Services Authority's (FSA) Sabre II dataset. The dataset includes all transactions by observable HFTs in the top 250 UK-listed stocks. To be in the dataset, a HFT firm must either be directly regulated in the European Economic Area (EEA) or must trade through a broker.[5] We do not observe the trades of unregulated HFT firms that are placed directly on trading venues.

The Sabre II database covers all transactions of EEA-regulated firms in all debt, equity and debt and equity derivative instruments listed in the UK. Sabre II, and its successor Zen, constitute a rich record of trading in one of the world's major financial centres, though they have only been used in two previous research studies (Gondat-Larralde and James, 2008, and Benos and Sagade, 2012).

This FSA dataset captures most HFT activity in the UK until July 2010. For 2010, we have an external dataset to corroborate our HFT activity measure and the FSA dataset covers between 70% and 80% of total trading volume of HFTs until July. However, from August 2010 coverage falls to 40%. The decline is attributed to some HFTs becoming direct members of a trading venue, and no longer being obliged to report.[6] Consequently we focus our analysis on the period prior to August 2010.

To study the role of HFT in institutional execution costs, first, we regress the level of HFT activity at the stock-day level on the speed of the LSE system. The regression includes a set of relevant control variables to control for long-term trends in our HFT activity measure and isolate the short-run impact of infrastructure changes on HFT activity. We find that, for two of the four LSE system changes before August 2010, HFT increases. We also examine execution costs around these technology changes, however, and find no measurable change. These results together show little measurable effect of HFT on institutional execution costs. However, intraday prices are volatile, which causes execution cost measures to be highly variable and makes it difficult to identify small changes in costs.

Second, under the maintained assumption of no effect on institutional execution costs through other channels, we can use an empirical method called two-stage least squares to examine the impact of HFT on institutional investors' execution costs. This method estimates the influence of one variable on another. We believe that a negligible effect through other channels is plausible as HFTs are the most capable to quickly utilize lower latencies. We regress our execution cost measure on our estimated proxy for HFT from the first stage regression, continuing to find, however, no evidence that HFT has an effect on institutional execution costs.

[5] All brokers are regulated and must report the transactions of their clients.

[6] While we recognise that a factor in the fall in coverage is because some HFTs became direct members, we note that the FSA supervises and ensures that automated market abuse detection techniques are performed by the platforms on all of these trades, and indeed all of their Members' orders and transactions. The transaction reports for the other side of each trade are also collected and monitored.

Future work will attempt to extend our analysis to additional events that cause variation in HFT activity, to include further measures of execution costs and to examine other outcome measures including other components of transaction costs.

The rest of the paper is structured as follows. Section 2 covers the relevant background information. Section 3 discusses the data and shows the time series of HFT and execution costs. Section 4 lays out the empirical methodology and presents the results. Section 5 discusses the ten-year implications. Section 6 concludes.

2. Background

2.1. HFT literature

A variety of papers have been written about computerised trading, many of them coming from the United Kingdom government's Foresight Project, 'The Future of Computer Trading in Financial Markets.' This section covers the most relevant Driver Reports for this article.

Below we summarise the key findings of the Foresight Driver Reports on HFT so as to provide the reader an overview of the main issues and clarify where our paper fits in the project.

Hendershott (2011) argues that HFTs are similar to other speculators in the market and, while they often improve price efficiency, they could also decrease it, possibly having a negative effect on institutional investors. He defines four main HFT strategies – passive market-making, arbitrage, structural, and directional – and how they can affect market quality. While passive HFT market-making should generally reduce transitory pricing errors, a structural strategy that takes advantage of stale non-HFT orders could cause non-HFT liquidity providers to withdraw from the market, potentially increasing transitory pricing errors. Generally, however, such competition should lead to more efficient markets. Arbitrage strategies can be viewed as an efficient way to provide liquidity, but could also lead to non-HFT withdrawal. And while directional HFT strategies based on identifying and trading against transitory pricing errors lead to more efficient prices, momentum ignition would increase transitory volatility.

Rather than focus on trading strategies, Brogaard (2011) decomposes the activity of HFTs based on the type of information they observe, and how it relates to their profitability. The speed at which HFTs participate in financial markets implies their trading decision depends on information that changes rapidly. This information includes "order book dynamics, trade dynamics, past stock returns, cross stock correlations, cross asset correlations, and cross exchange information delays." It can also include "information that may be illegitimately obtained or created ... from front running, quote stuffing, or layering." Hendershott (2011) and Brogaard (2011) suggest there is a limited amount of information that can drive the trading activity and market participation of HFTs. While the information is limited, the way in which it is interpreted can vary from firm to firm. For instance, each firm may have a different parameter in its trading algorithm on how to handle different variables of interest, such as the slope of the order book or the depth at the best bid and offer. While both papers provide a useful overview of how HFTs might interact with data, their conclusions on the impact of HFT are ambiguous. They do not provide empirical evidence on how it can benefit or harm institutional traders' trading activities.

It could be that information on trade and order book dynamics is used in such a way to push prices away from their true value temporarily when an institutional trader is detected in the market. On the other hand, institutions also use algorithms to implement their buying and selling and so can actively try to 'hide' in the order flow. HFTs likely use order book and trade

dynamics to manage risk. A better managed market maker could benefit institutional investors by being a more stable market participant and being relatively less likely to suddenly reverse their normal activities and unload equity positions onto the market.

Menkveld (2011) hypothesises on the growth of HFT and its influence on transaction costs. He concludes similarly that HFTs could improve the trading outcome for investors by providing quotes and linking markets. On the other hand, HFTs might also negatively affect trading:

> If HFTs trade aggressively on quickly-processed public information they effectively increase adverse selection on investors' price quotes. This essentially eliminates the ability for investors to earn the bid-ask spread rather than pay it which might make some trading strategies prohibitively expensive (e.g., option replicating strategies). Also, speed might trigger a socially wasteful arms race among high-frequency traders. Finally, electronic markets might be particularly vulnerable to new manipulative or socially destructive trading strategies.

Other Foresight papers examine market quality over time. Friederich and Payne (2011) show the trends in market conditions, though they are not able to infer causes for the trends. They find "evidence consistent with CBT [Computer-Based Trading] generating order flow that is more 'continuous', with smaller trade sizes, more frequent trades." However, the impact of CBT is not necessarily the same for all types of stocks: "the increase in CBT that has affected larger stocks may have diminished trading interest in the stocks of hundreds of smaller listed companies, with attendant negative impact on liquidity." While they find that CBT in liquid and large market capitalisation stocks may have positive effects, they also observe that smaller stocks appear to have had a decrease in their liquidity. They suggest this may be a result of attention being taken away from these smaller stocks and instead being focused on the large stocks.

Linton (2011) studies how the UK markets have evolved over the last decade and finds few changes that are statistically significant. He finds no clear evidence for a change in price volatility, the frequency of large price changes, liquidity, market efficiency or the contribution to volatility from the intraday period. While volume has declined, "it is hard to make a firm conclusion about the statistical significance of the recent declines in volume for the index."

Linton and O'Hara (2011) survey the literature and conclude "the evidence suggests that computerised trading (whether in the guise of high frequency trading or algorithmic trading) has generally improved market quality. Liquidity, as measured by bid/ask spreads and other metrics, has improved over the last decade. During this period, transaction costs have also fallen for both retail and institutional traders. These liquidity and transaction cost effects have been particularly pronounced for large stocks."

However, they do raise concerns. In particular, the concern that while on average, computerised trading may result in improved market quality, it can create episodic periods of poor market quality. "Unlike traditional designated specialists, high frequency traders typically operate with little capital, hold small inventory positions, and have no obligations to provide liquidity during periods of market stress. The speed of trading as well as the interconnectedness of markets made possible by HFT can transmit disruptions almost instantaneously across markets."

While Friederich and Payne (2011), Linton (2011), and Linton and O'Hara (2011) provide useful insights into the time series properties of market quality, they do not directly examine

institutional investors' trading costs, or unravel the relationship between HFT and trading costs. This paper examines whether there is a link between institutional investors' execution costs and HFT.

2.2. Execution costs literature

As there is an extant execution cost literature, we primarily focus on introducing the execution cost measure and describing the Ancerno dataset. We use the Ancerno dataset to measure institutional investors' trading costs in UK securities.

Bessembinder (2003) succinctly summarizes why trading costs matter: "Obtaining accurate measures of trade execution costs and assessing the reasons for their systematic variation is important to individual investors, portfolio managers, those evaluating brokerage firm or financial market performance, and corporate managers considering where to list their shares."

We focus on the effective spread costs as the measure of interest with respect to execution costs. It is a measure widely used by academics as well as practitioners (Bacidore and Sofianos, 2002, Grullon, Kanatas, and Weston, 2004, Hendershott, Jones, and Menkveld, 2011). In addition, the fact that it assumes a price away from the time during which the institution traded is the right measure of the 'true' price is particularly valuable in our context.

The dataset we use to measure execution costs comes from Ancerno. The Ancerno dataset has been used in other academic papers. For instance, Anand, Irvine, Puckett, and Venkataraman (2010a) and Anand, Irvine, Puckett, and Venkataraman (2010b) use the dataset. The former analyses institutional trading costs during the financial crisis, while the latter examines the persistence of different institutional traders' execution costs. Anand et al. (2010b) provides a thorough description of the dataset in its appendix. The authors had several conversations with Ancerno to understand the nuances in how the data were gathered. Most important for this paper are survivorship bias and selection bias. Anand et al. (2010b) argue that survivorship bias is not a concern for two reasons that are likely to hold for the UK data. First, they were reassured by Ancerno representatives that there was no such bias. Second, they observe firms in the data that dropped out of the sample in the middle of the dataset time series.

They also test whether the institutions differ markedly from all 13F institutions using a set of 64 institutions in the Ancerno database. They "find that the characteristics of stocks held and traded by Ancerno institutions, including size, book-to-market, lagged returns, volatility, and liquidity attributes, are not significantly different from the characteristics of stocks held and traded by the average 13F institution. Ancerno institutions differ from the average 13F institution primarily in institution size."

The Ancerno dataset is one of only a handful of datasets applicable for measuring execution costs at the institutional level. While this study does not need to use all the depth of the Ancerno dataset, it does benefit from its thoroughness as outlined by Anand et al. (2010a).

3. Data

We use two datasets to study the influence of HFT on execution costs. The first is the Ancerno database on the execution costs of institutional traders described above. The second captures HFT activity in the UK equity market.

The first source gives us institutional investors' trades. Ancerno is a leading provider of data for trading cost analysis and, as discussed in the previous section, its data has been used in previous academic research.[7] Ancerno collects institutional trading costs at the trade-by-trade level. The dataset contains the date and time of trades by institutions that report to Ancerno. In the report, the trade price, number of shares, and the direction of the trade are disclosed. The dataset also includes benchmark price measures, such as the value-weighted asset price over the previous trading day and the current day, the end of day price, and the beginning of day price.

The HFT data is from the Financial Services Authority (FSA) Transaction Reporting System (the FSA dataset). European legislation, the Markets in Financial Instruments Directive (MiFID), and Chapter 17 of the FSA Handbook define the reportable securities and authorised firms have to report transactions on those securities to the FSA.[8] While reportable instruments include debt, equity, and derivative instruments listed on EEA-regulated markets, we focus on equities. Not every entity must report, only those entities subject to FSA regulation. Thus, given current regulation, not all HFTs are required to file transaction reports.[9]

The FSA dataset provides many variables of interest. It includes a date and time stamp (to the second) of when a trade occurs, the number of shares traded, the counterparty, whether it is a buy or sell trade, and the price at which the trade occurred. Importantly, it includes the user identification (at the firm level) carrying out the trade. Thus, we are able to see precisely which firm is engaging in which trades. The result of the FSA dataset is that we have an accurate measure of HFT activity from EEA-authorised firms in the UK equities asset class. Our sample of the FSA dataset is from 5 November 2007 to 5 August 2011.

HFT activity mainly concentrates in the most liquid stocks, and therefore we restrict our analysis to the 250 stocks with the largest market capitalisation from the FTSE (the FTSE 100 stocks plus the 150 stocks from the FTSE 250 with the highest market capitalisation). For methodological reasons explained in the next section, we group these stocks into seven groups. The seven categories are based on the market capitalisation of the stocks as of 1 November 2007 from Bloomberg and are as follows:

[7] See http://www.Ancerno.com/OurProducts/Research.aspx

[8] The Transaction Report User Pack gives full details of the content http://www.fsa.gov.uk/pubs/other/trup.pdf. In addition, more details are described in Appendix 1.

[9] "…not all high frequency traders are currently required to be authorised under MiFID as the exemption in Article 2.1(d) of the framework directive for persons who are only dealing on own account can be used by such traders." http://ec.europa.eu/internal_market/consultations/docs/2010/mifid/consultation_paper_en.pdf

Stock size category	Market capitalisation (1 = largest)
1	1 – 10
2	11 – 30
3	31 – 50
4	51 – 100
5	101 – 150
6	151 – 200
7	201 – 250

Finally, to collect the total daily trading volume for each of the stocks and its market capitalisation we rely on Bloomberg data.

3.1. Execution costs

We want to look at execution costs because even though raw measures of market liquidity improve over time, this does not necessarily show how institutional investors are affected. Some claim execution costs could be increasing because of HFTs. Possible reasons include HFTs' ability to more quickly react to public information or that HFTs can detect large orders being worked in the market.

Using the Ancerno dataset, we measure the execution cost with the effective spread of daily institutional traders. The interpretation of our measure is the volume-weighted average price institutional investors pay for a share compared to its true price, the price that prevailed in the market when the sell-side broker received the order. The daily institutional traders' cost of trading for each stock is TC_{jt},

$$TC_{jt} = \sum_{n=1}^{N} \omega_{jtn} \left[buy_{jtn} \left(\frac{P_{jtn} - P_{j,t-}}{P_{j,t-}} \right) - 1 \right] \qquad (1)$$

where n identifies a specific share traded, buy_{jtn} takes the value one if on day t, for stock j, share n was bought by the institutional investor, and negative one if the institutional investor sold share n; P_{jtn} is the price at which the share n for stock j was traded on day t; and $P_{j,t-}$ is the price of stock j at the time the broker received the order; ω_{jtn} is the volume weight. Following Keim and Madhavan (1995) and Anand, Irvine, Puckett, and Venkataraman (2011), we control

for market movements by subtracting the daily return on the FTSE 100 index from an order's execution cost after accounting for an order's direction.

Our measure of execution cost focuses solely on market costs (capturing the bid-ask spread, market impact and price drift while executing the order) and does not attempt to account for other trade-related costs, such as brokerage commissions. Note that the execution cost measure can be negative. Negative trading costs have previously been documented in the literature (Keim, 1999). For instance, if an institution desires to buy shares of stock *j* and *j*'s stock price decreases between the time the institution gives its broker the order and the time the broker carries out the trade, the transaction would be recorded as having a negative execution cost. An institution that uses limit orders or follows a contrarian strategy should have negative trading costs. While our *TC* measure is on average positive, it does occasionally take on negative values.

Table 1 gives summary statistics of our average daily execution cost measure for the 250 stocks and for each of the seven categories. The variance of the daily series is high relative to the mean for all the groups and the variation increases for less liquid stocks. The variation is large enough that we are unable to conclude statistical significance between the means of the different groups.

	Mean	Median	Standard deviation
FTSE 1 - 250	0.0026	0.0023	0.0074
FTSE 1 – 10	0.0019	0.0015	0.0086
FTSE 11 – 30	0.0028	0.0023	0.0093
FTSE 31 - 50	0.0022	0.0018	0.0095
FTSE 51 - 100	0.0019	0.0019	0.0090
FTSE 101 - 150	0.0017	0.0017	0.0104
FTSE 151 - 200	0.0020	0.0017	0.0100
FTSE 201 - 250	0.0015	0.0008	0.0120

Table 1 gives summary statistics of the daily average institutional traders' execution cost for the FTSE top 250 and each of the seven categories. The daily measure of execution cost is the effective spread of daily institutional traders (Equation 1).

To identify if there is a time trend we smooth the time series. The quarterly averages of the execution costs for all the stocks (Figure 1, left) together with their one year moving average trend (two quarters before and two quarters after) show that the execution costs for institutional investors in UK equities have a decreasing long-term trend between 2003 - 2011. This downward trend was temporally interrupted during the financial crisis with costs increasing. Visually, it is easier to identify the downward trend when excluding the financial crisis (Figure 1, right).[10,11] Appendix 2 shows the smoothed time series for each of the seven categories of stocks.

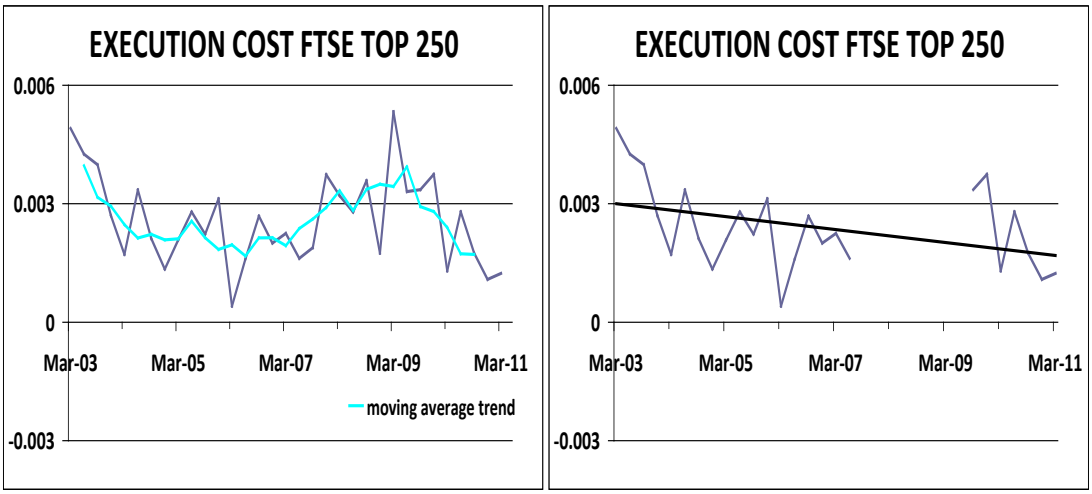


Figure 1 shows the quarterly average of the execution cost for the FTSE top 250 together with its one year moving average trend. In the right hand plot we excluded the financial crisis to better identify the execution costs' downward trend. The daily measure of execution cost is the effective spread of daily institutional traders (Equation 1).

3.2. HFT activity

As HFT still does not have a common definition, we define which firms primarily engage in HFT based on a couple of criteria. Our primary criterion is based on a definition that HFTs are a subset of Algorithmic Trading (AT) participants that use proprietary capital to generate returns using computer algorithms and low latency infrastructure. Using these criteria, it was agreed which participants of the three major trading venues for UK stocks—LSE, BATS and Chi-X—were HFT firms, based on the platforms' understanding of the business of the participant. 62 participants were classified as HFTs this way.

[10] We excluded a two-year period, from July 2007 to June 2009.

[11] The downward trend is significant at the 5% level.

An alternative list, for robustness checks, is based on a list of firms that the supervision division of the FSA has identified as firms that engage in HFT. To this list we added firms that the authors know trade frequently, rapidly, and tend to have very tight inventory controls. We also identify HFTs in the FSA dataset based on observed trading patterns, and analyse the reported trading activities of these traders. This second list consists of 25 firms. The results presented in this document are based on the first classification procedure.

Neither approach captures all HFTs. First, if a firm engages in multiple trading activities and HFT is not its primary function (such as large investment banks), we do not consider the firm to be a HFT firm. Second, we miss HFT activity coming from non-authorised firms that are not subject to FSA regulation based on MiFID and do not trade through a broker (who would be regulated).

To corroborate the extent of the coverage of HFT activity in the FSA dataset, we compare the FSA dataset level of HFT activity to the level of HFT activity in a dataset containing trade data from the London Stock Exchange, BATS, and Chi-X for all constituent stocks in the FTSE 100 for 30 trading days in 2010.[12] This dataset was obtained from the FSA.[13] While at the beginning of 2010 the FSA data cover between 70% - 80% of HFT trading volume, in August 2010 there is a drop in coverage, and by the end of 2010 our data includes only 40% of HFT volume. The fall can in part be attributed to the fact that we do not observe unregulated HFTs that are direct members of trading venues and some HFTs became direct members at this time. As mentioned before, while we recognise that a factor in the fall in coverage is because some HFTs became direct members, the FSA supervises and ensures that automated market abuse detection techniques are performed by the platforms on all of these trades, and indeed all of their Members' orders and transactions. The transaction reports for the other side of each trade are also collected and monitored. More information about the FSA dataset and the reasons for the discrepancy between the two datasets is given in Appendix 1.

Using the FSA dataset and Bloomberg data, we measure HFT activity (H_{jt}) by their degree of daily volume participation for each stock,

$$H_{jt} = \frac{HFTVol_{jt}}{Vol_{jt}} \qquad (2)$$

where $HFTVol_{jt}$ is the daily volume traded by the HFTs in stock *j* on day *t* and Vol_{jt} is twice the total daily volume traded (once for the buyer and once for the seller) in that same stock *j* on day *t*.

Note that another measure of HFT daily volume participation, the volume of transactions (or shares traded) where a HFT is on at least one side of the trade divided by the daily total volume, will by definition yield higher estimates than the one used in this paper. If, for example, there were no transactions where HFTs were in both sides of the trade, this measure will be

[12] The three trading venues are the primary places on which UK equities trade.

[13] This data has been anonymised at the platform level and aggregated into two categories of trading firms: HFT – Other.

twice our measure. Overall, as shown in Figure 2 – using data from the Tabb Group - measures of HFT activity for Europe have increased during the period.[14]

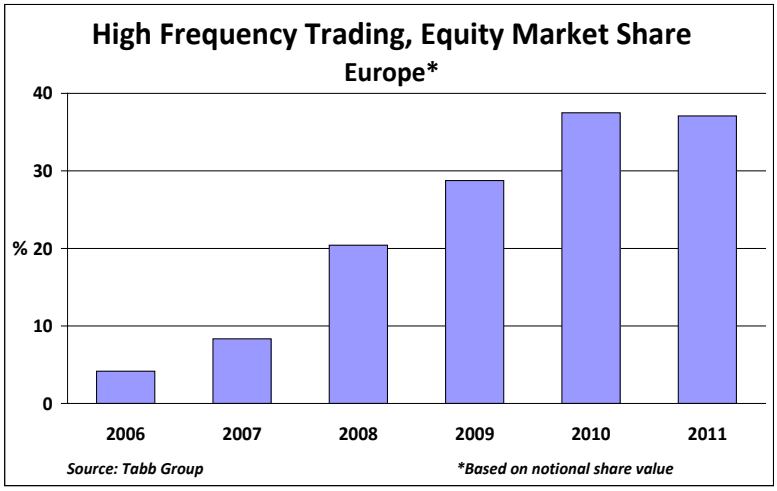


Figure 2 shows that measures of HFT for Europe have increased from 2006 to 2011.

We graph the average HFT activity for each of the seven size categories of FTSE stocks. As we can see in Figure 3, HFT activity increases steadily from the beginning of our sample until the first months of 2009 for the FTSE top 10. During 2009 there is a small reduction in HFT participation followed by a strong increase, particularly from May 2010, which may be partly due to the European debt crisis. At the beginning of August 2010, there is a sudden drop in HFT activity. This drop is mainly caused by HFT participants that changed from sponsored access through regulated brokers to direct access to the trading venues. Afterwards we see a constant level of participation with a slight increase at the end of our sample. We find a similar pattern for the most liquid stocks (FTSE 100), while HFT activity seems to be more constant through time for the less liquid stocks (see Appendix 2).

We do not believe that HFT activity decreased after August 2010 and instead believe the decline is a result of the nature of the FSA dataset. Our measure of HFT activity has sample selection and is understated over time. So the longer term trend of our HFT activity measure, especially after the middle of 2010, is not a fair representation of true HFT activity. We look at short-run changes in HFT activity around technology improvements, as explained in detail in the next section. These short-run changes, over a few weeks, should not be affected by the sample selection. In the statistical analysis we focus only on the data prior to May 2010.

[14] http://www.ft.com/cms/s/0/74ace24a-ac00-11e0-b85c-00144feabdc0.html#axzz1oX6Spjlj.

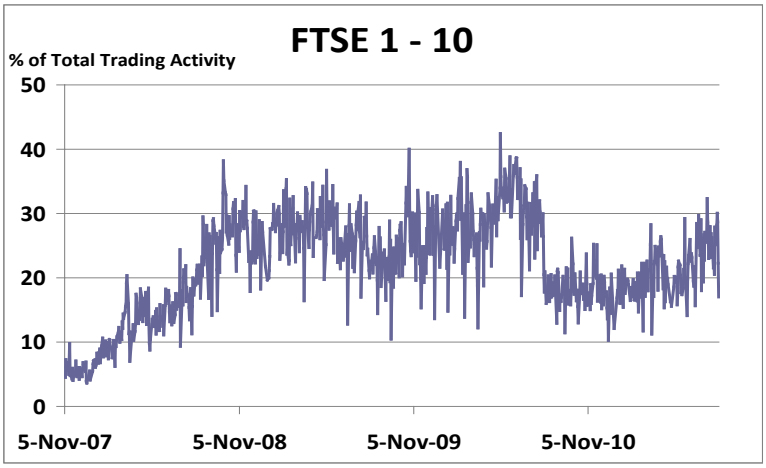


Figure 3 shows the HFT activity as percentage of the total trading volume for the FTSE top 10.

When comparing our measure of HFT activity between the different categories (see Table 2), we find that, as expected, HFTs are most active in the most liquid stocks.

	Mean (%)	Median (%)	Standard Deviation	Largest observation (%)
FTSE 1 – 250	11.25	11.40	4.05	22.90
FTSE 1 – 10	22.09	17.06	7.56	42.42
FTSE 11 – 30	14.55	11.18	4.87	26.48
FTSE 31 – 50	12.83	8.54	5.29	25.98
FTSE 51 – 100	10.63	7.32	4.27	20.11
FTSE 101 - 150	8.15	5.98	3.63	19.33
FTSE 151 - 200	5.70	4.16	2.59	17.12

	Mean (%)	Median (%)	Standard Deviation	Largest observation (%)
FTSE 201 - 250	4.45	2.60	2.65	15.58

Table 2 gives summary statistics of our measure of HFT activity, daily volume participation, for the FTSE top 250 and each of the seven categories.

4. Empirical methodology and results

In this section we examine the impact of technological upgrades in the London Stock Exchange on HFT activity and the execution cost of institutional investors separately. We present figures of HFT trading and execution costs around the technology change events and we also run panel regressions to examine the effect. Finally, to assess the link between HFT and institutional investors' execution costs we implement a two-stage least squares regression, though we caution that there are assumptions underpinning this technique which may not hold.

4.1. The technological change: exchange latency changes

The technological upgrades we base our analysis on are latency changes by the London Stock Exchange. Given that the changes in network speeds are in milliseconds, the changes will only have a direct impact on computer-based traders. Non-HFT algorithmic traders may be marginally impacted by millisecond latency changes, but arguably those who depend most on speed, HFTs, are the most likely to be affected. In addition, while HFTs may lobby the exchange to decrease its latency, HFTs do not determine exactly when such changes are implemented. As a result, network latency may provide a reasonable shock to HFT activity while having little direct impact on the trading activity of institutional investors. Wagener, Kundisch, Riordan, Rabhi, Herrmann, and Weinhardt (2010) follow a similar approach.

Since 2007 there have been a variety of technology changes at the LSE reducing latency. From the 2011 Annual Report of the London Stock Exchange we collect a list of five technology upgrades during the sample that decrease the latency of the fastest traders from 11 milliseconds to 0.113 milliseconds. The upgrades include the major changes of the TradElect system and the introduction of the Millennium system:

System	Implementation date	Latency (Milliseconds)
TradElect 2	October 31, 2007 (before the sample)	11
TradElect 3	September 1, 2008	6
TradElect 4	May 2, 2009	5

System	Implementation date	Latency (Milliseconds)
TradElect 4.1	July 20, 2009	3.7
TradElect 5	March 20, 2010	3
Millennium	February 14, 2011	.113

We study the implementations of TradElect 3 to 5. Originally, the research design called for the use of all five latency improvements during the sample; however, due to data limitations and market conditions, we are limited to the TradElect 3 to 5 upgrades. The TradElect 2 implementation occurs prior to our dataset and gives us the baseline latency at the onset of our sample. The Millennium implementation occurs during a time in which the fraction of HFT activity captured in the FSA dataset has declined, and the measure may not accurately capture changes in HFT activity. However, there are issues with TradElect 3 and 4.1 as well. TradElect 3 occurred shortly after the Lehman collapse and TradElect 4.1 occurred only a few weeks after TradElect 4.

4.2. The impact of the technology change on HFT activity and execution costs

To analyse the impact of technology changes on HFT activity and long-term investors' execution costs, we examine trading around the technology change events. We graphically compare the levels of the daily average of our HFT activity measure for the FTSE top 250 before and after the latency upgrade implementation change. We plot the time series for a 20-day window, 10 days before and 10 days after the latency change. We use a narrow window to isolate the possible effect of latency changes from other effects due to prevalent market conditions. We then conduct the same analysis for the execution cost time series. Appendix 3 shows the graphs (the results are similar for a 40-day window, not shown).

The graphs suggest that HFT activity increases after some of the latency changes. The major impact is after TradElect 5. There is also an increase after TradElect 4; however, it is easier to distinguish it when looking at the time series of the seven groups of stocks separately instead of the graph of the FTSE top 250. From the graphs, we see no evidence that the execution cost is affected by the latency changes.

The visual analysis suggests that HFT activity increases after some of the technology changes, while execution costs are unaffected. Next we perform econometric analysis.

We start by studying the effect of exchange latency changes on HFT activity by running the following panel regression using a 20-day window and also a 40-day window around the TradElect implementation:

$$H_{jt} = a_j + \beta_1^i t + \beta_2^i L_t + \nu V_{jt} + \varepsilon_{jt} \qquad (3)$$

where H_{jt} is our measure of activity by HFTs for stock j on day t, L_t is the LSE trading system latency, which starts off at 11 milliseconds and reaches three milliseconds with TradElect 5. V_{jt}

is a control variable: log volume. We use fixed effects at the stock level (α_j) and we allow a linear time trend. Instead of including different coefficients for all the latency variable dependent coefficients, we group stocks into seven groups according to their market capitalisation as described in Section 3. The use of seven different betas coefficients (i.e. the superscript *i* on the betas take on a value 1 to 7) reduces the amount of variables to be estimated, which is necessary for the individual regressions using short windows around the TradElect implementations, but still allows us to capture the cross-sectional variation in the effect of latency decreases across stocks. We expect highly liquid stocks to be more affected by technology changes. The model is estimated by OLS, and standard errors are clustered in two dimensions, at the stock and day level. The latency variable used implies that, if the change in latency is one millisecond as with TradElect 4, a coefficient of -0.01 translates to an increase in HFT activity that is one percentage point of traded volume.

Additionally, a similar panel regression is run for the stock-day execution costs again using a 20-day window and a 40-day window around the TradElect implementation:

$$TC_{jt} = \alpha_j + \beta_1^i t + \beta_2^i L_t + \nu V_{jt} + \varepsilon_{jt} \qquad (4)$$

where TC_{jt} is the institutional investors' execution cost for stock *j* on day *t*, and L_t is the LSE trading system latency. We use the same control variable as in the previous regression (log trading volume (V_{jt})), allow for fixed effects at the stock level (α_j), and a linear time trend. Seven groups (*i*) are used to reduce the parameters estimated while capturing cross-sectional differences.

Results for the 20-day window panel regressions are given in Table 3 for the HFT activity and in Table 4 for the execution cost. Table 7 and 8 in Appendix 4 show the results for the 40-day window.

	TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
	10 days before and after 1 Sep 08		**10 days before and after 2 May 09**		**10 days before and after 20 Jul 09**		**10 days before and after 20 Mar 10**	
HFT fraction	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
LSE latency FTSE 1 - 10	0.0551	1.39	-0.0599	-1.31	-0.0016	-0.03	-0.0710	-2.82
LSE latency FTSE 11 - 30	0.0212	0.67	-0.0470	-1.52	0.0094	0.28	-0.0401	4.14
LSE latency FTSE 31 - 50	-0.012	-0.41	-0.038	-2.00	0.0107	0.36	-0.026	2.44

	TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
	9		3				0	
LSE latency FTSE 51 - 100	0.0122	0.58	-0.0269	-1.46	0.0311	0.73	-0.0129	-1.71
LSE latency FTSE 101 - 151	0.0237	2.48	-0.0251	-2.13	0.0254	0.91	0.0004	0.06
LSE latency FTSE 151 - 200	0.0060	0.56	-0.0155	-1.45	0.0242	1.14	0.0039	0.9
LSE latency FTSE 201 - 250	0.0103	2.21	-0.0211	-2.62	0.0331	1.87	0.0081	2.06
Total volume	-0.0100	-0.84	-0.0336	-7.38	-0.0316	-9.83	-0.0287	-8.77
Intercept	0.4569	1.28	1.7029	7.08	1.3611	6.64	1.4994	12.92
Adj-R Squared		0.767		0.809		0.723		0.850
N		4700		4546		4540		4432

Table 3 shows the results from a panel regression with stock fixed effects of the fraction of HFT volume on, (i) LSE latency for the seven groups, (ii) total volume traded, and, (iii) a linear time trend for each category (coefficients not shown). The TradElect 4 and 5 latency reductions increase HFT activity, while the results for the other sample periods are inconclusive. The models are estimated by OLS and standard errors are double clustered at the stock and day level.

	TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
	10 days before and after 1 Sep 08		**10 days before and after 2 May 09**		**10 days before and after 20 Jul 09**		**10 days before and after 20 Mar 10**	
Execution cost	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**

	TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
LSE latency FTSE 1 - 10	-0.00504	-0.87	-0.00332	-1.12	0.00047	0.1	3.08E-05	0.02
LSE latency FTSE 11 - 30	0.00175	0.32	0.00533	2.29	-0.00804	-1.74	-0.00088	-0.57
LSE latency FTSE 31 - 50	0.00295	0.88	-0.00062	-0.24	-0.00987	-1.97	0.00294	1.87
LSE latency FTSE 51 - 100	0.00681	1.43	-0.00268	-0.98	-0.00519	-1.96	0.00158	1.21
LSE latency FTSE 101 - 151	0.00541	1.16	-0.00273	-0.72	0.00285	0.7	-0.0009	-0.31
LSE latency FTSE 151 - 200	-0.00421	-0.82	-0.00093	-0.22	0.00618	1.17	0.00141	0.39
LSE latency FTSE 201 - 250	-0.00015	-0.03	-0.00899	-1.59	-0.00264	-0.25	0.00461	0.66
Total volume	0.00112	0.89	0.00107	0.91	0.00080	0.59	0.00087	0.86
Intercept	0.02353	0.66	-0.00365	-0.12	-0.01621	-0.54	-0.03596	-1.23
Adj-R Squared		0.104		0.126		0.110		0.231
N		2189		2433		2362		1159

Table 4 shows the results from a panel regression with stock fixed effects of the execution cost of institutional investors on, (i) LSE latency f0r the seven groups, (ii) total volume traded, and, (iii) a linear time trend for each category (coefficients not shown). The results for all the sample periods are inconclusive. The models are estimated by OLS and standard errors are double clustered at the stock and day level.

TradElect 4 and 5, the two upgrades that do not suffer from data issues, have an impact on HFT activity. For TradElect 4 all coefficients are negative; three groups are significantly negative at the 5% level. For TradElect 5 the three groups covering the largest stocks are significantly smaller than zero (at the 5% level). This indicates that HFT activity increases after a technology upgrade on the London Stock Exchange. After the one millisecond improvement in minimum latency by TradElect 4, HFT activity jumps up by two to four percentage points. The 0.7 millisecond improvement of TradElect 5 increases the share of HFT activity by two to seven percentage points. TradElect 3 and 4.1 do not impact HFT activity, likely due to the issues described above. Note that the control, total trading volume, is significantly different from zero for TradElect 4, 4.1 and 5. A larger total trading volume decreases the share of HFT participation. The estimates for the time trends are not shown in Table 3, but are available upon request from the authors.

Table 4 confirms that there is no statistically significant relationship between the technology changes and execution costs for the 20-day sample windows. As mentioned previously, our lack of result is driven in part because the noisiness of our execution cost variable makes it difficult to measure changes in this variable. To ensure that our results are not an artefact of our particular variable and model specifications, we run a set of robustness checks that are available from the authors upon request. Similar results are obtained using a 40-day window. The results are robust to different specifications of latency like log latency and 1/latency. Excluding volume or the linear trend from the regression specification reduces the R^2 and significance, but has no impact on signs of our variables of interest. Additionally, we run placebo tests using four randomly chosen dates away from our current dates and none show equivalent results to what we find for the four actual changes.

4.3. The two-stage least square approach

Our aim is to assess the impact HFT has on institutional investors' execution costs. Correlation is not causation, of course, and so we cannot draw conclusions about the causal impact of HFT by simply looking at the association between HFT activity and execution costs. First, some third factor could drive both HFT activity and execution costs. For example, during the period in question, fundamentals, or aspects of the financial crisis, could cause greater uncertainty that both causes HFT market makers to trade more and causes spreads and execution costs to increase. Second, a correlation between HFT activity and execution costs could be because execution costs affect HFT participation and not the other way around.

One well-known approach to get around the endogeneity problem is to find a variable that satisfies two conditions. First, the variable is correlated with HFT activity. Second, the variable is not correlated with execution costs except through its relationship to HFT activity. Such a variable is known as an 'instrument.' The second condition, which is important, is known as the 'exclusion restriction.'

The economic method for conducting an instrumental variables regression is known as 'two-stage least squares' (2SLS) and consists of two regressions. In the first stage, HFT activity is regressed on the instrument and some control variables to find the relationship between the instrument and HFT activity. In the second stage, we use the outcomes of the first stage to

isolate a component of HFT activity that is independent of execution costs. We regress execution costs on this predicted component of HFT activity and control variables. The coefficient on HFT activity in the second stage is the causal effect of HFT activity on execution costs. If the exclusion restriction holds, this coefficient is an unbiased estimator for this causal effect.

To conduct our 2SLS regression, in the first stage we regress the level of HFT activity at the stock-day level on a variable capturing the speed change in the London Stock Exchange's systems to handle electronic messages and relevant control variables, as defined in equation 3. The model is first estimated from November 2007, the beginning of our sample period, until the end of April 2010, when sudden changes in our HFT activity measure begin. For this regression, we use a cubic trend and an additional control variable: a dummy for the short sale ban (b_t). Additionally, the model is estimated for 20-day windows around the four TradElect upgrades separately.

In the second stage, we use the estimated proxy for HFT activity from the first stage in a regression with the dependent variable being the stock-day execution costs of a set of firms included in the Ancerno dataset, i.e.

$$TC_{jt} = \alpha_j + \beta_1^i t + \theta \hat{H}_{jt} + \nu V_{jt} + \delta b_t + \varepsilon_{jt} \qquad (5)$$

where TC_{jt} is the institutional investors' execution cost for stock j on day t, and $\hat{H}_{jt}$ is the predicted measure of HFT activity from the first stage regression. We use the same control variables as in the first stage regression (log trading volume (V_{jt}), and a dummy for the short sale ban (b_t) put into place shortly after the Lehman collapse for the regression on the whole sample period), allow for fixed effects at the stock level (α_j), and a linear time trend (a cubic trend on the whole sample regression). Again seven groups (i) are distinguished to reduce the number of variables to be estimated, but still are able to capture differences in the cross-section.

Table 5 summarises the results from our first stage regressions. The point estimates for reductions in LSE latency are negative for all seven groups for the sample covering TradElect 3 to 5. However, statistically the coefficients for most groups are not different from zero even at a ten percent significance level. As indicated above, this is most likely due to the unusual market conditions during the period from mid-2008 to mid-2009 combined with the relatively minor changes introduced with TradElect 4.1. To get around these problems we focus on 20-day windows around the technology upgrades and estimate them independently. Table 5 also shows the results for 20-day windows around the TradElect implementations. The instrument appears to be valid for TradElect 4 and TradElect 5 as expected from the results of the previous section.

First stage										
	All 4 upgrades		TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
	5 Nov 07 to 30 Apr 10		10 days before and after 1 Sep 08		10 days before and after 2 May 09		10 days before and after 20 Jul 09		10 days before and after 20 Mar 10	
HFT fraction	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
LSE latency FTSE 1 - 10	-0.0036	-0.60	0.0314	1.36	-0.0268	-1.07	-0.0137	-0.32	-0.0554	-5.00
LSE latency FTSE 11 - 30	-0.0038	-0.80	0.0114	0.40	-0.0537	-1.78	0.0101	0.30	-0.0368	-3.60
LSE latency FTSE 31 - 50	-0.0154	-2.88	-0.0185	-0.63	-0.0422	-2.03	0.0076	0.25	-0.0362	-2.77
LSE latency FTSE 51 - 100	-0.0118	-2.74	0.0078	0.41	-0.0338	-1.47	0.0374	0.84	-0.0146	-1.64
LSE latency FTSE 101 - 151	-0.0076	-1.73	0.0259	2.93	-0.0306	-2.44	0.0296	1.13	-0.0115	-1.45

First stage										
LSE latency FTSE 151 - 200	0.0017	0.41	0.0024	0.39	-0.0142	-1.36	0.0481	1.32	0.0054	0.61
LSE latency FTSE 201 - 250	-0.0013	-0.32	0.0109	1.94	-0.0266	-2.15	0.0840	2.57	0.0164	1.41
Total volume	-0.0165	-8.94	-0.0080	-0.46	-0.0399	-6.20	-0.0415	-9.15	-0.0419	-8.66
Ban dummy	0.0002	0.05								
Intercept	0.5003	13.29	0.0653	0.30	1.1519	7.25	1.0406	5.10	0.6029	7.03
Adj-R Squared		0.586		0.771		0.721		0.626		0.838
N		67115		2184		2433		2362		1159

Table 5 shows the results from a panel regression with stock fixed effects of the fraction of HFT volume on (i) LSE latency, (ii) total volume traded, (iii) a short sale ban dummy, and, (iv) linear time trends for each category and a cubic trend for the whole sample regression (coefficients not shown). The TradElect 4 and 5 latency reductions increase HFT activity, while the results for the other sample periods are inconclusive. The models are estimated by OLS, and standard errors are double clustered at the stock and day level.

The results for the second stage are presented in Table 6. There is no relation between the instrumented HFT activity variable and execution costs for the 20-day sample windows around TradElect 4 and 5. The coefficient on $\hat{H}_{jt}$ activity is not statistically significant.[15]

Here we also run a set of robustness checks as in the previous section regressions. We also estimate the model for 40-day windows obtaining similar results (available upon request). Pooling the four windows around the technology changes in a single regression does not materially change our findings. And forming seven portfolios based on size with the same definition as the groups above and performing the analysis on seven portfolios instead of 250 stocks does not change our findings. Details about the pooled and portfolio regressions and the results can be found in Appendix 5.

In summary, we find no clear evidence that HFT has been beneficial or detrimental for institutional investors. However, these results should be interpreted with some caution. They are based on two events. To improve the robustness of these results it would be useful to include data on HFT activity from other sources so that the analysis covers non-regulated HFT firms as well. We would also like to include alternative sources of variation in HFT activity that might be argued to be exogenous, at least in the short run, such as further changes in trading venue technology, changes in HFT firms' hardware, HFT firms' software outages or HFT firm co-location dates. Our analysis would also benefit from additional measures of institutional investor trading costs.

Second stage				
	TradElect 4		**TradElect 5**	
T-Cost	**Coef.**	**z-stat**	**Coef.**	**z-stat**
Predicted HFT	0.0151	0.44	-0.0054	-0.23
Total volume	0.0019	1.37	0.0008	0.63
Intercept	-0.0281	-1.12	-0.0109	-0.49
Adj-R Squared		0.0018		0.0058
N		2433		1159

[15] Since the instrument is contaminated for TradElect 3 and 4.1, the second stage results are omitted from Table 6.

Table 6 shows the result from the second stage a 2SLS regression of execution costs on instrumented HFT activity, total trading volume, and linear time trends (omitted from the table). There is no relation between execution costs and HFT activity.

5. Looking forward

What does the future hold for high-frequency trading, financial markets, and institutional investors? Of course there is a great deal of uncertainty as to how markets and market participants will evolve over the next ten years. Here, we try to provide some insights into possible developments in the future.

Execution costs are comprised of bid-ask spreads, commissions, and clearing and settlement costs. These reflect other indirect costs such as human capital and infrastructure. Evidence in the Foresight driver reviews indicates that the advent of HFT has coincided with a decline in bid ask spreads and commissions (see for example Friederich and Payne, 2011) and a decline in clearing fees (Menkveld, 2011).

It is possible execution costs in ten years could continue to decrease for a variety of reasons. First, new entries may trigger price wars that end up with further reductions in clearing fees, as was the case when Chi-X began operations in the Dutch market (Friederich and Payne, 2011). Second, advances in technology such as the ones described in Cliff, Brown, and Treleaven (2011) might imply lower transactions costs. Cliff (2011) states:

> The primary impact of cloud computing on activities in the financial markets in the next ten years will not be in the provision of computing facilities that automate execution, but rather in the ability of the cloud to provide cheap, elastically scalable, high-performance computing (HPC). … The convergence of cheap computer-power, statistically sophisticated and computationally intensive trading strategies, fast automated execution via STP, and DMA, means that in the last two or three years it has become commonplace for market participants to seek counterparties to a transaction electronically, identify a counterparty, and then execute the transaction, all within a small number of seconds.

These changes in technology might lower barriers to entry. Additional market players could appear in the sector, which may offer lower commissions than those that currently exist. Moreover, new market players established in countries with lower operating costs could offer the same level of services at a lower price. In addition, new regulations fostering competition could increase competition for volume on a global scale, with a potential reduction of commissions charged by exchanges as more and more participants enter the market and compete for volume.

On the other hand, there are factors that may increase execution costs in the future.

First, "…[n]ew forms of manipulation, such as algorithms programmed to take advantage of other algorithms, can raise trading costs and move prices away from efficient levels" (Linton and O'Hara, 2011). Also, as observed in Menkveld (2011), "High-frequency traders overinvest in technology relative to a social optimum if the main motivation is to be ahead of rival HFTs when trading on a publicly observed signal." Some advances in technology (e.g. silicon encrypted algorithms and adaptive algorithms) might lead to a costly race to acquire the latest technology.

Government intervention may also increase execution costs for institutional investors. A new tax is one example, for instance. There could be an implementation of a tax on trading that would increase the costs of commissions. In addition, regulation could be passed that fosters single-exchange markets as opposed to more competitive, but fragmented, markets. Under such a scenario, commissions could be higher in the future, at least for transactions in assets with a lower possibility of price comparison by market participants and end-customers.

The continual development of technology in financial markets could increase technology requirements, and hence costs, of trading. In addition, new technological developments might involve big upfront fixed costs, generating higher barriers to entry. Moreover, companies within the sector with enough financial means could prevent their competitors from accessing cutting-edge technology, either by entering into exclusivity contracts with technology providers or by directly acquiring players within the technology space, to the extent economically possible. This might increase execution costs. However, execution costs could still trend downwards, particularly in a market environment where competition is still fostered.

Technological advances and a potential increase in the level of automation of trading in financial markets could result in positive effects of these advances being captured by a particular subset of asset classes, or even different components of the same asset class, e.g. large-caps vs. small-caps within the equities asset class. If this is the case, liquidity provision could shift away from certain assets to others, leading to a potential increase in bid-ask spreads for some assets but a decrease for others.

An increased level of competition and entry could increase the risk of financial instability in the system. The potential higher risk for market makers and systemic risk in general could be translated into higher returns expected by market makers for their services, and hence higher bid ask spreads.

Transaction costs in general might also increase if liquidity spikes become more prevalent. According to Menkveld (2011),

> A recent study shows that investors increasingly care about tail risk in liquidity supply as evidenced by higher required returns. Menkveld and Wang (2012) measure such 'tail risk' by identifying liquidity leaks, or short, liquileaks, through estimation of a regime-switching model. A liquileak is defined as the event that an investor finds the stock in a very poor liquidity state for more than a week. In the cross-section, a one standard deviation increase in liquileak probability commands an additional annual premium of 1.33% based on a 1963-2008 sample of 2147 U.S. equities. More importantly, the liquileak premium increased over time. … Transaction cost declined on average yet appears to feature more frequent and more extreme spikes when migrating to electronic trading. The 2010 May 6 flash crash is a dramatic example. A recent study shows that tail risk in liquidity supply seems to increasingly matter to investors. They command a premium for it which raises the cost of capital for the issuing firm.

These liquidity famines might lead to bigger bid-ask spreads.

6. Conclusion

One of the more important questions regarding any new development in financial market microstructure is its impact on transaction costs. If transaction costs are low, market

participants are better able to hold the assets most suited to them, and informed participants are more able to trade on their private information and impound it into asset prices.

High-frequency trading has quickly become a term known to the general public. The idea of computers running financial markets has raised concerns among other market participants, the media, regulators, academics, and the general public. One of these concerns is that HFT has increased execution costs, a component of transaction costs, at least for some market participants.

There is a small, but growing literature trying to understand the role HFTs are having on financial markets. This paper adds to that literature by addressing an open question, whether HFT activity increases or decreases institutional investor execution costs.

We show that in the UK, like in the US, there has broadly been a decrease in institutional execution costs over the last decade. This trend, however, was interrupted by the financial crisis, which caused execution costs to increase between mid-2007 and mid-2009.

The rise of HFT could be associated with an increase, decrease, or no change in institutional investor trading costs. We find an association between the major latency changes made by the London Stock Exchange and HFT activity but no measurable association between these latency changes and our measure of execution costs. Assuming that these technology changes only affect institutional trading costs through their effect on HFT, we use the technology shocks as an instrumental variable and find again that as HFT increases, execution costs are unchanged. The two stage least squares approach we implement reduces the problem of endogeneity, if the assumption on the exclusion restriction is valid, that makes answering the question of causation difficult. As we do see an increase in HFT activity but no concomitant change in execution costs, we fail to observe a relationship between HFT and institutional execution costs.

In part this lack of finding is driven by the noisiness of our (standard) measure of execution costs and the fact that we examine two events where HFT increases for a short time. Our work is ongoing and, with more time, we will develop the study. We will attempt to extend the analysis to additional measures of trading cost and to cover more events. And we will examine other outcome measures such as market spreads and HFT profits.

References

Anand, Amber, Paul Irvine, Andy Puckett and Kumar Venkataraman, 2010a, Market crashes and institutional trading, Working Paper.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2010b, Performance of institutional trading desks: An analysis of persistence in trading costs, Working Paper.

Benos, Evangelos and Satchit Sagade, 2012, A snapshot of high frequency trading activity in the UK equity market and its impact on market quality, Working Paper.

Bacidore, Jeffrey and George Sofianos, 2002, Liquidity provision and specialist trading in NYSE-listed non-US stocks, *Journal of Financial Economics* 63, 133-158.

Bessembinder, Hendrik, 2003, Issues in assessing trade execution costs, *Journal of Financial Markets* 6, 233-257.

Brogaard, Jonathan, 2011, High Frequency Trading, information, and profits, Foresight Driver Review 3.

Cliff, Dave, 2011, The impact of technology developments, Foresight Working Paper 3.

Cliff, Dave, Dan Brown, and Philip Treleaven, 2011, The future of computer trading in financial markets, Foresight Driver Review 3.

Friederich, Sylvian and Richard Payne, 2011, Computer based trading, liquidity, and trading costs, Foresight Driver Review 5.

Financial Services Authority, Transaction Reporting User Pack (TRUP). http://www.fsa.gov.uk/pubs/other/trup.pdf .

Gondat-Larralde, Celine and Kevin James, 2008, IPO pricing and share allocation: The importance of being ignorant, *Journal of Finance* 63, 449-478.

Grullon, Gustavo, George Kanatas, and James Weston, 2004, Advertising, breadth of ownership, and liquidity, *Review of Financial Studies* 17, 439- 461.

Hendershott, Terrence, 2011, High frequency trading and price efficiency, Foresight Driver Review 12.

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

Keim, Donald, and Ananth Madhavan, 1995, Anatomy of the trading process: Empirical evidence on the behavior of institutional traders, *Journal of Financial Economics* 37, 371-398.

Keim, Donald, 1999, An analysis of mutual fund design: The case of investing in small-cap stocks, *Journal of Financial Economics* 51, 173-194.

Linton, Oliver, 2011, What has happened to UK Equity Market Quality in the last decade? An analysis of the daily data, Foresight Driver Review 1.

Linton, Oliver and Maureen O'Hara, 2011, The impact of computer trading on liquidity, price efficiency/discovery and transaction costs, Foresight Working Paper 2.

Menkveld, Albert, 2011, Electronic trading and market structure, Foresight Driver Review 16.

Menkveld, Albert and Ting Wang, 2012, Liquileaks, Working paper.

Wagener, Martin, Dennis Kundisch, Ryan Riordan, Fethi Rabhi, Philipp Herrmann, and Christof Weinhardt, 2010, Price efficiency in futures and spot trading: The role of information technology, *Electronic Commerce Research and Applications* 9, 400-409.

Appendix 1:The FSA data

The data and the cleaning procedure

We use a propreitary dataset to identify and measure HFT activity. The dataset, 'Sabre II', is held at the Financial Services Authority (FSA) and consists of all the transaction reports from 5 November 2007 to 5 August 2011. By regulation, firms must report (European legislation (MiFID) and Chapter 17 of the FSA Handbook) certain details of their executed transactions by the end of the following business day.

Each report gives the date and trading time of the transaction, the name of the instrument and its type, price, currency, quantity and whether it was a buy or a sell. It also indicates who conducted the transaction (the reporting firm), with whom (counterparty 1), and in the case of an agency trade, on behalf of whom (counterparty 2). It also discloses the name of the trading platform on which the transaction was made or whether it was off-exchange. Sabre II contains transactions on different types of financial instruments (e.g. equity, bonds, futures, options). We focus on cash equity. A detailed description of the content of the transaction reports can be found in the FSA's Transaction Reporting User Pack.

To calculate our measure of HFT activity we consider all the transaction reports where a HFT firm reports a principal transaction or is reported as counterparty 1 or counterparty 2. Given current regulation, not all HFT firms are required to report. By considering counterparties reported with known codes (Business Identifier Codes (BIC)), we can identify part of this unregulated HFT activity. If counterparties are reported with a reporting firm's own internal codes we are unable to distinguish them.[16]

From the sample period, we remove bank holidays, Christmas Eve, and New Year's Eve. We also restrict the sample to the trading venues' opening hours (8am – 4:30pm) and exclude transactions done off-market.

In the dataset one transaction may be reported twice, once from the buyer side and once from the seller side, so we clean the data to avoid double counting. If a HFT firm reports a trade and the same HFT firm is also reported as counterparty in the opposite trade, we only keep one of the reports. We also disregard any duplicated reports.

Finally we focus on the reported volumes and remove extreme values. We use three different methods to define extreme values: 1) trimming trades in the top 5% of size for a given stock, 2) trimming all trades with a size larger than four times the standard deviation above the mean, and, 3) trimming all trades with a size larger than one and one-half times the inter-quartile range above the third quartile. The results presented in the paper are based on the first method.

Matching the FSA data with external datasets

To corroborate the extent of the coverage of HFT activity in the FSA dataset, we compare the FSA dataset level of HFT activity to the level of HFT activity in a dataset containing trade data from the London Stock Exchange, BATS, and Chi-X for all constituent stocks in the FTSE 100

[16] Further work could be done to identify the internal codes referring to HFT firms to improve our measure of HFT activity.

for 30 trading days in 2010. This dataset was obtained from the FSA.[17] In the days we have available at the beginning of 2010 between 70% - 80% of HFT trading volume is covered in the FSA dataset; by the end of 2010 it captures only 40% of it. The coverage falls as several unregulated HFT firms moved from trading through regulated brokers to becoming direct members of trading venues.

We believe the reason for the remaining discrepancy may be due to the reporting differences between the two datasets, misreporting or/and the increase in activity by non-UK-regulated HFTs.

The London Stock Exchange, BATS, and Chi-X dataset gives all trades executed from 11:30 to 15:30. In the FSA dataset firms must strive to capture the trading time correctly but there are several reasons why the reported time may not be the traded time.[18] The volume captured in the FSA dataset compared to the platforms' data may be underestimated because of this fact.

Misreporting can also be a cause of discrepancy between both datasets. The comparison will be affected if the firm code, the counterparty code, the instrument, the venue, the transaction time or the quantity is misreported.

Finally, not all HFTs must be authorised under MiFID, and those that are not do not need to report. We can capture part of their activity if they are reported as counterparties using their BIC. However, we will miss all the trades where these firms are reported with the reporting firm's internal codes or when they trade directly with the trading venues.

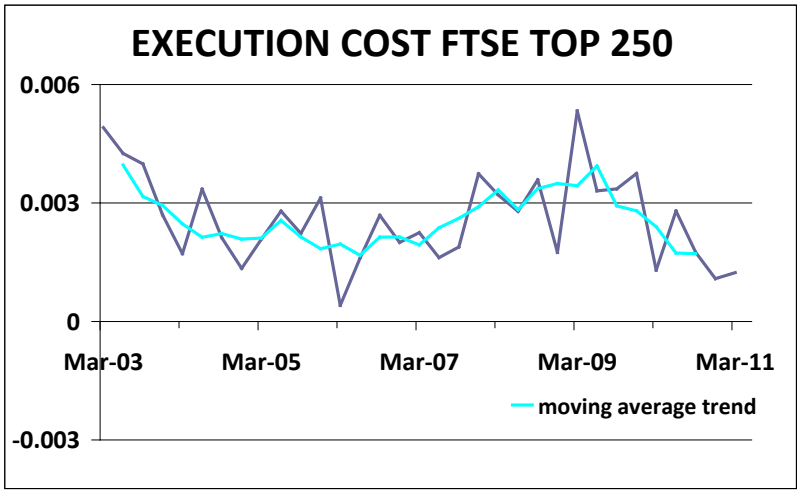
[17] This data has been anonymised at the platform level and aggregated into two categories of trading firms: HFT – Other.

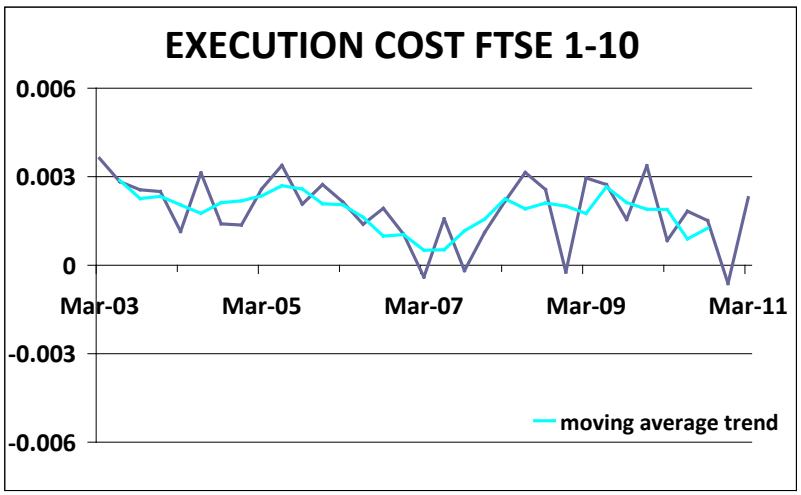
[18] For example, if the trading time is unknown, the default time is 00:01:00. If an external broker fills one order in several transactions, the time reflects the time when the firm becomes the beneficial owner. If the reporting firm executes a client's order as a riskless principal and holds the order on its book until the order is complete, the transaction is booked at the time agreed with the client. More details can be found in the FSA's Transaction Reporting User Pack.

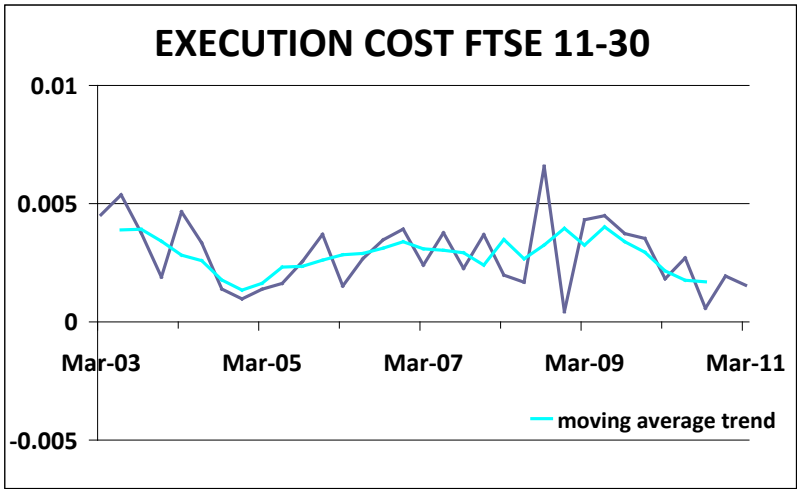
Appendix 2: Execution cost and HFT activity graphs

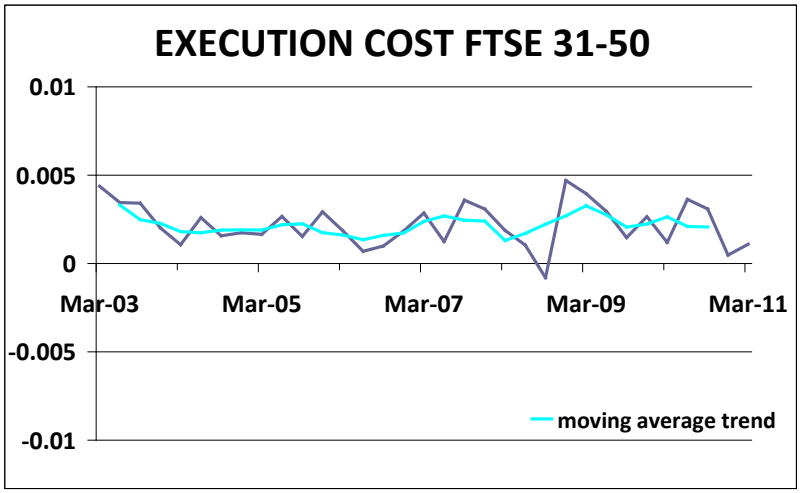
Here we plot the quarterly averages of the execution costs for each category in which we divided the 250 FTSE stocks together with their one year moving average trend (two quarters before and two quarters after). The quarterly averages are based on the daily measure of execution cost - the effective spread of daily institutional traders (Equation 1). We also graph the average HFT activity for each of these seven categories.

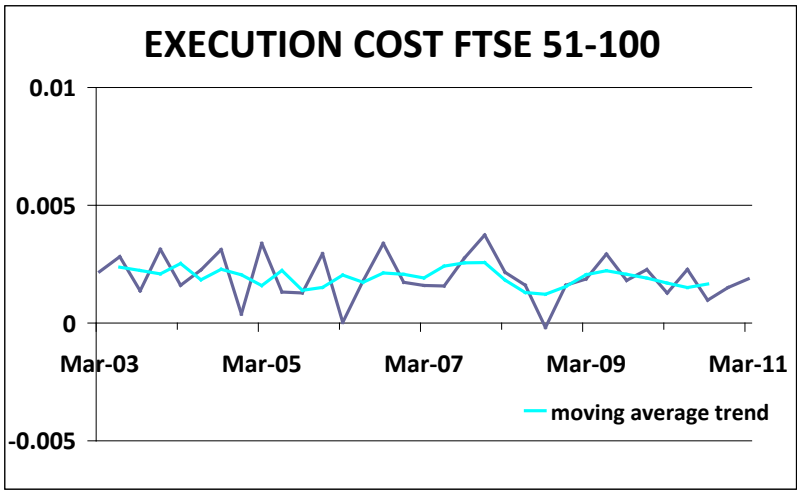
Execution costs

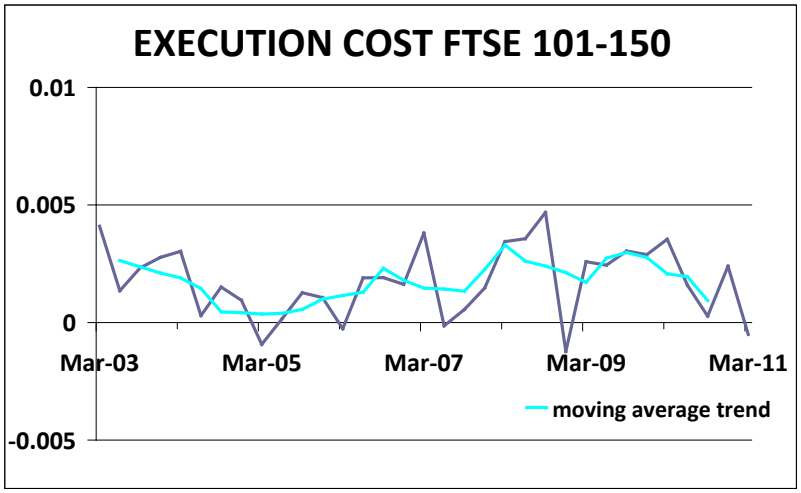


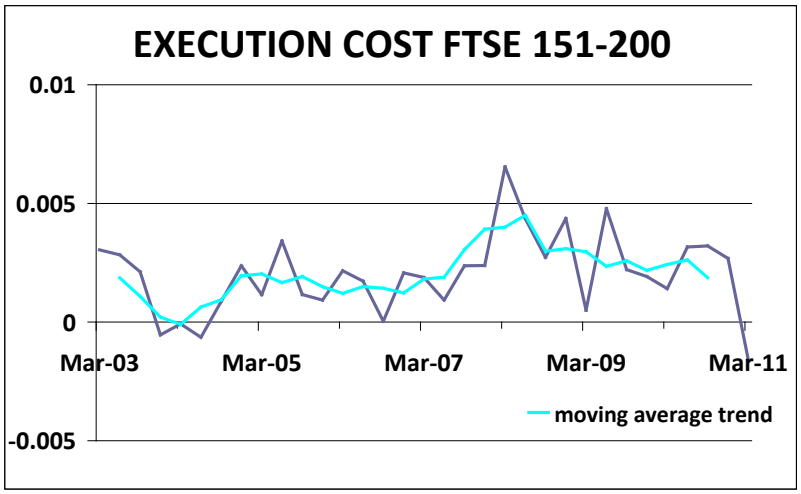


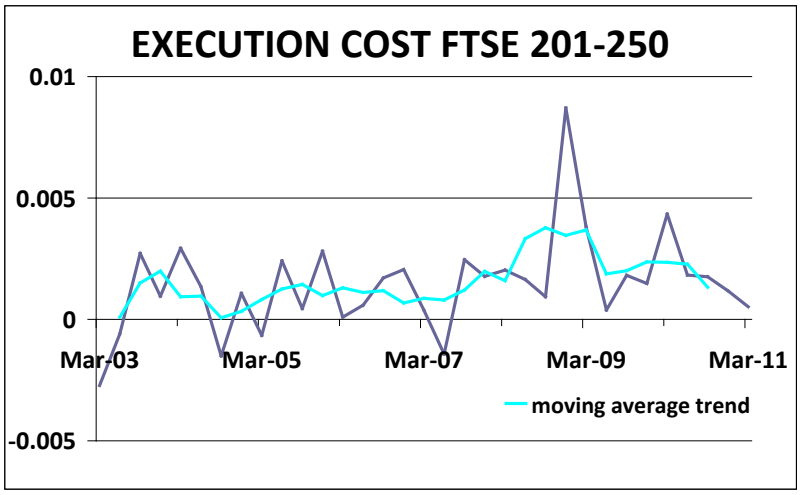

EXECUTION COST FTSE 11-30
0.01
0.005
0
-0.005
Mar-03
Mar-05
Mar-07
Mar-09
Mar-11
moving average trend


EXECUTION COST FTSE 31-50
0.01
0.005
0
-0.005
-0.01
Mar-03
Mar-05
Mar-07
Mar-09
Mar-11
moving average trend

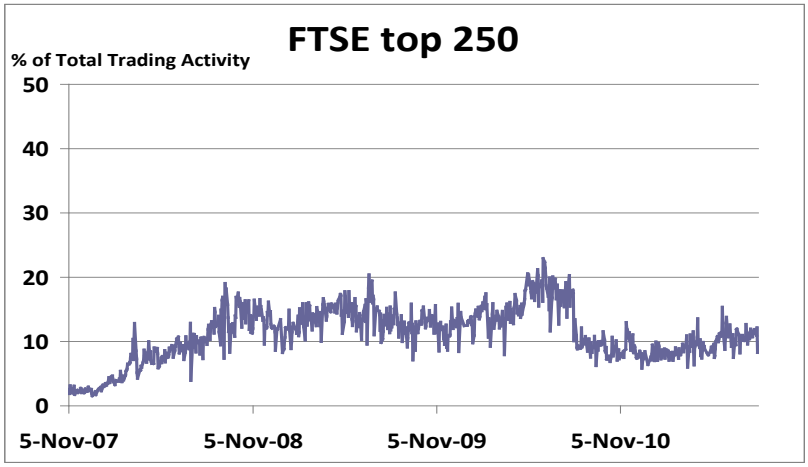

EXECUTION COST FTSE 51-100
0.01
0.005
0
-0.005
Mar-03
Mar-05
Mar-07
Mar-09
Mar-11
moving average trend

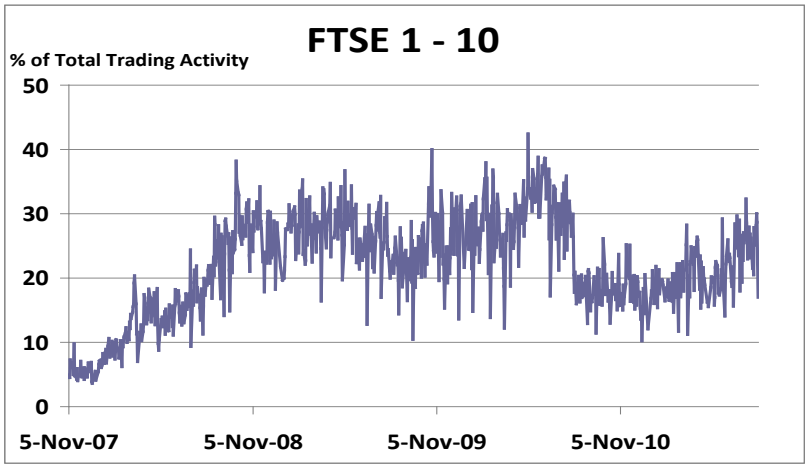

EXECUTION COST FTSE 101-150
0.01
0.005
0
-0.005
Mar-03
Mar-05
Mar-07
Mar-09
Mar-11
moving average trend

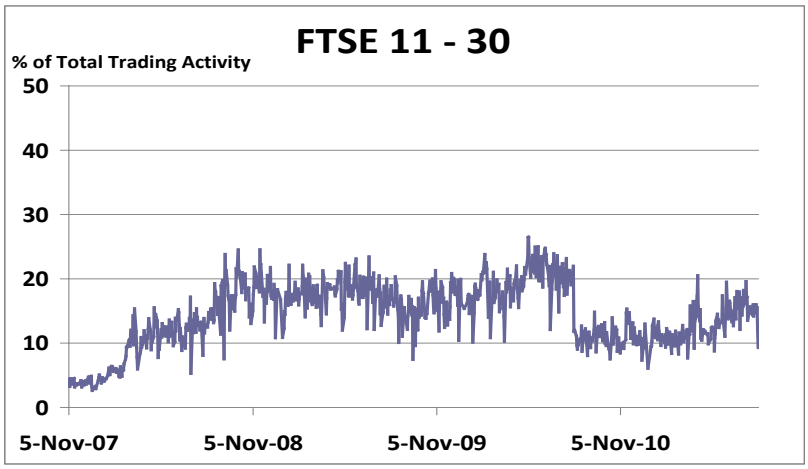

EXECUTION COST FTSE 151-200
0.01
0.005
0
-0.005
Mar-03
Mar-05
Mar-07
Mar-09
Mar-11
moving average trend

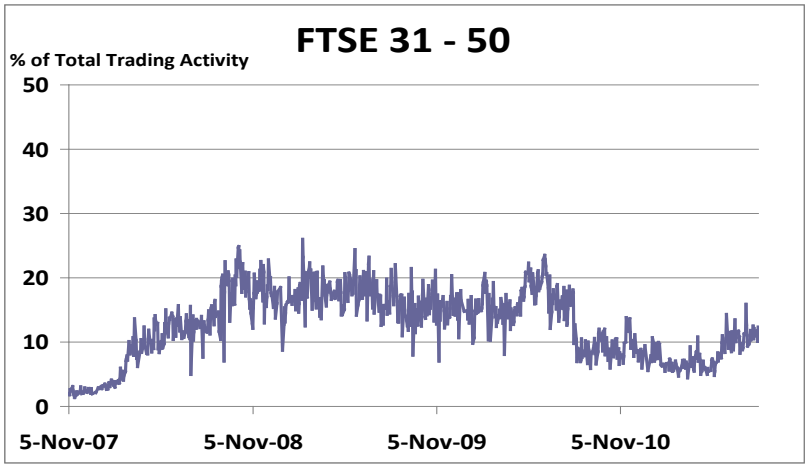

EXECUTION COST FTSE 201-250
0.01
0.005
0
-0.005
Mar-03
Mar-05
Mar-07
Mar-09
Mar-11
moving average trend

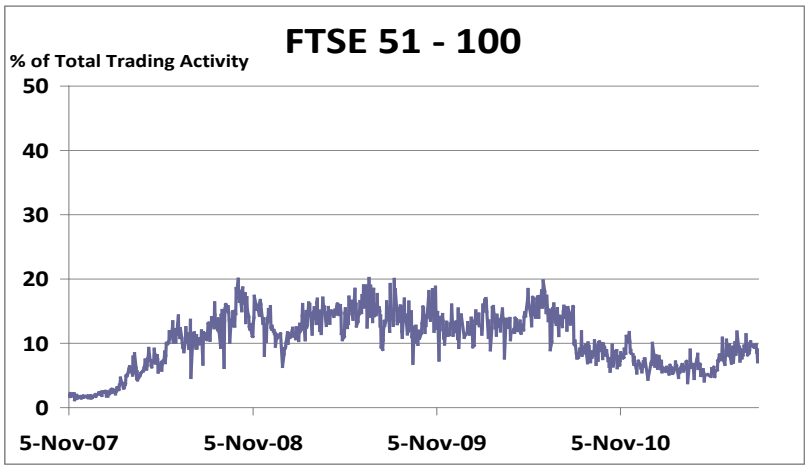
HFT activity

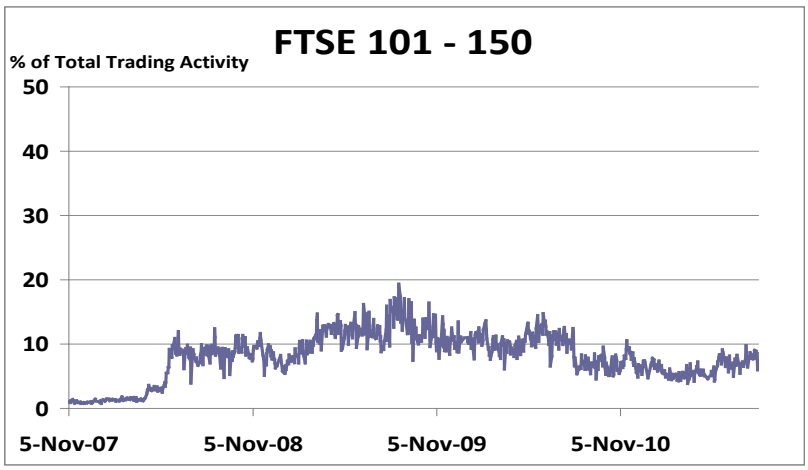


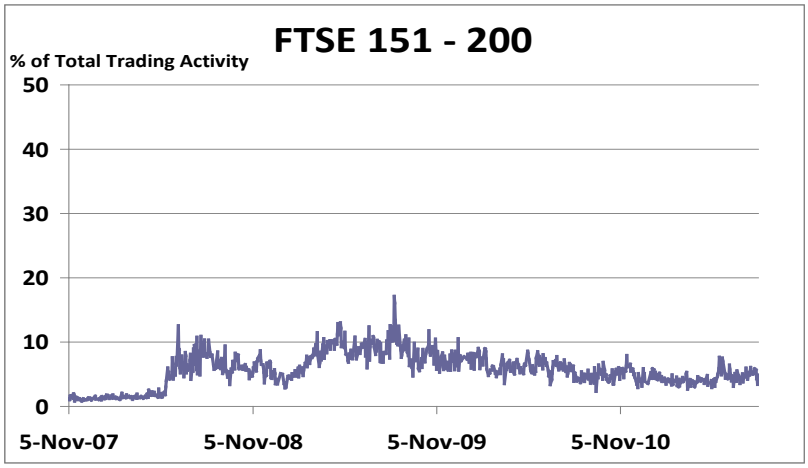


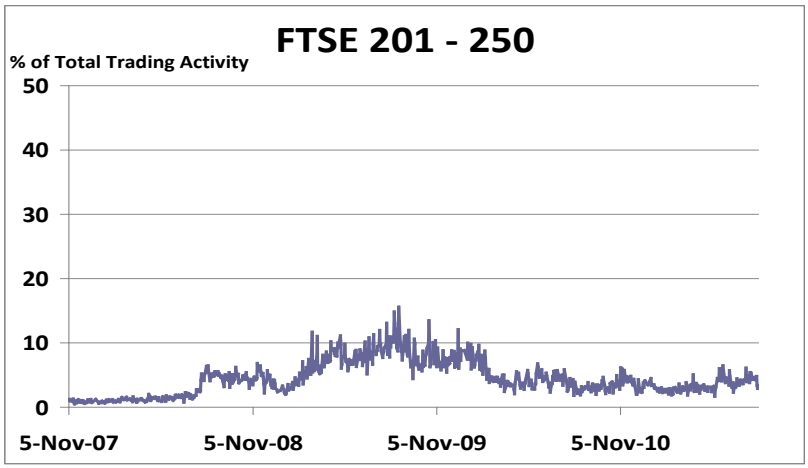



FTSE 31 - 50
% of Total Trading Activity
50
40
30
20
10
0
5-Nov-07
5-Nov-08
5-Nov-09
5-Nov-10


FTSE 51 - 100
% of Total Trading Activity
50
40
30
20
10
0
5-Nov-07
5-Nov-08
5-Nov-09
5-Nov-10

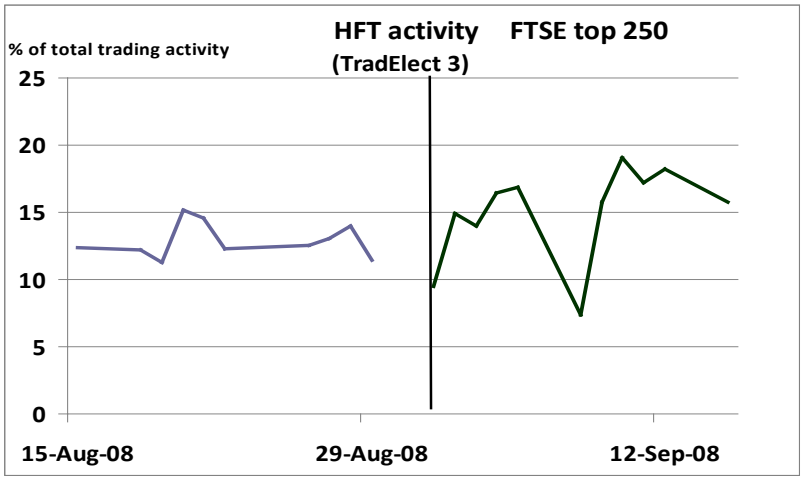

FTSE 101 - 150
% of Total Trading Activity
50
40
30
20
10
0
5-Nov-07
5-Nov-08
5-Nov-09
5-Nov-10

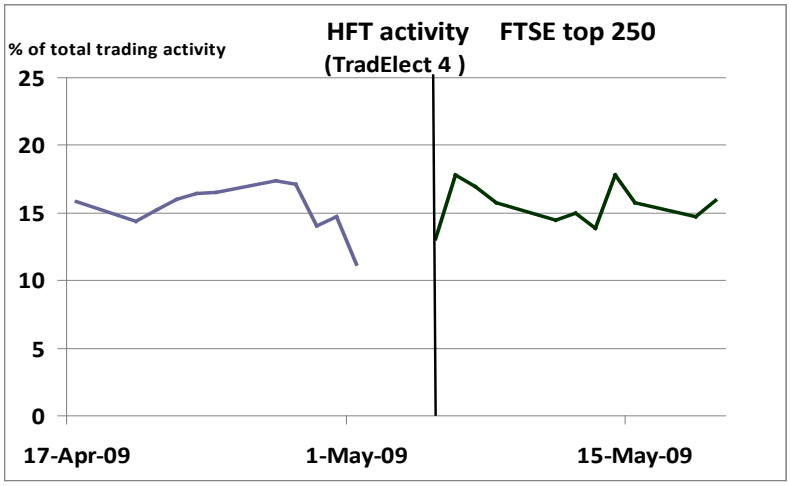

FTSE 151 - 200
% of Total Trading Activity
50
40
30
20
10
0
5-Nov-07
5-Nov-08
5-Nov-09
5-Nov-10

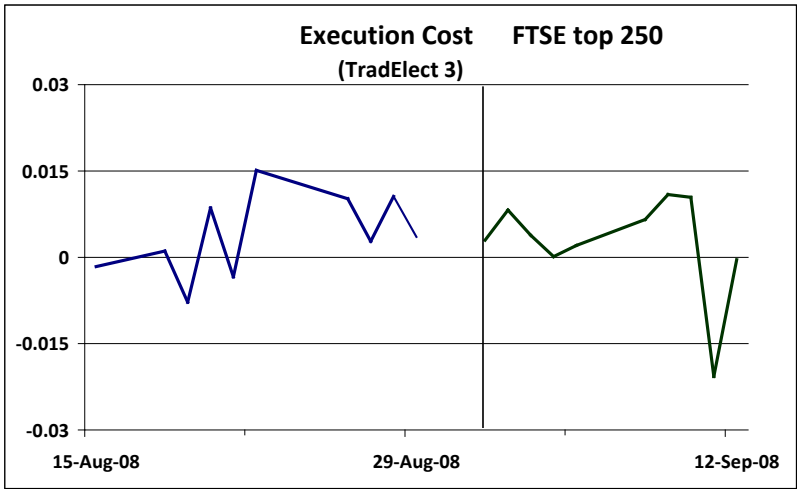

FTSE 201 - 250
% of Total Trading Activity
50
40
30
20
10
0
5-Nov-07
5-Nov-08
5-Nov-09
5-Nov-10

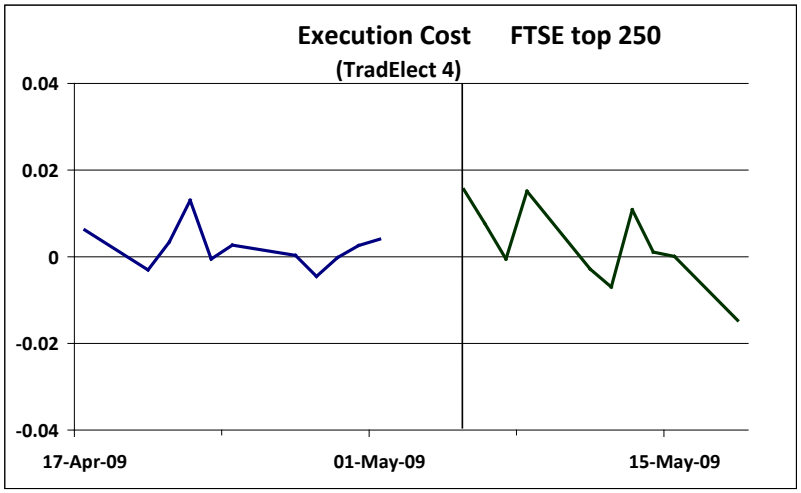
Appendix 3: Graphs around the technology change events

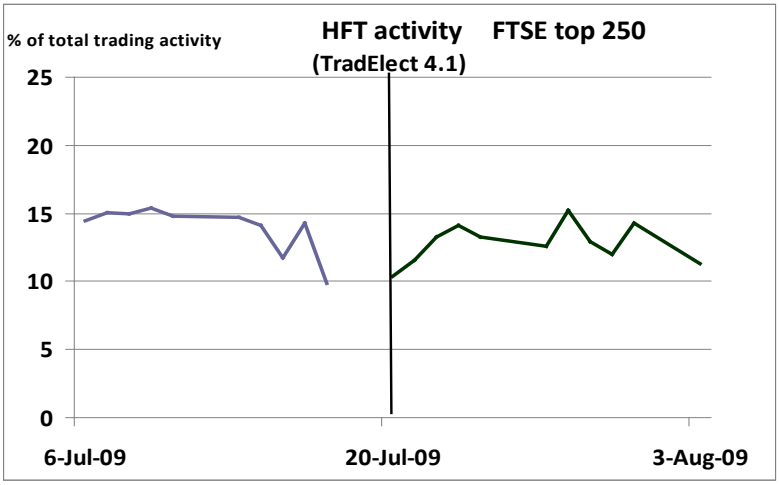
Here we compare the levels of the volume-weighted HFT activity measure for the FTSE top 250 before and after each of the four latency changes. We plot the time series 10 days before and 10 days after the latency change. We look at a narrow window to try to isolate the possible effect of latency changes from other effects due to prevalent market conditions. The same analysis is done for the execution cost time series.

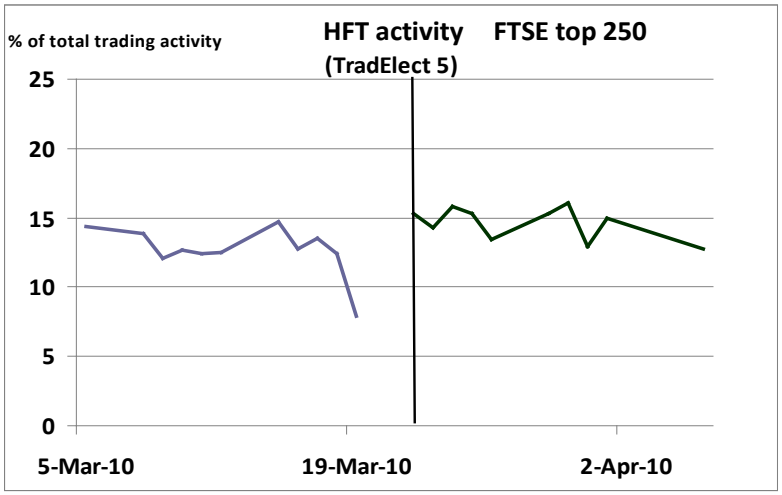


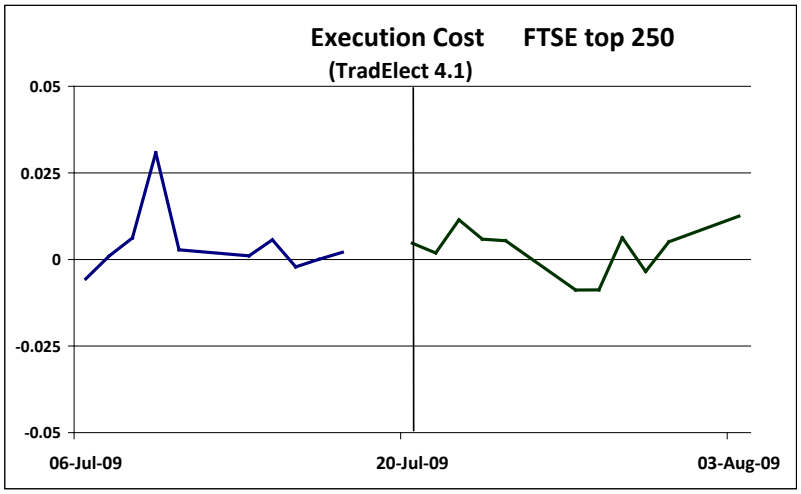


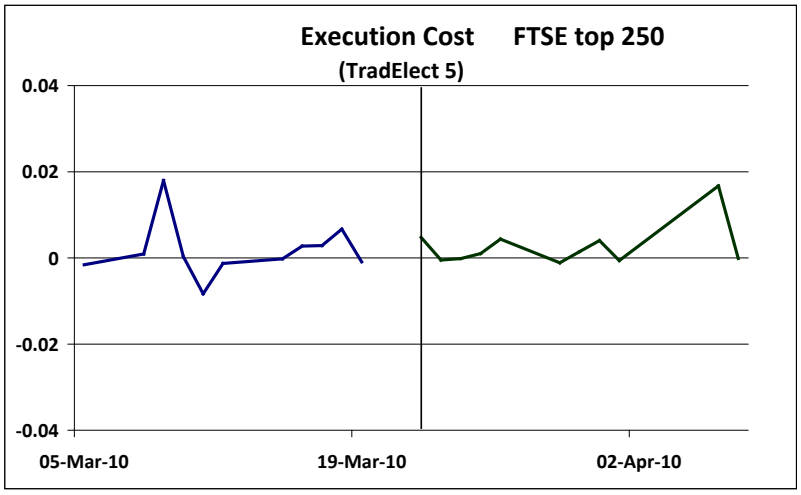

Execution Cost FTSE top 250
(TradElect 3)
0.03
0.015
0
-0.015
-0.03
15-Aug-08
29-Aug-08
12-Sep-08


Execution Cost FTSE top 250
(TradElect 4)
0.04
0.02
0
-0.02
-0.04
17-Apr-09
01-May-09
15-May-09

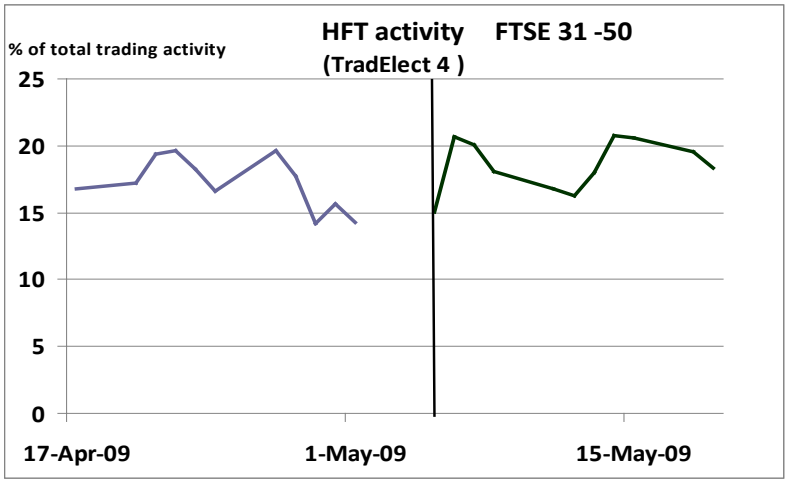

% of total trading activity
HFT activity FTSE top 250
(TradElect 4.1)
25
20
15
10
5
0
6-Jul-09
20-Jul-09
3-Aug-09







The major impact on HFT activity seems to be after TradElect 5. There is also an increase after TradElect 4, however, it is easier to distinguish it when looking at the time series of the seven groups of stocks instead of just the FTSE top 250. Here we present the graph for group three, FTSE 31-50, where the impact is more evident. The other graphs can be requested from the authors.


HFT activity
(TradElect 4)
FTSE 31 -50
% of total trading activity
25
20
15
10
5
0
17-Apr-09
1-May-09
15-May-09

Appendix 4: 40-day window regressions

The results of the 40-day window panel regression are shown here. Table 7 studies the impact of the technology changes on the HFT activity and Table 8 the impact of the technology changes on institutional investors' execution costs.

	TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
	20 days before and after 1 Sep 08		20 days before and after 2 May 09		20 days before and after 20 Jul 09		20 days before and after 20 Mar 10	
HFT fraction	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
LSE latency FTSE 1 - 10	0.01555	0.45	-0.0442	-2.11	0.00077	0.02	-0.0437	-2.63
LSE latency FTSE 11 - 30	-0.01076	-0.42	-0.02161	-1.2	-0.00658	-0.26	-0.0279	-3.48
LSE latency FTSE 31 - 50	-0.0253	-1.04	-0.02879	-2.38	0.00931	0.33	-0.01593	-1.9
LSE latency FTSE 51 - 100	0.00492	0.29	-0.01447	-1.27	0.01926	0.59	-0.00883	-1.45
LSE latency FTSE 101 - 151	0.01524	1.55	-0.01465	-1.86	0.01645	0.7	-0.00277	-0.62
LSE latency FTSE 151 - 200	0.00544	0.6	0.00561	0.66	0.01682	0.77	-0.00115	-0.27
LSE latency FTSE 201 - 250	0.00396	0.67	-0.00097	-0.13	0.01778	1.33	0.00624	2.07
Total volume	-0.01455	-2.25	-0.03342	-9.77	-0.03107	-9.37	-0.02923	11.06
Intercept	0.74764	2.65	1.60962	14.73	1.23752	10.63	1.52802	19.04
Adj-R Squared		0.748		0.8		0.625		0.838
N		9397		9096		9080		8872

Table 7 shows the results from a panel regression with stock fixed effects of the fraction of HFT volume on, (i) LSE latency for the seven groups, (ii) total volume traded, and, (iii) a linear time trends for each category (coefficients not shown). The TradElect 4 and 5 latency reductions increase HFT activity, while the results for the other sample periods are inconclusive. The models are estimated by OLS and standard errors are double clustered at the stock and day level.

	TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
	20 days before and after 1 Sep 08		20 days before and after 2 May 09		20 days before and after 20 Jul 09		20 days before and after 20 Mar 10	
Execution cost	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
LSE latency FTSE 1 - 10	-0.0014	-0.39	-0.0045	-1.55	0.0014	0.3	-0.0006	-0.35
LSE latency FTSE 11 - 30	0.0043	1.14	0.0028	1.11	-0.0036	-1.08	-0.001	-0.82
LSE latency FTSE 31 - 50	0.0053	1.77	-0.0012	-0.57	-0.0039	-0.86	0.0023	2.15
LSE latency FTSE 51 - 100	0.0047	1.42	-0.0014	-0.59	-0.0055	-2.79	0.0006	0.69
LSE latency FTSE 101 - 151	0.004	1	-0.0018	-0.68	-0.0013	-0.42	-0.0006	-0.3
LSE latency FTSE 151 - 200	-0.0005	-0.12	0.0008	0.25	0.0042	1.28	-0.0002	-0.07
LSE latency FTSE 201 - 250	0.0021	0.75	-0.0032	-0.88	0.0038	0.8	-0.0004	-0.09
Total volume	0.0023	1.3	0.0009	1.13	0.001	1.72	0.0008	0.98
Intercept	-0.0216	-0.63	-0.0041	-0.18	-0.0248	-1.29	-0.0111	-0.51
Adj-R Squared		0.0548		0.0595		0.0485		0.0998
N		4220		4859		4892		2562

Table 8 shows the results from a panel regression with stock fixed effects of the execution cost of institutional investors on, (i) LSE latency f0r the seven groups, (ii) total volume traded, and, (iii) a linear time trends for each category (coefficients not shown). The results for all the sample periods are inconclusive. The models are estimated by OLS and standard errors are double-clustered at the stock and day level.

Appendix 5: Alternative regression specifications

Pooled regression

To increase the power of the event-based regressions, we pool the four events into a single regression. The basic 2SLS regression setup as described in Equations 3 and 5 does not change. However, to limit the number of coefficients that need to be estimated we restrict the slope coefficients on the seven different groups to be the same. Instead we allow for different levels of HFT activity for the four windows,

$$H_{jt} = \alpha_j + w_k + \beta_4^i L_{tk} + \nu V_{jt} + \varepsilon_{jt} \qquad (6)$$

$$TC_{jt} = \alpha_j + w_k + \theta \hat{H}_{jt} + \nu V_{jt} + \varepsilon_{jt} \qquad (7)$$

where *k* runs from one to four, *w* is a window dummy, *t* is defined in event time (one to 20) and L_{tk} refers to changes in latency in event time (not calendar time). The results are summarized in Tables 9 and 10.

First stage				Second stage		
HFT fraction	**Coef.**	**z-stat**		**T-Cost**	**Coef.**	**z-stat**
LSE latency FTSE 1 - 10	-0.0116	-1.45		HFT-hat	-0.0203	-0.69
LSE latency FTSE 11 - 30	-0.0112	-1.72		Total volume	0.0004	0.51
LSE latency FTSE 31 - 50	-0.0032	-0.46		Intercept	0.0199	1.06
LSE latency FTSE 51 - 100	-0.0012	-0.19		Adj-R Squared		0.0306
LSE latency FTSE 101 - 151	-0.0082	-1.3		N		8138
LSE latency FTSE 151 - 200	-0.0008	-0.11		**Table 10: Pooled regression second stage**		
LSE latency FTSE 201 - 250	-0.0051	-0.74				
Total volume	-0.0259	-6.34				

First stage				Second stage		
Intercept	0.3416	4.67				
Adj-R Squared		0.506				
N		8138				

Table 9: Pooled regression first stage

Portfolio regressions

To reduce noise in the HFT and execution cost measures, the 2SLS regressions are run on seven portfolios instead of 250 stocks. The portfolios are formed by size as defined in Section 3. Within each portfolio HFT activity and execution costs are weighted by volume. The regression setup, as described in Equations 3 and 5, does not change with the exception that variables are now defined on the portfolio level (i=1,…7) rather than on the individual stock level (j=1,…,250):

$$H_{it} = \alpha_i + \beta_1^i t + \beta_4^i L_t + \nu V_{it} + \delta b_t + \varepsilon_{it} \qquad (8)$$

$$TC_{it} = \alpha_i + \beta_1^i t + \theta \hat{H}_{it} + \nu V_{it} + \delta b_t + \varepsilon_{it}. \qquad (9)$$

The results are summarized in Tables 11 and 12.

First stage										
	4 upgrades		TradElect 3		TradElect 4		TradElect 4.1		TradElect 5	
HFT fraction	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**
LSE latency FTSE 1 - 10	-0.0033	-1.59	0.0423	3.65	0.0059	0.45	-0.0350	-1.09	-0.0249	-3.80
LSE latency FTSE 11 - 30	-0.0096	-3.06	0.0013	0.06	-0.0251	-1.53	0.0171	0.54	-0.0466	-9.90
LSE latency FTSE 31 - 50	-0.0120	-5.15	-0.0413	-2.08	-0.0252	-1.27	0.0083	0.20	-0.0155	-1.91
LSE latency FTSE 51 - 100	-0.0073	-3.92	0.0151	1.18	-0.0269	-1.70	0.0212	0.52	-0.0112	-1.94
LSE latency FTSE 101 - 151	-0.0081	-5.72	0.0042	0.32	-0.0331	-2.30	0.0085	0.26	-0.0080	-1.88
LSE latency FTSE 151 - 200	-0.0023	-1.71	0.0130	1.21	-0.0047	-0.36	0.0215	0.63	-0.0101	-1.30
LSE latency FTSE 201 - 250	-0.0010	-0.70	-0.0028	-0.23	-0.0189	-1.52	0.0396	1.81	-0.0002	-0.06
Total volume	-0.0016	-0.42	0.0353	1.67	-0.0170	-1.10	-0.0398	-3.30	-0.0432	-4.35

First stage										
Ban dummy	-0.0032	-0.78								
Intercept	0.0601	0.79	-0.7804	-1.86	0.5373	1.69	1.1450	3.88	1.1202	5.93
Adj-R Squared		0.838		0.871		0.898		0.828		0.940
N		4337		140		140		140		139

Table 11: Portfolio regressions first stage

Second stage						
	4 upgrades		**TradElect 4**		**TradElect 5**	
T-Cost	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**
HFT-hat	-0.0136	-0.28	-0.0314	-0.31	0.0526	0.71
Total volume	0.0010	2.53	0.0042	1.64	0.0021	0.48
Ban dummy	0.0006	1.37				
Intercept	-0.0185	-2.42	-0.0757	-1.4	-0.0441	-0.50
Adj-R Squared		0.0011		0.0325		0.0075

Second stage						
N		4337		140		139

Table 12: Portfolio regressions second

Foresight
1 Victoria Street
London SW1H 0ET
www.foresight.gov.uk

URN 12/1054