High-Frequency Trading and Fundamental Price Efficiency*

Jasmin Gider†
jasmin.gider@uni-bonn.de

Simon N. M. Schmickler‡
snms@princeton.edu

Christian Westheide§
westheide@uni-mannheim.de

This version: March 2016

Abstract

We study the impact of high-frequency trading (HFT) on fundamental price efficiency, using the measure proposed by Bai et al. (2015). This measure captures how well current stock market valuations predict earnings in future years. We estimate the effect by exploiting the staggered start of HFT participation in a panel of international exchanges. Our results document a negative impact of the presence of HFT on fundamental price efficiency. These findings are consistent with theoretical models of HFTs' ability to anticipate informed order flow resulting in decreased incentives to acquire fundamental information. Supporting this interpretation, we also find evidence suggesting a decrease in demand for and hence increased competition among sell-side analysts. According to these findings the effect of HFT may be detrimental to welfare because the quality of real resource allocation is likely to be lowered by less efficient prices.

JEL classification: G10, G14
Keywords: High-Frequency Trading, Price Efficiency, Information Acquisition, Information Production

*We thank Mark Van Achter, Tobias Berg, Dion Bongaerts, Jonathan Brogaard, Gökhan Cebiroglu, François Degeorge, Sarah Draus, Thomas Gehrig, Hendrik Hakenes, Nikolaus Hautsch, Satchit Sagade, Erik Theissen, Brian Weller, Sergey Zhuk, and seminar participants at the University of Mannheim and the University of Vienna for helpful comments and discussions. Christian Westheide gratefully acknowledges research support from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE.All remaining errors are our own.

†Institute for Financial Economics and Statistics, University of Bonn , Adenauerallee 24-42, 53113 Bonn, Germany, +49 228 73 7994.

‡Princeton University, Department of Economics.

§University of Mannheim and Research Center SAFE (Goethe University Frankurt) , L9,1-2, 68131 Mannheim, Germany, +49 621 181 1822.

1 Introduction

Modern technology has brought about many changes to the nature of financial marketplaces, the ascent of high-frequency trading (HFT) being one of the major ones. While there exists no formal definition of HFT, common characteristics considered representative of HFT are trading activities through very fast computer systems, short holding periods and frequent order cancellations.[1]

The effects of HFT on certain aspects of market quality, more specifically, on liquidity and price efficiency, have been actively studied in recent years. Almost the entirety of research that studies the relation between HFT activity and price efficiency is concerned with effects at very short horizons, typically at an intra-day or day-to-day level. While this literature has yielded important results, it, by definition, has been restricted to studying the efficiency of the process of impounding *existing* information into prices, i.e., information which has already been obtained by some market participants. However, it has long been known that price efficiency also depends on another process: that of market participants acquiring information which can then be impounded into prices. The implications of HFT on this dimension of price efficiency has been neglected in the empirical literature and this paper is one of the first to make an attempt at tackling this topic.

The predominant view of the existent literature is that HFT enhances liquidity and price discovery, in the sense of improving the process of impounding information into prices. This view is backed by a number of theoretical and empirical studies. E.g., Jovanovic and Menkveld (2015) show theoretically that HFT may reduce adverse

[1] HFT is a subset of algorithmic trading. Algorithmic trading refers to the general class of trading strategies generated by computer-based algorithms, which determine the strategies based on a selection of certain market variables.

selection costs.[2] Furthermore, it is theoretically argued that HFT accelerate the incorporation of information into prices because of their superior signal processing abilities (e.g., Foucault et al. (2015)).

This view is supported by empirical studies: Investigating an optional colocation upgrade at NASDAQ OMX Stockholm, Brogaard et al. (2015) find that increased speed available to market makers improves liquidity. Furthermore, there is evidence showing that HFT incorporate information into prices at higher speed (Zhang (2013)). HFT may be able to improve efficiency because they overcome cognitive limitations that human traders are subject to. Chakrabarty et al. (2015) document that HFT seem to reduce inefficiencies around low-attention announcements. Brogaard et al. (2014) find that HFT enhances price efficiency: HFTs trade in the direction of permanent price changes while trading against pricing errors. Carrion (2013) documents that prices are more efficient when HFT participation is high, in particular with respect to aggressive trades. Riordan and Storkenmaier (2012) arrive at a similar conclusion, studying the consequences of an update to Deutsche Börse's trading system that reduced latency. They find a positive link between the speed of trading and the contribution of quotes to price discovery which they use to capture price efficiency. Conrad et al. (2015) study quoting activity in the U.S. and, exploiting an exogenous technological change, on the Tokyo Stock exchange. They find higher quoting activity to be positively associated with liquidity, measured as the effective spread, and price efficiency, measured using variance ratio tests. Boehmer et al. (2014) study the effect of algorithmic trading on price efficiency, defined as the absence of short-term return predictability; liquidity, defined as spreads and price impacts; and short-term volatility in an international

[2] This notion is also supported empirically in their study.

setting. Overall, their findings indicate beneficial effects on efficiency and liquidity, whereas volatility is increased.

Though the bulk of the early evidence supports the view that HFT improves market quality, there is also empirical evidence which documents a more nuanced picture. Brogaard et al. (2015) exploit the differential impact of a short sale ban on HFT and non-HFT traders and find that HFT may reduce liquidity because of adverse selection of limit orders. Because of certain exceptions for market makers, the set of high-frequency traders (HFTs) affected by the short sale ban may be rather characterized by liquidity taking as opposed to liquidity providing activities. If liquidity is improved by liquidity providing HFTs but suffers from liquidity taking ones, this would explain the estimated negative effect in Brogaard et al. (2015). Chakrabarty et al. (2015) study the effects of the U.S. SEC naked access ban that reduced the participation of aggressive HFTs. They find that the event led to reduced adverse selection costs and a corresponding reduction in bid-ask spreads, such that liquidity takers benefit. Price efficiency is reduced subsequent to the event, though Chakrabarty et al. (2015) suggest that the trade-off in favor of increased liquidity may, overall, be beneficial.

While most of the literature concerned with the effect of HFT on liquidity looks at liquidity for small, individual trade executions, there is some recent evidence on the effects on large institutional trades. Institutional orders usually consist of a large number of small child orders, making it difficult to study effects on costs at the parent order level with standard data sets. Tong (2015) studies the impact of HFT activities on the execution costs of institutional investors in the U.S. and finds that HFT significantly increase execution costs, and in particular so when they trade directionally. However, Brogaard et al. (2014), in a study of the U.K. equity market, do not find significant effects of HFT on institutional transaction costs. They use technology upgrades lowering the latency of the trading system on the London Stock Exchange as shocks to the participation of HFT and find that, while the amount of HFT is affected, institutional execution costs remain unchanged. Van Kervel and Menkveld (2015) analyze data from NASDAQ-OMX Stockholm and find that HFTs trade against large institutional orders during their first hour though turn around and trade in the direction of the order when institutional orders last longer. Transaction costs incurred by the large institutions are higher when HFTs trade in the same direction than when the converse is true. Studying the Canadian market, Korajczyk and Murphy (2015) limit their sample of HFTs to those trading predominantly passively. They find that these HFTs provide liquidity to large trades though they reduce their liquidity provision for the largest ones.

The studies of institutional transaction costs suggest that there may be negative effects of HFT on the implementation of institutional trades. If their use of private information becomes more costly, this in turn may lead to an endogenously lower level of costly information acquisition. This view is corroborated by theoretical models such as those of Yang and Zhu (2015) and Baldauf and Mollner (2015). The model by Baldauf and Mollner (2015) is most closely related to our empirical study as they model a potential tension between the bid-ask spread and the information acquisition incentives. The authors analyze a setting with several types of traders (liquidity providing market maker, investors trading for exogenous reasons, investor who engages in costly information acquisition and infinitely many "front-running" market makers), costly information acquisition and multiple trading venues. When front running market makers, i.e., HFT, become faster, the bid-ask spread decreases, but information acquisition is crowded out. Yang and Zhu (2015) develop a two-period Kyle (1985) model containing a so-called "back-runner" who trades based on observed past order flow. They conclude that "[a] lower cost of acquiring order-flow information reduces the fundamental investor's incentive to acquire fundamental information." Stiglitz (2014) also argues that HFTs may anticipate informed order flow and appropriate the information rents that would otherwise have accrued to the investors that have incurred information acquisition costs. As the rents from investing in real information decrease, real information production by investors decreases accordingly. As a result, less fundamental information is impounded into prices and resource allocation deteriorates, as it is based on less efficient market prices.

With the exception of one paper contemporaneous to ours, Weller (2015), the question of how HFT affects information acquisition has not been empirically addressed yet. Weller (2015) finds that algorithmic trading decreases the amount of information that is impounded into prices before earnings announcements. His evidence supports the existence of a trade-off between the incorporation of existing and new information in prices.

We seek to add to the empirical understanding of the implications of HFT for information acquisition. We consider a more general measure of fundamental price informativeness. This measure is put forward by Bai et al. (2015) and based on a simple, yet intuitively appealing idea: How well do prices today predict earnings in the future? Our analysis is, hence, complementary to Weller's approach. We consider a broader measure of fundamental informativeness which relates to the information content, whereas Weller considers the acquisition of information with respect to the information content that is revealed in the next earnings announcement.[3]

[3]We note a certain trade-off between the degree of precision of the empirical analysis and the

Using an international panel of stock markets, we estimate the impact of HFT on fundamental price efficiency. Following Bai et al. (2015), for each year and each exchange we regress future earnings of varying horizons on current market valuations. Fundamental price efficiency captures the extent to which variation in market valuations predicts variation in future earnings. The HFT "starting" dates from Aitken et al. (2015) capture the beginning of HFT presence.[4] Broadly, HFT started in the United States at the turn of the century, then spilled over to Europe and other developed countries until it finally reached the emerging market of India in 2009. China and South Korea serve as counterfactuals as they have not experienced HFT yet. We identify the impact of HFT in a multi-event difference-in-differences (DiD)analysis. The staggered introduction of HFT makes our identification strategy appealing because it makes it unlikely that a simultaneous unrelated event drives the results.

The start of HFT is associated with a substantial reduction in fundamental price efficiency that amounts to about 75% to 100% of a standard deviation for horizons of one and three years. We confirm the robustness of these results with respect to alternative definitions of HFT start dates. Furthermore, our results are robust to the inclusion of additional control variables, including the introduction of electronic trading that could be considered a potential explanation correlated with HFT. These findings are consistent with the notion that HFT reduces rents to information acquisition.

Among suppliers of fundamental information, financial analysts play an important

relevance with respect to welfare implications. Based on the reasoning by Hirshleifer (1971), longer-term information is more relevant for allocative efficiency as compared to information which is latent but will be revealed with certainty in the short-run. However, the amount of information incorporated into prices before earnings announcements is more likely to be measured precisely as compared to the variation in earnings that can be explained by today's market prices.

[4] Aitken et al. (2015) analyze the effect of HFT on end-of-day market price manipulation in international stock exchanges using their estimated HFT starting dates. Their findings suggest that HFT reduces price manipulation.

role. The decrease in information acquisition activities is likely to affect their business model and the quality of their estimates. We hypothesize that a drop in the demand for information increases competition, and thus raises analyst effort and decreases the relative importance of systematic biases caused by agency problems, resulting in better forecast quality. We find that HFT tends to increase the quality of analyst forecasts: Analyst forecast errors are halved and the dispersion of their estimates is reduced by approximately one third.

Fundamentally informative prices matter from a social welfare perspective because they lead to an efficient allocation of real resources. Prices should reveal the attractiveness of the future investment opportunity set and funds should flow accordingly. Information acquisition might also matter from a social welfare perspective if the information that market participants acquire feeds into real decision making, e.g., through learning or incentive channels. If market participants acquire information that is not known to decision makers at the firm, then the revelation of this information leads to more efficient investment decisions as discussed by Hirshleifer (1971), or more recently, by the market feedback loop literature, e.g., Dow et al. (2015) or Edmans et al. (2015).

Our findings lend support to the notion that HFT reduces information acquisition activities. Our results, hence, provide additional support for the existence of a tension between the incorporation of existing information in prices and incentives to acquire new information that appears to be aggravated by HFT. It also helps reconcile the opposing views of most of the existing academic literature on HFT and the opinions expressed by institutional investors who base their investment decisions on fundamental information, and who indeed appear to be the group of market participants negatively affected by HFT.

The remainder of this paper is structured as follows: Section 2 explains the data, main variables and the empirical strategy, while Section 3 presents the results regarding fundamental price efficiency. Sections 4 shows the results regarding analyst forecasts and Section 5 concludes.

2 Data and research design

2.1 Data sources

Our analysis is based on annual data from 1990 to 2014. Accounting data are obtained from Compustat North America and Compustat Global. CRSP and Compustat Global provide stock return data for U.S. and international exchanges, respectively. We source data on sell-side analyst forecasts from IBES. We use the GDP deflator from the World Bank to turn nominal into real values. We also convert all values to U.S. dollars using exchange rates from the Federal Reserve System. We obtain estimates of the years in which high-frequency traders become active on different exchanges, explained in more detail in section 2.4, from Aitken et al. (2015). Gorham and Singh (2009) provide information on the timing of exchanges' conversion from traditional floor to electronic trading.

2.2 Variable construction

2.2.1 Fundamental price efficiency

We proceed similarly to Bai et al. (2015) in the construction of our measure of fundamental price efficiency. This variable, estimated separately for each exchange and year, measures how well market prices predict earnings realized in future years. Market values are measured at the next end of March after the end of firms' fiscal year. As usual, we use the book value of debt as a proxy for its market value, and we eliminate financial firms from our sample. Similar to Bai et al. (2015), we estimate:

$$\frac{E_{i,t+k}}{A_{i,t}} = a_t \ln\left(\frac{M_{i,t}}{A_{i,t}}\right) \times 1_t + b_t \ln\left(\frac{E_{i,t}}{A_{i,t}}\right) \times 1_t + c_{s(i,t),t}(1_{SIC1}) \times 1_t + \varepsilon_{i,t} \tag{1}$$

where i identifies each firm, t the year, E is EBIT, A is total assets, M is market value of assets, $SIC1$ is the first digit of the SIC code and $k = 1, ..., 5$. All ratios entering equation 1 are winsorized at the 1% level.

Different from Bai et al. (2015), we use firms' total market value rather than market value of equity, so as to prevent effects of financial leverage on our results. For example, even if future EBIT were certain, a firm with great prospects but high debt might have a lower ratio of market value of equity to total assets than a firm with poor prospects and no debt. Thus, this ratio would be a poor predictor of future earnings and falsely indicate a low level of price efficiency.[5] Total market value is defined as the sum of the market value of equity and debt.

Bai et al. (2015) use predicted variation$_t$ $= a_t \times \sigma_t[ln(M_{i,t}/A_{i,t})]$ as their main dependent variable. However, whether more predicted variation means higher or lower price informativeness depends on total variation $= \sigma_t[E_{i,t+k}/A_{i,t}]$. Only if the ratio of predicted variation to total variation increases, we can say that a larger share of the existent information is reflected in prices, i.e., price informativeness increases. We thus

[5]The data confirm this intuition. For the part of our sample period that overlaps with that of Bai et al. (2015), we find our measure of price informativeness to be slightly higher than that reported in their paper.

choose the share of predicted variation

$$\mathrm{Eff}^k = \frac{\text{predicted variation}^k}{\text{total variation}^k} \tag{2}$$

as the dependent variable for our analysis.

The noise in the estimates of our measure will necessarily increase the fewer observations we use for its estimation.

While, on average, 861 data points are available in our regressions, the number is far lower for some exchange-years, leading to less precise estimates. As we do not want outliers computed from few observations to confound our results, we winsorize price informativeness at the 2.5% level.

2.2.2 Sell-side analyst measures

We employ two measures of the informativeness of sell-side analysts' earnings forecasts that have been used earlier, e.g., by Kang and Liu (2008). Forecast error and forecast dispersion are defined as

$$\mathrm{Error} = \frac{|\text{mean} - \text{actual}|}{\mathrm{S}} \tag{3}$$

$$\mathrm{Disp} = \frac{\sigma(\text{forecasts})}{\mathrm{S}} \times 100 \tag{4}$$

where mean is analysts' mean earnings forecast, actual is actual earnings, S is the stock price and $\sigma_{i,t}$(forecasts) is the standard deviation of analysts' earnings forecasts.

2.3 Summary statistics

Table 1 gives an overview of our data. Since our dataset comprises all stocks available in the major databases, the size of sample firms spans a wide range from few million dollars to the largest global firms. The measures of price efficiency and analyst accuracy are of particular interest. The average firm is traded on an exchange where price efficiency is relatively low though there is a wide dispersion in efficiency, and the efficiency for longer horizons of three or five years is somewhat lower than that for a one-year horizon. Forecast dispersion and forecast error are both heavily skewed, though for the typical firm the standard deviation of forecasts amounts to about 0.2% of the stock price, while the absolute error of the consensus forecast is about 0.3%.

2.4 Empirical model and identification strategy

To be able to identify the effects of high-frequency trading, we make use of the estimates of HFT market entry times by Aitken et al. (2015). Aitken et al. (2015) use two approaches to identify likely start dates of HFT participation on diverse international stock exchanges. HFT is generally considered to be associated with small trade sizes and a large amount of order cancellations relative to the trading volume. Thus, using order book and trade data from Thomson Reuters Tick History (TRTH), Aitken et al. (2015) identify times with a sudden decrease in trade size and increase in order cancellations to trade ratio, respectively.[6] Note that while TRTH allows perfect observations of trade sizes, exchanges send only periodic order book snapshots rather than information on each individual order to Thomson Reuters such that the measure based on estimated order to trade ratios can be expected to be somewhat noisy. Thus, our preferred proxy

[6]For the precise definition of the dates, see the appendix of Aitken et al. (2015).

for HFT market entry is based on a reduction in trade size.

The use of colocation, i.e., the housing of trading firms' computer servers within an exchange's data center, is closely related to HFT activity. While colocation today is used also by other major market participants, HFTs have originally been the primary clientele of exchanges' colocation offerings. It is important to note, however, that colocation is not a necessary condition as HFTs may house their servers near exchanges without the latter offering colocation services. In fact, it is likely that exchanges begin to offer colocation as a response to the demand by HFTs. Colocation does facilitate HFT and likely results in a larger amount of HFT, even though the first HFTs might have traded on an exchange before colocation has been offered. Aitken et al. (2015) identify the dates when exchanges offered colocation for the first time and we use these dates as a third alternative definition for HFT "start" dates.[7]

Based on these proxies for HFT start dates, we employ a multi-event difference-in-differences model using annual stock exchange level observations:

$$\mathrm{Eff}^k_{e,t} = \beta_0 + \beta_1 \mathrm{HFT}_{e,t} + \delta X_{e,t} + \eta_t + \eta_e + \varepsilon_{e,t} \tag{5}$$

e indicates the stock exchange and t the year. Eff^k represents the price efficiency measure for the time horizons $k = 1, ..., 5$. HFT is zero prior to the HFT starting date and one for all following years. The variable is set to the fraction of time in which HFT was present for the year in which it started. X is a vector of control variables which consists of logarithmic total market size, logarithmic average market capitalization and *electronic*, a dummy variable capturing the effect of the transition

[7]For ease of exposition, we subsequently refer to our proxies for HFT market entry as the "HFT starting dates".

from floor to electronic trading. η_t are year fixed effects. η_e are stock exchange fixed effects and $\epsilon_{e,t}$ is the error term.

Following the same approach, we also analyze changes in the sell-side analyst forecast error and dispersion. More concretely, we estimate

$$\text{IBES}_{i,t} = \beta_0 + \beta_1 \text{HFT}_{i,t} + \delta X_{i,t} + \eta_t + \eta_i + \epsilon_{i,t} \tag{6}$$

i indicates the firm. IBES stands for the analyst based measures analyst dispersion and analyst error. The vector of control variables $X_{i,t}$ contains the logarithmic market value, Tobin's Q, a dummy indicating an electronic market, and the rolling five year volatility of EBIT. η_i are firm fixed effects.

Both models feature year and panel variable fixed effects. The former flexibly eliminates common trends. The latter eliminates the impact of unobservable firm or stock exchange specific characteristics. The results are thus driven by within firm and within stock exchange variation. The key to our identification strategy is the staggered start of HFT across markets that we illustrate in figure 1. This staggered introduction of HFT mitigates the concern that our results could be driven by unrelated macroeconomic shocks. Because of the use of fixed effects, they can only drive our results if they occur in the same staggered way as our HFT starting dates, which appears very unlikely. As in other applications of the multi-event difference-in-differences methodology, e.g. Christensen et al. (2015), this fact and the concrete specification of our empirical model also imply that it is not crucial to establish parallel trends.

Empirical work often faces the threats of reverse causality and endogeneity. HFT clearly is not a direct consequence of lower price efficiency or higher analyst report quality. Nevertheless, HFTs may self-select into markets where they anticipate lower price efficiency or better analyst reports. We cannot rule out that HFTs make larger profits in inefficient markets or with better reports. This would bias our results. However, the decision by an HFT to establish a presence in a given market, which is a decision that cannot be implemented from one day to the next, cannot plausibly be made in anticipation of a permanently lowered fundamental price efficiency. Furthermore, an increase in price inefficiency would exacerbate the risk of facing better informed counterparties, whereas HFTs would prefer the exploitation of large but ultimately less informed order flow. HFT originated in large cap stocks traded on exchanges in the U.S. (Aitken et al., 2015), which is a market segment where prices are among the most efficient worldwide. Nowadays, HFTs still are most active in large, liquid stocks (Brogaard et al., 2015; Aldridge, 2013; Tong, 2015) whose price efficiency is higher than that of smaller and less liquid ones. Thus, while HFT obviously does not fall from the blue sky, the mechanisms driving it suggest that any estimate would rather be upward biased, such that estimates of negative effects on price efficiency likely underestimate their true magnitude.

Besides HFT, there are other factors that can be expected to cause variation in the level of price efficiency. We control for these factors to eliminate those concerns. First, there may be several time and stock exchange specific characteristics influencing the informativeness of prices. Thus, we include year and exchange fixed effects. This already captures many factors driving price informativeness. Further, we control for time-varying market characteristics.

Price efficiency is higher for large firms. It is thus higher on exchanges where investors trade shares of high value firms. Therefore, we control for average market capitalization. Price efficiency is also generally higher on large exchanges. For example, prices at the stock exchange in Oslo have relatively poor price efficiency, even though Norway is one of the most developed countries in the world. Price informativeness with horizon 1 is three times as high at the NYSE. Hence, we control for the sum of all firms' market value at an exchange. Stock markets need to have made the transition from open outcry to electronic trading before HFT can operate. At the same time this transition reduces trading costs and thus affects price efficiency. To eliminate this potential confounding effect, we add a dummy variable to control for the transition to electronic trading systems. Lastly, liquidity and volatility at an exchange may impact on price efficiency.These variables might also be channels via which HFT affects the dependent variable. Therefore, we only include average Amihud illiquidity and average volatility in a robustness check.

There are also other factors causing differences in analyst forecast error and dispersion. If earnings are more volatile, it is more difficult to forecast them correctly. Thus, we control for the five-year rolling earnings volatility. The estimation of earnings is easier for large firms than for small ones. Hence, we control for firm size. We further include Tobin's Q, because earnings of growth firms are harder to predict. Moreover, in 2003, under pressure from regulators, ten of the largest American investment firms agreed to mitigate the influence of investment bankers on research analysts under the "Global Analyst Research Settlement". Empirical evidence indicates that this agreement reduced analysts' optimism but also decreased the informativeness of their reports (Clarke et al., 2011; Kadan et al., 2009). In our robustness checks, we include a binary variable which is one for all American firms following the agreement to control for this event. Finally, for the same reasons as in the paragraph above, we control for the transition to electronic exchanges, year and panel variable fixed effects and include volatility and Amihud illiquidity in a robustness check.

3 Fundamental price efficiency

3.1 Empirical results

If the presence of HFT leads to an increase in information acquisition, we would expect a positive coefficient when regressing fundamental price efficiency on the variable HFT. Conversely, if HFT leads to an erosion of the rents to and thereby amount of information acquisition, we expect a negative coefficient. Table 2 shows the results of a regression of price efficiency on exchange fixed effects, time fixed effects, a set of control variables and the time-varying variable HFT as the main variable of interest. We estimate this empirical model with price efficiency as dependent variable for horizons 1 to 5. Using normal ordinary least squares standard errors may lead to an overestimation of the t-statistic because of correlation between residuals. Robust standard errors need not be consistent in fixed effects models. In this subsection, we consequently use cluster-robust standard errors as described in each table. Hence, heteroscedasticity and correlation between residuals in the specific clusters do not introduce a bias (Cameron and Miller, 2015). The robustness checks in subsection 3.2.3 address further concerns about error correlation.

The coefficients of HFT are negative for all horizons, consistent with the notion that HFT may decrease fundamental price efficiency, and statistically significant at least at the 5% level for time horizons 1 to 4. The effects are not only statistically, but also economically significant: The size of the HFT coefficient in column (1) (-0.091) corresponds to approximately 76% of the standard deviation of price efficiency for this horizon and the coefficient for horizon 3 (column (3)) presents 100% of the standard deviation. We note that the variation in price efficiency that can be explained by the dummy variable HFT, exchange and time fixed effects and control variables decreases with larger horizons: Adjusted R^2 is relatively high at 44% for horizon 1 and decreases gradually to 27% for horizon 5.

3.2 Robustness

3.2.1 HFT start dates

To explore the robustness of our findings, we first demonstrate the robustness with respect to the alternative proxies for HFT market entry from Aitken et al. (2015) that we defined in subsection 2.4. We estimate our main empirical models with the explanatory variable HFT defined by order cancellation and colocation instead of trade size. Table 3 shows the results of this robustness check for a forecast horizon of three years.

The results are robust to alternative definitions of the HFT start dates. Defining HFT based on cancelled orders yields a smaller estimate of -0.098 which is statistically significant at the 10% level. The coefficient of HFT according to colocation dates is -0.15 and, hence, identical to our base case. This coefficient is statistically significant at the 5% level.

3.2.2 Confounding staggered events

Interpreting the coefficient of the variable HFT as a causal effect rests upon the assumption that there is no other confounding factor correlated with HFT affecting fundamental price efficiency. The introduction of electronic trading platforms is such an alternative type of staggered events that affect financial markets and may gave occurred in a similar sequencing (see subsection 2.4).

There is a substantial gap between the transition to electronic markets and the start of HFT. While the former happened mostly during the 1990s, the latter mainly occurred during the last decade. One might be concerned that the dummy variable *electronic* could cause the drop in the dependent variable, but if the two variables are highly correlated, this effect could be falsely attributed to HFT. The two dummy variables are in fact correlated: The raw correlation between *electronic* and HFT is relatively large with a value of 0.31 and statistically significant at the 1% level. We directly test the impact of *electronic* on price efficiency in a model where we exclude HFT. The results are depicted in Table 4. For all horizons, the coefficient of *electronic* fails to be statistically significant at conventional levels. For horizons 1, 2 and 3, the coefficient is even positive and only negative for horizons 4 and 5. The empirical evidence is inconsistent with the notion that the introduction of electronic trading is responsible for a decrease in fundamental information efficiency. Based on these results, we can reject the objection that we misattribute a potential impact of *electronic* to the introduction of HFT.

3.2.3 Standard errors

The seminal paper by Bertrand et al. (2004) highlights potential problems of serial correlation in panel datasets that may cause biased standard errors in difference-in-differences analyses. This leads to a potential overrejection of the null hypothesis that a coefficient is equal to 0. We use cluster-robust standard errors in the subsection above and subsequently further address possible issues resulting from correlation of errors which threaten to introduce a bias.

Bertrand et al. (2004) suggest collapsing the data for the time periods before and after the treatment as a solution. In Table 5 we compare the means before and after the treatment for all observations and for each stock exchange individually using a t-test. This simple test confirms our results. The change is negative with only one exception and statistically significant in most cases.

We follow Petersen (2009) to further address the concern. Table 6 summarizes the estimation results from the following robustness checks. When analyzing Eff^k, residuals of a stock exchange may be correlated across years because of unobservable stock exchange characteristics and residuals of a certain year may be correlated across stock exchanges because of unobservable year characteristics. Row 1 of table 6 shows the coefficients of HFT from estimating the main model and using normal ordinary least squares standard errors. The coefficients of HFT are negative and highly significant. When we include year-fixed effects instead of a simple time trend and also stock exchange fixed effects, unobservable year or stock exchange characteristics cannot introduce a bias. But both forms of dependence can be temporary. The fixed effects estimator would then be inconsistent. Clustering standard errors on both dimensions instead of using fixed effects then yields consistent and potentially even more efficient estimates (Thompson, 2011; Cameron et al., 2011). The second row of table 6 shows that this procedure leaves our results unchanged. All relevant coefficients remain negative and statistically significant.

When both forms of dependence are jointly present, another solution is to include dummies for the first and to cluster standard errors by the second dimension. In the third rows of table 6 we cluster standard errors by stock exchange and by firm respectively and include year fixed effects. Again, our findings are robust to these specifications. In row (4), we cluster standard errors two-dimensionally and include time fixed effects. Our results are again robust to this alternative specification. We now obtain significant results also for horizon 5.

3.2.4 Placebo tests

Additionally, we run a placebo test. When Bertrand et al. (2004) generate placebo treatments in serially correlated data, they found that 45% of the coefficients of placebo treatment variables were significant when estimating simple DiD models. Similarly, we generate random HFT starting dates and use them to construct a placebo for HFT presence. We then use this placebo as an independent variable in our main model. We repeat this 1,000 times for each of our main regressions 5 with price efficiency with time horizons 1 to 5. We count how many coefficients of the randomly generated placebo are statistically significant at the 1, 5 and 10% level and report the resulting rejection rates in table 7. For each of the 7 regressions, all three rejection rates are lower than the respective significance level. For example, in the regression with price efficiency of horizon 4 as the dependent variable, the coefficient of the placebo is only significant at the 1% level in 0.5% of all cases. This test shows that our empirical strategy rejects the null hypothesis in cases where we expect it to, i.e, it does not reject the null in cases where there is no significant impact by design.

3.2.5 Further robustness checks

In a further robustness check, we estimate the main empirical model with additional control variables in 8. We excluded them previously because these variables could serve as potential channels for the effect. For the regressions with price efficiency as a regressand we add average Amihud illiquidity and average return volatility.

All coefficients of HFT, except for horizon 5, are negative and statistically significant. Similarly, economic significance decreases only slightly. This has two implications. First, our results are robust to this specification of the empirical model. Second, liquidity and volatility are not the dominant channels for the effect of HFT on price efficiency. We would otherwise have expected a sharp decrease in the statistical significance of the coefficients of HFT.

4 Implications for financial analysts

4.1 Hypotheses

Our empirical results from Section 3 lend support to the notion that HFT decreases the demand for fundamental information. HFT and fundamental price efficiency are linked by market participants who typically invest real resources to engage in information acquisition. Financial analysts represent an important supplier of fundamental information (see, e.g., Derrien and Kecskés (2013)). If HFT leads to a drop in the demand for fundamental information, we would expect this to be reflected in the competitive behavior of suppliers of fundamental information. In the present section, we derive empirically testable hypotheses corresponding to this notion.

In the short to medium term we expect that a drop in the demand creates incentives for analysts to increase their effort. This is because a drop in demand for information is likely to have a similar effect as an increase in competition. Sun (2011) provides empirical evidence for the notion that competition increases the quality of analyst reports. In particular, it decreases the noise and dispersion in earnings forecasts. Furthermore, previous literature (see, e.g., Barber et al. (2007), O'Brien et al. (2005) or Michaely and Womack (1999)) shows that analysts issue overly optimistic reports to please firms' executives so as to ensure that they keep providing access to the analysts and possibly generate investment banking revenue.[8] According to empirical evidence presented by Hong and Kacperczyk (2010), an increase in competition reduces this bias in analyst opinion.

However, this relationship may change over the longer-term, where the number of financial analysts working for a certain investment bank or boutique, can be adjusted to the new lower level of demand for information acquisition. In the short run, analysts' individual forecast quality increases as a response to the start of HFT. Hence, we expect that analyst error and dispersion decrease subsequent to the start of HFT in the short run. Yet, in the long run, as a new long-term equilibrium between information supply and demand is established, analysts' forecast quality should partially reverse as competition among suppliers eases. This implies that from the short to the long run

[8]This refers to all but very short-term forecasts where there is evidence that analysts low-ball their estimates to allow firms to beat the "expectations".

analyst measures deteriorate.

4.2 Empirical results

We estimate the main empirical model of this paper with earnings forecast error and dispersion as regressands as defined in subsection 2.2.2. The forecasts used refer to annual earnings and are issued one year ahead. The results of column (1) of table 9 show that HFT reduces the error of analysts' earnings forecasts as predicted. Forecast errors almost halve their mean value when HFT starts. In order to investigate whether forecasts are not only more precise but also less biased, we test whether the reduction in error is symmetric or stems from the reduction in overly optimistic forecasts. To this end, we construct two variants of analyst error: *Pos. Analyst Bias* (*Neg. Analyst Bias*) is equal to analyst error if the forecast is larger (smaller) than the actual value, and set to 0 otherwise. *Pos. Analyst Bias* seeks to capture overly optimistic forecasts. The coefficient of HFT for the *Pos. Analyst Bias* in column (2) is negative and statistically significant. However, the HFT coefficient in column (3) is negative, but fails to be statistically significant at conventional levels. These findings suggest that the improvement in the analyst error is mainly stemming from the reduction of overly optimistic forecasts.

Column (4) presents the estimate with respect to analyst dispersion. On average, the introduction of HFT reduces dispersion by approximately one third compared to its mean value. The effects are statistically significant at the 1% level. When HFTs enter the market, analysts appear to predict earnings better.

Our earlier line of reasoning predicts that error and dispersion increase from the short to the longer run. To test this claim, we drop the observations where HFT is absent. Using our main model, we then test whether error and dispersion actually increase 2, 3 and 4 years after the start of HFT. The results are presented in Table 10 and support this notion. The coefficient of the dummy indicating whether HFT is present for at least 4 years is positive and significant for analyst dispersion and error. When taking 3 instead of 4 years it is still positive and significant for analyst dispersion. When using only 2 years, the coefficients are positive, but fail to be statistically significant at conventional levels. This is consistent with our prediction that forecast quality reverses after some time, when demand and supply for analyst reports has balanced. However, we note that the HFT impact does not seem to *fully* revert back to its previous level.

4.3 Robustness

Table 11 depicts the estimates based on alternative definitions of HFT start dates. Columns (1) and (2) show the results based on order cancellation, while columns (3) and (4) use the colocation definition. The results are comparable.

We examine the robustness of the results with respect to various alternatives of clustering the standard errors. Table 12 shows the coefficient of the HFT variable using analyst error and analyst dispersion as outcome variables. The results are robust to these alternative specifications. However, we note that the economic magnitudes become larger if we do not control for exchange fixed effects for both analyst error and analyst dispersion as shown by rows (2), (3) and (4).

5 Conclusion

The two principal functions of financial markets are risk-sharing and efficient resource allocation. Accordingly, market quality is generally defined as consisting of two dimensions: liquidity and price discovery. While these two dimensions are naturally interlinked, this paper addresses the latter. It has been known since Hirshleifer (1971) that the efficiency of prices depends on two activities, the incorporation of existing information into prices and the acquisition of new information.

Previous literature on high-frequency trading has primarily studied the former. This paper is one of the first to address the influence of HFT on aggregate price efficiency, related to earnings realized years into the future, and thus speaks to the latter.

Our empirical evidence suggests that the presence of HFT reduces price efficiency. With HFT, market valuations predict future earnings less precisely.

Results on the quality of analysts' forecasts corroborate the notion that there is less demand for fundamental information subsequent to HFTs' entry into the market. Increased competition on the supply side, i.e., among sell-side analysts, temporarily leads to an improvement in forecast accuracy and a reduction in forecast dispersion, before the effect attenuates slightly 3 to 4 years later, consistent with a new equilibrium between supply and demand for information.

In sum, our results provide empirical support for the the arguments of Stiglitz (2014), modeled theoretically by, e.g., Baldauf and Mollner (2015). The findings are consistent with the idea that HFT reduces the gains from information for institutional investors through order anticipation, i.e., the ability to use past order flow to predict future order flow by institutional investors in the same direction, making the execution of large informed trades more expensive. Hence, institutional investors acquire less information and as a consequence, market prices reflect less fundamental information. Thus, the basis for real resource allocation is distorted. This result of HFT unambiguously decreases total welfare, while the aggregate effect of HFT on welfare would have to consider the trade-off with effects on liquidity that are generally considered to be positive in the existing literature. Since different trading strategies are involved in beneficial liquidity provision and aggressive exploitation of order anticipation, market operators or regulators may reasonably consider potential mechanisms to rein in aggressive HFT.

References

Aitken, M., D. Cumming, and F. Zhan (2015). High frequency trading and end-of-day price dislocation. *Journal of Banking & Finance 59*, 330–349.

Aldridge, I. (2013). *High-frequency trading: A practical guide to algorithmic strategies and trading systems* (2nd ed ed.). Wiley trading series. Hoboken, N.J: Wiley.

Bai, J., T. Philippon, and A. Savov (2015). Have financial markets become more informative? *Journal of Financial Economics forthcoming*.

Baldauf, M. and J. Mollner (2015). High-frequency trade and market performance. working paper.

Barber, B., R. Lehavy, and B. Trueman (2007). Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics 85*, 490–517.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates. *Quarterly Journal of Economics 119*(1), 249–275.

Boehmer, E., K. Y. Fong, and J. J. Wu (2014). International evidence on algorithmic trading. *Working Paper*.

Brogaard, J., B. Hagströmer, and R. Riordan (2015). Trading fast and slow: Colocation and liquidity. *Review of Financial Studies 28*(12), 3407–3443.

Brogaard, J., T. Hendershott, S. Hunt, and C. Ysusi (2014). High-frequency trading and the execution costs of institutional investors. *Financial Review 49*(2), 345–369.

Brogaard, J., T. Hendershott, and R. Riordan (2014). High frequency trading and price discovery. *Review of Financial Studies 27*(8), 2267–2306.

Brogaard, J., T. Hendershott, and R. Riordan (2015). High frequency trading and the 2008 short sale ban. *Working Paper*.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics 29*(2), 238–249.

Cameron, A. C. and D. L. Miller (2015). A practitioners guide to cluster-robust inference. *Journal of Human Resources 50*(2), 317–372.

Carrion, A. (2013). Very fast money: High-frequency trading on the nasdaq. *Journal of Financial Markets 16*(4), 680–711.

Chakrabarty, B., P. K. Jain, A. Shkilko, and K. Sokolov (2015). Speed of market access and market quality: Evidence from the sec naked access ban. *Working Paper*.

Chakrabarty, B., P. C. Moulton, and X. F. Wang (2015). Attention effects in a high-frequency world. *Working Paper*.

Christensen, H. B., L. Hail, and C. Leuz (2015). Capital-market effects of securities regulation: The role of implementation and enforcement. *Working Paper*.

Clarke, J. E., A. Khorana, A. Patel, and P. R. Rau (2011). Independents' day? analyst behavior surrounding the global settlement. *Annals of Finance 7*(4), 529–547.

Conrad, J., S. Wahal, and J. Xiang (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics 116*(2), 271–291.

Derrien, F. and A. Kecskés (2013). The real effects of financial shocks: Evidence from exogenous changes in analyst coverage. *Journal of Finance 68*(4), 1407–1440.

Dow, J., I. Goldstein, and A. Guembel (2015). Incentives for information production in markets where price affect real investment. *Working Paper*.

Edmans, A., I. Goldstein, and W. Jiang (2015). Feedback effects, asymmetric trading, and the limits to arbitrage. *American Economic Review forthcoming*.

Foucault, T., J. Hombert, and I. Roşu (2015). News trading and speed. *Journal of Finance forthcoming*.

Gorham, M. and N. Singh (2009). *Electronic exchanges: The global transformation from pits to bits*. Elsevier and IIT Stuart Center for Financial Markets press. Amsterdam and Chicago: Elsevier and Stuart School of Business Center for Finacial Markets.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review 61*(4), 561–574.

Hong, H. and M. Kacperczyk (2010). Competition and bias. *Quarterly Journal of Economics 125*(4), 1683–1725.

Jovanovic, B. and A. J. Menkveld (2015). Middlemen in limit order markets. working paper.

Kadan, O., L. Madureira, R. Wang, and T. Zach (2009). Conflicts of interest and stock recommendations: The effects of the global settlement and related regulations. *Review of Financial Studies 22*(10), 4189–4217.

Kang, Q. and Q. Liu (2008). Stock trading, information production, and executive incentives. *Journal of Corporate Finance 14*(4), 484–498.

Korajczyk, R. A. and D. Murphy (2015). High frequency market making to large institutional trades. *Working Paper*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

Michaely, R. and K. L. Womack (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies 12*(4), 653–686.

O'Brien, P. C., M. F. McNichols, and H. W. Lin (2005). Analyst impartiality and investment banking relationships. *Journal of Accounting Research 43*(4), 623–650.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480.

Riordan, R. and A. Storkenmaier (2012). Latency, liquidity and price discovery. *Journal of Financial Markets 15*(4), 416–437.

Stiglitz, J. E. (2014). Tapping the brakes: Are less active markets safer and better for the economy? speech at the federal reserve bank of atlanta 2014 financial markets conference.

Sun, L. (2011). How does competition affect opinion dispersion? *Working Paper*.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics 99*(1), 1–10.

Tong, L. (2015). A blessing or a curse? the impact of high frequency trading on institutional investors. *Working Paper*.

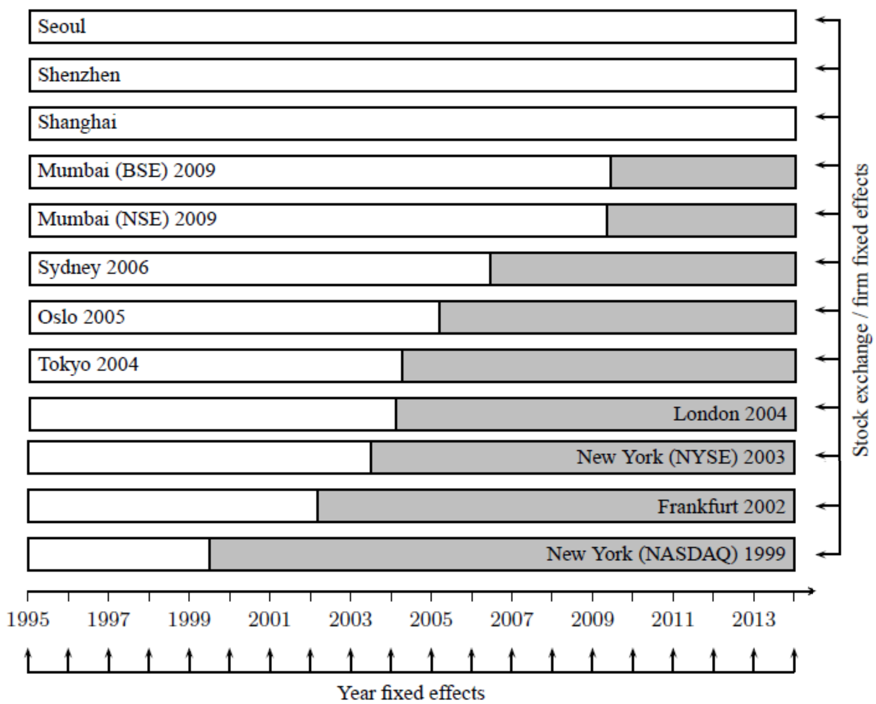
Van Kervel, V. and A. Menkveld (2015). High-frequency trading around large institutional orders. *Working Paper*.

Weller, B. M. (2015). Efficient prices at any cost: Does algorithmic trading deter information acquisition? *Working Paper*.

Yang, L. and H. Zhu (2015). Back-running: Seeking and hiding fundamental information in order flows. working paper.

Zhang, S. S. (2013). Need for speed: An empirical analysis of hard and soft information in a high frequency world. *Working Paper*.

Figure 1: Illustration of the identification strategy



This figure illustrates the identification strategy of this paper. The sample includes observations from global stock exchanges. Gray shades indicate HFT presence. The graphic shows the staggered start of HFT starting in North-America, then reaching Europe, other developed countries and finally the emerging market India, but not China and South Korea, which we include as counterfactuals in our analysis. This makes it very unlikely that simultaneous but unrelated events drive our results. It also allows us to include year and panel variable fixed effects. Year fixed effects model a flexible time trend. The panel variable is a stock exchange ID when we use price efficiency and a firm ID when we use analyst error or dispersion as dependent variable. After controlling for the transition from floor to electronic trading, firm or market level variables and clustering standard errors, we robustly identify the impact of HFT.

Table 1: Descriptive statistics

	Lower 5%	Median	Mean	Upper 5%	S.D.
Market Value	10.457	351.854	4504.537	18166.912	19626.43
Total Assets	3.375	198.661	2397.027	8928.243	12688.93
EBIT	-18.408	8.776	171.200	624.287	1129.93
R&D	0.000	3.991	75.199	216.584	465.65
CAPX	0.052	7.402	148.905	531.000	908.63
ln(MV/Total Assets)	-0.453	0.300	0.426	1.710	0.67
EBIT/Total Assets	-0.346	0.055	0.016	0.216	0.22
R&D/Total Assets	0.000	0.000	0.027	0.163	0.08
CAPX/Total Assets	0.000	0.031	0.057	0.215	0.08
Price Efficiency 1	-0.114	0.054	0.073	0.291	0.12
Price Efficiency 3	-0.169	0.028	0.062	0.354	0.15
Price Efficiency 5	-0.170	0.032	0.058	0.306	0.14
Analyst Dispersion	0.008	0.210	6.589	26.168	29.55
Analyst Error	0.007	0.337	16.930	56.426	82.44
Pos. forecast bias	0.000	0.000	8.916	16.753	52.13
Neg. forecast bias	0.000	0.007	3.642	8.440	19.04
Observations	300,366				

We use annual observations from 13 stock exchanges between 1990 and 2014. Data are from Compustat, CRSP, IBES, the World Bank and the Federal Reserve System. Accounting measures are real and in million USD. As in Bai et al. (2015), we measure stock market value at the next end of march after the close of the firm's fiscal year and drop financial firms from our sample. It then comprises 28,270 firms. Ratios and analyst measures are winsorized at the 1% and price efficiency at the 2.5% level.

Table 2: Price efficiency: main DiD model

	(1) Price Efficiency 1	(2) Price Efficiency 2	(3) Price Efficiency 3	(4) Price Efficiency 4	(5) Price Efficiency 5
HFT (trade size)	-0.091***	-0.10**	-0.15***	-0.10**	-0.016
	(-3.13)	(-2.55)	(-3.32)	(-2.23)	(-0.34)
Electronic	-0.0061	-0.0015	0.0040	-0.024	-0.038
	(-0.27)	(-0.04)	(0.10)	(-0.48)	(-0.81)
ln (Market Size)	0.0038	-0.0047	-0.011	0.0015	0.0099
	(0.45)	(-0.50)	(-1.19)	(0.17)	(0.66)
ln (Average Cap)	-0.0074	0.028	0.014	-0.0063	-0.039
	(-0.44)	(1.08)	(0.56)	(-0.23)	(-1.62)
Constant	0.085	0.25	0.10	-0.017	-0.12
	(0.77)	(1.41)	(0.57)	(-0.09)	(-0.74)
Year FE	yes	yes	yes	yes	yes
Exchange FE	yes	yes	yes	yes	yes
Adjusted R2	0.44	0.42	0.34	0.31	0.27
N	276	263	250	238	226

This Table shows the results of a regression price efficiency with horizon k on HFT, a set of control variables and year and stock exchange fixed effects:

$$\mathrm{Eff}^k_{e,t} = \beta_0 + \beta_1 \mathrm{HFT}_{e,t} + \delta X_{e,t} + \eta_t + \eta_e + \epsilon_{e,t}$$

We use annual stock exchange level observations from 13 stock exchanges between 1990 and 2014. All values are real and in USD when applicable. Market size and average capitalization are in trillion USD. This table presents the results of our main multi-event DiD model. All variables are defined as in Section 2.2. We cluster standard errors by year. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 3: Price efficiency DiD: alternative HFT start dates

	(1) Trade size	(2) Order cancellation	(3) Co-location
HFT (trade size)	-0.15***		
	(-3.32)		
HFT		-0.098*	
		(-2.04)	
HFT (colocation)			-0.15**
			(-2.68)
Electronic	0.0040	-0.0019	0.040
	(0.10)	(-0.05)	(0.92)
ln (Market Size)	-0.011	0.00059	-0.0016
	(-1.19)	(0.06)	(-0.15)
ln (Average Cap)	0.014	-0.0033	0.0039
	(0.56)	(-0.14)	(0.17)
Constant	0.10	0.029	0.067
	(0.57)	(0.17)	(0.42)
Year FE	yes	yes	yes
Exchange FE	yes	yes	yes
Adjusted R2	0.34	0.33	0.31
N	250	247	267

This table shows the results of the regression from Table 2, but uses alternative definitions of HFT presence. The HFT start date in column (1) is determined based on trade size, the definition in column (2) based on order cancellations and the one in column (3) according to colocation dates. All variables are defined as in section 2.2. We cluster standard errors by year. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 4: Price efficiency DiD: electronic as main explanatory variable

	(1) Price Efficiency 1	(2) Price Efficiency 2	(3) Price Efficiency 3	(4) Price Efficiency 4	(5) Price Efficiency 5
Electronic	0.014	0.018	0.018	-0.0093	-0.034
	(0.57)	(0.42)	(0.40)	(-0.20)	(-0.77)
ln (Market Size)	0.010	0.0052	-0.0017	0.013*	0.015
	(1.26)	(0.56)	(-0.15)	(1.89)	(1.29)
ln (Average Cap)	-0.023	0.0048	0.0060	-0.028	-0.048**
	(-1.25)	(0.19)	(0.26)	(-1.14)	(-2.22)
Constant	-0.0012	0.13	0.090	-0.13	-0.16
	(-0.01)	(0.72)	(0.59)	(-0.77)	(-1.08)
Year FE	yes	yes	yes	yes	yes
Exchange FE	yes	yes	yes	yes	yes
Adjusted R2	0.39	0.38	0.29	0.30	0.27
N	295	281	267	254	241

This table replicates the results of the regression from Table 2, but uses the variable *electronic* as the main independent variable and omits HFT. All variables are defined as in Section 2.2. We cluster standard errors by year. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 5: Price efficiency t-test

Stock exchange	μ_0	μ_1	$\mu_1 - \mu_0$	t statistic	p-value
All exchanges	0.10	0.01	-0.09	-5.88***	0.00
New York (NYSE)	0.23	0.14	-0.09	-3,01***	0.01
New York (NYSE Amex)	-0.05	-0.12	-0.07	-2,62**	0.02
New York (NASDAQ)	0.03	-0.01	-0.03	-2,70**	0.01
London	0.09	-0.05	-0.14	-3,28***	0.00
Frankfurt	0.06	0.02	-0.04	-0.71	0.49
Oslo	0.13	-0.04	-0.18	-3,07***	0.01
Tokyo	0.09	0.10	0.01	0.62	0.55
Sydney	0.00	-0.09	-0.08	-1,78*	0.09
Mumbai (NSE)	0.18	0.17	0.00	-0.05	0.96
Mumbai (BSE)	0.17	0.06	-0.11	-1.68	0.11

We use annual stock exchange level observations from 13 stock exchanges between 1990 and 2014. This table presents the results from a t-test comparing the means of price efficiency at the one year horizon when HFT is present to when it is not present. We do this for each stock exchange individually and pooling all exchanges. We cannot perform the test for China or South Korea as HFT has not yet started there. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively. we construct price efficiency similar to Bai et al. (2015). They estimate

$$\frac{E_{i,t+k}}{A_{i,t}} = a_t \ln\left(\frac{M_{i,t}}{A_{i,t}}\right) \times 1_t + b_t \ln\left(\frac{E_{i,t}}{A_{i,t}}\right) \times 1_t + c_{si,t,t}(1_{SIC1}) \times 1_t + \epsilon_{i,t}$$

where E is EBIT, A is total assets, $SIC1$ is the first digit of the SIC code, $k = 1, ...5$ and M is market value of equity. We use total market value for M instead. All ratios entering equation 1 are winsorized at the 1% level. With Predicted Variation$_t$ $= a_t \times \sigma_t[ln(M_{i,t}/A_{i,t})]$ and TotalVariation $= \sigma_t[E_{i,t+k}/A_{i,t}]$, we choose share of predicted variation, Price Efficieny k = PredictedVariation$_t$/TotalVariation, as main dependent variable for our analysis.

Table 6: Price efficiency: robustness checks

	(1) Price Efficiency 1	(2) Price Efficiency 2	(3) Price Efficiency 3	(4) Price Efficiency 4	(5) Price Efficiency 5
(1) Year and exchange FE, OLS	-0.091***	-0.10***	-0.15***	-0.10**	-0.016
	(-3.42)	(-3.07)	(-3.55)	(-2.37)	(-0.36)
(2) No FE, two-way cluster	-0.094***	-0.12***	-0.12***	-0.11***	-0.070**
	(-4.29)	(-6.02)	(-4.77)	(-5.66)	(-2.51)
(3) Year FE, cluster by exchange	-0.11***	-0.12***	-0.16***	-0.15***	-0.11**
	(-4.12)	(-4.54)	(-4.93)	(-3.95)	(-2.52)
(4) Year FE, two-way cluster	-0.11***	-0.12***	-0.16***	-0.15***	-0.11***
	(-4.57)	(-4.73)	(-4.56)	(-3.84)	(-2.97)

This Table presents the coefficients of HFT when we estimate four variations of the main empirical model as robustness checks. The four models differ with respect to the inclusion of fixed effects and clustering of standard errors. All variables are defined as in Section 2.2. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 7: Placebo test rejection rates (in %)

	Price Efficiency 1	Price Efficiency 2	Price Efficiency 3	Price Efficiency 4	Price Efficiency 5
1% level	1.0	0.5	0.7	0.5	0.5
5% level	4.9	3.6	4.3	3.3	3.9
10% level	9.3	7.4	8.7	7.4	8.3

This table reports the results from our placebo test. We generate random HFT starting dates and use them to construct a placebo for HFT presence. We then use this placebo as independent variable in our main models:

$$\mathrm{Eff}^{k}_{e,t} = \beta_0 + \beta_1 \mathrm{Placebo}_{e,t} + \delta X_{e,t} + \eta_t + \eta_e + \epsilon_{e,t}$$
$$\mathrm{IBES}_{i,t} = \beta_0 + \beta_1 \mathrm{Placebo}_{i,t} + \delta X_{i,t} + \eta_t + \eta_i + \epsilon_{i,t}$$

We repeat this 1,000 times for our main regressions 5 with price efficiency with time horizons one to five as dependent variables. We count how many coefficients of the randomly generated placebo are statistically significant at the 1, 5 and 10% level and report the resulting rejection rates of the null hypothesis that the coefficient is equal to zero in this table.

Table 8: Price efficiency DiD with further control variables

	(1) Price Efficiency 1	(2) Price Efficiency 2	(3) Price Efficiency 3	(4) Price Efficiency 4	(5) Price Efficiency 5
HFT (trade size)	-0.084***	-0.10**	-0.14***	-0.099**	-0.034
	(-3.31)	(-2.60)	(-3.56)	(-2.70)	(-0.75)
Electronic	-0.040*	-0.030	-0.046	-0.045	-0.038
	(-2.03)	(-0.93)	(-1.42)	(-1.34)	(-1.09)
ln (Market Size)	-0.00067	-0.017*	-0.023**	-0.0077	-0.0011
	(-0.07)	(-1.75)	(-2.61)	(-0.79)	(-0.07)
ln (Average Cap)	0.0025	0.030	0.017	-0.026	-0.044
	(0.14)	(1.03)	(0.51)	(-0.78)	(-1.38)
Amihud Market	-3.66**	-4.94***	-7.00***	-9.93***	-16.4***
	(-2.44)	(-2.93)	(-3.01)	(-3.48)	(-6.59)
Market Vola	1.25	-1.71	-0.14	-1.51	-4.20
	(0.67)	(-0.91)	(-0.07)	(-0.45)	(-1.71)
Constant	0.12	0.23	0.069	-0.18	-0.29
	(1.00)	(1.08)	(0.29)	(-0.77)	(-1.22)
Year FE	yes	yes	yes	yes	yes
Exchange FE	yes	yes	yes	yes	yes
Adjusted R2	0.48	0.48	0.44	0.39	0.32
N	264	251	238	226	214

This table replicates the regression from Table 2 and adds additional control variables. Market volatility is in thousands. Average Amihud illiquidity is in millions. We cluster standard errors by year. This table presents the results from adding further control variables to our main multi-event DiD model:

$$\text{Eff}^k_{e,t} = \beta_0 + \beta_1 \text{HFT}_{e,t} + \delta X_{e,t} + \eta_t + \eta_e + \epsilon_{e,t}$$

*, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 9: Analyst variables: main DiD model

	(1) Analyst Error	(2) Pos. Forecast Bias	(3) Neg. Forecast Bias	(4) Analyst Dispersion
HFT (trade size)	-8.53***	-4.86***	-0.54	-2.14***
	(-5.52)	(-5.13)	(-1.42)	(-3.51)
EBIT vola	-11.9	-7.57	-2.38	-6.43
	(-1.11)	(-1.08)	(-1.15)	(-1.10)
ln(MV)	-10.3***	-5.34***	-1.11***	-2.69***
	(-12.27)	(-9.55)	(-6.22)	(-10.97)
Tobin's Q	-1.37***	-0.71***	-0.27***	-0.68***
	(-4.77)	(-3.77)	(-4.08)	(-6.59)
Electronic	15.9***	9.37***	1.39***	5.61***
	(9.56)	(8.46)	(3.72)	(8.56)
Constant	63.0***	33.8***	7.39***	19.4***
	(14.23)	(11.27)	(7.75)	(13.27)
Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Adjusted R2	0.57	0.42	0.40	0.64
N	77631	77631	77631	61655

This Table presents the results from a regression of analyst-based measures on HFT, a set of control variables and year and firm fixed effects:

$$\text{IBES}_{i,t} = \beta_0 + \beta_1 \text{HFT}_{i,t} + \delta X_{i,t} + \eta_t + \eta_i + \epsilon_{i,t}$$

We use annual firm level observations from 13 stock exchanges between 1990 and 2014. All values are real and in USD when applicable. The five year rolling volatility of EBIT is in millions. IBES stands for analyst dispersion and analyst error. All variables are defined as in section 2.2. We cluster standard errors by firm. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 10: Analyst variable DiD: HFT present and shifted

	(1) Analyst Error	(2) Analyst Dispersion	(3) Analyst Error	(4) Analyst Dispersion	(5) Analyst Error	(6) Analyst Dispersion
Placebo +2y	0.35	0.57				
	(0.14)	(0.66)				
Placebo +3y			1.15	1.22**		
			(0.70)	(2.14)		
Placebo +4y					2.51*	1.48***
					(1.80)	(3.12)
EBIT vola	-1.36	-0.38	-1.58	-0.71	-1.73	-0.68
	(-0.08)	(-0.05)	(-0.09)	(-0.10)	(-0.10)	(-0.10)
ln(MV)	-8.88***	-2.01***	-8.87***	-1.99***	-8.82***	-1.97***
	(-13.72)	(-9.15)	(-13.69)	(-9.07)	(-13.62)	(-8.96)
Tobin's Q	-0.30	-0.21**	-0.30	-0.21**	-0.30	-0.21**
	(-1.07)	(-2.22)	(-1.07)	(-2.23)	(-1.05)	(-2.20)
Electronic	0.72	0.096	-0.085	-0.61	-1.16	-0.73
	(0.28)	(0.11)	(-0.04)	(-0.81)	(-0.57)	(-1.10)
Constant	69.5***	17.9***	69.5***	17.9***	69.6***	17.8***
	(17.02)	(12.33)	(17.02)	(12.32)	(17.03)	(12.30)
Year FE	yes	yes	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes	yes	yes
Adjusted R2	0.67	0.71	0.67	0.71	0.67	0.71
N	35983	29644	35983	29644	35983	29644

Variables, data, t statistics and significance stars are as in table 9. This table presents the results of the main multi-event DiD model when we shift HFT by x years and drop observations with HFT = 0. We cluster standard errors by firm.

$$\text{IBES}_{i,t} = \beta_0 + \beta_1(\text{Placebo +xy}_{i,t}) + \delta X_{i,t} + \eta_t + \eta_i + \epsilon_{i,t}$$

*, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 11: Analyst variables: alternative HFT definition

	(1) Analyst Error	(2) Analyst Dispersion	(3) Analyst Error	(4) Analyst Dispersion
HFT (order cancellation)	-9.09***	-2.60***		
	(-5.50)	(-4.26)		
HFT (colocation)			-3.69**	-2.15***
			(-2.08)	(-3.03)
EBIT vola	-12.1	-6.59	-10.4	-6.50
	(-1.13)	(-1.12)	(-1.00)	(-1.10)
ln(MV)	-10.5***	-2.75***	-10.6***	-2.79***
	(-12.43)	(-11.16)	(-12.55)	(-11.30)
Tobin's Q	-1.31***	-0.66***	-1.25***	-0.65***
	(-4.54)	(-6.38)	(-4.38)	(-6.34)
Electronic	16.1***	5.74***	16.3***	5.76***
	(9.58)	(8.63)	(9.85)	(8.92)
Constant	63.7***	19.7***	64.5***	19.8***
	(14.38)	(13.45)	(14.61)	(13.58)
Year FE	yes	yes	yes	yes
Firm FE	yes	yes	yes	yes
Adjusted R2	0.57	0.64	0.57	0.64
N	77953	61933	79107	62772

This table replicates the regressions in columns (1) and (4) of Table 9 and uses alternative definitions of HFT presence. Columns (1) and (2) use the definition based on order cancellation, while columns (3) and (4) use the definition based on colocation dates. All variables are defined as in section 2.2. We cluster standard errors by firm. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.

Table 12: Analyst variables: robustness checks

	(1) Analyst Error	(2) Analyst Dispersion
(1) Year and firm FE, no cluster	-8.53***	-2.14***
	(-8.00)	(-5.41)
(2) Two-way cluster by year and firm	-15.8***	-8.36***
	(-3.86)	(-4.64)
(3) Year FE, cluster by firm	-18.8***	-5.64***
	(-18.40)	(-15.11)
(4) Year FE, two-way cluster	-18.8***	-5.64***
	(-3.81)	(-3.16)

This table presents the coefficients of HFT when we estimate three variations of the main empirical model as robustness checks. The three models differ with respect to the inclusion of fixed effects and clustering of standard errors. All variables are defined as in section 2. T statistics are in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.