High Frequency Traders and the Price Process

by

Yacine Aït-Sahalia and Celso Brunetti

This version: November 29, 2019

Original: October 23, 2017

OCE Staff Papers and Reports, Number 2020-003

Office of the Chief Economist
Commodity Futures Trading Commission

High Frequency Traders and the Price Process[*][†]

Yacine Aït-Sahalia[‡]
Department of Economics
Bendheim Center for Finance
Princeton University
and NBER

Celso Brunetti[§]
Division of Research and Statistics
Board of Governors
of the Federal Reserve System

This version: November 29, 2019

Abstract

Using a dataset that uniquely identifies counterparties to each S&P500 eMini transaction, we classify each market participant as high or low frequency, and each transaction, by the speed of the traders involved. We investigate empirically the comparative influence of high and low frequency traders on the price process, and conversely the influence of the price process on the trading of high and low frequency traders. We find that high frequency traders have a particularly high success rate on each transaction, measured by the likelihood that the following price change will go in their direction as well as by the amount of time they have to wait to realize their gain, when trading against low frequency traders. Contrary to common wisdom, we find that high frequency traders' activity does not induce volatility or jumps. In fact, it is their absence that is problematic: volatility and jumps are more prevalent in periods when they trade less intensely. Conversely, we find that spikes in volatility and jumps cause high frequency traders to trade less intensely, decreasing their provision of liquidity. Finally, looking at the market microstructure noise component to the price model, we find that higher level of noise generates trading opportunities for high frequency traders and lead them to increase their trading activity.

Keywords: High frequency trading, liquidity, market impact, price process, volatility, jumps, market microstructure noise.
JEL Classification: G10, G12, G14.

---

[*]We are grateful to Bobak Moallemi for excellent research assistance. We are also grateful to two anonymous referees and the Editor for very helpful comments and suggestions.

[†]**Disclosure Statement:** The research presented in this paper was co-authored by Yacine Aït-Sahalia and Celso Brunetti in their official capacities as consultants within the Office of the Chief Economist of the Commodity Futures Trading Commission (CFTC). The Office of the Chief Economist and CFTC economists produce original research on a broad range of topics relevant to the CFTC's mandate to regulate commodity futures markets, commodity options markets, and the expanded mandate to regulate the swaps markets pursuant to the Dodd-Frank Wall Street Reform and Consumer Protection Act. These papers are often presented at conferences and many of these papers are later published by peer review and other scholarly outlets. The analyses and conclusions expressed in this paper are those of the authors and do not reflect the views of other members of the Office of Chief Economist, other Commission staff, or the Commission itself. The paper was reviewed by the Office of the General Counsel of the CFTC to ensure that no confidential information pertaining to individual trading accounts could be inferred from the results contained in this paper. Furthermore, the views in this paper should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. All errors and omissions, if any, are the authors' sole responsibility.

[‡]Email: yacine@princeton.edu
[§]Email: celso.brunetti@frb.gov

1 Introduction

Since the creation of NASDAQ in 1971 as the first electronic market, a consistent trend in the marketplace has been the increased role played by computers in the trading process (see, e.g., Litzenberger et al. (2012) for a history). As late as the mid 1990s, NYSE trading was for the most part taking place manually on its floor. While NASDAQ posted and matched orders electronically, its orders predominantly came by manual keyboard entry. In the late 1990s, the NYSE and NASDAQ saw their duopoly progressively eroded by fully electronic trading networks. As fully electronic means became available to access markets, technological developments made possible the implementation of high frequency trading algorithms, which analyze market data such as short term momentum or mean reversion, historical correlations with recent trades within or across markets, order book imbalance, and the predicted price response to electronically interpreted news. Depending upon the types of strategies followed, the algorithms make a strategic choice between market and limit orders, may decide to split them across time and/or trading venues, then submit or cancel orders, all without direct human intervention and all within a few milliseconds. Regulatory changes during the 2000s, including decimalization in 2001 and Reg NMS in 2007, were designed to increase competition among exchanges; they proved to be catalysts for the development of high frequency trading. Competition among exchanges indeed increased, driven in part by the desire to serve and benefit from the presence of high frequency traders (HFTs), who quickly became the largest customers of these for-profit trading venues. HFTs thrive on low latency, vying for computer locations colocated with the exchanges' servers, and timely access to information. The competition among exchanges took the form of increasing the speed at which they operate, and of providing other benefits afforded exclusively to HFTs, such as flashing orders to HFTs before sending them to the public market, providing them private data feeds, various colocation services, etc. Even if these benefits are in theory available to everyone willing to pay, they are in reality of value only to a small number of firms with the technology and inclination to exploit them in real time. Various estimates indicate that over half of all trading volume in US equity and futures markets is attributable to HFTs.[1]

This paper provides an empirical investigation, using a unique dataset that identifies the traders in each transaction, of the comparative influence of high and low frequency traders on the asset price process, and conversely of the influence of the price process on the trading of high and low frequency traders. The paper contributes to the existing literature in three different ways. First the data we use capture all transactions in the E-mini S&P500 futures contract. A large part of the empirical

[1] Estimates ranging from 50 to 70% are provided by Biais and Woolley (2011) and Sussman et al. (2009).

literature on high-frequency traders analyzes transactions (and/or quotes) referring to a specific venue–e.g. NASDAQ, see Brogaard et al. (2014b). It is reasonable to assume that HFTs may trade the same asset on more than one market and hence it might be difficult to generalize the behavior of HFTs based on the analysis of a single market only.[2] A related point is that most of the contributions mainly concentrate their analysis on the provision and supply of liquidity by HFTs. While we also look at this aspect of HFTs activity in our data, our main focus is on which side of the market HFTs are in terms of buying or selling. While the provision of liquidity may pertain more to the functioning of the market, our goal is to study the price process itself. In fact, our second contribution is to analyze all components of the price process. Finally, one appealing feature of our analysis is that it is completely nonparametric, or model-free. None of our conclusions are influenced by a prior hypothesis or a model.

A semimartingale model for asset prices consists of three components: drift, volatility, and jumps. We analyze which change(s) in the components of the price process lead different types of traders to act, and conversely analyze the influence of these traders' propensity to trade on these three components. Regarding the drift, we find that HFTs are not directional traders. However, they enjoy a high success rate on each transaction, measured both by the likelihood that the next price change following their transaction will go in their direction (e.g., an uptick following a buy) as well as by the amount of time they have to wait to realize their gain (shorter).[3] In particular, we document that HFTs do not win just 51% of the time and rely simply on the law of large numbers. Their winning percentage is substantially higher when they trade against low frequency traders (LFTs) but not when they trade between one another. They also display an uncanny ability to avoid each other as counterparties despite the anonymity of orders. Contrary to popular perception, we do not find that the presence of HFTs induce volatility or price jumps. In fact, we find the opposite: high volatility and jumps tend to follow periods where HFTs' share of trading decreases.[4] In other words, it is their absence that is problematic: HFTs induce volatility and price jumps when they choose to be less present in the market. Conversely, we study whether increases in volatility, and most importantly, jumps, cause HFTs to withdraw. There, we find that spikes in volatility and jumps cause HFTs to trade less intensely, which is self-reinforcing as the decreased provision of liquidity leads to more volatility. This specific question has important consequences for any potential regulation of HFTs, particularly as they have largely replaced the NYSE specialists and NASDAQ market makers of yesteryear: in exchange for the benefits they enjoy, should high frequency market makers be required to maintain a "fair and

[2] A similar dataset has been previously adopted by Kirilenko et al. (2017). Their analysis concentrates on the May 6, 2010 flash crash. Our analysis is much broader in scope and does not concentrate on a single, yet important, event.

[3] Similar results for NASDAQ are reported in Brogaard et al. (2014b). Our approach is more granular, as it will be explained below.

[4] Chaboud et al. (2014) and Hasbrouck and Saar (2013) find weak evidence that HFTs activity reduces volatility.

orderly market" and act in the interests of the marketplace at large in case of a disruption, even at a cost? This debate has taken a renewed importance following the "flash crash" of May 6, 2010, and the evidence uncovered in the aftermath of very many, although less striking and less reported, mini-crashes in numerous markets. When we add a market microstructure noise component to the price model in the form of an additive error, the noise generates short term autocorrelation of returns. We find that higher levels of noise are followed by higher trading activity by HFTs, consistent with the noise generating further trading opportunities; conversely, higher HFT activity does not lead to a change in the level of the noise.

The paper is organized as follows. Section 2 provides a brief review of the rapidly evolving literature on high frequency trading. Section 3 describes the data we use in the paper and the process by which we identify the HFTs on the basis of their frequency of trading, number of trades, and (lack of) carried inventory. Section 4 studies how HFTs trade as a function of the characteristics of the price process, compared to LFTs. Section 5 studies the reverse causality: we examine how the price process responds to trades that involve HFTs as opposed to LFTs. Section 6 concludes.

2 Literature Review

HFTs are at the center of many policy debates and controversies. As the latest type of market makers to enter the fray, are they benevolent providers of the liquidity and price discovery that the marketplace needs, or are they potentially destabilizing the market, increasing systemic risk, and creating a non-level playing field?

The empirical literature generally supports the view that HFTs play a role that is, on balance, beneficial for market quality, measured using standard metrics involving combinations of bid-ask spreads, liquidity, and transitory price impacts. Hendershott et al. (2011) find that HFTs improve liquidity and enhance the informativeness of quotes, using the automation of quote dissemination by the NYSE in 2003 as an exogenous change in market structure. They find that for large stocks in particular, HFTs narrow spreads, reduce adverse selection, and reduce trade-related price discovery. Biais et al. (2016), using data from Euronext and the French financial markets regulator (AMF), find that HFTs provide liquidity by leaving limit orders in the book thus helping the market absorb shocks. Hasbrouck and Saar (2013), using order-level NASDAQ data, show that increased HFTs' activity is associated with lower posted and effective spreads, increased depth, and lower short-term volatility. They nevertheless show that HFTs exhibit high turnover and high rates of order cancellation ("fleeting orders") relative to actual trade execution. Jovanovic and Menkveld (2010) show that the entry of HFTs in a market

reduces bid-ask spreads. Menkveld (2013) provides evidence that HFTs usefully act as market makers, particularly in new markets. Chaboud et al. (2014) establish that an increase in algorithmic trading is associated with a decrease in volatility levels in the foreign exchange market. Similarly, Hagströmer and Nordèn (2013) provide evidence that HFTs trading activity mitigates price volatility for stocks traded at the NASDAQ OMX Stockholm exchange.

Our results also indicate that HFTs trading activity does not increase volatility levels. Our approach, though, is different. Most of the literature looks at contemporaneous relationship between volatility and trading activity within a regression framework. Instead, we condition on past information. In particular, we condition trading activity on past volatility levels and, vice versa, volatility on past trading activity. Our approach allows us to study how HFTs respond to a period of high (low) volatility as well as how volatility behaves following periods of high (low) trading activity of HFTs.

Brogaard et al. (2014b) find that HFTs enhance price discovery and market efficiency on NASDAQ, with prices reflecting information more quickly. Decomposing price movements into permanent (interpreted as information-based) and temporary (interpreted as microstructure or pricing errors-based) components, they find that HFTs trade in the direction of permanent price changes and in the opposite direction of transitory pricing errors. We complement these results by documenting that HFTs trade more intensely when market microstructure noise is high than when the noise is lower. Brogaard et al. (2016) use Canadian regulatory data to study the contribution of HFTs' activity (trades and limit orders) to price discovery and find that HFTs are responsible for 60-80 percent of price discovery, mainly through their limit orders.[5]

Of course, these market quality statistics are typically assessed over relatively long horizons, of months or years, and as a result, are not designed to account for the temporary dislocations that appear in markets in the form of transitory mini-crashes that may last a few seconds: these are smoothed out over time and will not be reflected in longer-run measures. Recently Brogaard et al. (2018) analyze the provision of liquidity during extreme price movements. They find that HFTs provide liquidity during extreme events but only when these extreme events refer to single stocks. We do not examine specific episodes of extreme price movements and we adopt a broader definition of jumps: any price change exceeding three times the standard deviation. We find that HFTs shy away when there are jumps.

Other evidence suggests that HFTs can play a less benevolent role in the marketplace. Huh (2016)

[5] Also on NASDAQ, Brogaard (2011) finds that HFTs tend to trade more when systematic (index) volatility is higher but less when idiosyncratic (stock-specific) volatility is higher. Saliba (2019) uses AMF regulatory data and identifies three categories of market participants in the order flow: HFTs, agency and proprietary participants. She finds that HFTs aggressive orders contain more information than those of agency and proprietary participants.

distinguishes between liquidity-taking HFTs (those submitting market orders) and liquidity-providing HFTs (those submitting limit orders), and shows that the information asymmetry induced by liquidity-taking HFTs' use of machine-readable information reduces the supply of liquidity by HFTs. This effect is particularly strong when markets are volatile.[6] Brogaard (2011) and Hirschey (2011) find that HFTs on NASDAQ tend to predict future order flow. To the extent that the informational advantage of HFTs often takes the form of advance knowledge of order book imbalances or other traders' future actions, the case for benign and useful price discovery role is weaker (see Jarrow and Protter (2012) for a theoretical argument).

The ability to place large amounts of order, and cancel them before slower traders can take advantage of them, is an inherent part of most HFTs' strategies. Some strategies followed by HFTs may come close to market manipulation (see Biais and Woolley (2011)). Some HFTs "stuff" the order book by submitting a very large number of orders which they have no intention of executing, but have the effect of limiting for a short time the access to the market for other, slower, traders. "Smoking" consists in posting attractive orders to attract slow traders, orders that the HFT has no intention of executing. The HFT then rapidly cancels these orders, leaving in place only those to be executed profitably against the incoming flow of slow traders' market orders who reacted to the initial HFTs' inducement. "Spoofing" involves an HFT disguising its trading intention by stuffing the order book on the side opposite to its true trading direction, not at the best price so that they do not get executed against market orders. Slower traders will then react to this imbalance by hoping to get ahead of the HFT-induced imbalance, providing liquidity for the HFT's desired trade. Cartea et al. (2019) model trading strategies of an investor spoofing the limit order book. Of course, it is possible for a high rate of cancellation on the part of HFTs to simply reflect "market making," that is, providing liquidity to other traders by quoting two-sided prices, but rapidly revising their quotes as market conditions dictate to avoid being executed against in unfavorable circumstances. Hens et al. (2018) study front-running by HFTs and find that this strategy does not affect market quality.[7]

HFTs typically generate profits out of a large number of small size, small gain trades, all without accumulating any significant inventory. As a result of the small trade sizes they often have little

[6] Zhang (2010), using quarterly data, finds that HFTs activity is positively correlated with stock price volatility and negatively linked to market ability to incorporate firm's fundamental news into asset prices.

[7] Aït-Sahalia and Sağlam (2016b) and Aït-Sahalia and Sağlam (2016a) propose a model where liquidity is provided by HFTs who are both faster and better informed than their counterparties. Their model predicts that, left to their own devices, market makers should be expected to provide more liquidity as they get faster, but then shy away from it as volatility increases. In fact, Raman et al. (2014), for crude oil futures, and Brogaard et al. (2018), for NASDAQ stocks, document that provision of liquidity by HFTs diminishes when volatility is high. Breckenfelder (2019) finds that increasing competition among HFTs increases speculative trading by HFTs and, as consequence, market liquidity deteriorates and volatility increases. Menkveld and Zoican (2017) and Budish et al. (2015) also find that an increase in HFTs speculative trading deteriorates market quality.

price impact, and their small inventories mean that their trading risk is relatively limited (see Baron et al. (2019)). On the other hand, their strategies are often correlated among themselves (see Chaboud et al. (2014)). Operational risk is a concern, including software errors, rogue trades, or poorly executed algorithms: for example, Knight Capital in August 2012 and the Flash Crash of May 2010, which was a $4 billion sell order sent without a price limit. Another concern would be a liquidity or other common shocks affecting many HFTs, which are typically all proprietary trading firms relying on low capital levels. The large quantity of transactions involving HFTs, their large number of counterparties in futures markets, with different settlement and clearing, etc., could potentially trigger a cascade of defaults, leading to systemic problems not unlike those experienced without HFTs' involvement in the Fall of 2008.[8]

Many empirical studies of HFTs relied on the "NASDAQ HFT dataset" which covered 120 US stocks over 2008-2009 and identified 26 HFT firms as proprietary traders with colocated facilities, high turnover, and high rates of cancellations relative to executions. The profitability of HFTs taken collectively has been estimated by Brogaard (2011) at about $3 billion (gross of any operational expenses) over those two years. Using similar data as this paper, Baron et al. (2012) estimate the gross profitability of HFT firms in one E-mini S&P 500 futures contract during one month at about $30 million. Various evidence points towards a recent decline of HFT profitability in 2012,[9] due in part to HFTs becoming victims of their own success, the increased competition from other HFTs, the costs of the technological arms race, and declining trading volume after a peak in 2009.

Most empirical studies in the literature, with the notable exception of Easley et al. (2011), Kirilenko et al. (2017), and Brogaard et al. (2018) characterize the behavior of HFTs in typical times and market environments. In this paper, we do not focus on specific episodes of market crises, but rather decompose the asset price process into volatility and jump components, as is natural in the context of semimartingale price models (see, e.g., Aït-Sahalia and Jacod (2012)). We analyze HFTs' trading behavior conditional on small (volatility) and large (jump) price movements, and in the other direction analyze their impact on price volatility and jumps. The latter results address the question of the potential for HFTs to destabilize markets in which they operate and generate abnormal volatility and jumps. To this end, we consider two distinct samples: one characterized by low volatility levels and the other by high volatility levels. Moreover, our model free approach provides a unique perspective

[8]Other related contributions include Brogaard et al. (2015), who investigate how speed affects market liquidity, Brogaard et al. (2014a), who study the impact of HFTs on execution costs of institutional traders, and Brogaard et al. (2017), who analyze the effect of the short sale-ban of September 2008 on HFTs provision of liquidity. Li et al. (2019) model liquidity provision by HFTs and slower algo traders. For an excellent review of the theoretical as well as empirical literature on HFTs see Menkveld (2016).

[9]See "Citadel Securities Cuts Almost 10% of Employees After Review", Bloomberg News, March 6, 2013, and "Getco Discloses 90% 2012 Profit Decline As Knight Deal Tweaked", The Wall Street Journal, April 15, 2013.

on how HFTs affect the various components of the price process and how the price process influences the behavior of HFTs.

3 The Data

We analyze audit-trail[10], transaction level data that took place during the months of August 2010 in the September 2010 E-mini S&P 500 futures contract and August 2011 in the September 2011 E-mini S&P 500 futures contract (ticker ES). The E-mini S&P 500 futures contract is a fully electronic, cash-settled contract traded on the Chicago Mercantile Exchange's GLOBEX platform.[11] Initially designed to appeal to small investors, this contract was introduced in September 1997 but soon became the most actively traded stock index futures contract in the world and the de facto reference contract for the S&P 500 index.[12]

Contracts mature every quarter (March, June, September, and December) and cease trading on the third Friday of the expiration month.[13] Liquidity typically concentrates on the front month, or nearby, contract. In August, the front month is the September contract, which is the contract month we analyze in what follows. We concentrate only on two months of data to avoid the possible distortions induced by the expiration of the contract and by any roll-over strategies to move positions to the next contract month. The data are collected by the CME and provided to us by the Commodity Futures Trading Commission (CFTC) as part of its market surveillance program.

Each trade record contains a time stamp, a transaction type, the price, and the volume. The records also contain detailed information about counterparts' IDs and traders' IDs. We can therefore follow a trading account, using its unique identifier, throughout the sample period, and use counterparts' IDs and traders' IDs to uniquely identify counterparties. In addition, each trade record contains an aggressor flag which indicates which of the two parties to the transaction submitted a market order and took away liquidity from the electronic limit order book. The data is sequentially ordered, which provided the sequencing of transactions when several trades have the same time stamp. While the

[10]The data is reported to the CFTC as part of its regulatory mandate. The dataset is proprietary. Steps described below have been taken to ensure the anonymity of the individual trading accounts.

[11]See http://www.cmegroup.com/trading/equity-index/us-index/e-mini-sandp500_learn_more.html.

[12]In fact, Hasbrouck (2003) shows that E-mini exhibits the largest information share, accounting for 90% of the price discovery of the S&P 500 index. There are three main reasons for the success of the E-mini S&P 500. First, the small size (equal to a fifth of the regular S&P 500 futures contract) makes this contract more accessible to traders and investors who would like to gain exposure to large-cap U.S. stocks. Second, the E-mini contract is traded on an open electronic limit order book which is accessible by off-floor traders. Third, the speed of execution of the GLOBEX platform enhances price discovery and attracts HFTs, including statistical arbitrageurs who may be contemporaneously trading the spot underlying assets in New York in the form of S&P500 Depository Receipts (SPDRs).

[13]Several papers analyze the E-mini S&P 500 index, including Easley et al. (2012), Baron et al. (2012), and Adamic et al. (2017).

market is trading on a 24-hour basis with only short interruptions,[14] we analyze only transactions that occurred between 8:30 a.m. and 3:15 p.m. Central Standard Time. This is the time when the stocks in the underlying S&P 500 index are most actively traded in New York and elsewhere, and when liquidity concentrates as a result.[15] Overall, we analyze 8.1 and 15.5 million E-mini regular transactions in August 2010 and August 2011, respectively.[16] As a result of the high degree of trading activity evident in this market, it makes sense to sample time in different ways and check the results for robustness. We will sample in calendar time (n milliseconds), tick time (the amount of time it takes for the price to move by n ticks), volume time (the amount of time it takes for n contracts to be exchanged), and transaction time (the amount of time it takes for n transactions to be recorded). Table 1 reports the correspondence in the data between these different sampling mechanisms.

Figure 1 shows the closing price and the two-scale realized volatility (an estimator of the annualized standard deviation) for the data. While August 2010 is a relatively quiet month with low volatility, August 2011 is characterized by high volatility levels. In particular, August 4, 5, 8, and 9, 2011 exhibit extremely high variability following the S&P downgrade of the US credit rating, reinforced by negative news concerning the European sovereign debt crisis. The DJIA went up and down by over 400 points four days in a row, something never experienced before. This contrast between the two months allows us to look for, among other things, any differences in the behavior of high frequency traders in different market conditions. Volatility signature plots are reported in Figure 2 in transaction, volume, tick, and calendar time; they are fairly typical of high frequency data. Volatility at each sampling frequency is computed as the average over the month of the daily volatilities, and standard errors are the sampling standard deviations over that month. The top panel reports the volatility signature plot in transaction time, the second panel in volume time, the third in tick time, and the fourth in calendar time. The figure adjusts the range of number of contracts and number of ticks so that they are similar to the range of time intervals we used in calendar time, from 100ms to 1mn for August 2011, and from 1s to 1mn for August 2010.

[14] Trading hours (in Central Time) are 5:00 p.m.-3:15 p.m. & 3:30 p.m.-4:30 p.m. Monday through Thursday, and 5:00 p.m.-3:15 p.m. on Sundays.

[15] The results of the analysis that follow are largely consistent when computed on the two alternative trading periods consisting of 8:00 a.m.-3:15 p.m. and the full trading day, respectively.

[16] Not only is the size of each contract smaller than regular S&P 500 futures, but the median number of contract(s) exchanged in each transaction is 1. This was of course the idea behind the creation of the S&P 500 E-mini market — to allow small, infrequent traders access to S&P 500 futures. But this feature of the market is also beneficial to HFTs who need to be able to execute a large number of small trades to capture whatever tiny empirical advantage their algorithms target, relying on the law of large numbers for success. It is therefore not surprising that the E-mini has become one of their favorite markets to operate in. Consistent with this, there is a tendency in other markets for the volume exchanged on each transaction to go down: for example, Angel et al. (2015) document that the average trade size has been decreasing in US equity markets in recent years, down to about 300 shares in 2009, but it is still a far cry from the median trade size of 1 contract in our data.

3.1 Identifying the Cluster of High Frequency Traders

Our first objective is to classify traders into two groups, high frequency traders (HFTs) and low frequency traders (LFTs). We expect HFTs to share specific characteristics: they generate a large number of transactions, executed at a very high speed, and carry low inventory levels.[17] Hence, the features we employ to single out the HFTs are the number of trades and trade volume, as measures of trading activity (high), inter-trade duration (low), as a measure of the speed of execution, and inventory (low relative to volume traded). The unique characteristic of the data that makes this analysis possible is the fact that it contains unique account identifiers, so we can track all the trading done by a particular account over time. Of course, we do not report any results in the paper that would allow a specific firm's trading strategy to be reverse-engineered. All the results we report are aggregated and we have taken further steps to anonymize the presentation of the results.

To separate HFTs from LFTs in what is effectively a massive amount of data, we employ clustering analysis. Clustering analysis refers to the organization of points or patterns into clusters based on similar behavior: points or patterns within a cluster are more similar to each other than points or patterns belonging to a different cluster. The idea is to select HFTs as a specific cluster within the set of all traders, a cluster that becomes apparent, and fairly consistent over time, once we look in the four dimensions described above. Indeed, we find that a small subset of traders look markedly different from the others on these dimensions. We proceed as follows. For each trading account, we compute the following measures of trading activity: the number of transactions, the number of contracts traded, inventory, and (average) trade duration. We compute these measures at different sampling frequencies: 50,000; 100,000; 200,000 transactions and the full day. The number of transactions and the number of contracts traded (trading volume) measure account-specific activity related to the specific decision horizon and the execution strategy of different traders. We expect HFTs to be characterized by a large number of both transactions and volume. Inventory measures the net position of each account over the relevant sample. High frequency traders should have a tiny net inventory position relative to their volume traded. Trade duration for a given account is defined as the time between trades for that account. The data is time stamped to the second in 2010 and to the millisecond in 2011. Often, for several accounts in our data, the median trade duration is equal to 0. Therefore, we adopt the arithmetic mean of trade duration, which does vary with traders, and we expect highly active, low latency traders to have lower mean duration. Taken together, the four measures considered appraise

[17] The Sub-Committee on Automated and High Frequency Trading, part of the CFTC Technical Advisory Committee, provides the following definition: *High frequency trading is a form of automated trading that employs: a) algorithms for decision making (without human intervention); b) low-latency technology; c) high speed connections to markets; and d) high message rates.*

different aspects of how much, in which direction, and how quickly each trader transacts. We compute these measures of trading activity and cluster all variables each day and each month in the sample.

Clustering analysis then formalizes statistically the separation of the full universe of traders into different groups. It employs a pattern proximity measure which quantifies the distance between pairs of patterns and then groups the data into clusters. There are two main approaches in grouping. The first approach is referred to as hierarchical and implies either that patterns are merged (agglomerative clustering) or split (divisive clustering) based on the proximity measure to form a hierarchy of clusters. The second approach is partitional clustering which is based on the identification of a partition of patterns that optimizes some clustering criterion. Figure 3 show the results of the clustering analysis for August 2010 and 2011.[18] The different clustering methods produce consistent results, which are additionally consistent with what a simple visual partitioning of the data would suggest. We consistently identify a small number of traders who persistently stand out in the data as those having the highest account-specific daily trading volume, number of trades, lowest average trade duration, and lowest inventory: these same traders are identified as HFTs by all the clustering algorithms we employ. For confidentiality reasons, namely to avoid any potential reverse-engineering by an HFT of one of its competitors' trading patterns or strategies, we are only stating that the number of traders identified as HFTs is between 5 and 15, and only report aggregate statistics on HFTs. By contrast, there are over $30,000$ and $35,000$ LFTs in the same sample in 2010 and 2011, respectively.

In the two-dimensional plots, frequency (inverse of duration) is graphed against the average number of trades, daily volume, and daily inventory (inverse of inventory as a percentage of total daily volume) respectively; in the 3d plot, we include frequency (inverse of duration), number of trades, and inventory (inverse of inventory as percentage of total daily volume). Accounts we classify as LFTs are indicated as belonging in the blue shaded area, while the HFTs are indicated as belonging in the red shaded area; again, we do not report the positions of HFTs and LFTs individually, only the region in which they collectively lie in the graph. Accounts with fewer than 20 transactions per month were dismissed in the clustering analysis. We obviously classify these accounts as LFTs. To avoid cluttering the graphs to the point of illegibility, the plots only include LFTs with more than 1,000 transactions.

Most but not all accounts we identify as HFTs overlap between the two months in the sample, indicating some limited degree of entry and exit of HFTs in this market, or perhaps some variation

[18] Computations are done in Matlab. For robustness reasons, we adopt several clustering algorithms: the agglomerative hierarchical clustering which begins with each pattern in its own distinct cluster and clusters are then successively merged; k-means partitional clustering in which patterns are assigned to the cluster with the nearest mean (high proximity); Gaussian mixture partitional clustering in which patterns are assumed to be drawn from one of several distribution and patterns are assigned to the cluster with the nearest distribution; and the Plaid model introduced by Lazzeroni and Owen (2002) in biology to cluster genes by looking for patterns that deviate from average profiles.

in the trading strategies they follow over time. Not surprisingly, by employing a combination of criteria, our identification mechanism for HFTs is more conservative than what any single criterion would identify. For example, there are over 30 accounts that trade, on average, more than 5,000 times per day in August 2010 and August 2011, respectively. But some of these accounts do accumulate inventories and can be better classified as directional or fundamental traders. For instance, Figure 4 displays the clusters for August 2010 based on the two measures of trading frequency and inventory: we can distinguish among HFTs (again only indicated as a red shaded area instead of individually), LFTs who are fundamental buyers accumulating long positions, LFTs who are fundamental sellers accumulating short positions (both likely institutional investors), and the residual mass of LFTs who trade relatively low volumes and do not accumulate much inventory. In the analysis that follows, we do not distinguish among these different types of LFTs.

Table 2 reports the summary statistics for HFTs vs. LFTs and shows the stark differences between the two groups in terms of the four criteria we employed: HFTs are responsible for more than 100 times the number of trades of LFTs and 300 times their volume; HFTs have average intratrade durations below 1 second compared to about 10 minutes for LFTs and carry inventory that is about 10 times smaller than that of LFTs as a percentage of their trading volume. The "aggressor flag" in the data allows us to determine the identity of the traders who are liquidity-demanding vs. liquidity-supplying. According to the CME, the aggressor flag indicates whether the order that generates the trade is incoming or resting on the limit order book. Figure 5 shows that LFTs are slightly more likely than HFTs to prove liquidity over time.

3.2 Classifying Transactions by the Type of Trader Involved: HH, HL, LH and LL

We now exploit the unique characteristics of the dataset to classify each transaction. Each transaction in the data contains markers that identify the two counterparties (trading accounts) to the transaction. Thus, given our classification of each trading account as HFT or LFT, we proceed to classify each transaction as belonging to one of four categories: "HH" (a HFT buying from another HFT), "HL" (a HFT buying from a LFT), "LH" (a LFT buying from a HFT), or "LL" (a LFT buying from another LFT). These being futures contracts, there is no real "buyer" and "seller". We call "buyer" the trader entering into the long position in the contract, and "seller" the counterparty entering into the short position. Note that this distinction is separate from the "aggressor" determination, which refers to the party taking away the liquidity by being responsible for the market order that led to the transaction, and could be either on the long or short side of the transaction. Most of the literature

mainly concentrates on the provision and supply of liquidity by HFTs. While we also look at this aspect of HFTs activity, our main focus is on which side of the market HFTs are in terms of long and short positions.

Table 3 reports the summary statistics based on the four types of transactions. Many of the questions that will be analyzed in detail in following sections can be glanced at from a simple inspection of the descriptive statistics separated by type of transaction. In August 2010, the average trading activities are 61% for LL, 18% for both HL and LH, and 3% for HH. So of the 8.1 million transaction in August 2010 only 3% involve HFTs trading with HFTs, despite the fact that HFTs participate in about 40% of transactions. The majority of transactions are between LFTs. The percentage of trades between HFTs is 6% in August 2011. In terms of the immediate success rates of different transactions, we notice in Table 3 that after LFTs buy from HFTs, prices tend to drop. The opposite happens when LFTs sell to HFTs. The return standard deviation is highest when LFTs trade with each other (perhaps capturing an information effect). Trades between HFTs generate the lowest standard deviation (perhaps due to their careful avoidance of adverse selection). Statistics for absolute returns confirm that trades that involve LFTs exhibit the highest average volatility. Moreover, these trades also have the highest volatility of volatility. Absolute returns are persistent when LFTs are involved in a trade. This is not the case when HFTs trade with each other. In August 2010, the average trading volume is highest when HFTs trade with HFTs and lowest when LFTs trade with one another. In August 2011, the average trading volume is about 4 contracts for all types of transactions. The average duration and standard deviation of the duration are highest when HFTs trade with each other indicating that this type of trades does not occur very often.

4 How Different Traders Respond to Characteristics of the Price Process

We start with the premise that the price process X is an Itô semimartingale, which is the standard hypothesis in light of the fundamental theorem of asset pricing that restricts no-arbitrage prices to be semimartingales. That is,

$$X_t \;=\; X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \text{JUMPS} \tag{4.1}$$

$$\text{where JUMPS} = \underbrace{\int_0^t \int_{\{|x|\leq a\}} x(\mu-\nu)(ds,dx)}_{\text{small jumps}} + \underbrace{\int_0^t \int_{\{|x|>a\}} x\mu(ds,dx)}_{\text{big jumps}} \tag{4.2}$$

Here, b denotes the drift process of X, σ its volatility process, W a Brownian motion, μ is the jump measure of X, and its predictable compensator (or Lévy measure) is ν (see, e.g., Aït-Sahalia and Jacod (2014) for details on the model and its use in high frequency financial econometrics). The size cutoff a between small and big jumps is fixed.

In this Section, we examine how changes in $\{b, \sigma, \mu\}$ affect the actions of the different types of traders we have identified. We rely on nonparametric estimates of these quantities as conditioning variables. Specifically, we ask the following questions: Are HFTs vs. LFTs more or less likely to trade when the drift is positive or negative, when the volatility process is higher, when the jump measure is higher? Do the different market environments represented by different values of $\{b, \sigma, \mu\}$ predict which type(s) of transaction among HH, HL, LH and LL are more likely to follow? Further, if market microstructure noise is present, so that the observed price is $Y_t = X_t + \varepsilon_t$ rather than X_t, are HFTs or LFTs more or less likely to trade when the noise is high? For each of the same price environments as above, are different types of traders more or less likely to provide/take liquidity?

4.1 Drift: Response of Different Traders to Price Direction

We start by examining whether HFTs and LFTs have different trading behaviors in response to recent price movements. Are either type more likely to be momentum traders or contrarian traders? For this purpose, we introduce some notation. Define a given interval, a day, say, as $[\underline{\tau}, \overline{\tau}]$. Partition the interval in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. The length of the subinterval, $k = \tau_i - \tau_{i-1}$, is constant and τ may represent, transaction time, volume time, tick time or calendar time. We will often refer to the subintervals as "buckets." For each subinterval, $[\tau_{i-1}, \tau_i)$, we define trading activity of category $j = \{HH, HL, LH, LL\}$ as

$$TA_{j,i} = \frac{\sum_{\tau \in [\tau_{i-1}, i_i)} I_{j,\tau}}{\sum_j \sum_{\tau \in [\tau_{i-1}, \tau_i)} I_{j,\tau}} \tag{4.3}$$

where $I_{j,\tau} = 1$ for $j = j_s$ and zero otherwise. In other words, if $j = HH$, $I_{j,\tau}$ is equal to 1 for each trade between two HFTs in a given subinterval. The denominator is scaling the number of HH transactions by the total number of transactions in k.[19] The rate of return is computed as $R_i = Y_{\tau_i} - Y_{\tau_{i-1}}$. We examine how $TA_{j,i}$ vary conditionally on the return R_{i-1}. We partition our data in subintervals of

[19] In a similar way we define volume activity as

$$VA_{j,i} = \frac{\sum_{\tau \in [\tau_{i-1}, \tau_i)} I_{j,\tau} V_{j,\tau}}{\sum_j \sum_{\tau \in [\tau_{i-1}, \tau_i)} I_{j,\tau} V_{j,\tau}}$$

$k = 1,000$ trades, which is the size where the return R_{i-1} tends empirically to be most consistent within a bucket due to short run momentum. When k gets smaller, returns within that bucket cannot be reliably measured; if k gets too large, they become subject to mean reversion in prices and changes in returns within that bucket.

Figure 6[20] shows the results of this analysis. We employ a nonparametric kernel regression for this purpose, to avoid imposing any prior parametric form to the shape of the relationship, and compare the conditional expected value of $TA_{j,i}$ to their unconditional values. Horizontal lines on the plot indicate the unconditional average trading activity for the four types of transactions. We find that, on average, HFTs are neither trend followers nor trend chasers. In fact, they buy or sell to LFTs at about the same rate following positive or negative returns. The only slight difference that is apparent is that HFTs tend to sell more to LFTs following large negative returns than to buy from LFTs, but that difference is small. Rather consistently, however, HFTs tend to trade less, both among themselves and with LFTs, following large returns of either sign: the proportion of LL trading increases sharply at both the left and right end of the graph, away from the unconditional mean. This shows that HFTs tend to (temporarily) withdraw from the market following periods of large absolute returns. While there are undoubtedly some HFTs who follow directional strategies that attempt to capture mean reversion following transitory pricing errors or momentum exploiting trends, this does not appear to be a dominant characteristic in our data. Instead, HFTs seem to prefer trading in quiet periods following small returns, with no particular directional component (i.e., no drift). Brogaard et al. (2014b) using NASDAQ data find that HFTs' net positions are negatively correlated with past return, implying that HFTs follow contrarian strategies. While Brogaard et al. (2014b) compute HFTs' net positions, in our analysis we keep the distinction between long and short positions for each trade (or bucket of trades). Moreover, our results suggest that the size of the previous return matters–i.e., HFTs' behavior changes according to whether the price moves substantially or just a little.

4.2 Volatility and Jumps: Response of Different Traders to Price Variability

The previous results examined the propensity of HFTs to trade following periods of price drift. We now consider the relationship between trading activity by different types of traders conditional on past volatility and jumps. Figure 7 shows the results of this analysis. We again partition our sample in to subperiods (buckets) of k transactions and compute volatility within each bucket. Volatility is computed using the subsampled and averaged estimator RV_i^{avg} of Zhang et al. (2005) with subsampling

where $V_{j,i}$ represents the traded volume of each trade of category j in a given subinterval. All the analysis reported below has been conducted for both $TA_{j,i}$ and $VA_{j,i}$. Results are very similar and we only report results for $TA_{j,i}$.

[20]For some of the analysis, we omit results for August 2011 since they are very similar to those of August 2010.

grids of 100 transactions within each bucket. We set $k = 10,000$ so that we can have an accurate estimate of RV_i^{avg} for each bucket i.[21] The plot reports a kernel regression of $TA_{j,i}$ on volatility in bucket $i-1$, that is RV_{i-1}^{avg}, to be compared to the unconditional average trading activities marked by horizontal lines. We find that trading activity involving two LFTs, $TA_{LL,i}$, is increasing in volatility, while the transactions involving at least one HFT are all decreasing. We denote by σ the average volatility (RV^{avg}) across buckets for the entire month of August 2010 (or 2011). The plot indicates different integer levels of average volatility as vertical lines, and terms jumps buckets where $|R_{i-1}| \geq 3\sigma$. Observations are marked individually to avoid relying on the kernel regression for rare observations. By looking at the extremes of Figure 6, we already saw that HFTs tended to trade less following periods of large absolute returns. The pattern that emerges from Figure 7 is consistent with that finding: HFTs shy away from periods of price variability, especially larger ones involving jumps.

To measure the trading intensity of different traders following different levels of price jumps, we report in Table 4 the proportion of trades in bucket i following a price move of n ticks. In other words, we are computing $TA_{j,i}$ in tick time: when we observe a price change of n ticks, bucket i starts (τ_i) and ends when we observe the next change in price (which we convert in ticks). For larger price moves (larger n) we find that the proportion of LL transactions ($TA_{LL,i}$) increases. Looking at the longer horizon price impact of trading activity in bucket $(i+k)$, $k > 1$, given a price move of n ticks in bucket $i-1$, we find no evidence of a longer impact of the past price moves as a determinant of future trading intensity by different types of traders. On the other hand, if we look not at the average trading intensity in bucket i following a price move in bucket $i-1$, but rather at the next trade immediately following a given price move, we find an even more marked result, reported in the right side of the same table: HH transactions are concentrated following price moves of 1 or 2 ticks, whereas almost all large price moves are followed by LL transactions.

Taken together, these results show that HFTs appear predominantly to follow strategies that exploit small price movements rather than larger ones. In periods of market instability, LFTs see their share of the trading go up, with all the risks that entails, while the share of HFTs goes down. When volatility becomes extreme, as during the "flash crash", a brief period of extreme market volatility on May 6, 2010, Kirilenko et al. (2017) and Easley et al. (2011) concluded that although HFTs did not trigger the crash, their response to the unusually large selling pressure on that day exacerbated market volatility. Brogaard et al. (2018) also find that HFTs do not trigger extreme price movements on NASDAQ stocks.

[21] For each subperiod (bucket) i of length $k = 10,000$ trades, we compute realized volatility sampling every $w = 100$ trades. For each bucket we compute k/w measures of realized volatility and then we average them to obtain RV_i^{avg} for that specific bucket i (see Zhang et al. (2005)).

We looked so far at whether different types of traders were more likely to trade in different market conditions. We now look in Figure 8 at whether HFTs are more or less likely to provide vs. take liquidity in the same conditions. We again partition time in buckets (subperiods) of $k = 10,000$ transactions and compute the percentage of transactions in time bucket i where the aggressor (i.e., the taker of liquidity) is a LFT,

$$Agg_{LFT,i} = \frac{\sum_{\tau \in [\tau_{i-1}, i_i)} G_{LFT,\tau}}{\sum_{j=\{HL,LH\}} \sum_{\tau \in [\tau_{i-1}, \tau_i)} I_{j,\tau}} \tag{4.4}$$

where $G_{LFT,\tau}$ is equal to 1 when a LFT is an aggressor. Obviously, this applies to HL and LH transactions only. We find that the provision of liquidity by HFTs is stable for most of the volatility environments, consistent with the unconditional time series evidence presented in Figure 5, but tends to decrease in extreme volatility (i.e., jumps) ones.[22]

4.3 Response of Different Traders to Market Microstructure Noise

We now examine the propensity of HFTs and LFTs to trade following episodes of high market microstructure noise. Negative first order autocorrelations of log-returns are symptomatic of market microstructure noise, since if observed transaction log-prices are $Y_{\tau_i} = X_{\tau_i} + \varepsilon_{\tau_i}$ where X is the fundamental price in (4.1) expected to have uncorrelated increments by market efficiency arguments, and ε an i.i.d. noise component orthogonal to the X process, then observed log-returns R would follow the process

$$R_i = Y_{\tau_i} - Y_{\tau_{i-1}} = X_{\tau_i} - X_{\tau_{i-1}} + \varepsilon_{\tau_i} - \varepsilon_{\tau_{i-1}}. \tag{4.5}$$

Therefore the first-order autocorrelation ACF(1) of the observed market returns is expected to be negative in the presence of market microstructure noise, as

$$\mathrm{Cov}(R_i, R_{i+1}) = -\mathrm{Var}(\varepsilon_{\tau_i}) < 0 \tag{4.6}$$

and $\mathrm{Cov}(R_i, R_{i+j}) = 0$ for $j > 1$. The MA(1) structure of returns is the basis for the likelihood correction for noise in Aït-Sahalia et al. (2005). In this framework, ACF(1) closer to -1 is indicative of a higher level of noise (ε) relative to the signal (X). Pure bid-ask bounces are a particular example of this, with X nearly fixed and ε drawn from a binary distribution with values of $\pm$ half the spread.

[22] Brogaard et al. (2018) document that when extreme price movements affect contemporaneously several stocks, HFTs do not supply liquidity. Brogaard et al. (2019) looking at messages submitted on recognized Canadian equity markets, show that HFTs' behavior changes when volatility is high causing a reduction in the submission of limit orders and their contributions to price discovery.

Figure 9 reports trading activity, $TA_{j,i}$, attributable to the four types of transactions, $j = \{HH, HL, LH, LL\}$, conditional on different levels of ACF(1) in the previous bucket, $i-1$. We are sampling in tick time for this purpose since the autocorrelogram is most naturally defined in terms of price ticks. The length of the subperiods is $k = 500$ ticks. We find that HFTs trade more following periods where ACF(1) is close to -1, consistent with them trading more when the price is stuck in a corridor, doing not much other than bid ask bounces. In fact, a popular HFT trading strategy consists of capturing rebates offered by the exchanges to providers of liquidity; posting passive orders is always risky, but less so when these orders can be cancelled quickly and when the price does not move much. When the price actually moves, then the ACF(1) is higher (i.e., less negative or closer to 0) and we find that this is associated with higher LFT activity, as HFTs tend to shy away from markets that actually move, consistent with the evidence presented in Section 4.2. In other words, it appears that what an econometrician might call higher levels of market microstructure "noise" corresponds to more trading opportunities for the HFTs. From Panel B of Table 3, we see that only LL transactions lead to negative autocorrelation, as expected under (4.6). The other three types of transactions, each involving a HFT on at least one side, lead to positive first-order autocorrelations, indicating that they are not followed by much market microstructure noise. HH transactions lead to a near zero autocorrelation, suggesting no predictability – or as close to an even playing field as possible. HL and LH transactions lead to positive first order autocorrelation. As predicted, longer-lag auto-correlations are close to 0 following all types of transactions, which is the expected result with or without noise, provided that the noise term is not autocorrelated itself.[23]

5 How Characteristics of the Price Process Respond to Trading by High and Low Frequency Traders

In the previous Section we examined how changes in $\{b, \sigma, \mu\}$ and the presence of noise affect the actions of the different types of traders we have identified. We now study the reverse causality: whether the shares of traders involved in a transaction (HH, HL, LH and LL) affects the future realizations of $\{b, \sigma, \mu\}$, and whether they lead to different levels of market microstructure noise.

[23] Our findings are in line with Brogaard et al. (2014b) who find that HFTs reduce price errors and Chaboud et al. (2014) who show that the provision of liquidity by HFTs in Forex markets reduces the autocorrelation of returns. Our results add to these findings by showing that HFTs trade more intensely when pricing errors are high to take full advantage of bid ask bounces.

5.1 Price Direction Response to Different Traders

We saw above that HFTs tend to trade less (and LFTs more) following large returns of either sign. Is it also the case that transactions involving at least one HFT are less likely to be followed by large returns? The answer is yes: Figure 10 shows the results of kernel regressions relating returns in time bucket i to trading activity in time bucket $i-1$, $TA_{j,i-1}$ – equation (4.3). We employ non-overlapping buckets of $k = 50$ transactions to capture the immediate return following a transaction; these results are consistent to different choices of k within a reasonable range. We find that, on average, transactions involving at least one HFT are followed by a smaller absolute return than transactions involving two LFTs. In other words, HFT transactions have less market impact than those involving LFTs, and this effect holds controlling for the size of the transaction. Our results complement previous findings: Brogaard et al. (2014b) show that HFTs facilitate price efficiency and Brogaard et al. (2019) provide evidence that HFTs convey market information in to prices. Here we show that HFTs smooth the price process. In fact, when they trade more intensely the return process moves less.

Next, we examine how often HFTs tend to "win" on average. For this purpose, we investigate whether transactions of the type HL (where the HFT holds the long position) are more likely to be followed by a price up-tick (i.e. the next price change is a price increase) and whether those of the type LH (where the HFT holds the short position) are more likely to be followed by a price down-tick (i.e. the next price change is a decrease in price). The top panels of Figure 11 show kernel density estimators of the probability of an up-tick following the four types of transactions. The answer is striking: transactions LL and HH are basically fair games, with density symmetrical around 1/2, while HL transactions lead to densities centered at about 2/3 and LH lead to densities centered at about 1/3. In other words, when a HFT enters into a long position against a LFT, there is a 2/3 probability, on average, that the price will go up in the immediate future and vice versa when a HFT enters into a short position against a LFT. And these average winning probabilities are quite stable over time in the sample: the lower panels of Figure 11 show time series plots of the winning probability following the four types of transactions with confidence bands.[24] Furthermore, HFTs tend to wait less on average before realizing the gain following a trade. Table 5 reports the up-tick and down-tick wait times conditional on the type of transaction for each month in transaction, volume, and calendar times. HL transactions are followed by the shortest wait times for an up-tick, and LH transactions are followed by the shortest wait times for a down-tick. These results are consistent across all three sampling mechanisms, transaction, volume, and calendar, and provide further evidence of the winning

[24] In light of these results, it is perhaps not too surprising that Virtu, Inc., a major HFT firm, managed to report in its IPO S-1 Prospectus only one down day in four years of trading.

ability of HFTs beyond the winning probabilities reported above: not only do they win more often, but they wait less to see their win realized.

One legitimate question then becomes whether these gains are simply compensation for risk, perhaps hidden, that the HFTs would be somehow assuming in the process – the proverbial nickels picked up in front of a steamroller. While recognizing that our sample period is not long enough to show evidence of a potential peso problem, we find no evidence of rare disasters in the sample: Figure 12 shows the distribution of losses conditional on a HFT losing money in an HL or LH transaction, which happens about 1/3 of the time. The figure shows that in over 99.95% of the cases the conditional loss is of only one tick with most of the few remaining ones consisting of losses of two ticks.

These results suggest that HFTs find it profitable to trade with LFTs as opposed to trading with another HFT. How are HFTs able to avoid each other? One possibility might be that HFTs recognize their potential counterparty by using observable order characteristics. One such characteristic is the lot size. O'Hara et al. (2014) show that, on NASDAQ, HFTs are more likely to rely on odd lots, which helps hide their trading as odd lots do not appear in consolidated market data feeds and as such are not included in the TAQ database, for instance. Figure 13 shows the probability density of lot size for the different types of transactions. Unlike O'Hara et al. (2014), we find no evidence in this market of HFTs (or any other trader) systematically employing different trade sizes: in our data there are no statistically significant differences between the lot size in transactions involving HFTs or LFTs. Another possibility might be the type of order employed. Are HFTs more likely to be liquidity providers or liquidity takers? Figure 5 above showed that HFTs are slightly less likely to provide liquidity than to take it when trading with a LFT, but HFTs do not seem to privilege one type of order over another.

Here is another way of quantifying the extent to which HFTs manage to avoid each other in transactions. Suppose that HFTs are involved in a percentage p_H of the transactions and LFTs in $p_L = 1 - p_H$ percent of the transactions in the sample. If the pairing between counterparties were random, we would expect the transactions pairings to satisfy $p_{HH} = p_H^2$, $p_{LL} = p_L^2$ and $p_{HL} = p_{LH} = p_H p_L$. The data in Table 3 tell us the empirical frequencies of the four types of transactions, p_{HH}, p_{LL}, p_{HL} and p_{LH} so we can compare the empirical distribution to the predictions if the allocation of counterparties were truly random. Inferring p_H and p_L from $p_{HL} = p_{LH} = p_H p_L$, we find that the empirical frequency p_{HH} is lower than what p_H^2 would suggest: $p_{HH} = 3\%$ and $p_H^2 = 6\%$ in 2010 and $p_{HH} = 6\%$ and $p_H^2 = 9\%$ in 2011. So HFTs do not trade as much between themselves as purely random pairing would suggest.

One possibility, discussed above, is that HFTs are somehow able to recognize each other's orders

despite operating in an anonymous market. Another possible explanation is that HFTs are more likely to be trading on the same side of the market at each point in time, thereby limiting the possible transactions among themselves. To test this hypothesis, we compute the percentage of time that HFTs are on the same side of the market. We partition each day in time tick subsamples based on price changes (i.e., if the price of trade s does not move, that price belongs to subsample i, if the price moves, that price belongs to a new subsample, $i+1$). Figure 16 shows the proportion of same-side trading for HFTs and LFTs for August 2010 and August 2011, respectively. This proportion is high – about 95% in August 2010 and 93% in August 2011. This suggests that HFTs are hunting like a pack, perhaps employing algorithms that are reacting to the same signals in the same way.[25]

Finally, we look at whether the types of traders involved in a transaction are more likely to give rise to price momentum or mean reversion. Figure 14 reports the probability that the second tick is also an up-tick following the first up-tick. We find no evidence whatsoever of persistence of the direction taken by the first tick and in fact quite strong evidence of mean-reversion: given that the first tick is up, the second one is very likely to be down, and vice versa. In tick time, the first order autocorrelation is very strongly negative, of the order of -0.75. This is not to be confused with the autocorrelation indicative of market microstructure noise, which is based on calendar time or other forms of non-tick sampling. This strong evidence of mean-reversion in tick time suggests that there is indeed little room for error if HFTs are to capture the (small) gains associated with each transaction: it would likely necessitate a limit order at the next tick up following a buy, exploiting their latency advantage to insure their priority positioning at that place in the limit order book, or even a pre-positioning of that order.

5.2 Price Variability Response to Different Traders

We now examine how volatility and jumps respond to trading activity by high and low frequency traders. Figure 15 reports kernel regression estimates of the volatility and jumps of the returns process in trading bucket i conditional on the trading activity of the four types of transactions in bucket $i-1$, $TA_{j,i-1}$ – equation (4.3).

The plot notes different integer levels of average volatility as horizontal lines, with jumps corresponding by convention to returns larger than 3 times the average volatility in the data. Observations are again marked individually to avoid relying exclusively on the kernel regression for these relatively rare observations. We find that LFTs' increased share of trading, on the whole, tends to precede large

[25] Chaboud et al. (2014) also find that HFTs do not trade with each other as much as random matching would predict and interpret this result as evidence that HFTs' trading strategies are less diverse than trading strategies used by LFTs.

price movements: when the share of LL trading increases beyond its average range, higher volatility and more jumps tend to follow. So we find that HFTs's trading does not appear to be "causing" jumps in the sense that jumps do not disproportionately follow periods when their share of trading is higher, to the contrary. Our results confirm findings in Hasbrouck and Saar (2013) and Hagströmer and Nordèn (2013), among others. Most of the previous literature concentrates on the contemporaneous relation between HFTs' activity and volatility levels while we condition on past trading activity and study whether a high (low) level of participation of HFTs at time $i-1$ "causes" volatility in the next time period, i. Table 6 reports the distribution of the price impact of the different types of transactions: we find that most transactions with high price impact involve two LFTs. Similar results hold in transaction time over longer horizons, in volume time and in calendar time. For instance, in calendar time over 100 milliseconds, we see in the right side of the same table that the accumulation of mass in the tails of the price impact distribution is primarily due to LL transactions.

These results do not mean that there is no potential for HFTs to be involved in, or even cause market incidents. Herding by HFTs provides a potential mechanism by which an a priori trivial market event could translate into a flash crash. Figure 16 indeed shows that HFTs have a high propensity to appear on the same side of the market at any given point in time. So even though HFT-participating transactions may not individually be predictive of jumps, herding by HFTs creates a risk, with liquidity potentially withdrawn by a large number of HFTs together if the circumstances warrant it from the point of view of their optimization algorithms. In terms of liquidity provision (equation (4.4)), we find in Figure 17 that higher volatility tends to follow a decreased provision of liquidity (i.e., an increased use of market orders) by HFTs.

5.3 Noise Response to Different Traders

Figure 18 shows different levels of negative first order autocorrelation of returns following different types of traders having transacted.[26] We find that higher-than-normal levels of LL activity are followed by less negative autocorrelation. Negative autocorrelation levels following transactions involving at least one HFT are consistently high and do significantly vary in response to HFTs trading activity.

6 Conclusions

We systematically analyze the influence of the price process on the trading propensity of HFTs in the S&P500 E-mini market, and the reverse influence of their trading on the characteristics of the price

[26]Trading activity, $TA_{j,i}$ is defined in equation 4.3. We are sampling in tick time since the autocorrelogram is most naturally defined in terms of price ticks. The length of the subperiods is $k = 500$ ticks.

process. We document that HFTs are not predominantly directional traders, but have a substantially higher than even chance of winning following each transaction when they trade against LFTs but not when they trade between one another. HFTs also appear to have devised mechanisms for avoiding trading with each other as often as their unconditional importance in the market would predict. Contrary to popular perception, we do not find that the trading propensity of HFTs induce volatility or price jumps. In fact, we find the opposite: it is their withdrawal from the market that leads to high volatility and jumps, whether the decrease of HFT trading intensity is intentional or exogenous. Conversely, we find that spikes in volatility and jumps cause HFTs to trade less intensely. This specific finding has important consequences for the resilience of the markets. Our results are largely consistent across the two samples of August 2010 and August 2011, which was characterized by high volatility.

Given the large trading volume generated by HFTs, there have been concerns that their speed advantage may result in unfair competition and disruptive trading behaviors. HFT activity has attracted regulatory attention.[27][28] Some European countries have either implemented or proposed to implement transaction taxes to discourage high-frequency trading activity. In addition, some trading venues have adopted rules in order to limit the speed advantage of HFTs. Aït-Sahalia and Sağlam (2016a) find that most of these policies are unable to increase the provision of liquidity by high frequency market makers when volatility is high. While HFTs generate increased activity and narrower spreads in normal times, the evidence presented in this paper suggests that they may reduce the resilience of the system as a whole during stressed times, possibly due to having displaced traditional market participants (such as, for example, major market-maker banks) who may have otherwise been an important stabilizing presence in volatile environments.

[27] The Dodd-Frank Wall Street Reform and Consumer Protection Act called for an in depth study of HFT (Section 967 (2)(D)). The CFTC and SEC have proposed to require more transparency from HFTs such as disclosing their algorithms to regulators in the event of severe market disruptions and requiring certain HFT broker-dealers to register with the Financial Industry Regulatory Authority (FINRA), which oversees broker-dealers. The SEC and CFTC have taken steps to bring some HFTs under closer scrutiny through enforcement actions.

[28] For example, in 2016, the SEC announced settlements with Barclays and Credit Suisse over allegations that Barclays had misled its investors on HFT practices permitted on its private trading platforms, and that Credit Suisse failed to operate its trading systems as advertised. The CFTC has cracked down on spoofing, using the anti-spoofing authority granted in the Dodd-Frank Act in a number of recent enforcement actions involving algorithmic trading.

References

Adamic, L., Brunetti, C., Harris, J. H., Kirilenko, A., 2017. Trading networks. Econometrics Journal 20, S126–S149.

Aït-Sahalia, Y., Jacod, J., 2012. Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. Journal of Economic Literature 50, 1005–1048.

Aït-Sahalia, Y., Jacod, J., 2014. High Frequency Financial Econometrics. Princeton University Press.

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Sağlam, M., 2016a. High frequency market making: Implications for liquidity. Tech. rep., Princeton University.

Aït-Sahalia, Y., Sağlam, M., 2016b. High frequency market making: Optimal quoting. Tech. rep., Princeton University.

Angel, J., Harris, L., Spatt, C. S., 2015. Equity trading in the 21st century: An update. Quarterly Journal of Finance 5, 1–39.

Baron, M., Brogaard, J. A., Hagströmer, B., Kirilenko, A., 2019. Risk and return in high frequency trading. Journal of Financial and Quantitative Analysis 54, 993–1024.

Baron, M., Brogaard, J. A., Kirilenko, A., 2012. The trading profits of high frequency traders. Tech. rep., University of Washington Foster School of Business.

Biais, B., Declerck, F., Moinas, S., 2016. Who supplies liquidity, how and when? Working paper 563, Bank of International Settlement.

Biais, B., Woolley, P., 2011. High frequency trading. Tech. rep., Toulouse School of Economics and London School of Economics.

Breckenfelder, J., 2019. Competition among high frequency traders, and market quality. Tech. rep., European Central Bank Working Paper.

Brogaard, J. A., 2011. High frequency trading and volatility. Tech. rep., University of Washington.

Brogaard, J. A., Hagströmer, B., Nordèn, L., Riordan, R., 2015. Trading fast and slow: Colocation and liquidity. Review of Financial Studies 28, 3407–3443.

Brogaard, J. A., Hendershott, T., Hunt, S., Ysusi, C., 2014a. High frequency trading and execution costs of institutional investors. Financial Review 49, 345–369.

Brogaard, J. A., Hendershott, T., Riordan, R., 2014b. High frequency trading and price discovery. Review of Financial Studies 27, 2267–2306.

Brogaard, J. A., Hendershott, T., Riordan, R., 2016. Price discovery without trading. Tech. rep., University of California - Berkeley.

Brogaard, J. A., Hendershott, T., Riordan, R., 2017. High frequency trading and the 2008 short sale ban. Journal of Financial Economics 124, 22–42.

Brogaard, J. A., Hendershott, T., Riordan, R., 2019. Price discovery without trading: Evidence from limit orders. The Journal of Finance 74, 1621–1658.

Brogaard, J. A., Riordan, R., Shkilko, A., Sokolov, K., Carrion, A., Moyaert, T., 2018. High frequency trading and extreme price movements. Journal of Financial Economics 128, 253–265.

Budish, E., Cramton, P., Shim, J., 2015. The high-frequency trading arms race: Frequent batch auctions as a market design response. Quarterly Journal of Economics 130, 1547–1621.

Cartea, À., Jaimungal, S., Wang, Y., 2019. Spoofing and price manipulation in order driven markets. Tech. rep., University of Oxford Working Paper.

Chaboud, A., Chiquoine, B., Hjalmarsson, E., Vega, C., 2014. Rise of the machines: Algorithmic trading in the foreign exchange market. The Journal of Finance 69, 2045–2084.

Easley, D., de Prado, M. M. L., O'Hara, M., 2011. The microstructure of the "flash crash": Flow toxicity, liquidity crashes and the probability of informed trading. Journal of Portfolio Management 37, 118–128.

Easley, D., de Prado, M. M. L., O'Hara, M., 2012. Bulk classification of trading activity. Tech. rep., Cornell University.

Hagströmer, B., Nordèn, L., 2013. The diversity of high frequency traders. Journal of Financial Markets 16, 741–770.

Hasbrouck, J., 2003. Intraday price formation in the US equity index markets. The Journal of Finance 58, 2375–2400.

Hasbrouck, J., Saar, G., 2013. Low-latency trading. Journal of Financial Markets 16, 646–679.

Hendershott, T., Jones, C. M., Menkveld, A. J., 2011. Does algorithmic trading improve liquidity? The Journal of Finance 66, 1–33.

Hens, T., Lensberg, T., Schenk-Hoppé, K., 2018. Front-running and market quality: An evolutionary perspective on high frequency trading. International Review fo Finance 18, 727–741.

Hirschey, N. H., 2011. Do high-frequency traders anticipate buying and selling pressure? Tech. rep., London Business School.

Huh, Y., 2016. Machines vs. machines: High frequency trading and hard information. Tech. rep., Stanford University.

Jarrow, R. A., Protter, P., 2012. A dysfunctional role of high frequency trading in electronic markets. International Journal of Theoretical and Applied Finance 15 (3), 1–15.

Jovanovic, B., Menkveld, A. J., 2010. Middlemen in limit-order markets. Tech. rep., New York University and VU University Amsterdam.

Kirilenko, A., Kyle, A. P., Samadi, M., Tuzun, T., 2017. The flash crash: High-frequency trading in an electronic market. The Journal of Finance 72, 967–998.

Lazzeroni, L., Owen, A., 2002. Plaid models for gene expression data. Statistica Sinica 12, 61–86.

Li, S., Wang, X., Ye, M., 2019. Who provides liquidity and when? Tech. rep., NBER Working Paper.

Litzenberger, R., Castura, J., Gorelic, R., 2012. The impacts of automation and high frequency trading on market quality. Annual Review of Financial Economics 4.

Menkveld, A. J., 2013. High frequency trading and the new-market makers. Journal of Financial Markets 16, 712–740.

Menkveld, A. J., 2016. The economics of high-frequency trading: Taking stock. Annual Review of Financial Economics 8, 1–24.

Menkveld, A. J., Zoican, M. A., 2017. Need for speed? Exchange latency and liquidity. Review of Financial Studies 4, 11881228.

O'Hara, M., Yao, C., Ye, M., 2014. What's not there: Odd lots and market data. The Journal of Finance 69, 2199–2236.

Raman, V., Robe, M., Yadav, P. K., 2014. Electronic market makers, trade anonymity and market fragility. Tech. rep., University of Warwick, School of Business.

Saliba, P., 2019. The information content of high frequency traders aggressive orders: Recent evidence. Tech. rep., École Polytechnique Working Paper.

Sussman, A., Tabb, L., Iati, R., 2009. US equity high frequency trading: Strategies, sizing and market structure. Tech. rep., The TABB Group.

Zhang, F. X., 2010. High-frequency trading, stock volatility, and price discovery. Tech. rep., Yale University, School of Management.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Time	August 2010			August 2011		
Trans	Vol	Tick	Cal	Vol	Tick	Cal
1	4	0.070	71	4.192	0.078	38.77
10	41	0.671	709	41.91	0.776	387.7
50	204	3.350	3,543	209.5	3.880	1,939
100	407	6.701	7,086	419.0	7.759	3,878
1,000	4,065	66.92	70,700	4,182	77.54	38,832
10,000	40,076	663.1	693,630	41,399	772.7	391,740
Vol	Trans	Tick	Cal	Trans	Tick	Cal
1	0.250	0.018	17.75	0.239	0.019	9.249
10	2.439	0.163	172.9	2.386	0.185	92.50
50	12.25	0.821	868.4	11.93	0.926	462.6
100	24.57	1.646	1,741	23.86	1.852	925.4
1,000	246.0	16.46	17,392	239.1	18.54	9,287
10,000	2,495	165.5	173,079	2,416	186.6	94,625
Tick	Trans	Vol	Cal	Trans	Vol	Cal
1	14.29	57.14	1,014	12.89	54.02	499.6
10	149.3	611.9	10,582	128.9	540.1	4,996
50	746.3	3,045	52,881	644.4	2,700	24,983
100	1,493	6,075	105,761	1,289	5,400	49,973
1,000	14,943	60,744	1,056,485	12,897	53,930	500,832
10,000	150,807	604,373	10,460,413	129,421	535,791	5,069,952
Cal	Trans	Vol	Tick	Trans	Vol	Tick
1	0.014	0.056	0.001	0.026	0.108	0.002
10	0.141	0.578	0.009	0.258	1.081	0.020
50	0.706	2.879	0.047	1.290	5.404	0.100
100	1.411	5.744	0.095	2.579	10.81	0.200
1,000	14.14	57.50	0.947	25.75	107.7	1.997
10,000	144.2	577.8	9.560	255.3	1,057	19.72

Table 1: Relationship among transaction, contract, tick, and calendar time sampling in the data.

Notes: the table translates the four different ways we measure time. "Trans" refers to the number of transactions, "Vol" indicates the number of contracts (volume), "Tick" denotes the number of price moves in ticks and "Cal" refers to number of milliseconds.

	August 2010		August 2011	
	LFTs	HFTs	LFTs	HFTs
	Panel A: Number of Trades			
Mean	106	25,456	113	49,640
Median	21	27,573	24	35,261
St. Dev.	637	11,358	747	36,046
	Panel B: Volume			
Mean	462	109,133	498	202,500
Median	63	103,883	59	207,379
St. Dev.	2,457	93,005	3,262	56,302
	Panel C: Intertrade Duration (milliseconds)			
Mean	680,164	849	597,628	659
Median	508,252	876	434,936	677
St. Dev.	692,050	147	612,521	276
	Panel D: abs(Inventory) (%)			
Mean	16.73	0.0005	0.1908	0.0139
Median	2.080	0.046	0.0743	0.0078
St. Dev.	30.28	0.041	28.15	1.666

Table 2: Summary Statistics: HFTs vs. LFTs.

Notes: For each trading account we compute, at a daily frequency, number of trades, volume traded, average intertrade duration and end-of-day inventory as percentage of volume traded. LFTs refers to trading accounts which traded at least 10 times per day.

	August 2010				August 2011			
	LL	HL	LH	HH	LL	HL	LH	HH
				Panel A: Number of Trades				
Mean (%)	61.14	17.71	17.89	3.233	52.78	20.62	20.31	6.283
Mean	205,717	72,845	72,309	16,891	352,369	140,980	138,829	43,298
Median	194,285	68,555	68,546	16,110	314,460	121,241	119,467	40,101
St. Dev.	53,881	16,408	15,200	3,679	133,429	50,773	49,592	17,181
Min	103,814	39,577	42,277	9,743	163,393	75,593	74,944	22,339
Max	287,838	99,465	97,209	22,999	678,427	259,402	256,283	84,022
Skew	-0.401	-0.464	-0.321	-0.264	0.284	0.361	0.323	0.262
Kurt	2.581	2.910	2.758	2.835	2.623	2.165	2.161	2.070
				Panel B: Log-Returns				
Mean	0.000	0.137	-0.134	0.001	-0.001	0.144	-0.140	0.000
Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
St. Dev.	0.630	0.565	0.564	0.231	0.631	0.564	0.558	0.251
Min	-4.692	-2.350	-4.601	-2.343	-8.258	-4.323	-4.449	-4.139
Max	4.774	4.655	2.391	2.336	8.512	6.593	4.318	4.157
Skew	0.001	3.269	-3.263	0.339	-0.015	2.955	-2.941	-0.103
Kurt	13.77	14.43	14.38	100.4	12.67	12.92	12.88	83.41
ACF(1) (%)	-16.10	22.39	22.98	1.896	-15.33	18.18	18.08	1.424
ACF(2) (%)	1.842	18.21	18.26	1.653	0.990	12.06	11.53	0.620
ACF(3) (%)	2.844	14.74	14.23	0.665	1.983	8.910	8.131	0.477
				Panel C: Absolute Log-Returns				
Mean	0.173	0.148	0.146	0.023	0.186	0.158	0.154	0.029
Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
St. Dev.	0.606	0.563	0.562	0.229	0.604	0.561	0.555	0.249
Min	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Max	4.774	4.655	4.601	2.343	8.512	6.657	5.976	4.231
Skew	3.279	3.604	3.602	9.799	3.056	3.349	3.339	8.535
Kurt	11.88	14.15	14.02	98.62	10.78	12.48	12.46	81.48
ACF(1) (%)	25.66	24.46	24.99	2.700	22.72	19.66	19.13	1.855
ACF(2) (%)	18.13	18.75	19.33	1.609	15.20	12.17	11.65	0.413
ACF(3) (%)	15.27	15.11	15.71	1.399	11.99	8.950	8.336	0.591
				Panel D: Volume				
Mean	3.869	4.270	4.302	4.312	4.132	4.058	4.158	3.911
Median	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
St. Dev.	10.94	12.77	13.03	10.91	10.73	10.42	10.44	10.98
Min	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Max	785	650	779	375	841	786	742	500
Skew	1.685	1.756	1.776	1.772	1.415	1.554	1.571	1.733
Kurt	5.475	5.914	6.029	6.127	4.487	5.052	5.142	6.284
				Panel E: Intertrade Duration (milliseconds)				
Mean	120.7	341.4	341.4	1,453	74.41	193.0	195.9	583.5
Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
St. Dev.	653.2	2,058	2,003	7,282	366.8	907.8	936.5	2,722

Table 3: Summary Statistics: Transactions by Type of Traders Involved.

	Next Tick (%)								Next Transaction (%)							
	August 2010				August 2011				August 2010				August 2011			
Ticks_{i-1}	LL_i	HL_i	LH_i	HH_i	LL_i	HL_i	LH_i	HH_i	LL_i	HL_i	LH_i	HH_i	LL_i	HL_i	LH_i	HH_i
-12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-10	0.73	0.09	0.14	0.05	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
-9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
-8	0.00	0.00	0.00	0.00	0.40	0.40	0.00	0.20	0.00	0.00	0.00	0.00	0.60	0.40	0.00	0.00
-7	1.00	0.00	0.00	0.00	0.61	0.22	0.17	0.00	1.00	0.00	0.00	0.00	0.67	0.17	0.16	0.00
-6	0.91	0.03	0.05	0.01	0.80	0.12	0.03	0.05	0.86	0.00	0.14	0.00	0.67	0.22	0.06	0.05
-5	0.44	0.56	0.00	0.00	0.56	0.28	0.11	0.05	0.75	0.25	0.00	0.00	0.64	0.18	0.12	0.06
-4	0.72	0.28	0.00	0.00	0.73	0.16	0.09	0.02	0.90	0.10	0.00	0.00	0.85	0.08	0.07	0.00
-3	0.66	0.27	0.05	0.02	0.73	0.19	0.06	0.02	0.75	0.11	0.12	0.02	0.76	0.11	0.10	0.03
-2	0.69	0.24	0.06	0.01	0.67	0.24	0.06	0.03	0.63	0.20	0.14	0.03	0.66	0.19	0.12	0.03
-1	0.61	0.33	0.04	0.02	0.57	0.34	0.06	0.03	0.62	0.23	0.14	0.01	0.57	0.26	0.15	0.02
1	0.61	0.04	0.33	0.02	0.57	0.06	0.34	0.03	0.62	0.14	0.23	0.01	0.56	0.15	0.26	0.03
2	0.68	0.04	0.27	0.01	0.67	0.05	0.26	0.02	0.65	0.10	0.22	0.03	0.65	0.11	0.21	0.03
3	0.75	0.02	0.20	0.03	0.72	0.05	0.21	0.02	0.69	0.07	0.24	0.00	0.75	0.07	0.17	0.01
4	0.70	0.00	0.30	0.00	0.70	0.10	0.19	0.01	0.80	0.00	0.20	0.00	0.78	0.10	0.12	0.00
5	0.65	0.00	0.35	0.00	0.71	0.07	0.20	0.02	0.78	0.00	0.22	0.00	0.63	0.13	0.23	0.01
6	0.83	0.00	0.17	0.00	0.80	0.00	0.20	0.00	1.00	0.00	0.00	0.00	0.87	0.13	0.00	0.00
7	1.00	0.00	0.00	0.00	0.71	0.14	0.14	0.01	1.00	0.00	0.00	0.00	0.86	0.00	0.14	0.00
8	0.00	0.00	0.00	0.00	0.97	0.00	0.03	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
9	0.76	0.14	0.09	0.01	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
10	0.45	0.02	0.48	0.05	0.00	0.00	0.00	0.00	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
12	0.94	0.00	0.06	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4: Price Variability Response: Next Tick and Next Transaction (%).

Notes: The left side of the table reports trading activity in bucket i following a price move of n ticks in bucket $i-1$. Trading activity, $TA_{j,i}$, is defined in equation (4.3). We are sampling in tick time: when we observe a price change of n ticks, bucket i starts (τ_i) and it ends when we observe the next change in price (which we convert in ticks). The right side of the table reports the proportion of trades immediately following a price move of $n-$ticks. This is analogous to the left side, except that buckets consist of one transaction

	LL	HL	LH	HH
	Transaction Time: August 2010			
Uptick	2,182	1,283	2,859	2,061
Downtick	1,809	2,594	1,077	1,773
	Transaction Time: August 2011			
Uptick	2,257	1,536	2,867	2,130
Downtick	2,232	2,671	1,474	1,895
	Volume Time: August 2010			
Uptick	8,466	4,958	11,138	7,957
Downtick	7,231	10,484	4,291	7,065
	Volume Time: August 2011			
Uptick	9,167	6,273	11,803	8,737
Downtick	9,056	10,934	5,980	7,741
	Calendar Time (milliseconds): August 2010			
Uptick	146,980	84,646	191,440	132,791
Downtick	135,688	197,700	79,158	131,636
	Calendar Time (milliseconds): August 2011			
Uptick	81,758	55,384	109,489	83,144
Downtick	85,076	104,407	56,031	73,267

Table 5: Uptick and downtick average wait in transaction, volume and calendar time.

Notes: This table shows the average waiting time for an uptick and a downtick.

	Transaction Time Number of Trades					Calendar Time Number of Trades				
Ticks_i	Total_i	LL_i	HL_i	LH_i	HH_i	Total_i	LL_i	HL_i	LH_i	HH_i
-11	0	0	0	0	0	262	249	5	7	1
-10	1	0	1	0	0	0	0	0	0	0
-9	0	0	0	0	0	0	0	0	0	0
-8	0	0	0	0	0	1,029	958	63	7	1
-7	2	2	0	0	0	1,004	880	110	13	1
-6	7	7	0	0	0	1,806	1,574	170	60	2
-5	8	5	1	2	0	6,605	5,719	845	40	1
-4	21	17	1	3	0	13,809	11,730	1,915	147	17
-3	58	43	0	15	0	44,640	36,014	6,691	1,619	316
-2	500	372	7	117	4	234,620	160,050	40,363	26,890	7,317
-1	269,277	165,859	4,922	96,395	2,101	2,618,810	1,435,700	371,100	676,310	135,700
0	7,539,015	4,183,944	1,488,619	1,500,008	366,444	9,681,680	4,821,100	2,051,800	2,108,300	700,480
1	269,367	166,409	95,827	5,049	2,082	2,641,430	1,431,800	687,880	378,530	143,220
2	463	309	144	6	4	239,848	159,130	29,986	42,456	8,276
3	55	41	9	4	1	38,449	29,460	1,877	6,611	501
4	15	13	1	1	0	6,765	5,801	118	833	13
5	9	7	2	0	0	2,511	2,031	71	405	4
6	6	5	0	0	1	211	182	4	24	1
7	4	4	0	0	0	1,058	922	0	136	0
8	0	0	0	0	0	677	621	13	43	0
9	1	0	1	0	0	553	462	6	83	2
10	2	2	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0

Table 6: Contemporaneous Price Impact in Transaction and Calendar Time.

Notes: This table shows the distribution of the number of price change ticks following each type of transaction, in transaction and calendar time for August 2010. In transaction time, the sampling frequency is 1 transaction. In calendar time, the sampling frequency is 100 milliseconds.

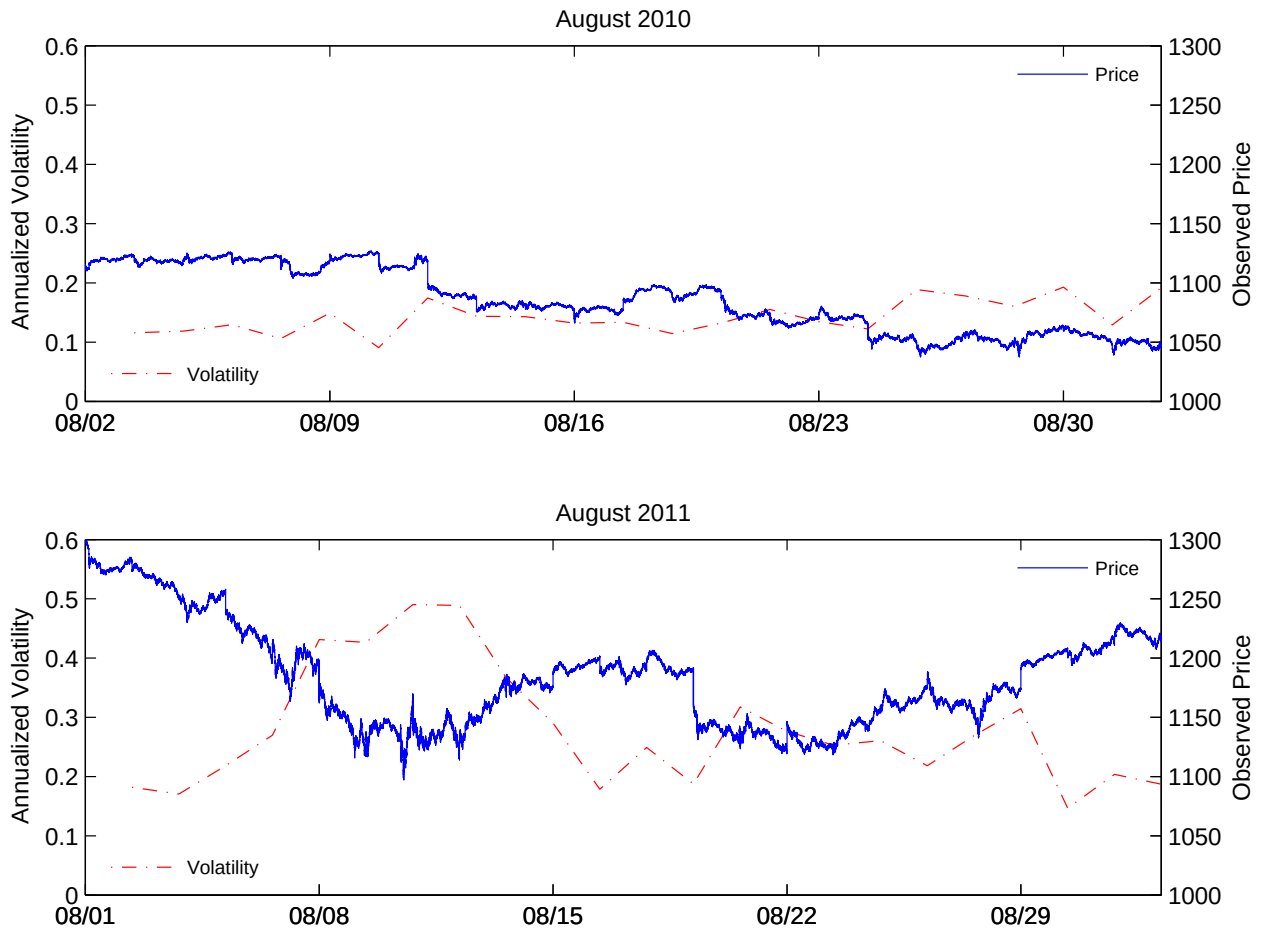

Figure 1: eMini Price and Two-Scale Realized Volatility

Notes: This figure depicts price levels and realized volatility measures for the E-mini September 2010 contract in August 2010 and for the E-mini September 2011 contract in August 2011. We see that the two months provide contrasted market environments, with low (resp. high) volatility in August 2010 (resp. 2011) and a significant leverage effect in August 2011.

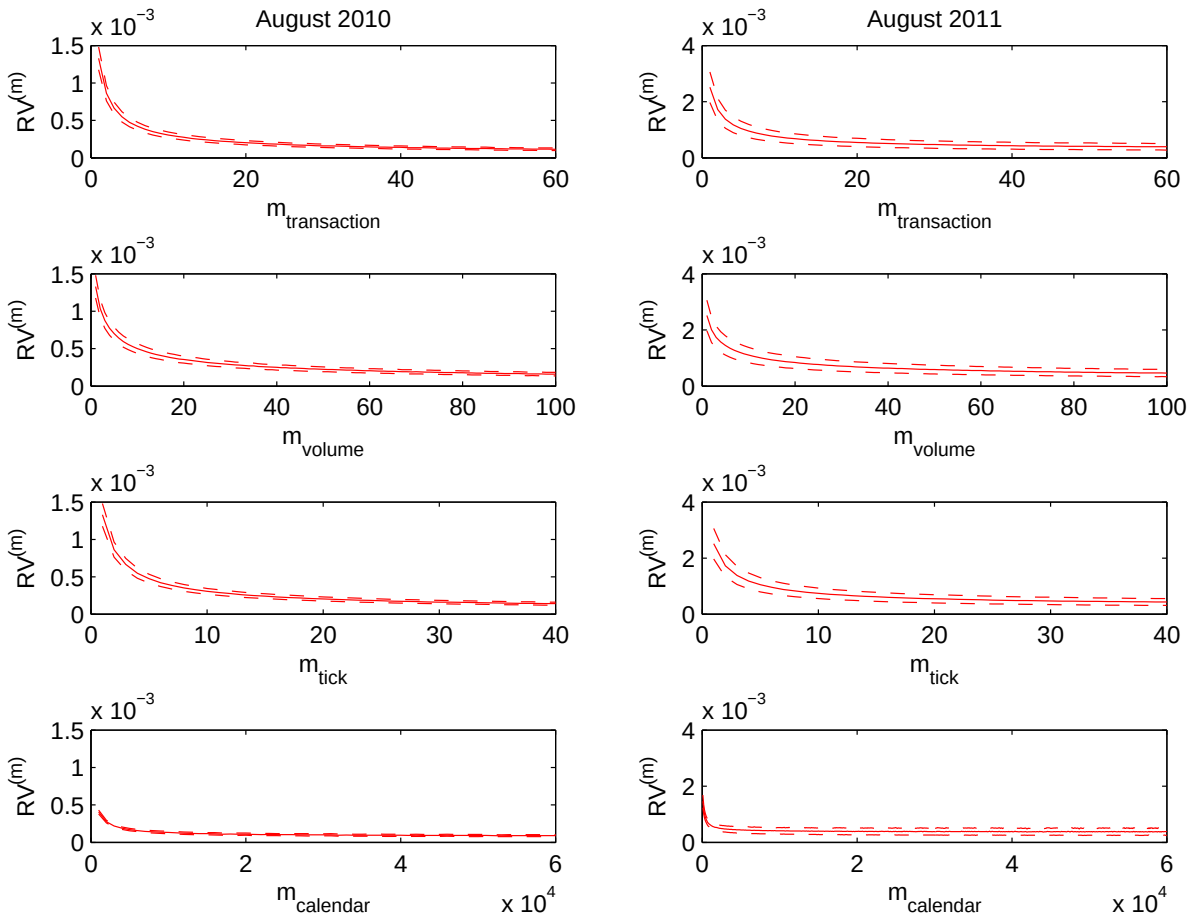

Figure 2: Volatility Signature Plots

Notes: This figure depicts volatility signature plots in transaction time, volume time, tick time and calendar time.

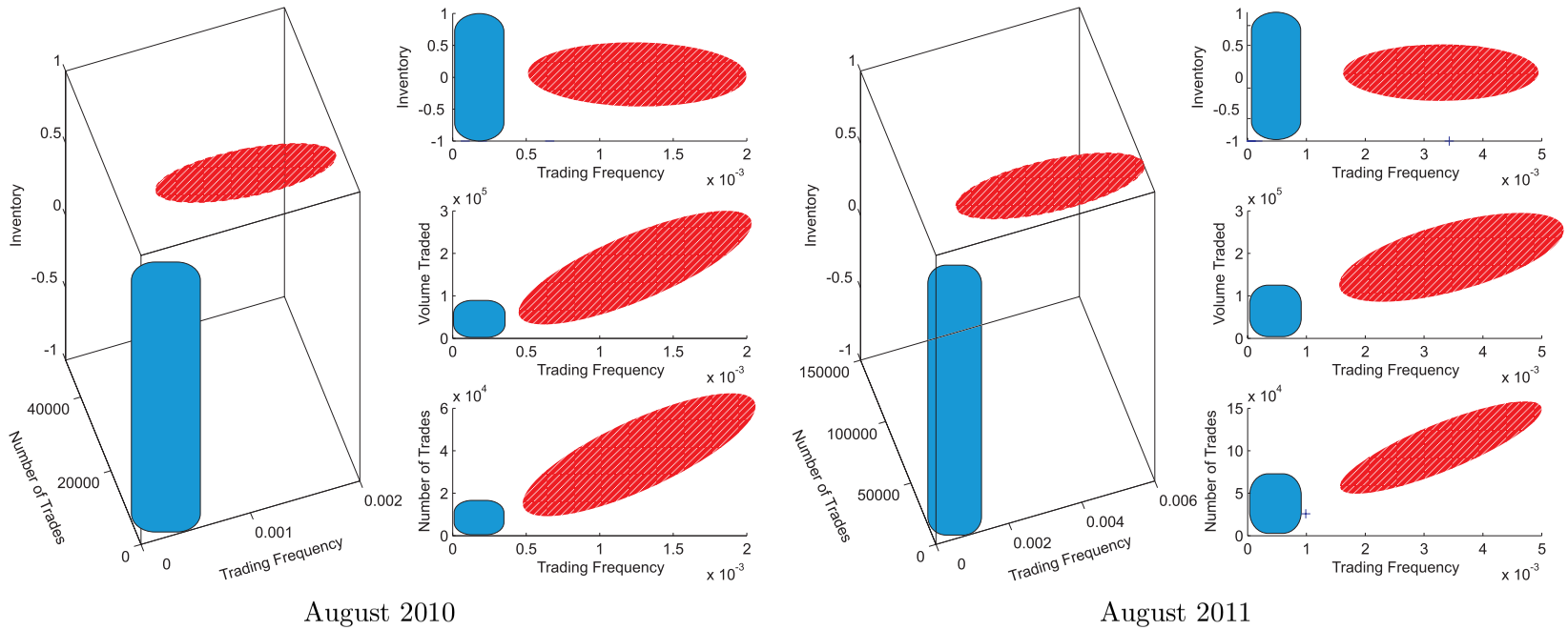

Figure 3: Clustering Analysis

Notes: This figure displays the (daily) average values of the four measures of trading activity used in the clustering analysis for August 2010. Trading frequency is the reciprocal of duration, number of trades refers to the number of transactions of each account, volume traded is the quantity of E-mini futures contract traded, and inventory refers to the ratio of positions accumulated over volume. LFTs are located in the blue shaded areas while the red shaded area marks the cluster of HFTs, with no individual accounts indicated to preserve their anonymity.

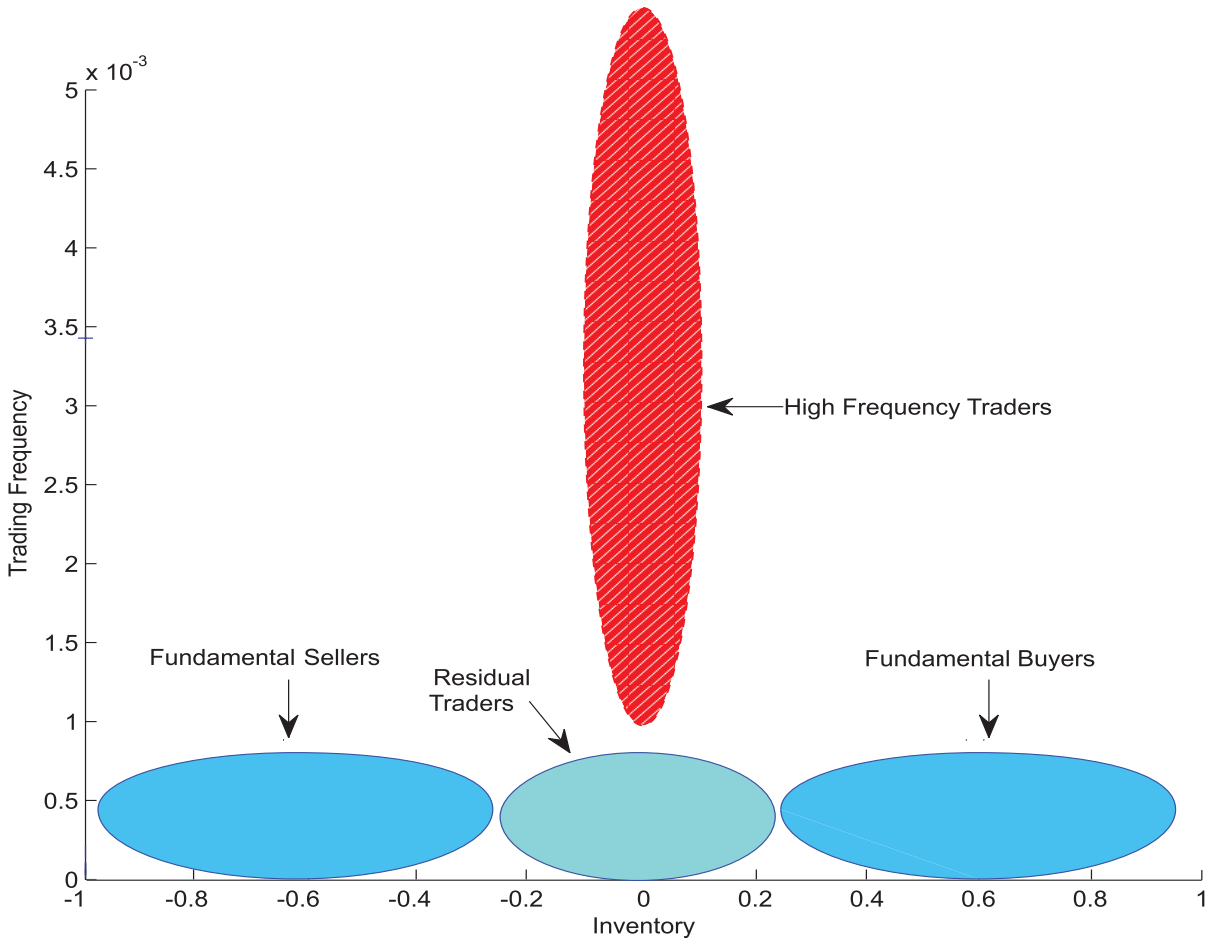

Figure 4: Clustering Analysis: Trading Frequency and Inventory Held by Different Types of Traders, August 2010.

Notes: This figure plots the clusters of traders identified using the two criteria of trading frequency and carried-over inventory. LFTs are located in the blue shaded areas, with a distinction between fundamental buyers and fundamental sellers, while the red shaded area marks the cluster of HFTs, with no individual accounts indicated to preserve their anonymity.

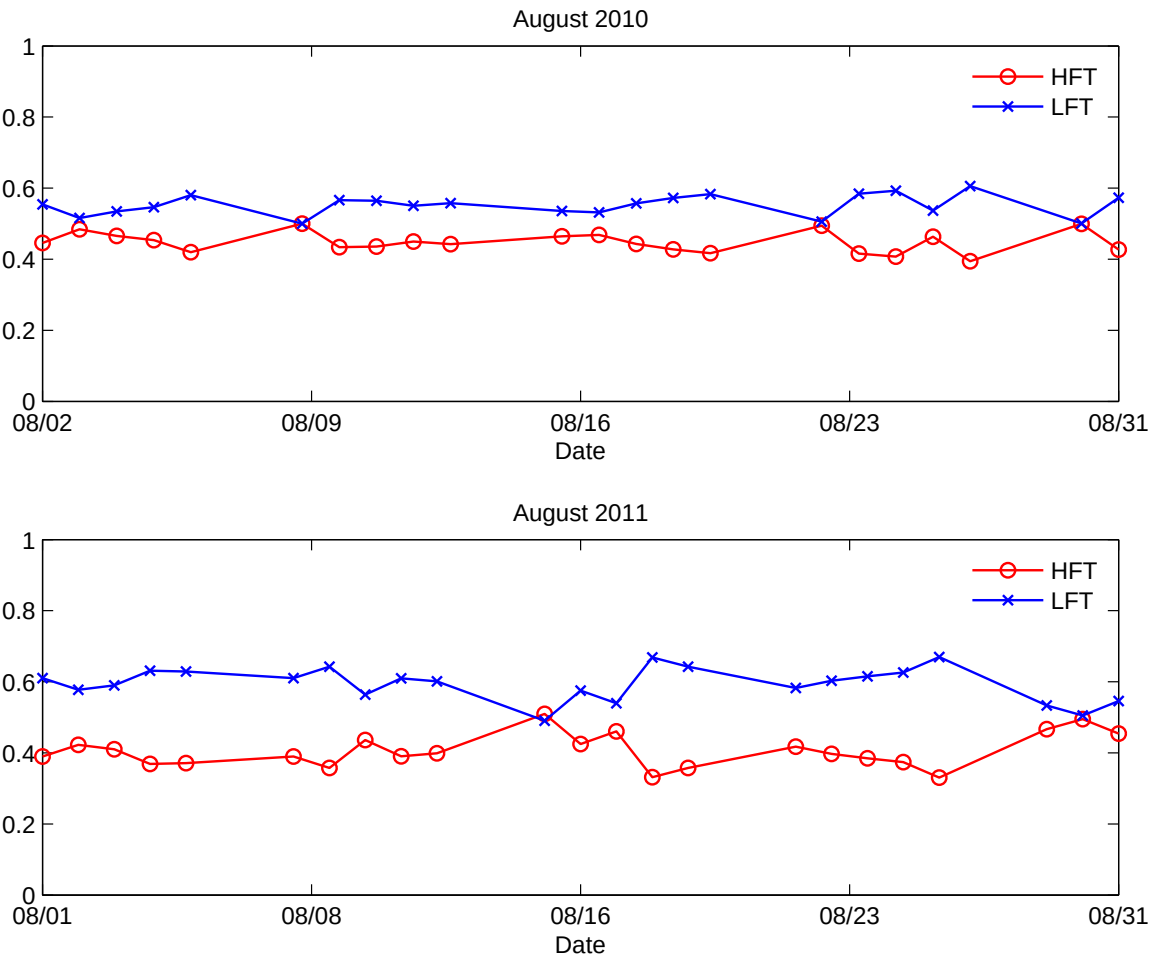

Figure 5: Daily Provision of Liquidity

Notes: This figure presents the daily average provision of liquidity which we determined using the aggressor flag. The aggressor flag indicates whether the order that generates the trade is incoming or resting on the limit order book. For example, if a high frequency trader is the aggressor in a trade, we consider this high frequency trader as the "taker" of liquidity. We only consider transactions HL and LH (high frequency trader buying from a low frequency trader and a low frequency trader buying from a high frequency trader).

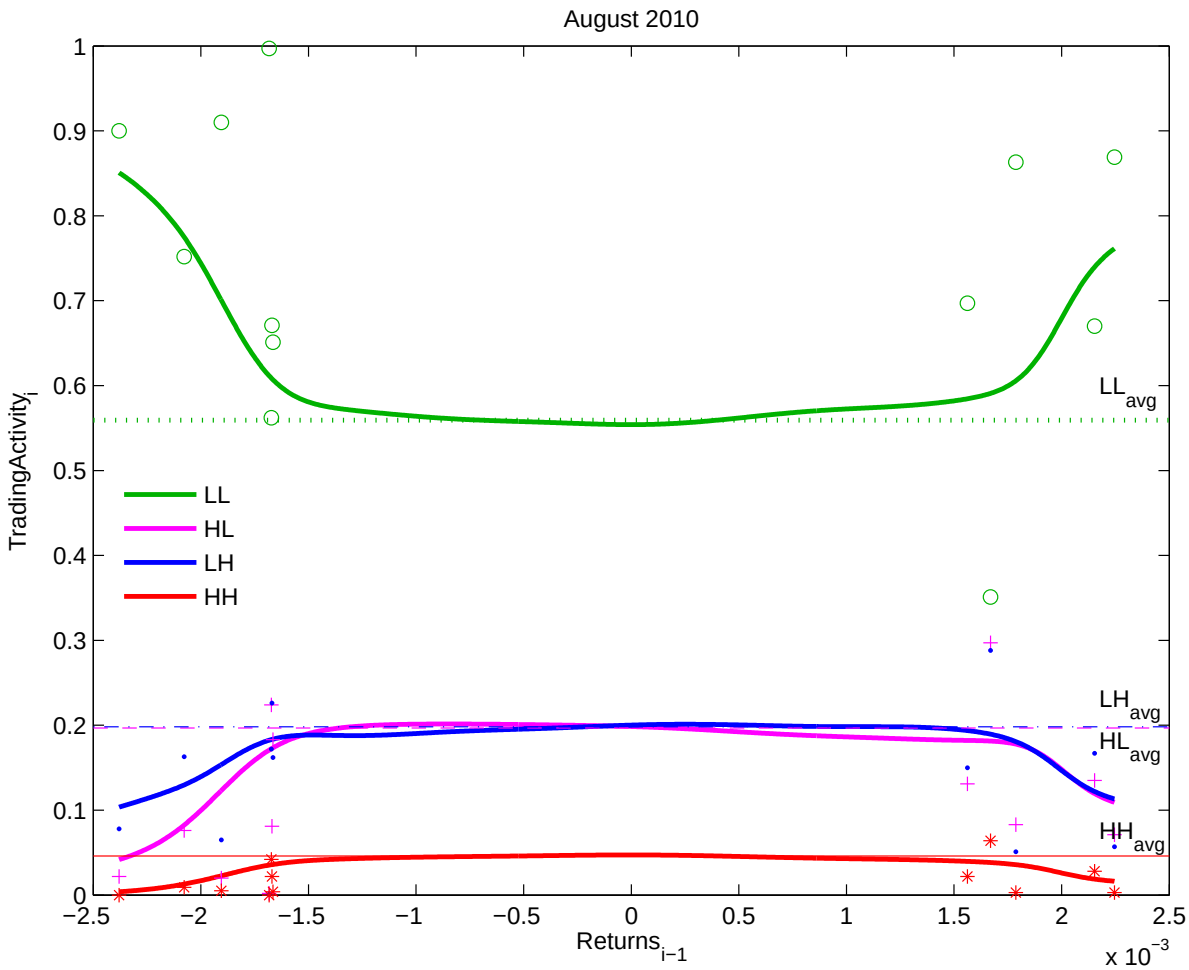

Figure 6: Trading Activity Conditional on Past Returns

Notes: This figure reports kernel smoothing regressions relating trading activity, $TA_{j,i}$, at time i, (see equation (4.3)), to rate of returns (%) at time $i-1$. Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. The length of the subinterval, $k = \tau_i - \tau_{i-1} = 1,000$ transactions. The rate of return is computed as $R_i = Y_{\tau_i} - Y_{\tau_{i-1}}$. The plot also reports individual extremal values (defined as $R_i \geq \pm 1.5\%$): Circles (in green) refer to transactions between two low frequency traders (LL); asterisks (in red) refer to transactions between two high frequency traders (HH); dots (in blue) indicate transactions where a low frequency trader is buying from a high frequency trader (LH); crosses (in purple) indicate transactions where a high frequency trader is buying from a low frequency trader (HL). Horizontal lines indicate the unconditional average trading activity for the four types of transactions.

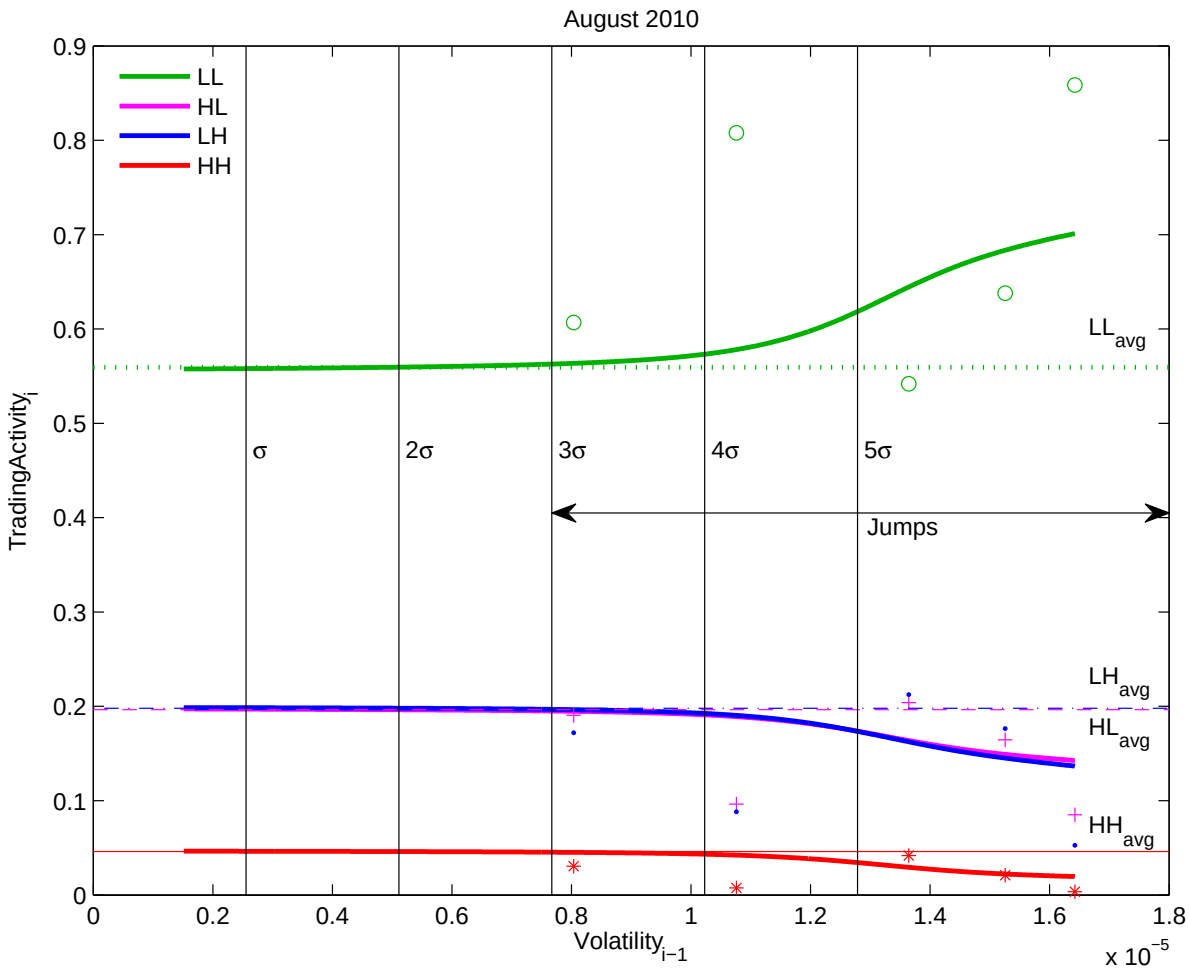

Figure 7: Trading Activity Conditional on Volatility and Jumps

Notes: This figure contains kernel smoothing regressions relating trading activity, $TA_{j,i}$, at time $\not{i}$ (see equation (4.3)), to volatility at time $i-1$. Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. For each bucket i of length $k = 10,000$ trades, we compute realized volatility sampling every $w = 100$ trades. For each bucket we compute k/w measures of realized volatility and then we average them to obtain RV_i^{avg} for that specific bucket i (see Zhang et al. (2005)). σ refers to the average volatility (RV^{avg}) across buckets for the entire month of August 2010. Vertical lines indicate different integer levels of average volatility. The plot also reports jumps (defined as $|R_i| \geq 3\sigma$): Circles (in green) refer to transactions between two low frequency traders (LL); asterisks (in red) refer to transactions between two high frequency traders (HH); dots (in blue) indicate transactions where a low frequency trader is buying from a high frequency trader (LH); crosses (in purple) indicate transactions where a high frequency trader is buying from a low frequency trader (HL).

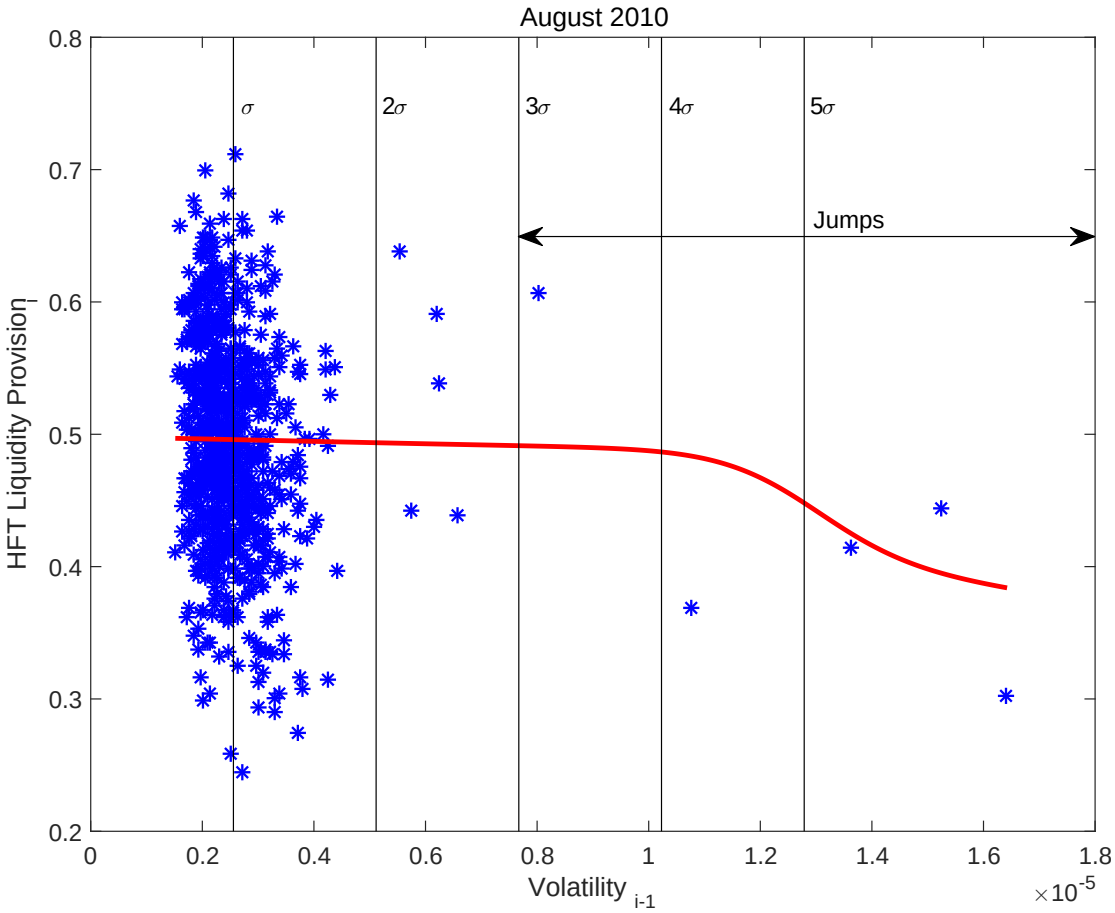

Figure 8: Liquidity Provision by HFTs Conditional on Volatility and Jumps

Notes: This figure reports a kernel smoothing regression relating liquidity provision by HFTs at time i (see equation (4.4)), to volatility and jumps at time $i-1$. Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. For each bucket i of length $k = 10,000$ trades, we compute realized volatility sampling every $w = 100$ trades. For each bucket we compute k/w measures of realized volatility and then we average them to obtain RV_i^{avg} for that specific bucket i (see Zhang et al. (2005)). σ refers to the average volatility (RV^{avg}) across buckets for the entire month of August 2010. Vertical lines indicate different integer levels of average volatility and jumps (defined as $|R_i| \geq 3\sigma$). HFT Liquidity provision is computed as the percentage of transactions in time bucket i where the aggressor (i.e., the taker of liquidity) is a LFT. We only consider transactions HL and LH (high frequency trader buying from a low frequency trader and a low frequency trader buying from a high frequency trader).

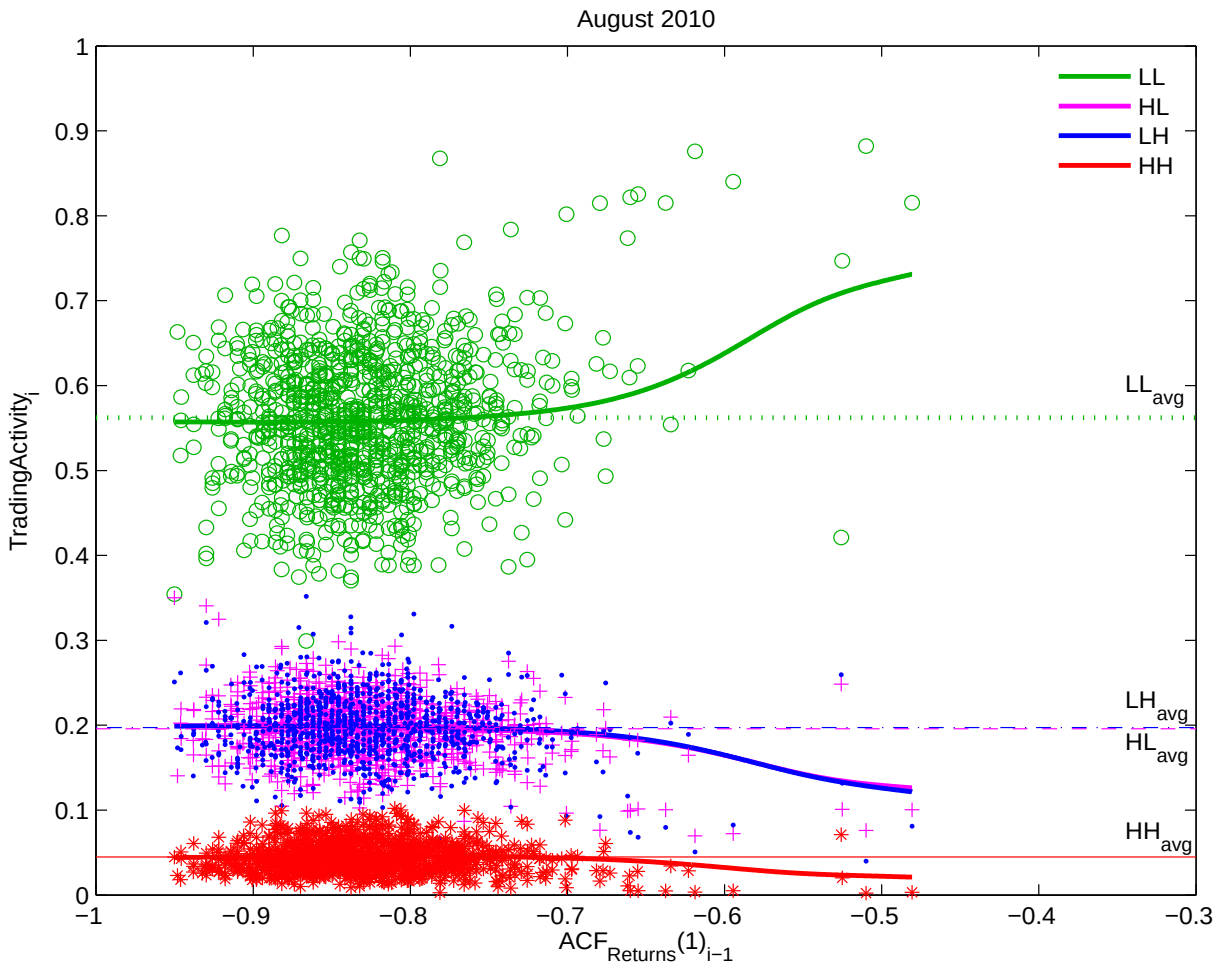

Figure 9: Trading Activity Conditional on First Order Returns Autocorrelation

Notes: This figure contains kernel smoothing regressions relating trading activity, $TA_{j,i}$, at time i (see equation (4.3)), to first order autocorrelation of returns at time $i-1$. Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. For each bucket i of length $k = 500$ ticks, we compute $\text{ACF}(1)_i$ of the rate of returns within that bucket. Circles (in green) refer to transactions between two high frequency traders (HH); asterisks (in red) refer to transactions between two low frequency traders (LL); dots (in blue) indicate transactions where a low frequency trader is buying from a high frequency trader (LH); crosses (in purple) indicate transactions where a high frequency trader is buying from a low frequency trader (HL).

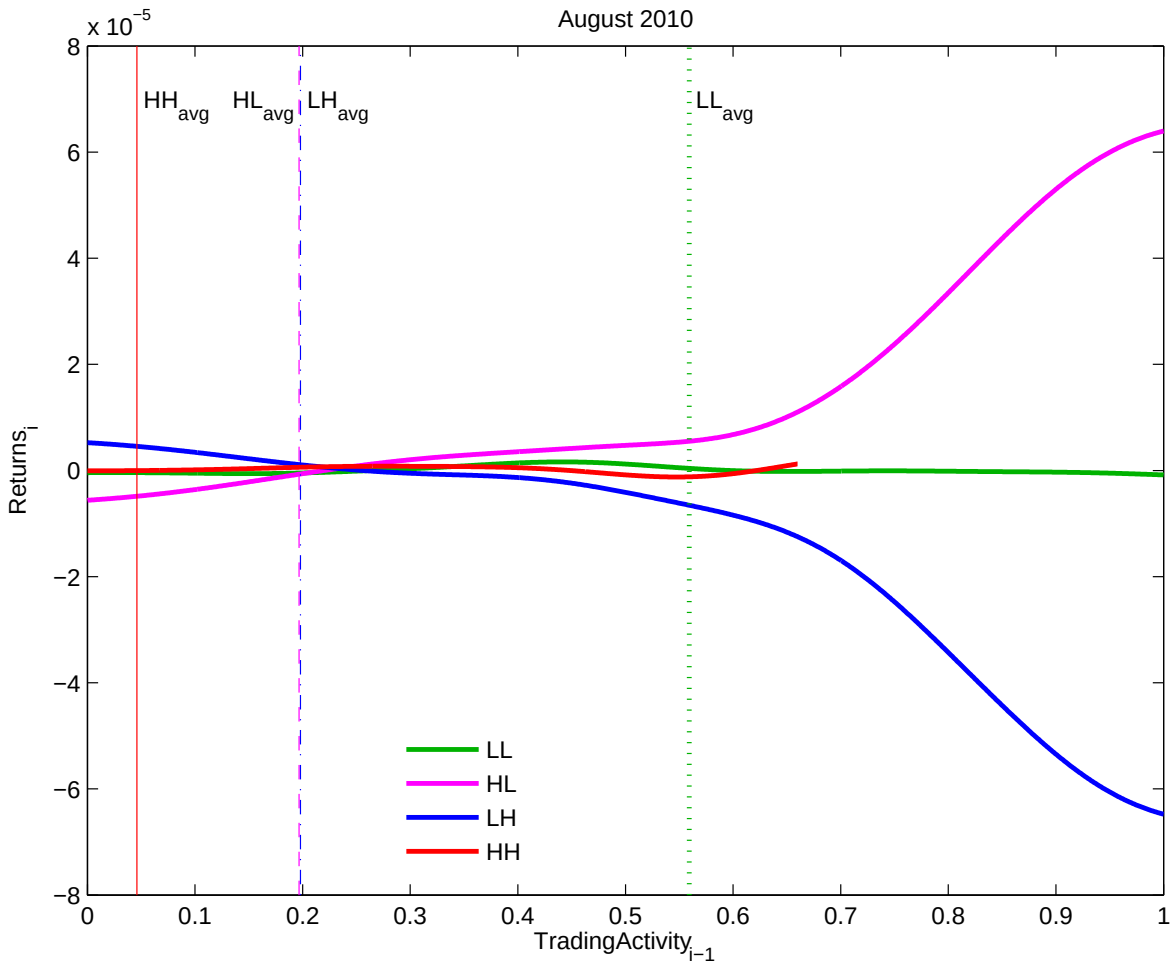

Figure 10: Returns Conditional on Trading Activity

Notes: This figure reports kernel smoothing regressions relating returns (%) at time i to trading activity, $TA_{j,i-1}$, at time $i-1$, (see equation (4.3)). Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. The length of the subinterval, $k = \tau_i - \tau_{i-1} = 50$ transactions. The rate of return is computed as $R_i = Y_{\tau_i} - Y_{\tau_{i-1}}$. Vertical lines indicate the unconditional average trading activity for the four types of transactions.

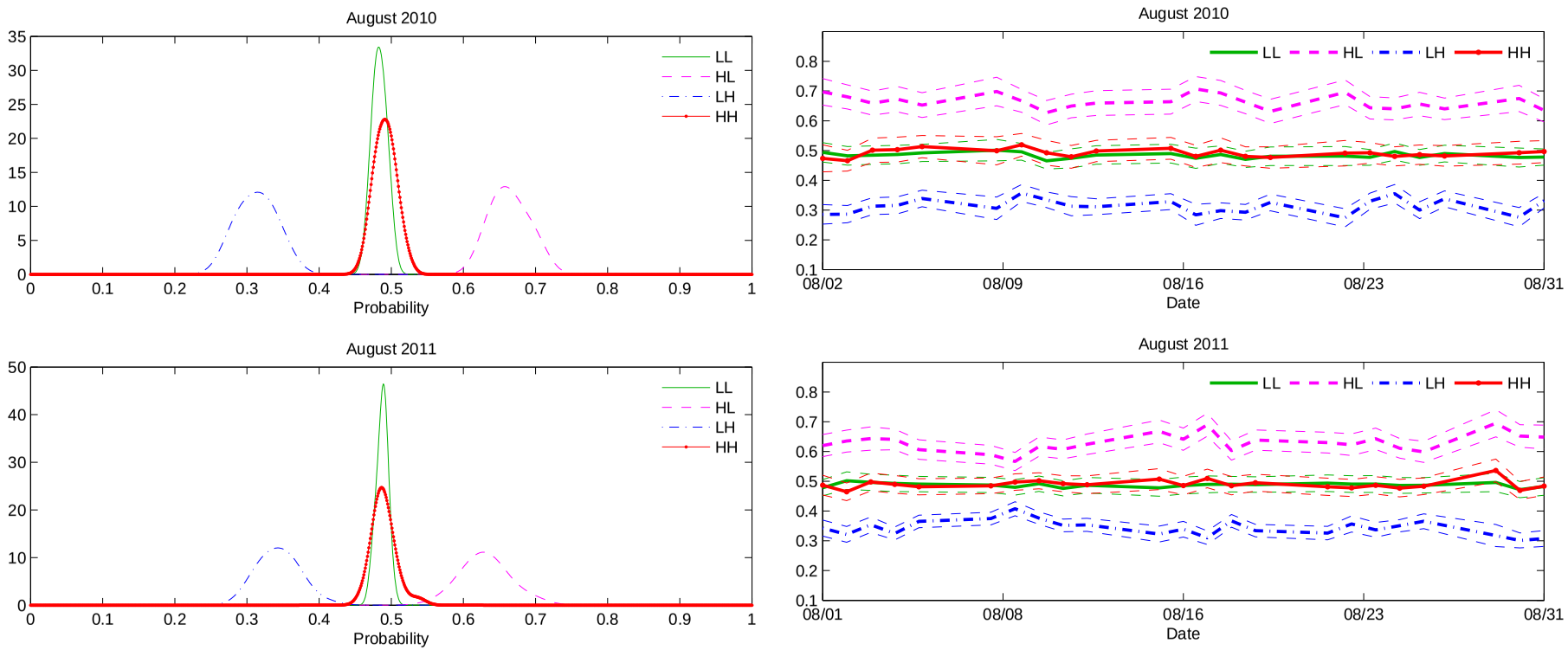

Figure 11: Up-tick Probability Distributions and Time Series

Notes: This figure depicts the kernel density of an up-tick conditional on trading activity (top panels) and the time series of up-tick probabilities (bottom panels) following each of the four transaction types LL, HL, LH and HH for August 2010 and August 2011. Standard errors (light, broken lines) are computed as intramonth variation of up-tick probabilities.

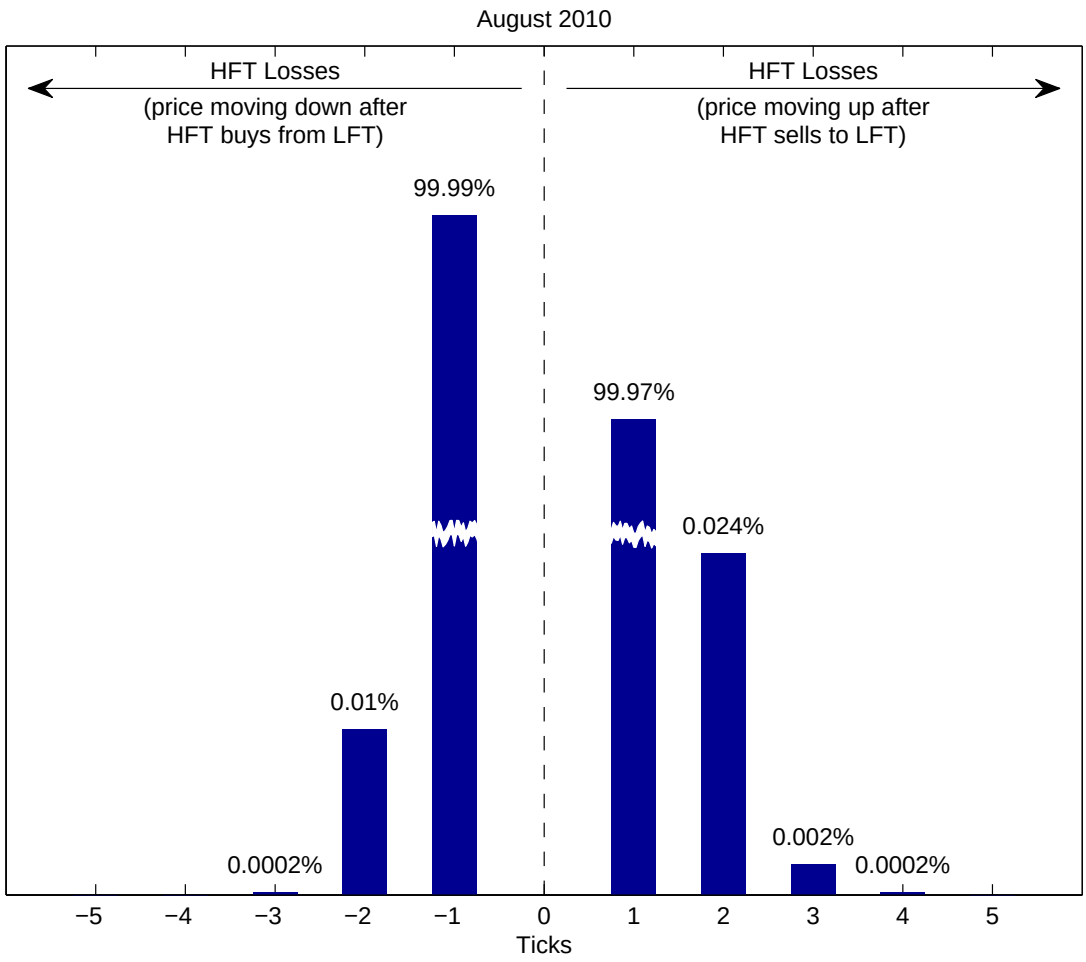

Figure 12: Distribution of High Frequency Traders' Losses

Notes: This figure plots the distribution of losses of high frequency traders. 99.9% of losses occur at ±1-*tick*.

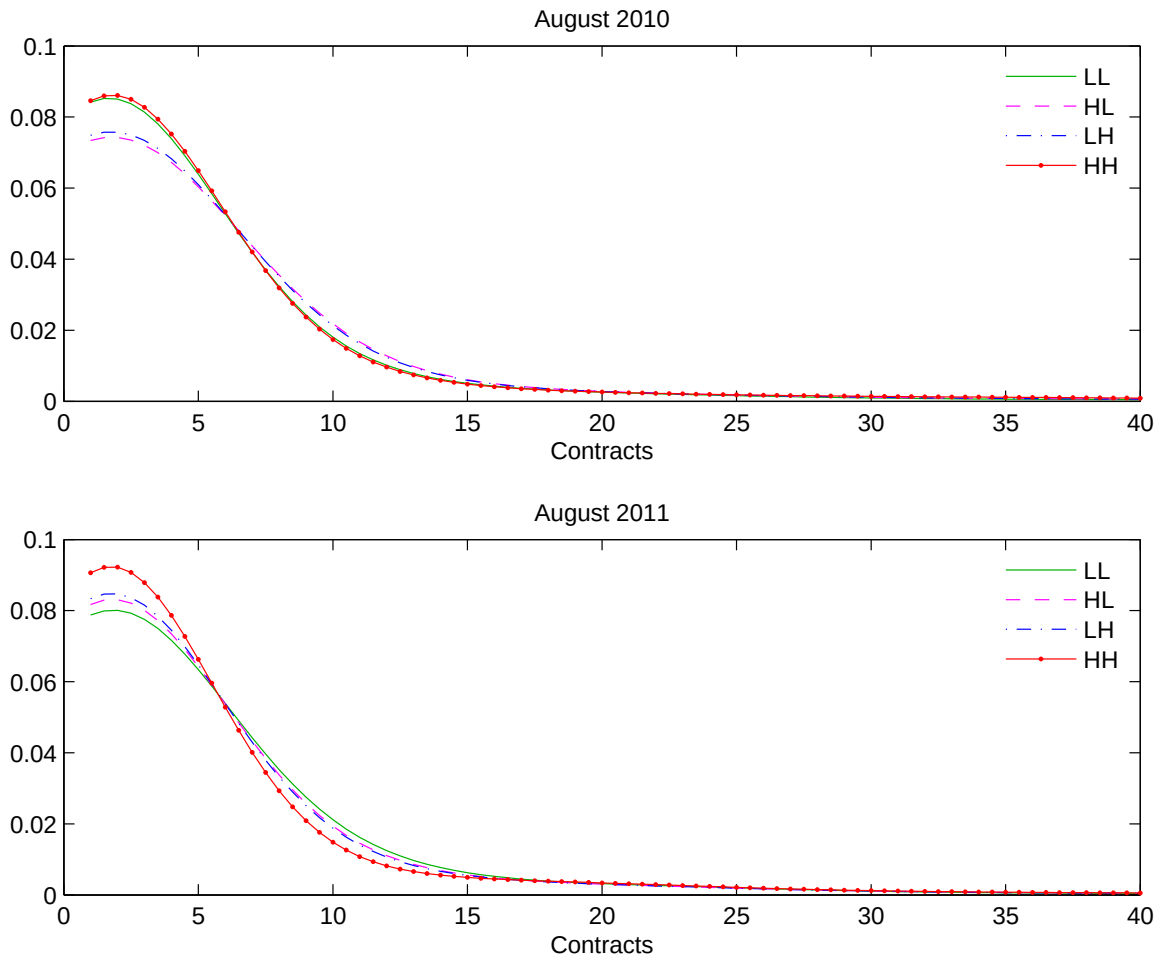

Figure 13: Distribution of Lot Size

Notes: This figure presents kernel densities for the lot size by trading activity.

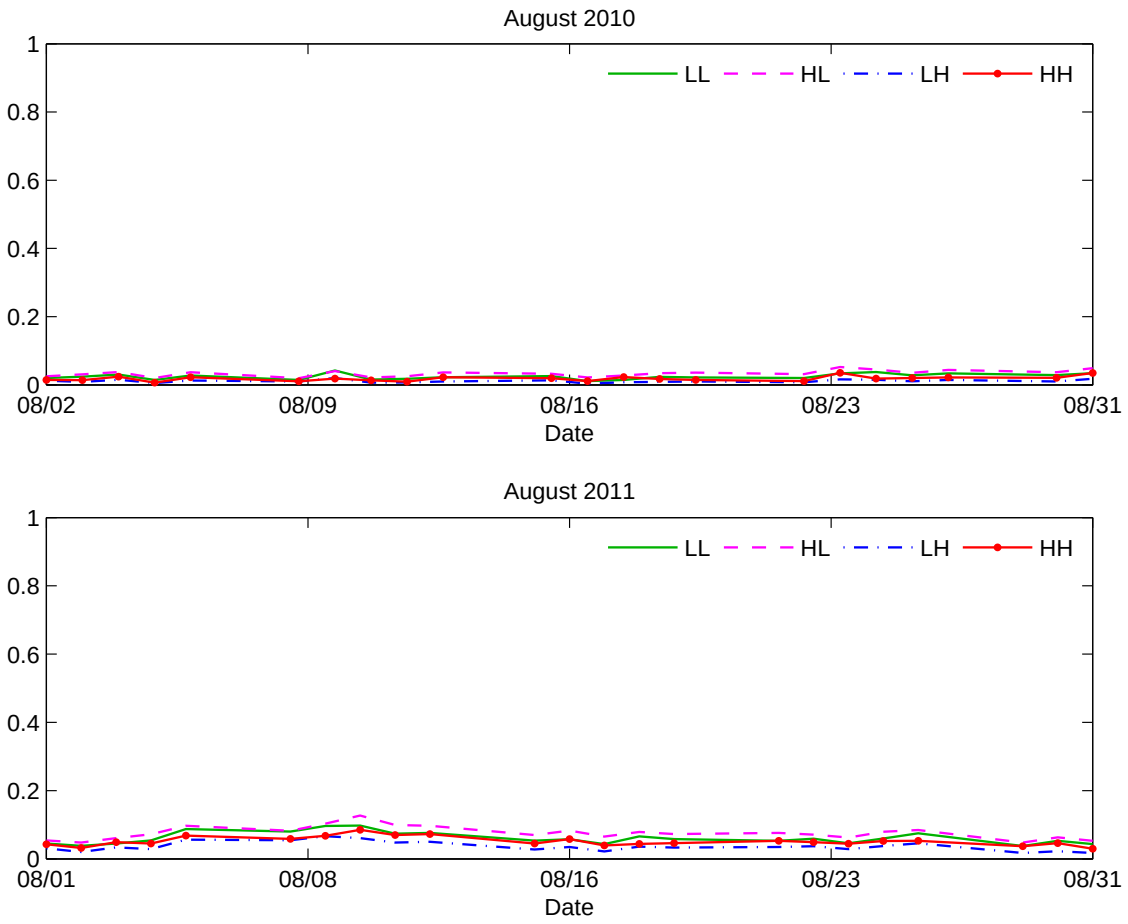

Figure 14: Probability that the Second Tick Will Be an Up-Tick, Given that the First One is an Up-Tick

Notes: This figure depicts the time series that the second tick will be an up-tick conditionally on the first tick being an up-tick, following each of the four transaction types LL, HL, LH and HH for August 2010 and August 2011. The plot shows strong evidence of mean-reversion in tick time.

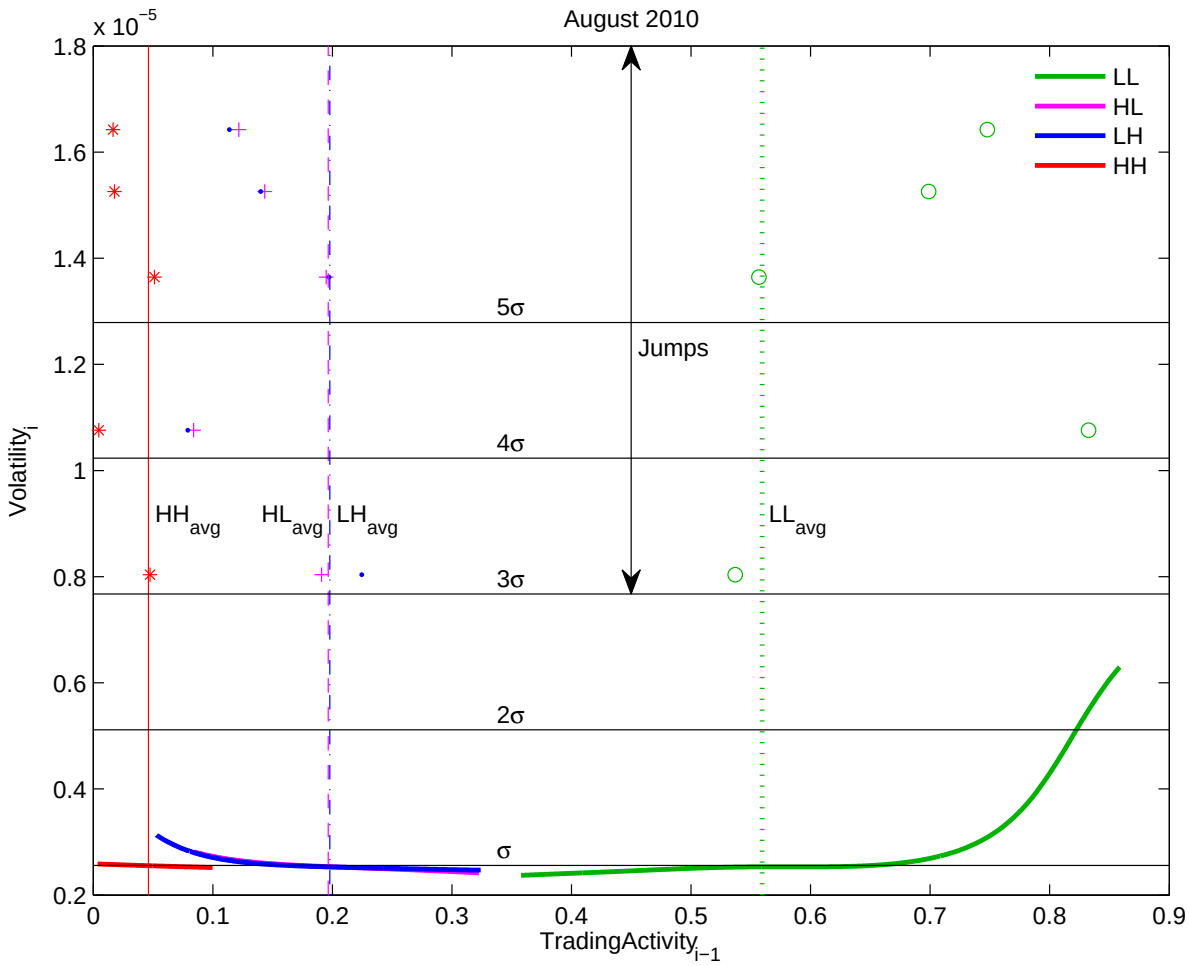

Figure 15: Volatility and Jumps Conditional on Trading Activity

Notes: This figure contains kernel smoothing regressions relating volatility at time i to trading activity, $TA_{j,i-1}$, at time $i-1$ (see equation (4.3)). Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. For each bucket i of length $k = 10,000$ trades, we compute realized volatility sampling every $w = 100$ trades. For each bucket we compute k/w measures of realized volatility and then we average them to obtain RV_i^{avg} for that specific bucket i (see Zhang et al. (2005)). σ refers to the average volatility (RV^{avg}) across buckets for the entire month of August 2010. Horizontal lines indicate different integer levels of average volatility. The plot also reports jumps (defined as $|R_i| \geq 3\sigma$): Circles (in green) refer to transactions between two low frequency traders (LL); asterisks (in red) refer to transactions between two high frequency traders (HH); dots (in blue) indicate transactions where a low frequency trader is buying from a high frequency trader (LH); crosses (in purple) indicate transactions where a high frequency trader is buying from a low frequency trader (HL). Vertical lines indicate the unconditional average trading activity for the four types of transactions.

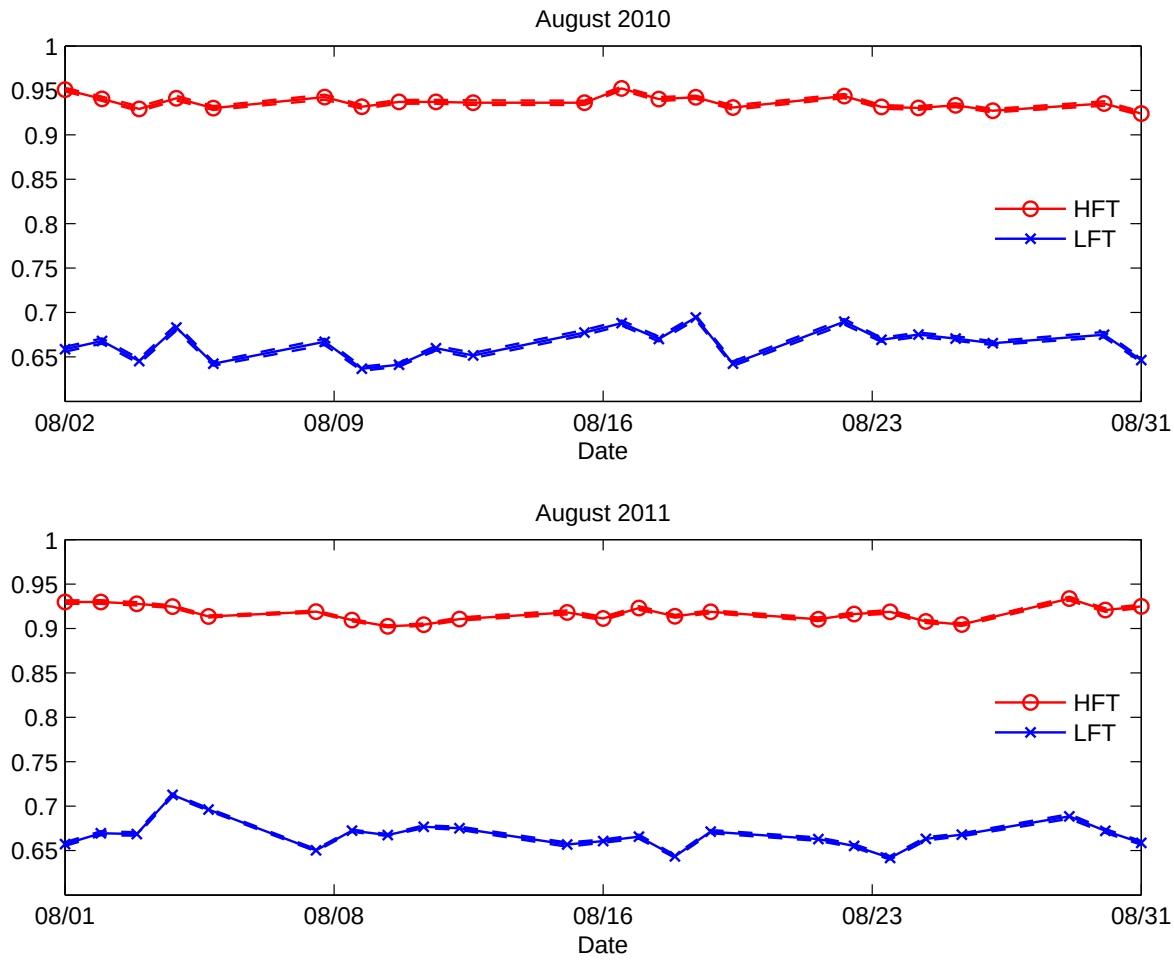

Figure 16: Same Side Trading by HFTs and LFTs

Notes: This figure plots the percentage of trading activity done by HFTs and LFTs on the same side of the market, buying or selling. Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. The length of each bucket i is $k \geqslant 1$ tick. We compute these percentages by counting the number of times HFTs and LFTs are respectively on the same side of the market within each tick-time bucket.

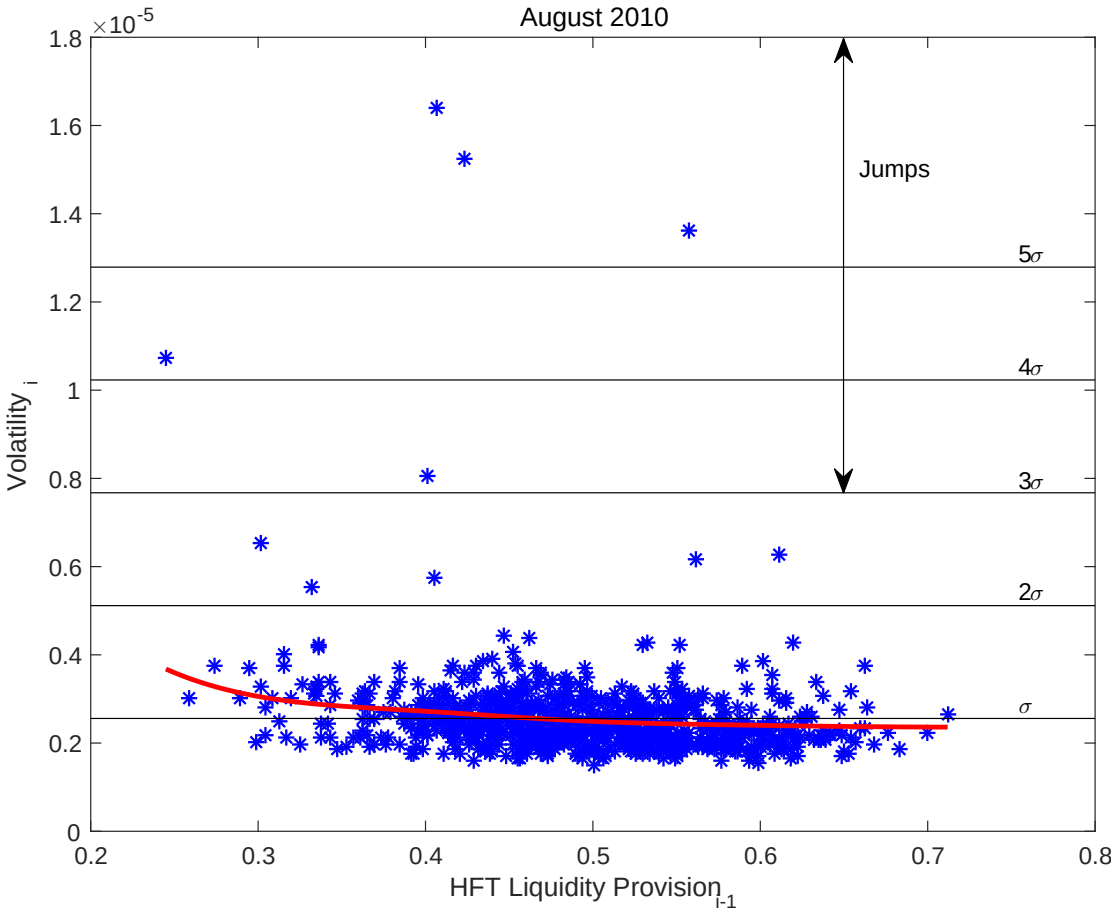

Figure 17: Volatility and Jumps Conditional on Provision of Liquidity by HFTs

Notes: This figure reports a kernel smoothing regression relating volatility and jumps at time i to liquidity provision by HFTs at time $i-1$ (see equation (4.4)). Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. For each bucket i of length $k = 10,000$ trades, we compute realized volatility sampling every $w = 100$ trades. For each bucket we compute k/w measures of realized volatility and then we average them to obtain RV_i^{avg} for that specific bucket i (see Zhang et al. (2005)). σ refers to the average volatility (RV^{avg}) across buckets for the entire month of August 2010. Vertical lines indicate different integer levels of average volatility and jumps (defined as $|R_i| \geq 3\sigma$). HFT Liquidity provision is computed as the percentage of transactions in time bucket i where the aggressor (i.e., the taker of liquidity) is a LFT. We only consider transactions HL and LH (high frequency trader buying from a low frequency trader and a low frequency trader buying from a high frequency trader).

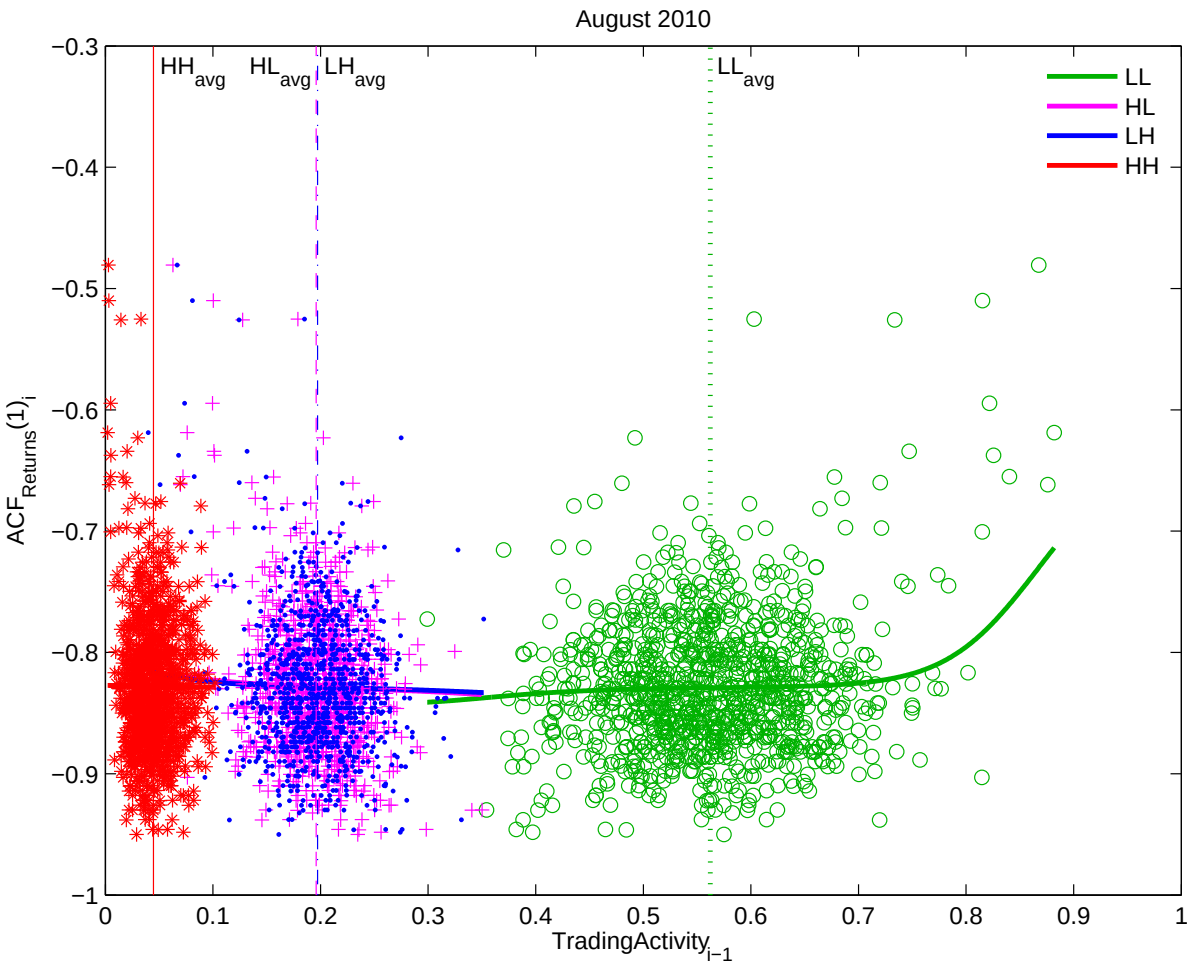

Figure 18: First Order Returns Autocorrelation Conditional on Trading Activity

Notes: This figure contains kernel smoothing regressions relating the first order autocorrelation at time i to trading activity $TA_{j,i-1}$, at time $i-1$ (see equation (4.3)). Time is partitioned in subintervals (buckets) $\underline{\tau} = \tau_0 < \tau_1 < \tau_2 < ... < \tau_m = \overline{\tau}$. For each bucket i of length $k = 500$ ticks, we compute $\text{ACF}(1)_i$ of the rate of returns within that bucket. Circles (in green) refer to transactions between two high frequency traders (HH); asterisks (in red) refer to transactions between two low frequency traders (LL); dots (in blue) indicate transactions where a low frequency trader is buying from a high frequency trader (LH); crosses (in purple) indicate transactions where a high frequency trader is buying from a low frequency trader (HL).