Title

The impact of high-speed quoting on execution risk dynamics: Evidence from interest derivatives markets

Accepted for Publication in the Journal of Futures Markets

Authors

Jing Nie
School of Banking and Finance, University of International Business and Economics, China Beijing,

Juliana Malagon
School of Management, Universidad de Los Andes, Bogotá, Colombia

Julian Williams
Durham Business School, Durham University, Durham, UK

The impact of high speed quoting on execution risk dynamics: Evidence from interest derivatives markets

Abstract

This paper intends to characterize the effect of high-frequency quoting (HFQ) on the execution risk of Eurodollar futures. We construct a unique dataset to capture the quoting and trading activities within the limit order book, which allows us to classify the realised fraction of HFQ activity within the market. We then estimate the marginal effect of the HFQ fraction on the execution risk through a novel semi-parametric regression. The results suggest that the effect of HFQ on market quality is non-linear with critical saturation levels. The HFQ effects on market quality seem to disappear once certain critical points are reached.

Keywords: Execution Risk, High-frequency Quoting, Market Quality, Semi-parametric Model
JEL Classification: G12, G14

High frequency quoting (HFQ), often thought of as automated quoting at speeds beyond human reaction times, is considered one of the greatest innovations in the financial markets. Given its current relevance and the lack of consensus about its impact on market quality, HFQ is the subject of considerable policy attention. Theoretically, the positive effects of HFQ derive from lower trading costs while the negative effects are attached to adverse selection on the part of slower investors induced by their HFQ counterpart. Scholars have not yet established whether these positive or negative effects dominate; therefore, empirical studies exploring the impacts of HFQ are divided in their conclusions. Among the studies supporting the goodness of HFQ, Hendershott et al. (2011) find that the introduction of the auto-quote facility for US equities has a significantly positive effect on a number of liquidity proxies. Using statistical experiments based on the introduction of technologies providing new avenues for HFQ, Hendershott and Moulton (2011) and Riordan and Storkenmaier (2012) draw similar conclusions about HFQ as a net provider of liquidity. Considering various markets and asset classes, Harris (1989), Stoll and Whaley (1990), Chan (1992), Huang and Stoll (1994), Engle et al. (2012), Frino and McKenzie (2002) and Menkveld and Zoican (2014) either find evidence to suggest that HFQ improves market characteristics or find no evidence that the liquidity is compromised by the presence or introduction of HFQ. Brogaard et al. (2018) propose that when the US equity market experiences extreme price movement, high-frequency traders (HFTs) provide liquidity, however, they are not the cause for extreme price movement. Employing a quote-to-trade ratio used by market makers, Rosu et al. (2021) suggests that a generous number of quotations reduces mispricing and lowers expected returns.

There is abundant evidence for the negative impact of high frequency quoting. In a recent empirical work using a random sample of intensely traded US equities, Hasbrouck (2018) find that HFQ generates a much higher quote volatility than if it was traced only to fundamentals. In addition, HFQ has a negative impact on the informativeness of the US stock market. In a similar vein, Kirilenko et al. (2017) illustrate that while high-frequency trading (HFT) did not trigger the 2010 Flash Crash, HFTs do exacerbate the degree of price volatility. Using the 2008 short-sale ban, Brogaard (2017) study the different impacts from HFTs and non-HFTs on market liquidity and find that some HFQ activities erode market liquidity during the highly volatile short-sale ban period. From a different perspective, Chakrabarty et al. (2014) find little evidence of a deleterious effect of the Securities and Exchange Commission (SEC) banning naked access to US equity exchanges (which hinders the supposed advantage of HFTs) and cite several positive outcomes from this regulatory action. Furthermore, theoretical contributions from Foucault et al. (2015) and Kyle and Obizhaeva (2016) show the negative effects of HFT by indicating that the benefits of HFT are outweighed by a subset of traders having substantially greater opportunities simply through their high-frequency access.

This paper contributes empirical evidence to the discussion on the effects of HFQ on market quality. Compared with earlier studies, our analysis focuses on the Eurodollar futures market because it offers characteristics that allow more general inferences. On the one hand, a markedly fast market populated by a relatively small number of high-frequency players, such as the Eurodollar futures market, provides a unique opportunity to study the potential effects of HFQ on market quality. On the other hand, HFT studies are often characterised by the use of proprietary data with a corresponding burden of data intensiveness. Consequently, most of the available literature on HFT are either based on random securities samples with biased results from the choice of samples or use data that other scholars cannot access with results that are difficult to replicate.[1] The methodology we propose overcomes this shortfall, which enables traders to both work with the population of trades using the CME venue and identify transactions generated at such high frequency that they are likely to be related to HFQ, which is not trivial when working with non-proprietary data.[2]

To develop our study, we compute a proxy for the fraction of HFQ using order flow data. We also run a semi-parametric instrumental variable regression to determine the impact of increasing the HFQ fraction over both calendar time and the term structure of the contracts. Our results provide rich detail for interpretation. First, this paper reveals that the marginal effect of HFQ order flow on market quality not only changes in direction, but also in order of magnitude over the term structure as the level of ambient quoting activity changes. In this sense, our results go some way to explaining why there is such a lack of consensus in the literature on the impact of HFQ, such as studies looking at ultra-high-frequency data over short time scales (e.g. all available messages for 1 day, such as Kirilenko et al. (2017)). Second, this paper also extracts the causal effects of exogenous HFQ adjustments on the market by instrumenting our regression system. Third, in the absence of the identifying account numbers, the findings suggest that public order flow data from the limit order can be used to obtain comparable results to studies based on proprietary data, such as CME Group (2010).Finally, our results suggest that there is no reason to doubt the methodological or theoretical integrity of either side of the HFQ debate. In fact, both sets of studies can be validated for both calendar time and term structure in the Eurodollar futures market.

The remainder of this paper is organised as follows. Section 1 outlines our theoretical development using an asymmetric information-based model for futures trading. In Section 2, we present

[1]In a relatively large study of the impact of HFQ order flow on their Globex platform during the aftermath of the May 2010 Flash Crash, for instance, the Chicago Mercantile Exchange (CME) find that high-frequency order flow actually increases liquidity (CME Group, 2010). In addition, CME uses their proprietary data to determine if an account connection to the Globex exchange is generating orders following instructions from an individual (i.e. a responsible party) or a computer algorithm (i.e. a developer group). These proprietary data are not available to public data users.

[2]It is important to note that the CME is not the only venue where Eurodollar futures are traded. For example, investors can also trade them at the Intercontinental Exchange (ICE) Europe Futures venue. In terms of the volume of Eurodollar futures traded at the CME, however, the ICE Europe Futures is a small venue. While it is possible that important spillovers occur from one venue to the other, studies of their impacts are beyond the scope of this paper.

our empirical methodology to calculate a set of market quality indicators and high-frequency order flow proxies. Then, we also introduce a novel semi-parametric partially linear regression to estimate the relationships between the fraction of HFQ order flow and execution risks in the markets. In Section 3, we describe our Eurodollar futures dataset and provide a general characterisation of the market. Section 4 includes an analysis of the marginal impact of HFQ on market quality, and Section 5 concludes. For robustness, we conduct a significant number of ancillary treatments, which are presented in an extensive online appendix.[3]

1. A Theoretical Model for Latency-advantaged Traders in Futures Markets

Considering that the Eurodollar futures market is naturally prone to the presence of informed trading, we begin our analysis by developing a simple linear noisy rational expectations model in the spirit of Admati and Pfleiderer (1988, 1989) and Watanabe (2008), by highlighting the impact of the small number of traders with significant trading advantages. Our model predicts that the fraction of informed trading should exhibit a non-linear relationship with our execution risk and other market quality indicators, whereby the impact of the small number of traders with significant trading advantages can have a deleterious effect on market quality. However, this pattern is reversed when more traders enter the market. Our model is based on two futures traded simultaneously in the market and $\tilde{N}$ traders, separated between informed traders, denoted by N_A, and uninformed traders, denoted by N_B, such that $\tilde{N} = N_A + N_B$. Although we only present the case of two futures, that is, a short maturity future denoted by the subscript S and a long maturity one denoted by the subscript L, our results are generalisable and analytically tractable for n_f futures. We model 1 day of trading; therefore, we set (t) as a continuous time index, such that $t \in [0, T]$ where $T = 1$ for 1 day.

Let $\tilde{f}(t) = [\tilde{f}_S(t), \tilde{f}_L(t)]'$ for $\tilde{f}(t) \in \mathbb{R}^2$ be the log prices of a pair of futures contracts with maturity dates $\bar{T}_S$ and $\bar{T}_L$, respectively. Both maturity dates are such that $\bar{T}_L > \bar{T}_S$ and $\bar{T}_{i\in\{L,S\}} > \bar{T}$. This implies that neither future is maturing within the day's trading and there is no final

[3] The online appendix is available at https://drive.google.com/file/d/1nINGyzDyMyKpAhCCsXQU4gN_sqXqAmmF/view?usp=sharing

settlement.[4] In the Eurodollar market, where the reference yield curve from the dollar London Interbank Offered Rate (LIBOR) is revealed several hours prior to the daily settlements, we assume that the reference rate is revealed such that the closing price for the day at mark to market, $\tilde{\delta}(t) - \tilde{f}(\bar{T})$, is the difference between the marked to market futures price, $\tilde{\delta}(t)$ and the fair valuation, $\tilde{f}(\bar{T})$.[5]

We presume that all trading takes place within a single day. Our model is slightly stylised in the sense that the traders, while being marked to market by the stated prices at closing, value their positions at the agreed equilibrium price, which follows the original Kyle-type model. However, we examine a bivariate long and short maturity model directly. We set the dynamics of the underlying valuation of a futures position as:

$$\mathrm{d}\tilde{\delta}(t) = \tilde{\Sigma}^{1/2} d\tilde{W}(t) \tag{1}$$

where $\tilde{W}(t)$ is a two-dimensional Brownian motion and subsequently $\tilde{\delta}(t + \Delta t) - \tilde{\delta}(t) \sim \mathcal{N}(0, \Delta t \tilde{\Sigma})$ over the interval $0 \leq t \leq T$. We presume that a maturity effect exists; therefore, we set $\Sigma = [\sigma_{ij}]_{ij \in \{S,L\}}$ such that $\sigma_{SS} > \sigma_{LL}$. Under the fair valuation assumption, the final true value of futures $\tilde{\delta}(T)$ is known and agreed on with certainty by all participants immediately after time T.

Uninformed traders submit a random aggregate order flow denoted by vector $\tilde{d}$ for each of the futures over the interval $[0, \bar{T}]$. This order flow is presumed to have unconditional moments $\mathbb{E}_0[\tilde{d}] = 0$ and covariance matrix covariance matrix $\mathbb{E}_0[\tilde{d}\tilde{d}'] = N_B \tilde{\Psi}$. In the approaches by Kyle (1985), Glosten and Milgrom (1985) and Admati (1985) these traders are noisy or liquidity traders who provide aggregate liquidity into the market.[6]

The N_A group comprises algorithmic HFTs who trade off information about order flows. These traders are presumed to have a technical advantage over institutional traders because they can track order flow directions and have an unbiased but noisy estimate of the terminal valuation $\tilde{\delta}(\bar{T})$.

[4] Recall that the mark to market prices in a futures market are assumed to be set at the end of a day's trading.

[5] Recall that the Eurodollar futures market is a 24-hr market with settlement at 20:00 CET.

[6] These institutional traders are forced to create specific positions in the market.

Therefore, they all receive a signal such that

$$\tilde{\delta}_A(\bar{T}) = \tilde{\delta}(\bar{T}) + \tilde{\zeta} + \tilde{\xi} \tag{2}$$

where $\tilde{\zeta} \in \mathcal{N}(0, \tilde{\Gamma})$ and $\tilde{\xi} \in \mathcal{N}(0, \tilde{\Phi})$ are the global-specific and trader-specific noise vectors, respectively.[7] For tractability, we do not model traders' instantaneous profit function; however, we model their objective in the form of their integrated profit function. Let $\tilde{a}_n$ be the aggregate order flow of the $n \in N_A$ trader, then the aggregate order flow is given by:

$$\tilde{a}_n^* = \arg\max_{\tilde{a}_n} \mathbb{E}[(\tilde{\delta}(\bar{T}) - \tilde{f}(\bar{T}))' \tilde{a}_n] \tag{3}$$

Finally, let $\tilde{e} = \sum_{n=1}^{N_A} \tilde{a}_n^* + \tilde{d}$ be the aggregate order flow across both futures contracts. Notice that in a centrally cleared trading platform, such as CME, the clearing house matches and clears all available trades and then places the remainder in the limit order book. This mechanism can be considered as a price-matching model to promote efficient information clearing, such that $\tilde{\delta}(t) - \tilde{f}(\bar{T}) \to 0$. Therefore, the pricing rules for market-clearing mechanisms follow the principle that $\mathbb{E}_0[\tilde{f}(\bar{T})] = \tilde{\delta}(t)$.

The model can be expressed as a simple one-period Kalman filter assuming that the positive semi-definite matrix $\tilde{\Theta} \in \{\tilde{\Sigma}, \tilde{\Gamma}, \tilde{\Phi}, N_B\tilde{\Psi}\}$, representing the structural parameters matrices, and the scalar N_A are known a-priori by all participants.[8] In this framework, it is possible to derive the market equilibrium under rational expectations stated on the following Theorem 1.

Theorem 1. *A Noisy Linear Rational Expectation Market Equilibrium. Let the observed price of the futures contracts be $\tilde{f}(t) = \tilde{\delta}(t) + \tilde{\Lambda}\tilde{e}$, where $\tilde{\Lambda}$ is the multivariate equivalent of "Kyle's lambda", following the definition in Admati (1985) and Watanabe (2008) and $\tilde{e}$ is the net order flow, with the algorithmic traders' linear noisy rational expectations equilibrium given by $\tilde{a}_n^* = \tilde{B}\tilde{\delta}(t)$. The market clearing equilibrium under noisy linear rational expectations (NLRE) is defined by the following*

[7] Here, we depart from the standard approach by Kyle (1985), Glosten and Milgrom (1985) and Admati (1985).

[8] Recall that $\tilde{\Sigma}$ refers to the variance covariance matrix of the true valuation price process, $\tilde{\Gamma}$ is the covariance matrix of the global noise across all A traders signals, $\tilde{\Phi}$ is the covariance matrix for each A traders signal, and $\tilde{\Psi}$ represents the variance of an individual B traders order flow.

equilibrium equations of state:

$$
\begin{aligned}
\tilde{\Lambda} &= \sqrt{N_A}(N_B\tilde{\Psi})^{-1/2}\tilde{M}^{-1/2}(N_B\tilde{\Psi})^{-1/2}, \quad \tilde{B} = \tilde{\Lambda}^{-1}\tilde{J}^{-1}\tilde{\Sigma}^{-1} && (4)\\
\tilde{M} &= \sqrt{N_B}\tilde{\Psi}^{1/2}\tilde{J}^{-1}\tilde{\Sigma}_\xi\tilde{J}'^{-1}\sqrt{N_B}\tilde{\Psi}^{1/2}, \quad \tilde{\Sigma}_\xi = \tilde{\Sigma}(\tilde{\Sigma}+\tilde{\Gamma}+\tilde{\Phi})^{-1}\tilde{\Sigma}' && (5)\\
\tilde{J} &= 2I + (N_A - 1)\tilde{\Sigma}(\tilde{\Sigma}+\tilde{\Gamma}+\tilde{\Phi})^{-1}(\tilde{\Sigma}+\tilde{\Gamma})\tilde{\Sigma}^{-1} && (6)
\end{aligned}
$$

Proof. Is provided in Appendix A.1. □

To guide the model calibration, it is useful to rearrange the equilibrium equations of state, the expected quadratic variation and the expected trading volume across the two assets, which are expressed in the following propositions:

Proposition 1. *Expected Price Variance. Let* $\Delta t\tilde{H} = \mathbb{E}[(f(t+\Delta t) - f(t))(f(t+\Delta t) - f(t))']$ *be the observed pricing variance-covariance matrix. From Theorem 1, this is determined from* $\tilde{\Theta} \in \{\tilde{\Sigma}, \tilde{\Gamma}, \tilde{\Phi}, N_B\tilde{\Psi}\}$ *using the following expression:*

$$\Delta t\tilde{H} = \tilde{\Sigma} + \tilde{\Sigma}_\Lambda, \textit{where} \quad \tilde{\Sigma}_\Lambda = N_A\tilde{\Sigma}((N_A+1)(\tilde{\Sigma}+\tilde{\Gamma}) + 2\tilde{\Phi})^{-1}\tilde{\Sigma} \tag{7}$$

Proof. is provided in Appendix A.2. □

Proposition 2. *Expected Market Volume. Following Admati (1985) and Admati and Pfleiderer (1988, 1989), let the aggregate volume be defined by* $\tilde{V} = \frac{1}{2}|\sum_{n=1}^{N_A}\tilde{a}_n| + |\tilde{d}| + |\tilde{e}|$ *in terms of the equilibrium in Theorem 1, which is given by:*

$$
\begin{aligned}
\mathbb{E}_0[\tilde{V}] &= \tfrac{1}{\sqrt{2\pi}}\mathbb{E}_0[(diag(N_A\tilde{B}(N_A(\tilde{\Sigma}+\tilde{\Gamma})+\tilde{\Phi})\tilde{B})^{1/2} + \\
&\quad diag(\Psi))^{1/2} + (diag(N_A\tilde{B}(N_A(\tilde{\Sigma}+\tilde{\Gamma})+\tilde{\Phi})\tilde{B}' + N_B\tilde{\Psi}))^{1/2}] && (8)
\end{aligned}
$$

Proof. is provided in Appendix A.3. □

We now move to describe the baseline specifications for our simulation and to demonstrate the existence of a non-linear response in the price impact matrix. Our simulation does not include specific point estimates for $\tilde{\Gamma}$, $\tilde{\Psi}$ and $\tilde{\Phi}$. However, if we set the first element of $\tilde{\Sigma}$ as the square of the approximate long run daily volatility of the reference LIBOR rate, and $\tilde{H}$ and $\tilde{V}$ as the observed price volatility and the volume of the futures, respectively, we generate reasonable domains for $\tilde{\Gamma}$, $\tilde{\Psi}$ and $\tilde{\Phi}$. The following list summarises the starting points and respective domains used for the simulation. For the variance-covariance matrix of the true valuation price process, $\tilde{\Sigma}$, $\sigma_1 = 1/4$, $\sigma_3 = \sigma_1/5$ and $\sigma_2 = 7/10\sigma_1\sigma_3$. For the covariance matrix of the global noise across all A traders' signals, $\tilde{\Gamma}$, $\gamma_1 = 1/10$, $\gamma_3 = 1/10$ and $\gamma_2 = 5/10\gamma_1\gamma_3$. For the covariance matrix for each A-type traders' signal, $\tilde{\Psi}$, $\psi_1 = \{1/4, 1/2, 1, 3/2\}$, $\psi_3 = \psi_1/2$ and $\psi_2 = 0.2\psi_1\psi_3$. Finally for the variance of an individual B-type trader's order flow, $\tilde{\Phi}$, $\phi_1 = 1$, $\phi_3 = \phi_1/2$ and $\phi_2 = 0$.

[Insert Figure 1 about here.]

Figure 1 exhibits the level and the partial derivative of the elements of the price impact matrix $\tilde{\Lambda}$, which determine the degree of misalignment between the observed futures price and the terminal valuation as a function of the aggregate order flow, $\tilde{e}$, with respect to the fraction, $N_A/\tilde{N} \in [0.1, 0.8]$, of type A traders in the market. Across a variety of values of ψ_1 and ψ_2, Figure 1 shows that the model predicts that a small number of type A traders will increase the magnitude of the misalignment between the true value and the price in the market. However, as the number of type A traders increases, this detrimental effect on the market decreases and reverses rapidly. Moreover, we predict this pattern for both the short and long maturity contracts.

Overall, our model predicts a non-linear relationship between the proportion of advanced traders and the market depth. Based on our theoretical model and the mixed results in the related literature, we predict that a non-linear HFQ influence on the quality indicators in the Eurodollar futures market is mostly likely and our algorithmic trading proxies may exhibit significant endogeneity. We now provide empirical evidence for this pattern in the Eurodollar futures market.

2. Market Quality Measures and Proxies for HFQ: Data and Methodologies

We begin our empirical analysis by presenting the most salient characteristics of our dataset and the different methodologies utilised to perform our empirical study. We obtain our data from the CME tapes via Thomson Reuters, which corresponds to all inside quotes, trades and the complete history of the limit order book from 1st July 2008 to 1st January 2014 for all 3-month Eurodollar futures contracts. These contracts are traded concurrently and extending up to ten years to maturity. The maturity dates are recorded on March, June, September and December annually. The dataset includes 40 contracts labelled using Reuters Instrument Codes (RIC), which comprise three elements, where ED represents the product code, while H, M, U or Z, correspond to the maturity month (i.e. for March, June, September and December, respectively) and the number between 0 and 9 represents the maturity year of the contract. Hence, the 40 contracts in our dataset are labelled EDH0, EDH1, ..., EDZ9, where EDH0 and EDZ9 represent contracts maturing in March 2010 and December 2019, respectively.[9]

For each of the contracts, we can reconstruct the limit order book and its evolution over time. Therefore, the total number of messages updated to the limit order book, K, together with the time stamp, t_k, are attached to each message $k \in \{1, \dots, K\}$, which correspond to either a quote addition, update or cancellation. We always presume that the time stamp is measured in fractions of a day, hence $t_k \in (0, 1]$.[10] Both sides of the limit order book are ordered by price, such that the level of the book is indexed by $j \in \{1, \dots, J\}$. Therefore, the best (highest) bid and best (lowest) ask are at level $j = 1$. In this framework, the tuples $(P_{b,J,K}, V_{b,J,K}, N_{b,J,K})$ and $(P_{a,J,K}, V_{a,J,K}, N_{a,J,K})$ describe the scalar prices, volumes, and number of active trading accounts by side of the order book, where a represents the ask side and b represents the bid side. The status of the entire limit order book at timescale t_k is represented by the vectors $P_{t_k} = \{P_{b,J,K}, P_{b,J-1,K}, \dots, P_{b,1,K}, P_{a,1,K}, P_{a,2,K}, \dots, P_{a,J,K}\}$.[11] Following the price ordering, the volume state and the number of active trading accounts state are given by: $V_{t_k} = \{V_{b,J,K}, V_{b,J-1,K}, \dots, V_{b,1,K}, V_{a,1,K}, V_{a,2,K}, \dots, V_{a,J,K}\}$ and $N_{t_k} = \{N_{b,J,K},\ N_{b,J-1,K}, \dots, N_{b,1,K}, N_{a,1,K}, N_{a,2,K}, \dots, N_{a,J,K}\}$.

The following section explains in detail how each market quality measure is constructed. The first set of measurements collects the volatility ratios based on Hasbrouck (2018)'s wavelet approach.

[9] All 3-month Eurodollar futures are 10-year maturity contracts; therefore, the labels corresponding to contracts maturing at the end of our dataset are situated between EDH0 and EDZ9.

[10] We use midnight Chicago time as the end of the day as the futures are traded continuously over 24 hrs.

[11] For brevity, we presume that J is the same for the bid and ask sides, but it is possible to add subscripts J_b and J_a for asymmetric level numbers.

The second corresponds to the liquidity indicators proposed by Hendershott et al. (2011). Towards the end of the section, we introduce our proxy for the presence of HFQ in the market, as well as a semi-parametric approach to quantify the marginal effect of HFQ on market quality.

2.1. Variance and covariance ratios based on Hasbrouck (2018)

These measures capture the effects derived from cases where some traders are faster than others. In this sense, the ratio between the price variance attributed to faster traders versus the price variance attributed to fundamentals is a measure of market quality. When the ratio is close to one, we postulate that the scaling of the variance and covariance represent the variance of the fundamental price process. Similarly, the ratio between the covariance of bid and ask prices attached to each trader type (i.e. slow or fast) and the covariance of the fundamentals should be close to unity if market quality does not depend on the presence of faster traders. The methodological challenge related to these measures is how to isolate the effects of fast and slow traders.

We use a wavelet multi-resolution analysis (WMRA) based on a maximal overlap discrete wavelet transform (MODWT) to differentiate traders. The methodology involves applying wavelet filters associated with different scales, such that $\{\tilde{\mathcal{H}}_{w,m} : m = 0, \ldots, M_{w-1}\}$ is the wth MODWT wavelet filter associated with scale η_w, where $M_w \equiv (2^w - 1)(M - 1) + 1$ is the width of the filter and M is the width of the $w = 1$ base filter. Following Hasbrouck (2018), we utilize a MODWT Haar wavelet basis with a wavelet filter $\{\frac{1}{2}, -\frac{1}{2}\}$ to maintain methodological consistency.

Intuitively, a WMRA separates an initial time-series dataset, such as the prices on each of the limit order book levels, into W equally sized vectors where each vector corresponds to a broader wavelet scale, $\eta_w = 2^w$, such that W corresponds to the slowest traders. We decompose the results into one time-series dataset for each trader type where $w = 2$ corresponds to slower traders compared with $w = 1$, and faster traders are represented by $w = 3$. To understand the information given by each scale, consider a simple example that assumes data are sampled every second. In this case, the first and second scales would be $\eta_1 = 2$ and $\eta_2 = 4$, respectively. In this context, η_1 would isolate the effects on prices over 0 to 2 seconds and η_2 the ones over 2 to 4 seconds. In turn, the information corresponding to the η_1 and η_2 are stored at vectors $w = 1$ and $w = 2$, respectively.

In this paper, we calculate the wavelet variance ($\mathcal{V}_\iota$), covariance ($\mathcal{V}_{ab}$) and correlation (ρ_{ab}) of the vectors for prices resulting from each wavelet scale η_w, which are defined as:

$$\mathcal{V}_{\iota,w} \equiv Var\{\overline{H}_{\iota,w}\}, \quad \iota \in \{a,b\} \qquad \text{Wavelet Variance} \tag{9}$$

$$\mathcal{V}_{ab,w} \equiv Cov\{\overline{H}_{a,w}, \overline{H}_{b,w}\} \qquad \text{Wavelet Covariance} \tag{10}$$

$$\rho_{ab,w} \equiv \frac{Cov\{\overline{H}_{a,w}, \overline{H}_{b,w}\}}{(Var\{\overline{H}_{a,w}\}Var\{\overline{H}_{b,w}\})^{1/2}} \qquad \text{Wavelet Correlation} \tag{11}$$

where $\overline{H}_{a,w}$ and $\overline{H}_{b,w}$ are the wavelet coefficients corresponding to scale η_w of the ask and bid vectors of prices, and $Var\{.\}$ and $Cov\{.\}$ are the variance and covariance of the wavelet coefficients.

Analysing the degree of decoupling between each of the wavelet scales η_w allows us to capture the influence of HFQ on market quality. To do so, this paper estimates the wavelet variance and covariance ratios by comparing the $\mathcal{V}_\iota$ and $\mathcal{V}_{ab}$ at relatively short scales η_w to the longest scale η_W. We define the wavelet variance and covariance ratio operators as: $\mathrm{R}_\iota[w,W] = 2^{W-w}\frac{\mathcal{V}_{\iota,w}}{\mathcal{V}_{\iota,W}}$, for $\iota \in \{a,b\}$ and $\mathrm{R}_{ab}[w,W] = 2^{W-w}\frac{\mathcal{V}_{ab,w}}{\mathcal{V}_{ab,W}}$. If the covariance ratio decreases (i.e. the covariation at a high frequency is less than the covariation at a lower frequency), then the bid and ask prices are decoupled.[12] If HFQ increases liquidity and decreases latency, then the degree of decoupling should be less when there are more HFTs in the market. However, if high-frequency order flows are designed to deliberately obfuscate to enable manipulative market trading strategies, then we might expect that the degree of decoupling will be higher and $\mathrm{R}_{ab}[w,W]$ will subsequently decrease. In terms of variance, the ratios $\mathrm{R}_a[w,W]$ and $\mathrm{R}_b[w,W]$ denote the excess volatility at the wavelet scale η_w relative to η_W. Therefore, higher levels of $\mathrm{R}_a[w,W]$ and $\mathrm{R}_b[w,W]$ represent a higher degree of risk for buyers and sellers at the point of execution.

2.2. Liquidity Spreads Measurements

Our second empirical approach estimates additional market quality measures based on the salient study by Hendershott et al. (2011), which provides an overview of the liquidity measures in the equity literature using inside quotes and trades. The measures in Hendershott et al. (2011)

[12] Some degree of decoupling is to be expected as order flow updates are discrete at the highest frequency. Therefore, one price (e.g. bid or ask) will be expected to lead the other at different points. However, over a day this will even itself out so, as the frequency decreases the degree of coupling will naturally increase.

do not provide a complete picture of order flows because inside quotes only include the highest bid prices and lowest ask prices. However, our data include the whole limited order book and trades. We also extend the depth of our empirical analysis by calculating the liquidity spreads for each book level.[13] In total, we measure eight liquidity spreads, namely, bid–ask spreads (S^{Q}), quoted half-spreads ($\mathrm{S}^{Q^{1/2}}$), quoted depth (S^{D}), effective half-spreads (S^{E}), realised spreads ($\mathrm{S}^{R,5}$ for 5 mins and $\mathrm{S}^{R,30}$ for 30 mins) and price impacts ($\mathrm{S}^{AS,5}$ for 5 mins and $\mathrm{S}^{AS,30}$ for 30 mins). Each spread is calculated as follows:

$$\mathrm{S}^{Q}_{j,k} = 100(p_{a,j,k} - p_{b,j,k})/(p_{m,j,k}) \qquad \text{Bid–ask spreads} \tag{12}$$

$$\mathrm{S}^{D}_{j,k} = p_{a,j,k}v_{a,j,k} + p_{b,j,k}v_{b,j,k} \qquad \text{Quoted depth} \tag{13}$$

$$\mathrm{S}^{E}_{j,k} = q_k(p_l - p_{m,j,k})/p_{m,j,k} \qquad \text{Effective spreads} \tag{14}$$

$$\mathrm{S}^{R,\Delta t}_{j,k} = q_k(p_l - p_{m,jk+\Delta t})/p_{m,j,k} \qquad \text{Realised spreads} \tag{15}$$

$$\mathrm{S}^{AS,\Delta t}_{j,k} = q_k(p_{m,j,k+\Delta t} - p_{m,j,k})/p_{m,j,k} \qquad \text{Adverse selection} \tag{16}$$

where $p_{m,j,k}$ is the quoted mid-price; q_k indicates the trade direction, that is, +1 for buyer-initiated trades, 0 for no trades, and -1 for seller-initiated trades; and Δt represents how many minutes after transactions are considered (i.e. either 5 or 30 mins).[14]

2.3. A Semi-parametric Model for Market Quality and HFQ

After outlining the construction of our market quality measurements, we turn to the primary objective of this paper, which is to understand the impact of HFQ on market quality. To construct an HFQ proxy, that is, the fraction of HFQ to the total number of traders acting in the market,

[13] To construct these liquidity measurements, Hendershott et al. (2011) use the Lee and Ready (1991) algorithm to calculate trade direction; however, this method is not very accurate for HFQ data. We found that about 85% of the data can be classified as 'no trades'. Therefore, we alter the Lee–Ready Algorithm using the volume-weighted average price within a five-level order book.

[14] We need to sample the vector of message updates at 5- and 30-min intervals. We denote this sampling by $t_{k(t_0+r\Delta t=5/1440)}$ and $t_{k(t_0+r\Delta t=30/1440)}$ for 5- and 30-min sampling respectively, for $r \in \{0,\dots,R\}$, where $R \equiv (1 \bmod \Delta t)$ and t_0 is the first available time stamp of the day. Let $\tau = \{t_0+\Delta t, t_0+2\Delta t,\dots,1\}$ be the order collection of time intervals representing the desired sampling grid, with Δt measured in fractions of a single trading day. We use the nearest past neighbour rule to match the timestamp of trades and quotes to the sampling grid, that is, the index $k(\tau\{r\}) = \min[|t_0+r\Delta t - t_k|; t_k \leq t_0 + r\Delta t]$ where the indexed vectors are denoted as $k(\tau)$. The market orders (i.e. trades) are similarly indexed by $l \in \{1,\dots,L\}$ and recorded by the tuple (P_l, V_l), with time stamp t_l.

we take the information for each contract per day and consider the time elapsed between two consecutive quoting messages. Any message separated from the earlier message by less than 25 ms is classified as having been generated by HFTs. Our HFQ proxies by contract, book level, and day are denoted by:

$$\mathrm{A}_{j,\mathcal{T},\Delta t} = \mathrm{M}^{\Delta t}_{j,\mathcal{T}}/\mathrm{M}_{j,\mathcal{T}} \tag{17}$$

where $\mathrm{A}_{j,\mathcal{T},\Delta t}$ denotes the fraction of HFQ quotes for day $\mathcal{T}$, Δt denotes the timestamp (e.g. less than 25 ms) for order book level j, and $\mathrm{M}_{j,\mathcal{T}}$ is the total number of messages for order book level j for the day $\mathcal{T}$.[15]

We use this information to derive two distinct datasets. We construct the first dataset, which includes 35,491 contract-day observations, by taking the volume-weighted average of book levels by contract as follows:

$$\mathrm{A}_{i,\mathcal{T},\Delta t} = (\mathrm{A}_{j,\mathcal{T},\Delta t} \times V_{j,\mathcal{T}})/\sum_{j=1}^{5} V_{j,\mathcal{T}} \tag{18}$$

where $\mathrm{A}_{i,\mathcal{T},\Delta t}$ denotes the volume-weighted fraction of HFQ over the whole book per contract i and day $\mathcal{T}$. The second is a time-series dataset that includes 2,339 day observations to study the effects of HFQ on market quality based on the maturity effect of Samuelson (1965). This dataset results from averaging the fraction of HFQ, $\mathrm{A}_{\mathcal{T},\Delta t}$, for all contracts sharing the same days to maturity as:

$$\mathrm{A}_{\mathcal{D},\Delta t} = \frac{1}{40}\sum_{i=1}^{40} \mathrm{A}_{i,\mathcal{D},\Delta t} \tag{19}$$

where $\mathrm{A}_{\mathcal{D},\Delta t}$ denotes the average fraction of HFQ over all contracts with the same days to maturity $\mathcal{D}$, while i indexes the contracts.

The empirical characterisation of the influence of HFQ on market quality is not a trivial task. As discussed earlier, there is no agreement about whether HFQ improves market quality or not.

[15] The physiology literature estimates that the average human reaction time is greater than 250 ms (Kosinski, 2008). The 25-ms thresholds in this paper yield a modal fraction of order flow messages close to the average quantity of order flows computed by CME with access to the algorithmic trading account prefixes. We acknowledge that algorithmic trading is not 100% related to HFQ. For robustness, we constructed proxies at between 50 ms and 200 ms. The time evolution of these proxies closely correlates to the 25 ms proxy. On the most active days, 90% of messages are below 200 ms. However, 75% are also below 25 ms, which indicates that the clustering is not caused by the number of traders.

Moreover, there is no clarity about the functional form that links HFQ with market quality. When we analyse the Eurodollar futures market characteristics, we find that changes in HFT quoting activities are not a simple linear relationship with both the liquidity and volatility in the market. In this context, a semi-parametric regression is the most suitable methodology to capture the complexity of this relationship. The flexibility offered by this approach is based on the assumed linear relationship between some independent variables and the dependent variable, whereas the functional form between the remaining independent variables and the dependent variable was not determined. We formally propose a semi-parametric regression:

$$Y_{i,\mathcal{T}} = X_{i,\mathcal{T}}\beta + \mathscr{G}[Z_{i,\mathcal{T}}] + \varepsilon_{i,\mathcal{T}}, \quad \mathbb{E}[\varepsilon_{i,\mathcal{T}}|W_{i,\mathcal{T}}] = 0 \tag{20}$$

where $Y_{i,\mathcal{T}}$ represents each of our market quality measures for the i contract for day $\mathcal{T}$, $X_{i,\mathcal{T}}$ represents the linear regressors, namely, the lagged dependent variable and a constant $Z_{i,\mathcal{T}}$ is our HFQ order flow proxy and $W_{i,\mathcal{T}}$ is a set of instrumental variables. Flattening the dataset to an index $i \in \{1, \ldots, N\}$, where N is the total number of contract days and the index i represents the set of N tuples of $(i, \mathcal{T})$ contract days. The list of dependent variables in the regression includes the variance and covariance ratios and the spreads collected in the vector $Y_{i,\mathcal{T}} = \{\mathrm{S}^{Q}_{i,\mathcal{T}}, \mathrm{S}^{Q^{1/2}}_{i,\mathcal{T}}, \mathrm{S}^{D}_{i,\mathcal{T}}, \mathrm{S}^{E}_{i,\mathcal{T}}, \mathrm{S}^{R,5}_{i,\mathcal{T}}, \mathrm{S}^{R,30}_{i,\mathcal{T}}, \mathrm{S}^{AS,5}_{i,\mathcal{T}}, \mathrm{S}^{AS,30}_{i,\mathcal{T}}, \mathrm{R}_{a1,i,\mathcal{T}}, \mathrm{R}_{a4,i,\mathcal{T}}, \mathrm{R}_{a9,i,\mathcal{T}}, \mathrm{R}_{b1,i,\mathcal{T}}, \mathrm{R}_{b4,i,\mathcal{T}}, \mathrm{R}_{b9,i,\mathcal{T}}, \mathrm{R}_{ab1,i,\mathcal{T}}, \mathrm{R}_{ab4,i,\mathcal{T}}, \mathrm{R}_{ab9,i,\mathcal{T}}\}$. Hence, the non-linear semiparametric nature of the model is reflected in that β is a vector of unknown linear parameters, and $\mathscr{G}[.]$ is a nonparametric function.

This type of semi-parametric estimator has not been used previously in the context outlined in this paper. Following Robinson (1988), Newey and Powell (2003) and Florens et al. (2012) on semi-parametric and non-parametric estimation, we briefly review this type of estimator together with our new approach for addressing the endogeneity issue and generating confidence intervals and marginal effects. In Appendix B, we briefly summarise our semi-parametric estimator and implementation.

Before estimating the semi-parametric regression, using an instrumental variable becomes necessary as HFQ proxies may exhibit significant endogeneity; therefore, we need to focus on the endogeneity problems between independent variables X_i (i.e. the lagged execution risk measurements) and Z_i (i.e. HFQ proxies) with the error terms ε_i during the semi-parametric regression. For example, the level of HFQ could be caused by the level of market quality proxied by our dependent variables in addition to the effect we are trying to identify in the opposite direction. Therefore, we use the log of days to maturity and the log of total messages as instrumental variables because they are related to both the quality measure and the degree of HFQ activity, but the possibility of being caused by them is low to none.

To address the endogeneity issue, we employ the logarithms of time to maturity and total number of messages per day as instrumental variables, where denoted as W_i. As valid instrumental variables, W_i should satisfy two conditions. The first is that W_i should be partially correlated with the endogenous variables Z_i, $cov(W_i, Z_i) \neq 0$. In our case, both the log total number of messages per day and the log time to maturity are correlated with the HFQ proxy. The second condition requires that $cov(W_i, \varepsilon_i) = 0$; however, this condition cannot be tested because of the unobservable error term ε_i.

For the robustness check, we conduct a number of ancillary treatments. We employ both the number of traders (i.e. open trading accounts) and the volume-to-message ratio as the instrumental variables to proxy for high-frequency algorithmic traders. According to the CME technical trading manual, activating a trading account takes more than a day; therefore, the number of active accounts could not be caused by a shock to any of the market quality proxies in the dependent variable. The results are qualitatively and quantitatively similar. To avoid unnecessary replication, we report these results in an online appendix, including marginal effects plots with algorithmic trading proxies on market quality indicators with different instrumental variables.[16]

3. Market Characterisation

With well over 100 quadrillion dollars of depth on both sides of the order book, around a quarter of it at the best-bid and best-ask prices, the Eurodollar futures market has considerable average liquidity (see Panels A and B of Table 1). Throughout this section, we demonstrate that this

[16]The online appendix is available at https://drive.google.com/file/d/1nINGyzDyMyKpAhCCsXQU4gN_sqXqAmmF/view?usp=sharing

liquidity is not evenly spread across the year or across the term structure of the Eurodollar futures and that all liquidity indicators we propose exhibit an acute maturity effect.

[Insert Table 1 about here.]

Figure 2 presents the term structure of the wavelet variance ratios for the best bids and best asks of the 40 Eurodollar futures contracts. To construct Figure 2, we first perform a nine-scale wavelet multi-resolution analysis by considering only the days when the number of quotes is at least 5,000.[17] We then compute the sum of the bids and asks in the wavelet variances across all the contracts based on the time to maturity instead of the actual transaction dates used by Hendershott et al. (2011) and Hasbrouck (2018). Given that each Eurodollar future has a 10-year tenor, our procedure implies plotting the term structure from 10 to 0 years to maturity.[18] As observed in Figure 2, the quoting variance ratio is relatively stable at all timescale levels during the first 2 years, indicating that for every contract, the price variability is less vigorous at the beginning of the life cycle. From 8 years before maturity, the variance and covariance ratios gradually increase to reach their maximum value around a 2-year tenor. This change is consistent with the maturity effect, which is characterised by monotone increases in price variability as the contract approaches maturity (Samuelson, 1965). Interestingly, the quoting variance suddenly drops around 1 week before maturity. Furthermore, the price volatility of Eurodollar futures increases when contracts are close to expiring at all wavelet scales.

[Insert Figure 2 and Table 2 about here.]

To offer a more detailed look at the maturity effect, we construct a set of forward rates in Eurodollar futures that map to the forward curve. We sort Eurodollar futures by the days to maturity and measure the wavelet variance and covariance ratios of inside quotes for all Eurodollar

[17] The choice of 5,000 is relatively arbitrary because it is the minimum number of updates for the variation in quotes needed to perform a nine-scale wavelet decomposition. Eurodollar futures quotes tend to be less frequent than an actively traded stock; however, they have a far larger volume attached to each individual quote. Therefore, 5,000 updates result in no singularities in our dataset.

[18] Wavelet variance plots for the first, second, third and fourth quarter maturity contract groups and plots for each single contract can be found in our online appendix presenting the supplementary data analysis.

futures over 30 days, 60 days, 3 months, and 1 year before the contracts' maturity date. We then separate Eurodollar contracts according to their prices and average the variances and covariances across price quartiles using each of the four time horizons listed above. The results allow us to compare the Eurodollar futures market with the US equity market in terms of fundamental and transient volatility (Table 2) and therefore illustrate the decoupling of bid and ask sides of the limit order book over time.

Several elements are worth mentioning. First, Table 2 reveals that the bid variances for all pricing groups at all wavelet scales are relatively similar to the ask variances. Therefore, unlike Hasbrouck's findings for the US equity market, both best-bid and best-ask effects are comparable for Eurodollar futures. Our results also suggest that the Eurodollar futures market differs from the US equity market in that the higher price quartiles tend to have higher variance and wavelet bid–ask correlations. For example, at wavelet scale level 1, the average median asks variance ratio in the 30-day tenor increases from 1.5223 in the lowest price group to 2.6613 in the highest price group. At the same wavelet scale level, the average median bid variance ratio in the 30-day period also increases from 1.5157 to 2.6380 when moving to higher price groups. Both bid and ask variances increase with shorter wavelet scales. Second, the results show that the volatilities of both the best bid and best ask increase before approaching maturity. For the 76%–100% price group at timescale level 1, the average median ask variance for 1-year tenor Eurodollar contracts is 1.7641, while the ask variance for 30 days before maturity increases to 2.6613. Third, Table 2 also reveals that the volatilities of the bids and asks of every maturity-type contract maintain positive covariances and correlations at all wavelet scales. The average of the best-bid and best-ask covariance ratios increase from a 1-year tenor to a 30-day tenor. Meanwhile, the average best-bid and best-ask correlations decrease for four price quartiles and nine wavelet scales. This suggests that best bid and ask prices are departing from each other as the Eurodollar contracts approach their maturity date. As expected, the correlation level drops with frequency; therefore, the coupling dynamics of the bid–ask spread are not constant over wavelet scales.

The degree of decoupling is significant, with the highest frequency (shortest timescale) at around one third of the daily correlations. However, there are very few comparators in the extant literature showing the variation in correlation across time scales for the bid–ask spread. Therefore, a broad survey of markets is needed. Considering traders, the ability to make market orders in the knowledge that the dynamics of the order book on the opposite side of the market are only 70% correlated with their initiation side (i.e. buy or sell) is quite striking given the very large order sizes within the Eurodollar futures market in general.[19]

[Insert Figure 3 about here.]

For the remaining Eurodollar futures market quality indicators, Figure 3 summarises the weekly average of liquidity spreads for all the contracts by book level and time to maturity. The daily bid–ask spread at each level is quite volatile; therefore, we plot each liquidity spread on a weekly basis. In general, the patterns of the different liquidity measures are consistent. Both the bid–ask spreads and quoted half-spreads are quite volatile and slowly become narrower and more consistent over time among the five levels of the limit order book. Few bid–ask prices decouple at the early stage of the Eurodollar futures life cycle; therefore, there are few spikes in the liquidity spreads shown before the 5-year tenor. The quoted spreads become narrower nearly 2 years before maturity, which indicates greater market liquidity. In addition, the liquidity suddenly drops 1 week before settlement. In terms of quoted depth, there is a clearly increasing trend from 6 years to maturity, which is broken only few days before maturity. The findings also suggest that the largest liquidity in this market is provided in level 2, which exhibits the highest depth during these 6 years, although it decreases 1 week before maturity.

This maturity effect is also observed in the plots for effective half-spread, realised half-spreads (5 mins and 30 mins) and adverse selection indicators (5 mins and 30 mins) for all levels, which shows a monotonic decreasing trend during the last 5 years of the contracts' maturity. The narrowest effective half-spread, realised half-spreads and adverse selection indicators are observed in level 1. However, levels 1 and 2 tend to converge, which highlights the importance of level 2 of the limit order book.

[19]For robustness, we also consider both the top 10% most active days and only the most active day for each contract. The results are materially similar although the timescales are generally much smaller. These results are available in the online appendix.

Before discussing the final objective of this paper, we explore the behaviour of our proxy for the fraction of HFQ order flow in the market. In particular, we provide a general idea on how much HFQ activity is observed in the Eurodollar futures market, such as where in the book this order flow sits (i.e. are the ask or bid sides more active? Do they trade more on a specific order book level?) and the accuracy of the proposed proxy. Table 3 presents some descriptive statistics on the order book updates used to compute our HFQ order flow proxy.[20] Table 3 reports that although the 200 ms limit is lower than the average human reaction time reported in the physiology literature, it is not really a good classification limit because the message updates happen much quicker than this on average. The majority of quote updates to the order flow under 200 ms are also under 100 ms.

[Insert Table 3 and Figure 4 about here.]

Because physiology does not provide an accurate reaction threshold and the lack of identification of each trader in the market, we are left with the problem of how to identify HFQ. However, the 25 ms threshold (i.e. ten times shorter than the human reaction time) provides a credible proxy for HFQ in the Eurodollar futures market based on the Figure 4. Figure 4 (a) and (b) shows that the distribution of the bid and ask HFQ order flow intensity proxies for both the panel and the volume-weighted average datasets for the 25 ms threshold match the proportion of HFQ order flow in the market, that is measured by the CME Group which has access to the actual trading account number classifications (CME Group, 2010). In particular, the kernel distribution of the proxy for the volume-weighted average across the whole market is strongly bimodal and potentially trimodal, with density peaks at around 15% and 30% and the main spike at around 65%. In the whole cross-section of the market, the sub-peaks at 15% and 30% disappear, leaving only the main peak, which means that just below 70% of the total order flow is classified as HFQ activity.[21]

[Insert Figure 5 and Figure 6 about here.]

[20] These can be compared with the timestamps of the variance ratios from the inside spread in Table 2.
[21] The small tail in the negative domain is caused by a small trimming error.

A graphical analysis of the activity of HFTs in each level of the order book portrays that HFQ activity changes as the contracts approach maturity. Figure 5 exhibits the weekly average of HFQ order flows at different thresholds (200 ms, 150 ms, 100 ms, 75 ms, 50 ms and 25 ms) for both ask and bid sides. The fraction of HFQ order flow is quite high for all five levels, even when the fraction of the HFQ proxy is set to identify messages at or below 25 ms. Furthermore, the variation in the fraction of HFQ quoting updates follows a definite term structure and a day-of-the-week effect (Sundays are constantly lower). Considering the 25 ms threshold, the proportion of HFQ activities is relatively low during the first 2 years of the contracts' life cycle (i.e. between 8% and 18% for levels 1 and 2 and lower or does not exist for the rest of the levels) at both the bid and ask sides of the order book.

The proportion of HFQ quoting increases as maturity approaches, reaching a maximum of around 80% at levels 4 and 5 for the 2-year tenor contracts and 65% at level 2. This maximum involves all book levels, as HFQ activity gradually appears at different levels of the order book from 8 to 2 years to maturity. Figure 6 shows that level 2 of the order book has the highest number of total messages per minute (i.e. around 200 when 2 years before maturity remain). However, although level 2 is the predominant level, the fraction of HFQ is still very high at levels 4 and 5, which shows that level 2 is not the major driver of the HFQ proxy. Therefore, the total messages can be used as an instrumental variable to deal with the endogeneity effects of HFQ quoting activity in the semi-parametric regressions in the next section.

4. Determining the Marginal Effect of the HFQ Order Flow on Market Quality

As discussed earlier, the influence of the HFQ on the futures market quality did not achieve a uniform agreement in the literature. Moreover, the HFT quoting activities in the Eurodollar futures market do not show a linear relationship with the execution risk variables (e.g. the variance/covariance ratios and the battery of liquidity measures). Therefore, we conduct a semi-parametric regression to fit the complexity of the relationship between HFQ and execution risk.

This section analyses the results of our semi-parametric approach, which are separated into Table 4 showing the coefficients of the linear part of each regression and a collection of graphs depicting the marginal effects of HFQ on each of the measures of market quality we propose. Table 4 illustrates that there is a quite large persistence in market quality when quality is measured through the standard liquidity measurements inspired by Hendershott et al. (2011). The coefficients for each of these lagged measures is positive, ranging from 0.073 to 0.922 for the ask-side and from 0.069 to 0.921 for the bid-side.

Moreover, these coefficients are all significant at the 1% level and many are larger than 0.3. However, this persistence is not observed when market quality is measured through variance ratios. In those cases, most of the coefficients are not significant and virtually zero. The only two significant coefficients for lagged variance ratios are equal to about 0.12, which implies a much lower persistence than for the remaining market quality measures. Finally, it is also worth noting that the coefficients linked to the bid side are slightly higher than the coefficients linked to the ask side in the volume-weighted time-series dataset. However, this effect is not observed in the contract-day panel dataset where there is almost no difference between the persistence in market quality between both sides of the order book.

[Insert Table 4 about here.]

The most important part of our empirical analysis for the objectives of this paper is shown in the marginal effect graphs (Figures 7 to 9) related to the non-parametric component of our regressions. Given the large number of graphs derived from these regressions, we do not include them all in the paper because most of them are qualitatively similar. However, they are available in our online appendix, which presents our supplementary data analysis.

[Insert Figure 7 about here.]

We start by discussing the marginal effect of HFQ on the bid–ask spread, given that it is a common market quality measurement. Figure 7 includes four subplots showing this marginal effect, where subplots (a) and (b) correspond to the regression run over the time-series dataset and the other two subplots correspond to the contract-day panel dataset. In all of the subplots, the black line represents the marginal effects of HFT on the bid–ask spreads and the blue dotted line represents the bootstrapped confidence bounds.[22]

All of the subplots in Figure 7 indicate a number of interesting results. First, the marginal effects of HFQ on the bid–ask spreads are definitely non-linear. Second, even a small number of HFTs have significant effects on the bid–ask spreads. Third, once the proportion of HFQ has reached a certain saturation level, its effect on the bid–ask spreads disappears. Considering the time-series dataset, when the proportion of HFQ is between 20% and 52%, the marginal effects of both ask and bid HFQ change sign and level several times, which implies that a linear approach would be a poor approximation of the relationship between market quality and HFQ.

Considering the results of the contract-day panel dataset, the interpretation of the effects of HFQ on market quality is much more direct, especially for the ask side. Subplot (c) clearly shows that once the proportion of HFQ reaches 20%, the bid–ask spread increases, which suggests that market quality decreases sharply. This effect disappears once the proportion of HFQ increases to 25%, implying that market quality increases up to the moment when HFQ represents around 35% of the traders. Once the proportion of HFQ is higher than about 52%, it has virtually no effect on market quality. Although the effects on the bid side show higher volatility, their general evolution is similar to the effects on the ask side.

Overall, the evidence related to the bid–ask spreads suggests that HFQ initially act as informed traders receiving information before the other traders. The evidence is in line with front-running HFTs dominating quote-matching strategies to submit and cancel large amounts of orders. As the amount of HFQ in the market increases, they probably start cancelling each other out, which narrows the bid–ask spread and reduces transaction costs up to the point where their effect becomes negligible.

[Insert Figure 8 about here.]

The second market quality indicator that we consider is adverse selection at 5 mins and 30 mins. Subplots (a) and (b) in Figure 8 are related to the former and subplots (c) and (d) to the

[22]Regression models of this type may suffer from the 'trimming' problem (i.e. overfitting outside of the bulk of the data). The fitted function in several plots shows a substantial level of oscillation; however, the bootstrapped confidence bounds correctly identify this oscillation as being insignificant, but this result is subject to the usual caveats regarding confidence bounds versus parametric identification of significance.

latter. As with the bid–ask spreads, Figure 8 shows that HFQ activity has a non-linear effect on both 5-min and 30-min adverse selection in the time-series dataset. However, the effect is much less volatile than that observed for the bid–ask spreads. In this context, subplots (a) and (b) suggest that HFQ order flow has a more limited effect on adverse selection (5 mins) when the HFQ proxy is below 40%. However, once the HFQ order flow proxy increases over 40%, the effect on adverse selection (5 mins) rises dramatically, which indicates that HFQ increases trading costs, decreases market liquidity and damages the price efficiency, with less informative quotes in the market. At this stage, the degree of asymmetric information in the market rapidly increases to augment the transaction costs related to order execution. This marginal effect suddenly drops when both HFQ proxies on the ask and bid sides increase to about 46%–53%. From that point on until HFQ proportion reaches about 60% of the market, HFQ seems to provide liquidity to the market and increases price efficiency, with more informative quotes.

Subplots (c) and (d) in Figure 8 reveal that for the bid side, the marginal effects of HFQ are markedly more volatile than those computed from the ask side. However, the general results on the marginal effect of HFQ on market quality also hold when quality is measured through adverse selection metrics. In particular, in 30-min adverse selection measurement, when the proportion of HFTs is below around 36%, there is virtually no relevant effect. However, the marginal effects are positive when the proportion is between 36% and 44%. Once the fraction of algorithmic trading passes 44%, the impact of the ask side HFQ proxy on adverse selection (30 mins) drops until the fraction reaches around 55%. These effects appear more volatile in the bid side and for the 30-min measure than for the 5-min measure. This difference suggests that the effect of HFQ on adverse selection is better identified using shorter horizons.

Overall, our results suggest that the effect of HFQ is also non-linear in the adverse selection metrics. In particular, we observe that as the proportion of HFQ order flow increases, the HFTs act as liquidity demanders and their impacts on adverse selection increase. However, once the HFQ order flow proportion reaches a certain critical point, HFTs act as liquidity providers, which leads to the shrinkage of information asymmetry and increasing market liquidity. Moreover, when most of the traders in the market are HFTs, their marginal impacts effectively drop to zero.

For the remainder of the liquidity measures considered in this paper, the effects of both ask and bid HFQ proxies are less interesting to discuss in that they are either virtually non-existent or they do not appear to depend in an interpretable way on the quantity of HFQ in the market. Although for the sake of brevity we do not report the corresponding graphs, these results are available in our online appendix. A comparable situation is observed when market quality is measured through the variance and covariance ratios. However, the cases where market quality is measured through the variance and covariance ratios derived from wavelet scales η_1 and η_4 are two notable exceptions, which reveal that the variance-covariance measures tell a part of the story that should not be dismissed. Because these results are very similar, we only discuss the results for wavelet scale η_1 in detail.[23]

The analysis of our results for these variance and covariance ratios shows that the salient characteristic of these measures is their capacity to isolate the level of decoupling between both sides of the order book that solely affect HFQ.[24] Figure 9 represents the marginal effects of HFQ order flow on the variance and covariance ratios at timescale level 1 in the panel dataset, where the black continuous line depicts the marginal effects of HFQ order flows on the variance and covariance ratios and the blue dashed lines are their 95% lower and upper confidence bounds. Figure 9 allows us to describe in more detail how the market quality, namely, the decoupling between both sides of the order book, depends on the proportion of HFQ order flow that actively places orders in the market.

[Insert Figure 9 about here.]

As in the earlier cases, subplots (a) and (b) in Figure 9 suggest that the HFQ order flow marginal effects on ask variance ratio level 1 depend on the proportion of high-frequency order flow activities in the market. When this proportion is lower than approximately 20%, HFQ seems to have limited effect on the variance ratio. When the proportion of HFQ order flow is higher,

[23] This perspective is also supported by Budish et al. (2015), who present materially similar results framed in a different setting.

[24] In this paper, we report only the results for HFQ, which are defined as traders placing orders within a threshold equal to or lower than 25 ms. However, the results would be qualitatively similar for alternative thresholds, such as 50 ms, 100 ms, 150 ms or 200 ms, which are all below human reaction times.

the marginal effect on the variance ratio increases, which shows that HFQ order flows increase the high-frequency microstructure noise and volatility in the market. However, higher variance ratios are simultaneously symptomatic of less informative quotes and therefore a higher execution risk because the fundamental price of the contracts is more difficult to observe in the market. As the proportion of HFQ order flows increases further (to around 23%–25%), the effect on the variance ratio drops dramatically, which is most likely because the high proportion of HFQ order flows results in their orders cancelling each other out, shrinking liquidity and, in turn, the spread variance. Subplots (c) and (d) in Figure 9 show that these effects are similar on both sides of the order book. When the ask-side HFQ order flow proxy is below 20%, the impact of HFQ on the bid variance ratio is relatively weak. When the proportion of HFQ order flow is around 24%, its impacts on bid variance suddenly spike to the highest peak. Finally, the marginal effects drop dramatically when the fraction of HFQ order flow is higher than 24%. The marginal effects on the bid variance ratio level 1 increase when the ask-side HFQ order flow proxy passes around 27%. The effect eventually dies out when the proxy is higher than 30%. Although the sequence of events is the same, the point of inflection is lower for the bid variance ratio than for the ask side at around 10%–20% of messages being ascribed to be algorithmic in origin.

Subplots (e) and (f) in Figure 9 exhibit the decoupling of bids and asks. As in the earlier case, the marginal effect of HFQ order flow on the covariance ratio are virtually non-existent before the proportion of HFQ order flow reaches around 28%. Afterwards, it increases to approximately 48%–50%. Within a 28% to 33% proportion, the impact of the HFQ order flow is negative, which results in a significant decoupling of the bid–ask spreads. Up to a proportion of 42%, the HFQ order flow then increases the covariance ratio substantially, which reduces the decoupling. A key result is the shape of the transmission function for the covariance versus the shape of the variance. The deterioration in market quality (i.e. the reduction in covariation provides evidence for a decoupling) occurs at lower fractions of high speed messages than for the variance ratios. The reduction in the covariances for the ask side shows an impact at a substantially lower fraction of high speed messages than for the bid side. Moreover, the points of inflection for the bids versus the asks differ markedly from the inflection points for the variances. This separation of the volatility and correlation effects is quite fascinating because it indicates that the mix of strategies employed by HFTs have markedly different impacts. Evidence from a variety of legal actions suggest that HFTs' strategies fall into two broad categories: trading across markets, such as spot-future arbitrage, or trading within a single market. A plausible cause of the excess variance in the spread could be momentum ignition strategies, where algorithms generate quotes (usually at level 2, so they are no transacted) and uses this volume impact on the order flow to force the mid-price and inside quotes in a specific direction (e.g. adding a number of standard deviations worth of bid [ask] volume at or near the level 2 price, but less than the level 1 price to push the price up [down]). We consider this interpretation is further supported by the fact discussed some sections above that the volume of messages in the limit order book at level 2 is by far the highest, which approaches double that of the level 1 messages at or around 2 years from maturity. We acknowledge that the peaks for the bids and the asks are not strictly concurrent; however, we suggest that this aesthetic choice about the side of the order book in which the algorithms operate is a key driver of the asymmetries between the impacts of the bid and ask HFQ order flow proxies on the various market characteristics.

5. Conclusions

To the best of our knowledge, this paper is one of few studies to combine best-bid, best-ask and order book dataset and consider a proxy for the influence of HFQ on the Eurodollar futures market. Our study includes every trade, every inside quote, and every message update in the order book from 2008 to 2014 for the CME trading venue, where about 80% of the cumulative order book depth is observed (CME Group, 2011). In this sense, our study provides a relatively comprehensive empirical analysis of the microstructure of bid–ask volatility and its impact on the quality of the Eurodollar futures market. To study the effects of HFQ on market quality, we compute several high-frequency order flow metrics using every quote update across the limit order book. We then estimate common market quality measurements, such as bid–ask spreads, quoted half-spreads, quoted depth, effective half-spread, realised spreads, price impacts and a set of less common measures based on Haar wavelet multi-resolution analysis. Because there is no clarity in either the empirical or theoretical literature on the impact of HFQ on market quality, we propose the use of a semi-parametric regression that does not impose any assumption on the functional form that this relationship should have. Our results suggest a non-linear impact from high-frequency order flows on market quality measurements. This pattern reflects that HFQ is detrimental in the market when overall trading is sparse. Therefore, the HFTs present in the market appear to profit from their advantageous position, which deteriorates the overall market quality. However, as the fraction of HFQ order flow increases, the market quality recovers and then improves. This result is robust as we apply a large number of alternative specifications using different market quality proxies.

This paper offers several contributions. On the one hand, the Eurodollar futures market is a highly liquid market that trades trillions of dollars in value. Constructing a unique dataset comprising the quoting and trading activity within the limit order book in the Eurodollar futures market allows us to identify the impacts of the realised fraction of HFQ activity on market quality. Although some readers might think that the liquidity levels of the ED futures market is difficult to find in other markets, it is worth noting that Section 3 reveals that Eurodollar futures contracts are highly illiquid for large periods of time, which implies that our results incorporate different liquidity levels. Moreover, the capacity to work with every message update in the order book allows us to overcome plausible issues related to small samples, which is a typical limitation of studies on the effects of HFT in the market. [25] In this sense, our results could easily be extended to other markets.

On the other hand, unlike in an equity market where multiple overlapping exchanges create bottlenecks, the CME is centrally cleared with microsecond (0.1 ms in practice) time stamps and latency. Our results show that even when the updating speed approaches 100 ms, our variance ratios, although greater than one, are not at the levels found in the equity literature. In contrast, wavelet correlations between the bid and ask prices do drop considerably, which follows the equity literature. The combination of these two facts seem to indicate that the order books in the Eurodollar futures market have a great deal of noise, very likely from active HFT. Finally, although

[25] If market microstructure studies ignore higher levels of the order book than level 1 (e.g. best-bid and best-ask prices in the equity market), they will miss a great deal of activity. Our results show that level 2 prices in Eurodollar futures across most maturities are the most actively updated by a substantial margin. Besides, market microstructure studies in the futures market require a relatively long-period high-frequency sample because we must consider the maturity effects.

we cannot distinguish between quoting and trading activities, the results of our semi-parametric regression suggest that HFTs profit from their advantages only up to a point, when their arms race has saturated the market because each trader is sending out similar signals. Hence, if a linear model was fitted to these data, the results would be characterised by a very high degree of variation depending on the locations of the sampling points (based on the particular experiment performed in the paper).

Overall, the differences between our results and those in the prior literature are driven by the fact that earlier scholars only had access to short snippets of the order flow, such as 1 month or only a few days. We acknowledge that our results will not fully satisfy the detractors or backers of HFT. Nevertheless, we infer that if high-frequency market makers were to prop up trading for long maturity contracts, the market liquidity would increase very substantially (by up to an order of magnitude on most of our market quality measures), which may have significant implications for long maturity futures where the level of ambient liquidity is low, whether HFTs are absent or not.

References

Admati, A. R. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica 53*(3), 629–657.

Admati, A. R. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies 1*(1), 3–40.

Admati, A. R. and P. Pfleiderer (1989). Divide and conquer: A theory of intraday and day-of-the-week mean effects. *The Review of Financial Studies 2*(2), 189–223.

Brogaard, J., T. Hendershott, and R. Riordan. (2017). High frequency trading and the 2008 short-sale ban. *Journal of Financial Economics 124*(1), 22–42.

Brogaard, J., A. Carrion, T. Moyaert, R. Riordan, A. Shkilko, and K. Sokolov (2018). High frequency trading and extreme price movements. *Journal of Financial Economics 128*(2), 253–265.

Budish, E. B., P. Cramton, and J. J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics 130*(4), 1547–1621.

Chakrabarty, B., P. K. Jain, A. Shkilko, and K. Sokolov (2014). Speed of market access and market quality: Evidence from the SEC naked access ban. Working Paper, Western Finance Association 2014.

Chan, K. (1992). A further analysis of the lead-lag relationship between the cash market and stock index futures market. *The Review of Financial Studies 5*(1), 123–152.

CME Group (2010). Algorithmic trading and market dynamics. Technical report, CME Group.

CME Group (2011). Bottom line vs top of book: Eurodollar futures at CME group and elsewhere. Technical report, CME Group.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011). Nonparametric instrumental regression. *Econometrica 79*(5), 1541–1565.

Engle, R. F., R. Ferstenberg, and J. Russell (2012). Measuring and modeling execution cost and risk. *Journal of Portfolio Management 38*(2), 14–28.

Florens, J.-P., J. Johannes, and S. Van Bellegem (2012). Instrumental regression in partially linear models. *The Econometrics Journal 15*(2), 304–324.

Foucault, T., J. Hombert, and I. Roşu (2015). News trading and speed. *The Journal of Finance 71*(1), 335–382.

Frino, A. and M. D. McKenzie (2002). The impact of screen trading on the link between stock index and stock index futures prices: Evidence from UK markets. European Financial Management Association (EFMA) 2002 London Meetings.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Hall, P., and J. L. Horowitz. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics 33*(6), 2904–2929.

Harris, L. (1989). S&P 500 cash stock price volatilities. *The Journal of Finance 44*(5), 1155–1175.

Hasbrouck, J. (2018). High-frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis 53*(2), 613–641.

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

Hendershott, T. and P. C. Moulton (2011). Automation, speed, and stock market quality: The NYSE's hybrid. *Journal of Financial Markets 14*(4), 568–604.

Huang, R. D. and H. R. Stoll (1994). Market microstructure and stock return predictions. *The Review of Financial Studies 7*(1), 179–213.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun (2017). The flash crash: High-frequency trading in an electronic market. *The Journal of Finance 72*(3), 967–998.

Kosinski, R. J. (2008). A literature review on reaction time. Technical report, Clemson University.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Kyle, A. S. and A. A. Obizhaeva (2016). Market microstructure invariance: Empirical hypotheses. *Econometrica 84*(4), 1345 – 1404.

Lee, C. M. C. and M. J. Ready (1991). Inferring trade direction from intraday data. *The Journal of Finance 46*(2), 733–746.

Menkveld, A. J. and M. A. Zoican (2014). Need for speed? Exchange latency and market quality. *Journal of Financial Economics 14*, 71–100.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Percival, D. B. and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

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

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), 931–954.

Rosu, I., E. Sojli, and W. W. Tham (2021). Quoting activity and the cost of capital. *Journal of Financial and Quantitative Analysis 56*(8), 2764-2799.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review 6*(2), 41–49.

Stoll, H. R. and R. E. Whaley (1990). The dynamics of stock index and stock index futures returns. *Journal of Financial and Quantitative Analysis 25*(4), 441–468.

Watanabe, M. (2008). A model of stochastic liquidity. Technical report 03-18, Yale ICF Working Paper.

Yu, Y. and D. Ruppert (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association 97*(460), 1042–1054.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Data Citation:

[dataset] Nie, Jing; Malagon, Juliana; Williams, Julian (2022), Dataset for The impact of high speed quoting on execution risk dynamics: Evidence from interest derivatives markets, Harvard Dataverse, Version 1, https://doi.org/10.7910/DVN/SSZMGV

Table 1: Sample Characteristics and Descriptive Statistics

Panel A: Size data sample for best bid and best ask and the quote-to-trade volume ratio

Contracts	**Best Asks** Average Volume ($ quadrillion)	Average no. of obs. (million)	**Best Bids** Average Volume ($ quadrillion)	Average no. of obs. (million)	**Quotes/Trades Ratio** Average Ask Volume /Trades Volume	Average Bid Volume /Trades Volume
2010 Delivery	47.38	23.04	46.09	22.92	1,679.14	1,633.37
2011 Delivery	74.23	23.34	74.73	23.39	4,949.24	4,982.33
2012 Delivery	69.64	24.05	68.23	24.20	5,554.05	5,441.72
2013 Delivery	181.74	24.05	183.37	24.17	14,321.50	14,449.80
2014 Delivery	171.92	24.53	166.41	24.60	7,777.39	7,528.03
2005/15 Delivery	83.23	25.65	80.73	25.79	2,134.54	2,070.42
2016/16 Delivery	56.75	25.17	57.18	25.23	1,296.57	1,306.50
2017/17 Delivery	65.88	21.88	66.62	21.98	1,116.90	1,129.53
2008 Delivery	35.92	21.78	36.49	21.92	518.73	527.09
2009 Delivery	19.85	23.31	19.74	23.32	403.19	401.02

Panel B: Size data sample for order book data and the quote-to-trade volume ratio

Contracts	**Order Book Asks**		**Order Book Bids**		**Quotes/Trades Ratio**	
2010 Delivery	58.75	20.46	58.58	20.46	2,082.05	2,076.03
2011 Delivery	139.01	32.07	137.53	32.07	9,268.03	9,169.35
2012 Delivery	156.87	42.73	155.26	42.73	12,511.88	12,383.46
2013 Delivery	326.52	39.07	316.77	39.07	25,730.69	24,962.36
2014 Delivery	631.86	47.52	618.16	47.52	28,584.09	27,964.33
2005/15 Delivery	408.38	57.13	409.42	57.13	10,473.62	10,500.29
2016/16 Delivery	149.00	45.21	149.93	45.21	3,404.30	3,425.55
2017/17 Delivery	32.85	22.64	32.88	22.64	556.93	557.44
2008 Delivery	5.51	9.97	5.50	9.97	79.58	79.44
2009 Delivery	7.34	12.56	7.18	12.56	149.11	145.86

Notes: This table depicts data sample size for both best bid and ask, order book data and quote-to-trades volume ratio. Panel A represents the average volume, the number of observations, and the volume ratios of quotes to trades for the best bid and ask dataset. The best bid and best ask data are from the Thomson Reuters Tick History database for from 2008 to 2014. We calculate the average number of observations and the average quotes volume for ED contracts having their maturity date in the same year. Hence, the 2010 delivery contracts include four 10-year futures contracts from 2000 to 2010 – namely, the March 2010, June 2010, September 2010 and December 2010 contract. The minimum tick size on the exchange is 1/4 of a basis point for the nearest expiring contract and 1/2 otherwise. Panel B represents the average volume, the average number of observations, and the quotes to trades volume ratios for the entire limited order book. The order book data is from July 1, 2008 to January 1, 2014, and this yields 2,339 days times 40 contracts for our final daily frequency regression analysis.

Table 2: The Wavelet Variance and Covariance Ratio for the Inside Quotes

		Time scale	0% - 25%	26% - 50%	51% - 75%	76% - 100%	Time scale	0% - 25%	26% - 50%	51% - 75%	76% - 100%
				Asks Variance					**Bids Variance**		
	$w = 1$	14.3sec	1.5223	1.6642	1.8574	2.6613	14.1sec	1.5157	1.6654	1.8705	2.6380
30	$w = 3$	57sec	1.1933	1.3057	1.4947	1.9823	56.3sec	1.1913	1.2876	1.4552	1.8969
days	$w = 5$	3.8min	1.0755	1.1408	1.2482	1.4973	3.8min	1.0723	1.1372	1.2206	1.4300
before	$w = 7$	15.2min	1.0341	1.0664	1.1286	1.3061	15min	1.0324	1.0621	1.1311	1.2730
	$w = 9$	60.8min	1.0248	1.0447	1.0951	1.2243	60.1min	1.0251	1.0548	1.1111	1.2462
	$w = 1$	6.5sec	1.5054	1.6275	1.7590	2.2335	6.5sec	1.4972	1.6197	1.7574	2.2457
60	$w = 3$	26.2sec	1.1593	1.2737	1.4070	2.2219	26sec	1.1522	1.2675	1.3930	2.1705
days	$w = 5$	1.7min	1.0577	1.1160	1.1921	1.7756	1.7min	1.0559	1.1075	1.1896	1.7243
before	$w = 7$	7min	1.0244	1.0466	1.0871	1.2994	6.9min	1.0237	1.0485	1.0885	1.2854
	$w = 9$	27.9min	1.0155	1.0296	1.0546	1.1944	27.7min	1.0158	1.0331	1.0636	1.2085
	$w = 1$	4sec	1.5102	1.6067	1.7126	2.0749	4sec	1.5008	1.6018	1.7152	2.0695
3	$w = 3$	15.9sec	1.1660	1.2450	1.3593	1.6802	16sec	1.1617	1.2437	1.3392	1.6584
months	$w = 5$	1.1min	1.0634	1.1054	1.1662	1.3264	1.1min	1.0591	1.1020	1.1572	1.3139
before	$w = 7$	4.2min	1.0244	1.0457	1.0735	1.1632	4.3min	1.0248	1.0452	1.0719	1.1671
	$w = 9$	17min	1.0149	1.0260	1.0437	1.1113	17min	1.0159	1.0286	1.0475	1.1285
	$w = 1$	2.5sec	1.4876	1.5582	1.6287	1.7641	2.5sec	1.4867	1.5557	1.6232	1.7674
1	$w = 3$	9.9sec	1.1493	1.2022	1.2592	1.3925	10sec	1.1467	1.1989	1.2523	1.3812
Year	$w = 5$	39.8sec	1.0560	1.0871	1.1189	1.1909	39.9sec	1.0547	1.0842	1.1137	1.1797
before	$w = 7$	2.7min	1.0227	1.0379	1.0528	1.0876	2.7min	1.0224	1.0373	1.0517	1.0863
	$w = 9$	10.6min	1.0130	1.0213	1.0303	1.0527	10.6min	1.0138	1.0230	1.0323	1.0581
				Bid-Ask Covariance					**Wavelet Correlation**		
	$w = 1$	14.2sec	1.5191	1.6601	1.8598	2.6410	14.2sec	0.4274	0.5544	0.6050	0.6628
30	$w = 3$	56.7sec	1.1899	1.2870	1.4519	1.8888	56.7sec	0.5732	0.7083	0.7868	0.8495
days	$w = 5$	3.8min	1.0676	1.1215	1.1930	1.3660	3.8min	0.7242	0.8312	0.8890	0.9355
before	$w = 7$	15.1min	1.0254	1.0443	1.0813	1.1948	15.1min	0.8089	0.8975	0.9443	0.9700
	$w = 9$	60.4min	1.0128	1.0251	1.0489	1.1089	60.4min	0.8264	0.9110	0.9536	0.9765
	$w = 1$	6.5sec	1.4976	1.6178	1.7494	2.1800	6.5sec	0.4718	0.5733	0.6190	0.6681
60	$w = 3$	26.1sec	1.1515	1.2612	1.3716	1.7853	26.1sec	0.5834	0.7288	0.7928	0.8686
days	$w = 5$	1.7min	1.0521	1.0983	1.1555	1.3276	1.7min	0.7124	0.8466	0.9045	0.9474
before	$w = 7$	7min	1.0175	1.0338	1.0600	1.1283	7min	0.8108	0.9207	0.9548	0.9769
	$w = 9$	27.8min	1.0078	1.0151	1.0250	1.0695	27.8min	0.8411	0.9416	0.9687	0.9845
	$w = 1$	4sec	1.5013	1.5981	1.7043	2.0467	4sec	0.4976	0.5849	0.6256	0.6668
3	$w = 3$	15.9sec	1.1579	1.2349	1.3245	1.6043	15.9sec	0.6325	0.7524	0.8065	0.8625
months	$w = 5$	1.1min	1.0558	1.0919	1.1389	1.2673	1.1min	0.7697	0.8663	0.9087	0.9450
before	$w = 7$	4.2min	1.0186	1.0329	1.0512	1.1088	4.2min	0.8621	0.9343	0.9573	0.9760
	$w = 9$	17min	1.0082	1.0142	1.0222	1.0538	17min	0.8930	0.9550	0.9724	0.9844
	$w = 1$	2.5sec	1.4854	1.5521	1.6203	1.7561	2.5sec	0.5677	0.6169	0.6431	0.6728
1	$w = 3$	10sec	1.1423	1.1911	1.2432	1.3586	10sec	0.7292	0.8001	0.8350	0.8725
Year	$w = 5$	39.8sec	1.0493	1.0746	1.1010	1.1562	39.8sec	0.8495	0.8985	0.9227	0.9483
before	$w = 7$	2.7min	1.0171	1.0282	1.0383	1.0616	2.7min	0.9211	0.9510	0.9642	0.9782
	$w = 9$	10.6min	1.0070	1.0117	1.0164	1.0263	10.6min	0.9454	0.9687	0.9776	0.9864

Notes: This table illustrates the average of median wavelet variance, covariance and correlation ratios for best bids and best asks across the 40 Eurodollar contracts. Best bids and asks prices are divided into four different price quartiles: 0-25% prices, 26-50% prices, 51-75% prices and 76-100% prices. This table also reports the average of median wavelet variance, covariance and correlation ratios within different timezones, including 30 days, 60 days, 3 months and 1 year before Eurodollar contract maturity date.

Table 3: Order Book Update Frequency

		Mean (ms)	Median (ms)	Mode (ms)	Percent (%)	Mean (ms)	Median (ms)	Mode (ms)	Percent (%)	Mean (ms)	Median (ms)	Mode (ms)	Percent (%)
		Order-book Level $j = 1$				Order-book Level $j = 2$				Order-book Level $j = 3$			
Asks	HFT 200ms	33.17	10.94	4.00	69.01	37.22	13.67	5.00	66.45	33.57	10.53	5.00	81.21
	HFT 150ms	26.85	10.04	4.00	66.11	30.10	12.53	5.00	63.31	27.33	9.73	5.00	78.49
	HFT 100ms	20.36	8.94	4.00	62.07	22.70	10.99	5.00	58.87	20.75	8.59	5.00	74.52
	HFT 75ms	16.68	8.16	4.00	58.96	18.49	9.94	5.00	55.47	16.97	7.86	5.00	71.37
	HFT 50ms	12.69	7.24	4.40	54.54	13.94	8.48	5.00	50.61	12.81	6.99	5.00	66.64
	HFT 25ms	8.15	5.77	4.86	46.61	8.73	6.48	5.00	41.98	8.04	5.47	5.00	57.90
Bids	HFT 200ms	33.90	11.78	5.00	68.13	37.30	13.86	5.00	66.50	33.76	10.90	5.00	81.14
	HFT 150ms	27.42	10.79	5.00	65.20	30.20	12.76	5.00	63.32	27.48	10.05	5.00	78.42
	HFT 100ms	20.83	9.53	5.00	61.14	22.83	11.24	5.00	58.81	20.87	8.90	5.00	74.38
	HFT 75ms	17.12	8.66	5.00	58.03	18.63	10.18	5.00	55.37	17.08	8.09	5.00	71.19
	HFT 50ms	13.04	7.64	5.00	53.55	14.06	8.66	5.00	50.44	12.90	7.16	5.00	66.44
	HFT 25ms	8.37	6.05	5.00	45.49	8.82	6.59	5.00	41.68	8.10	5.59	5.00	57.57
		Order-book Level $j = 4$				Order-book Level $j = 5$				All Order-book Levels			
Asks	HFT 200ms	23.63	5.19	2.69	86.47	20.89	4.66	2.56	86.37	35.49	12.71	5.00	77.90
	HFT 150ms	19.33	4.99	2.69	84.71	17.33	4.54	2.56	84.78	28.69	11.75	5.00	75.48
	HFT 100ms	14.89	4.69	2.75	82.15	13.55	4.35	2.56	82.47	21.69	10.33	5.00	72.02
	HFT 75ms	12.41	4.44	2.75	80.20	11.45	4.21	2.56	80.69	17.70	9.28	5.00	69.34
	HFT 50ms	9.57	4.08	2.75	77.23	9.03	4.00	2.56	77.96	13.38	8.03	5.00	65.40
	HFT 25ms	6.31	3.55	2.69	71.46	6.17	3.57	3.02	72.56	8.45	6.21	5.00	58.10
Bids	HFT 200ms	24.80	5.38	2.87	84.21	21.86	4.90	2.81	84.85	35.66	13.10	5.00	76.97
	HFT 150ms	20.25	5.14	2.87	82.44	18.10	4.73	2.81	83.17	28.86	12.10	5.00	74.51
	HFT 100ms	15.52	4.90	2.87	79.84	14.10	4.52	2.81	80.71	21.86	10.66	5.00	70.98
	HFT 75ms	12.85	4.62	2.87	77.87	11.89	4.34	2.81	78.81	17.87	9.60	5.00	68.25
	HFT 50ms	9.90	4.27	2.87	74.86	9.29	4.11	2.81	75.95	13.53	8.26	5.00	64.25
	HFT 25ms	6.48	3.63	2.87	69.15	6.28	3.63	3.21	70.42	8.55	6.37	5.00	56.86

Notes: This table reports the central tendency for both the daily average update frequency across 40 contracts. Here HFT proxies are computed with different thresholds (200ms, 150ms, 100ms, 75ms, 50ms and 25ms) for both ask side and bid side. The unit of order book update speed is in milliseconds (ms). For the bid and ask side of the order book, we take the daily average value of updating speed below thresholds across all 5 order book levels and across all 40 Eurodollar contracts. See Figure 5 for more details.

Table 4: The Semi-parametric Regressions Results

		Liquidity Measures								Variance Ratios								
		S^Q	$S^{Q^{1/2}}$	S^D	S^E	$S^{R,5}$	$S^{R,30}$	$S^{AS,5}$	$S^{AS,30}$	R_{a1}	R_{b1}	R_{ab1}	R_{a4}	R_{b4}	R_{ab4}	R_{a9}	R_{b9}	R_{ab9}
Volume weighted average																		
	$L(Y_{i,\mathcal{T}})$	0.396	0.388	0.922	0.096	0.073	0.573	0.132	0.620	-0.009	-0.002	0.000	0.000	0.013	0.000	-0.004	0.000	0.005
Ask		(20.79)***	(20.68)***	(136.82)***	(4.65)***	(3.53)***	(33.85)***	(6.45)***	(38.30)***	(-0.42)	(-0.07)	(-0.02)	(-0.01)	(-0.60)	(-0.01)	(-0.19)	(-0.02)	(-0.23)
HFT Proxy	R^2	0.61	0.61	0.53	0.27	0.19	0.18	0.19	0.20	0.14	0.14	0.12	0.07	0.11	0.12	0.06	0.12	0.04
	N	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339
	$L(Y_{i,\mathcal{T}})$	0.419	0.413	0.921	0.101	0.069	0.579	0.129	0.616	-0.010	-0.002	0.000	-0.001	0.013	0.000	-0.005	0.001	0.004
Bid		(21.87)***	(21.76)***	(136.70)***	(4.18)***	(3.35)***	(34.05)***	(6.37)***	(38.05)***	(-0.49)	(-0.12)	(0.00)	(-0.03)	(-0.58)	(-0.01)	(-0.23)	(-0.03)	(-0.2)
HFT Proxy	R^2	0.59	0.59	0.53	0.29	0.21	0.18	0.22	0.22	0.16	0.14	0.12	0.11	0.15	0.12	0.07	0.12	0.05
	N	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339	2,339
Panel																		
	$L(Y_{i,\mathcal{T}})$	0.737	0.732	0.907	0.655	0.075	0.591	0.131	0.613	0.004	0.002	0.000	0.047	0.124	0.000	0.001	0.003	0.006
Ask		(207.20)***	(205.53)***	(412.92)***	(163.67)***	(14.23)***	(138.01)***	(24.94)***	(146.29)***	(-0.76)	(-0.43)	(-0.09)	(8.90)***	(23.48)***	(-0.09)	(-0.21)	(-0.58)	(-1.09)
HFT Proxy	R^2	0.22	0.22	0.22	0.16	0.03	0.03	0.03	0.03	0.07	0.06	0.04	0.08	0.08	0.04	0.02	0.03	0.01
	N	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,490	35,491
	$L(Y_{i,\mathcal{T}})$	0.739	0.733	0.907	0.656	0.075	0.591	0.131	0.613	0.004	0.002	0.000	0.047	0.123	0.000	0.001	0.003	0.006
Bid		(207.42)***	(205.80)***	(412.95)***	(163.74)***	(14.23)***	(137.96)***	(24.97)***	(146.24)***	(-0.66)	(-0.33)	(0.00)	(8.82)***	(23.42)***	(-0.01)	(-0.20)	(-0.58)	(-1.12)
HFT Proxy	R^2	0.22	0.22	0.22	0.15	0.03	0.03	0.03	0.04	0.07	0.06	0.02	0.06	0.09	0.02	0.03	0.03	0.02
	N	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491	35,491

T statistics in parentheses

*** p<0.01

Notes: This table illustrates coefficients estimate results from the semi-parametric regression: $Y_{i,\mathcal{T}} = X_{i,\mathcal{T}}\beta + \mathscr{G}[Z_{i,\mathcal{T}}] + \varepsilon_{i,\mathcal{T}}$, where $Y_{i,\mathcal{T}}$ is our measure of execution risk/market quality for the i contract type for day $\mathcal{T}$, $X_{i,\mathcal{T}}$ are our $k = 2$ linear regressors, $Z_{i,\mathcal{T}}$ is our $p = 1$ high-frequency order flow proxy thresholded at 25 milliseconds, and $W_{i,\mathcal{T}}$ is a set of instrumental variables, where $\mathbb{E}[\varepsilon_{i,\mathcal{T}}|W_{i,\mathcal{T}}] = 0$. Flattening the data set to an index $i \in \{1, \ldots, N\}$, where N is the total number of day-contracts and the index i represents the set of N tuples of $(i, \mathcal{T})$ contract-days. In this table, the dependent variables measuring the execution risk $Y_{i,\mathcal{T}}$ are reported into two groups: liquidity measures and variance/covariance ratios. Liquidity measurements contain the bid-ask spreads (S^Q), quoted half spreads ($S^{Q^{1/2}}$), quoted depth (S^D), effective spreads (S^E), realized spreads 5 minutes ($S^{R,5}$), realized spreads 30 minutes ($S^{R,30}$), adverse selection 5 minutes ($S^{AS,5}$) and adverse selection 30 minutes ($S^{AS,30}$). Variance/covariance ratios include ask variance ratios (R_a), bid variance ratios (R_b) and bid-ask covariance ratios (R_{ab}) at level 1, level 4 and level 9. The lagged one dependent variable is denoted as $L(Y_{i,\mathcal{T}})$. Instrumental variables are adopted to deal with the endogeneity problems, named as $W_{i,\mathcal{T}}$. Volume weighted average dataset uses HFT order flow proxy instrumented on the log of total messages, and the panel dataset adopts HFT order flow proxy instrumented on both the log of total messages and the log of time to maturity.

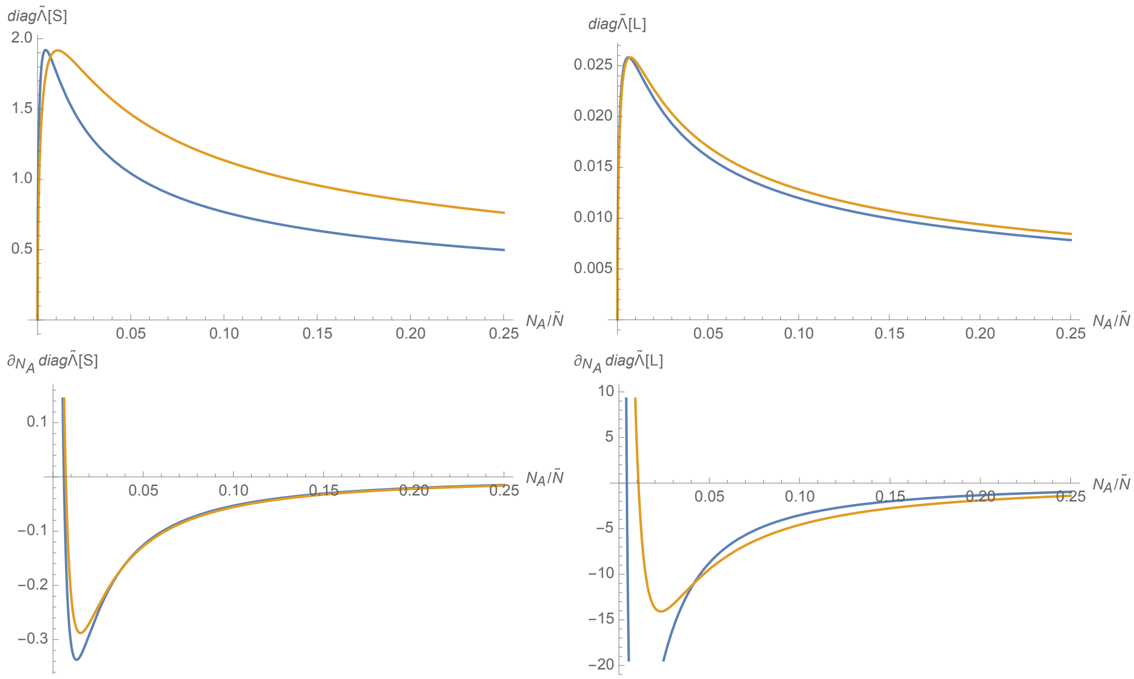


Figure 1: The Effect in Level and Derivative of a Change in the Fraction of Type A Traders versus Type B Traders

Notes: We do not model explicitly the advantage that speed gives the traders other than we presume that the ability to rapidly anticipate the direction of the underlying value process from the order flow provides a systematic advantage. Hence, Type A have a noisy but unbiased expectation of the terminal valuation $\tilde{\delta}(T)$. The model postulates four sources of quadratic variation in the market. $\tilde{\Sigma}$ is the quadratic variation in the underlying asset. $\tilde{\Gamma}$ is the global noise disturbing all type Type A traders forward looking signals. $\tilde{\Psi}$ represents the quadratic variation of the idiosyncratic noise disturbing each of the N_A Type A traders signal over and above the global noise. Finally, $\tilde{\Phi}$ is the variance-covariance matrix describing the quadratic variation of each of the N_B traders random submissions. In the four quadrants we plot the diagonal elements of the resulting $\tilde{\Lambda}$ matrix from the market clearing equilibrium in Theorem 1 and its derivative with respect to $N_A/\tilde{N}$ where $\tilde{N} = N_A + N_B$. Recalling that $\tilde{f}(T) = \tilde{\delta}(T) + \tilde{\Lambda}\tilde{e}$, where $\tilde{e}$ is the aggregate net order flow.

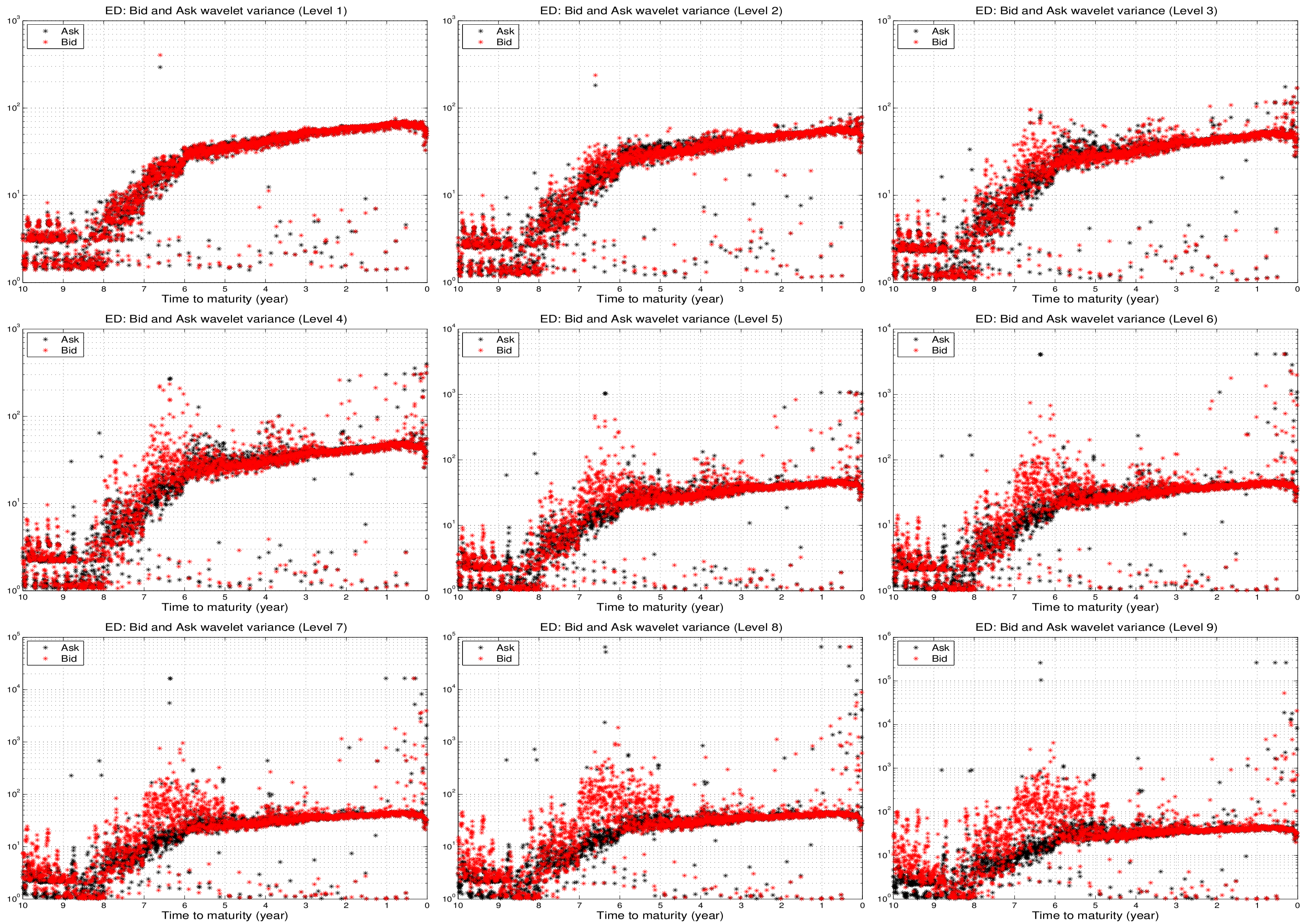


Figure 2: Term structure of variance ratios for Eurodollar futures contracts from MODWT analysis

Notes: The black markers reports the best ask variance ratios and red markets report the best bid variance ratios. We eliminate days with less than 5,000 bid or ask quotes. Level 1 refers to the highest frequency timestamp ($w = 1$) and level 9 is the lowest frequency timestamp ($w = 9$).

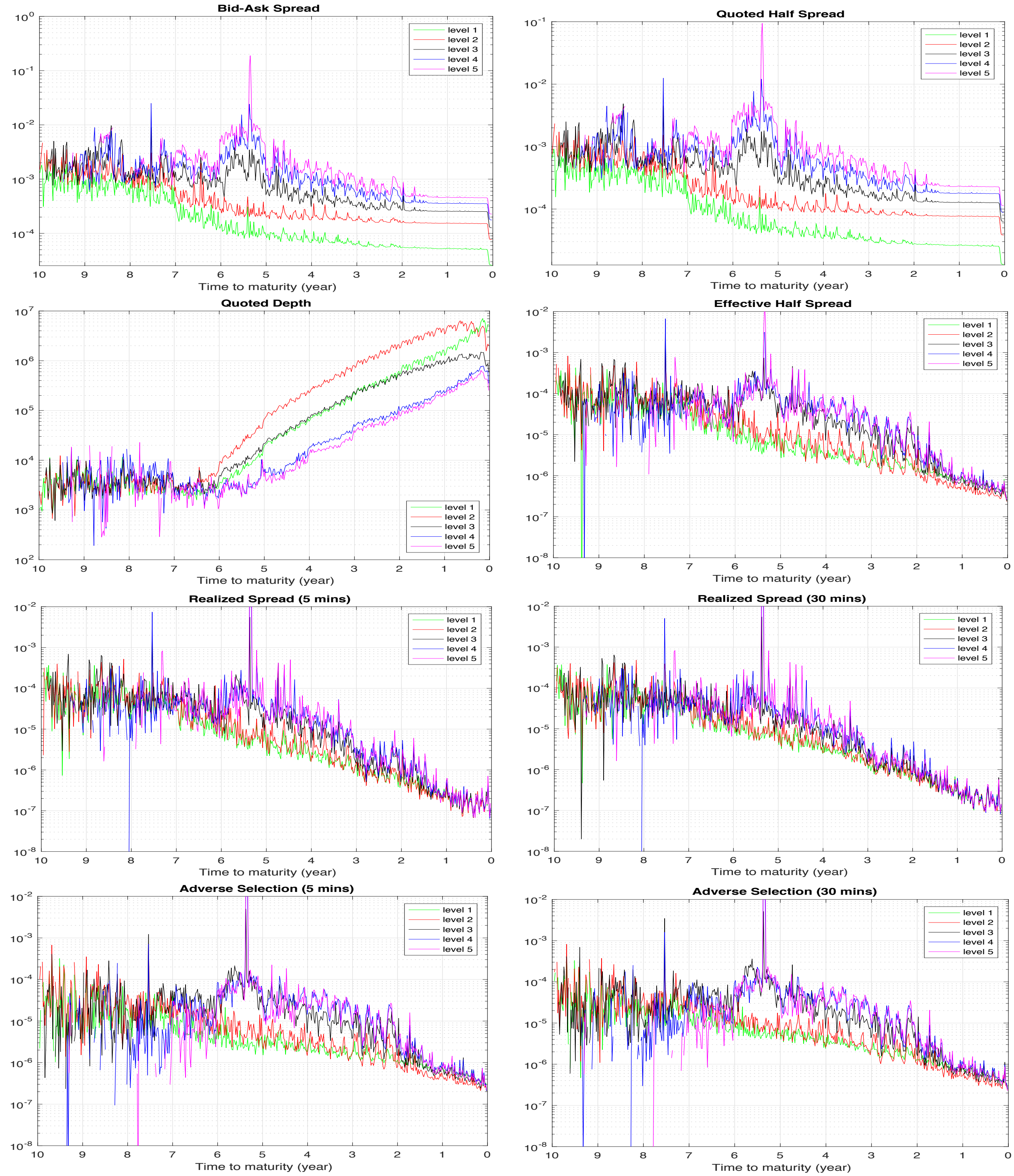
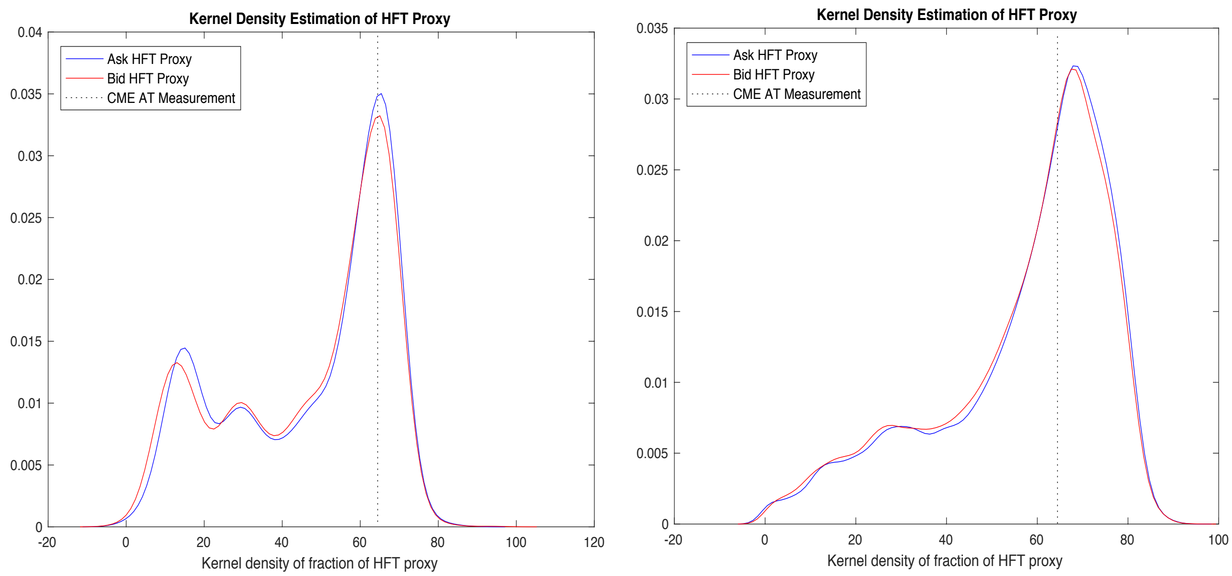


Figure 3: Liquidity Spreads

Notes: This figure displays the weekly average of liquidity measures at the five levels for all the contracts including bid-ask spreads, quoted half spreads, quoted depth, effective half spreads, realized spreads (5 mins and 30 mins) and adverse selection (5 mins and 30 mins). Theses spreads are averaged by time to maturity across 40 contracts, hence the x axis is the year to maturity.



(a) Volume weighted average

(b) Panel

Figure 4: Kernel Density Estimation of HFQ Proxies

Notes: Graph (a) shows the kernel density of the bids and asks daily fraction of high-frequency traders with 2,339 observations in the volume weighted average dataset; and graph (b) is the density of daily HFQ proxies with 35,491 observations in the panel dataset. The black dotted line represents the point proportion of computerized algorithmic traders order flow from the CME Group (2010) with knowledge of the connection type from the account numbers, which they report that the proportion of algorithmic activities is 64.46% on Eurodollar futures markets.

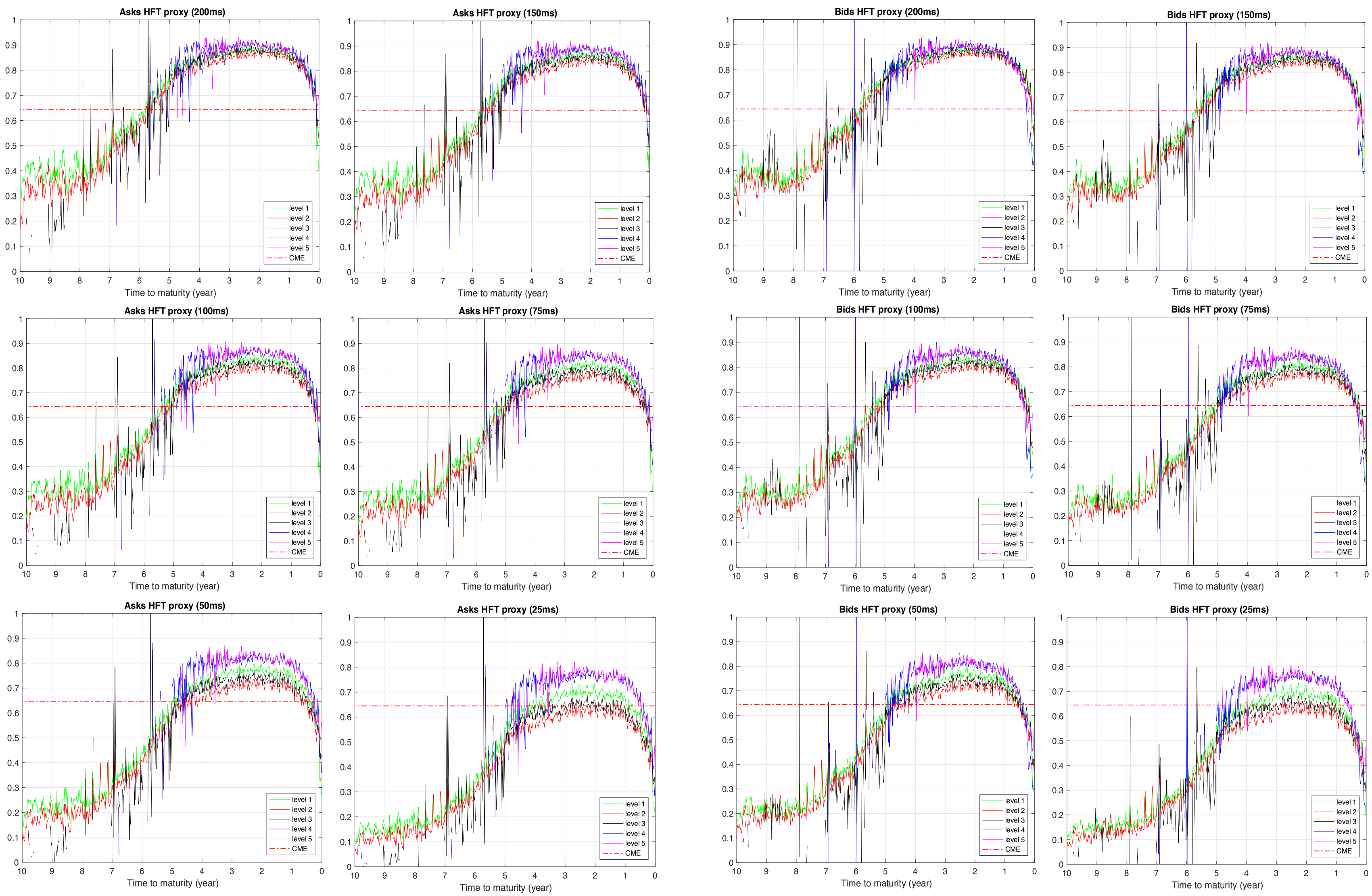


Figure 5: Order Book Quotes High-frequency Trading Fraction

Notes: This figure portrays the weekly average proportion of high-frequency trading in both ask and bid sides from order book level 1 to level 5 across all the contracts. The fraction of high-frequency trading A is calculated as $\mathrm{A}_{j,\mathcal{T},\Delta t} = \mathrm{M}^{\Delta t}_{j,\mathcal{T}}/\mathrm{M}_{j,\mathcal{T}}$, where $\mathrm{M}_{j,\mathcal{T}}$ is the total number of messages for ask and bid side level j for the day $\mathcal{T}$; and $\mathrm{M}^{\Delta t}_{j,\mathcal{T}}$ is the high-frequency trading messages, defined as a message with time stamp updates Δt of less than the different thresholds (200ms, 150ms, 100ms, 75ms, 50ms and 25ms) for ask and bid side level j for the day $\mathcal{T}$. The value of x axis is the years to maturity. The red dashed line represents the proportion of HFQ proxy from the CME Group (2010), which they report that the proportion of high-frequency activities is 64.46% on Eurodollar futures markets.

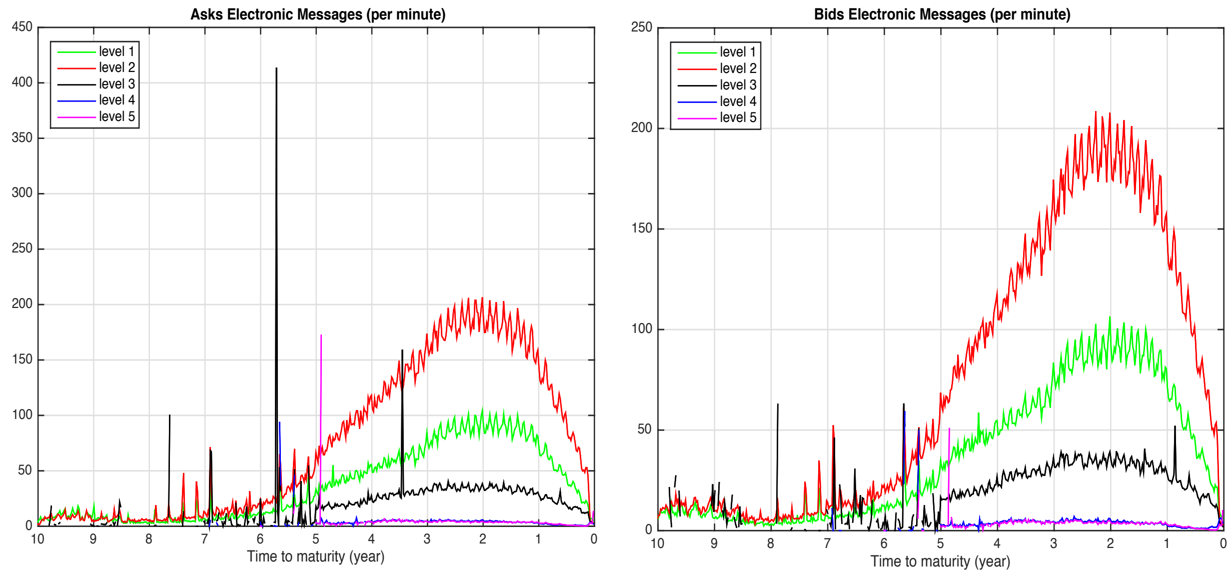
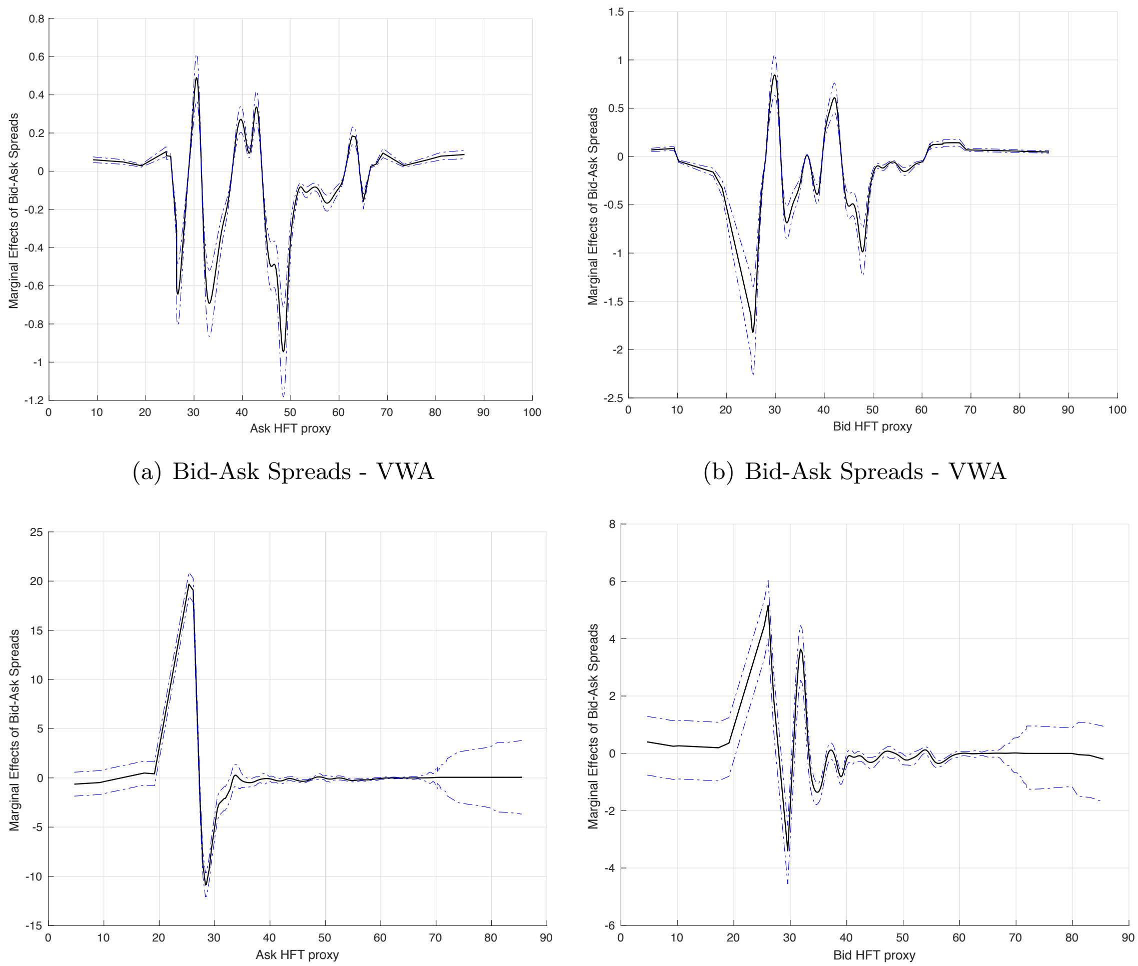


Figure 6: Total Messages By Quote Depth Level

Notes: These two graphs capture the weekly average of total messages (per minute) in both ask and bid sides at five different levels. The total messages per minute, $\mathrm{M}_{j,\mathcal{T},\Delta}$, can be calculated as $\mathrm{M}_{j,\mathcal{T},\Delta} = \mathrm{M}_{j,\mathcal{T}}/T^*_{\mathcal{T}}$, where $\mathrm{M}_{j,\mathcal{T}}$ is the total number of messages for bid- or ask-side order book level j for the day $\mathcal{T}$; and trading time, $T^*_{\mathcal{T}}$, is the length of the first order to the last order on an individual day $\mathcal{T}$, which is measured in minutes. The value of x axis is the years to maturity.



(a) Bid-Ask Spreads - VWA

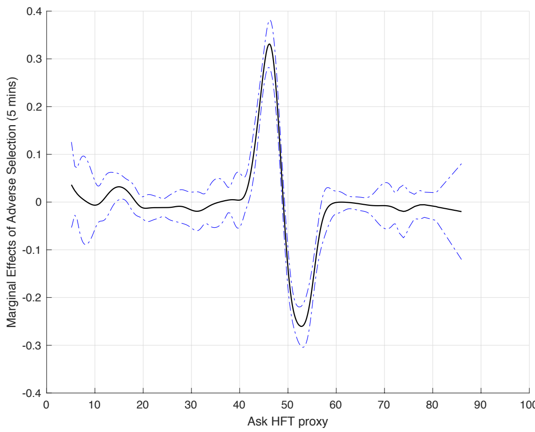
(b) Bid-Ask Spreads - VWA

(c) Bid-Ask Spreads - Panel

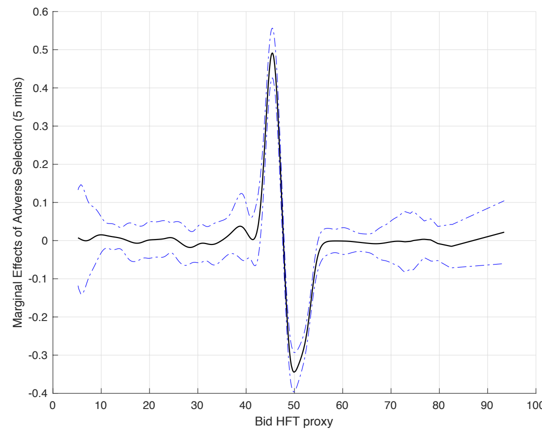
(d) Bid-Ask Spreads - Panel

Figure 7: Marginal Effects of Bid-Ask Spreads

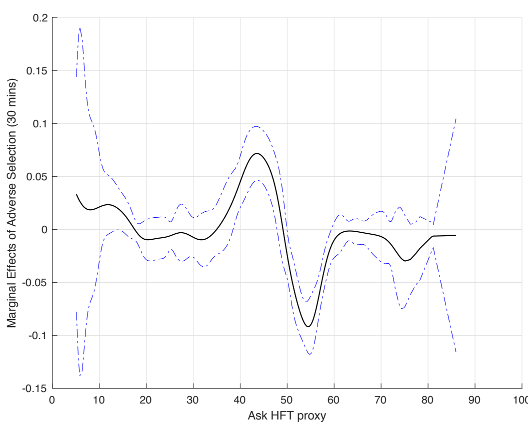
Notes: This figure represents the marginal effects of high-frequency trading proxy on bid-ask spreads. The black continuous line depicts the marginal effects of HFQ on bid-ask spreads in Subplots (a) to (b) with the volume weighted average dataset and Subplots (c) and (d) with the panel dataset. The blue dashed lines are the 95% lower and upper confidence bounds of bid-ask spread marginal effects.



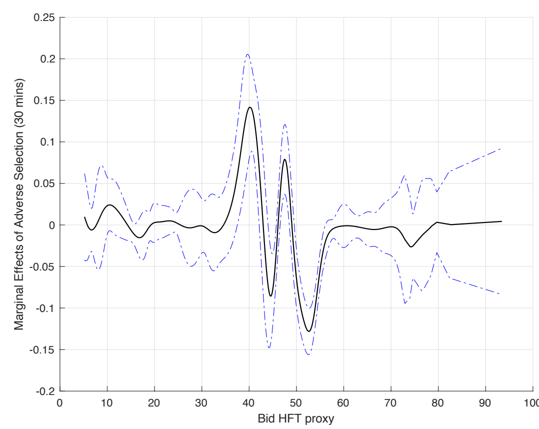
(a) Adverse Selection (5 mins) - VWA



(b) Adverse Selection (5 mins) - VWA



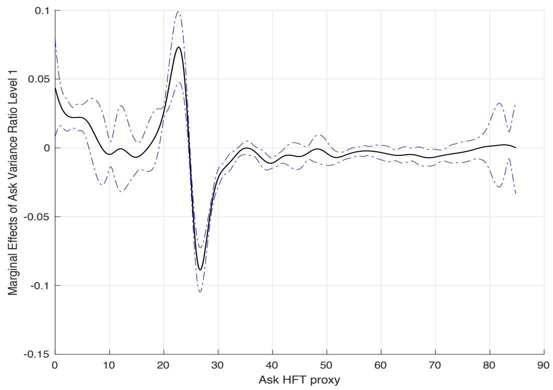
(c) Adverse Selection (30 mins) - VWA



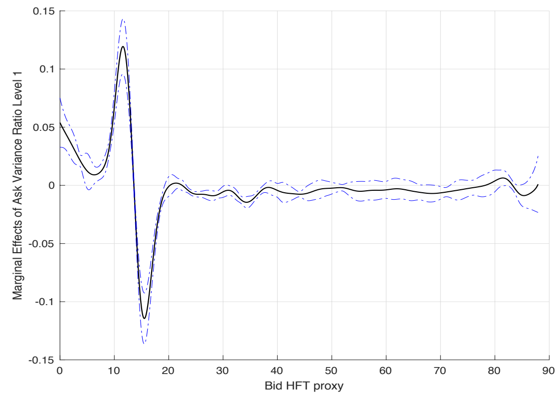
(d) Adverse Selection (30 mins) - VWA

Figure 8: Marginal Effects of Adverse Selection

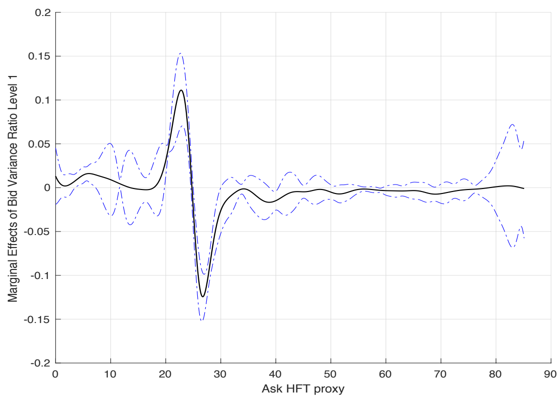
Notes: This figure represents the marginal effects of high-frequency trading proxy on adverse selection (5 mins) in Subplots (a) to (b) and, on adverse selection (30 mins) in (c) and (d), both with the volume weighted average dataset. The blue dashed lines are the 95% lower and upper confidence bounds of bid-ask spread marginal effects.



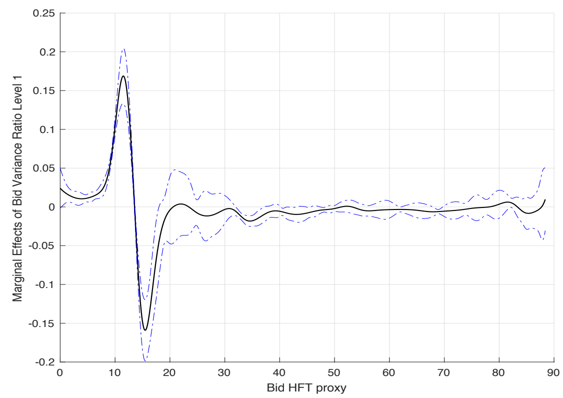
(a) Ask Variance Level 1 - Panel



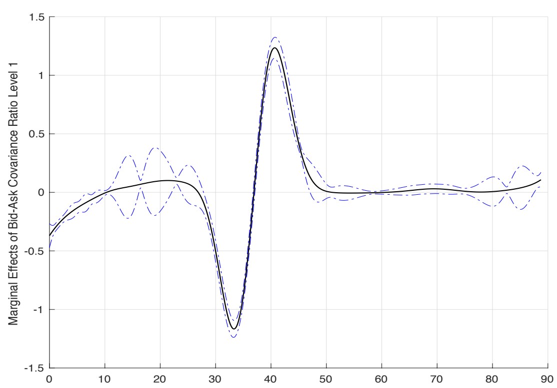
(b) Ask Variance Level 1 - Panel



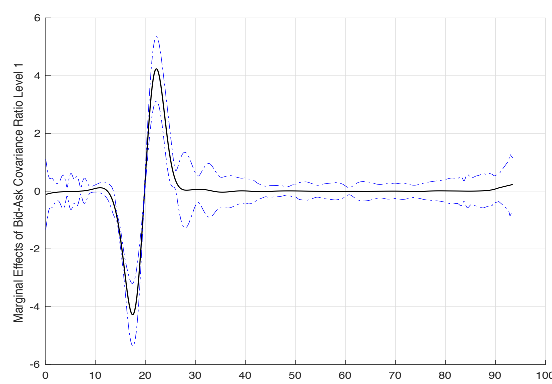
(c) Bid Variance Level 1 - Panel



(d) Bid Variance Level 1 - Panel



(e) Bid Ask Covariance Level 1 - Panel



(f) Bid Ask Covariance Level 1 - Panel

Figure 9: Marginal Effects of Variance and Covariance at Timescale Level 1

Notes: This figure represents the marginal effects of HFQ on variance and covariance ratio at timescale level 1 with the panel dataset. The black continuous line depicts the marginal effects of HFQ on variance ratio in Subplots (a) to (d), and on covariance ratio in Subplots (e) and (f). The blue dashed lines are the 95% lower and upper confidence bounds of marginal effects are by bootstrap resampling.

Appendix A. Extended Theorems and Proofs

Appendix A.1. Proof of Theorem 1

Our proof proceeds in effectively the same manner as the proof in Watanabe (2008), pages 33-39, however we generalize it to the case where the $\tilde{N} = [N_i]$ is a diagonal matrix of traders within each futures class and we augment the details of the proof with the linear multivariate rational expectations form presented in Admati (1985) and Admati and Pfleiderer (1988, 1989).

The initial part of this proof illustrates how our model is essentially the same as that contained in Watanabe (2008), page 34-39, in turn an adaptation of Admati and Pfleiderer (1989). We specifically concentrate on the fact that our addition does not lead to the loss of tractability to the analytic solution. Fortunately, the original linear equilibrium for the situation when $\tilde{N}$ is a scalar holds under this restrictive assumption, while it would not hold for the case when $\tilde{N}$ is a PD integer matrix. For this application the restriction is useful and defensible, however future work will focus on this treatment. First, we set up the general form of the asset price progression,[26] which is given by $\tilde{f}(t) = \tilde{\Lambda}_0 + \tilde{\Lambda}_1 \tilde{e}$, following our notation scheme the indexed demand side scheme is $\tilde{a}_n = \tilde{B}_0 + \tilde{B}_1 \tilde{\xi}_n$ where the matrices $\tilde{\Lambda}_1 \tilde{e}$ and $\tilde{B}_1$ and the vectors $\tilde{\Lambda}_0 \tilde{e}$ and $\tilde{B}_0$ are computed from the model structure. Using the first filtration

$$\tilde{\xi}_n \equiv \tilde{\mathbb{E}}[\hat{\delta}|\mathcal{F}_n] = \mathrm{cov}(\tilde{\delta}, \tilde{b}_n')\mathrm{var}_t^{-1}(\tilde{b}_n)\tilde{b}_n = \tilde{\Sigma}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})^{-1}\tilde{b}_n \tag{A.1}$$

the profit maximization condition for each trader is

$$\max_{\tilde{a}_n} \tilde{\mathbb{E}}[(\tilde{\delta}_T - \tilde{\Lambda}_0 - \tilde{\Lambda}_1(\tilde{a}_n + \sum_{i \neq n} \tilde{a}_i + \tilde{d}))'\tilde{a}_n|\mathcal{F}_n]\tilde{d} \tag{A.2}$$

with system consistent net order flow $\tilde{b} = \tilde{a}_n + \sum\limits_{i \neq n} \tilde{a}_i + \tilde{d}$, hence, the asymmetric information in trading process yields a first order condition[27] of $0 = \tilde{\delta} + \tilde{\xi}_n - \tilde{\Lambda}_0 - \tilde{\Lambda}_1(2\tilde{a}_n + \sum_{i \neq n} [\tilde{a}_i|\mathcal{F}_n])$ so that

[26] The interested reader is directed to Watanabe (2008) for a fuller description of the simpler equilibrium conditions as we concentrate on the adjustments needed with the inclusion of segmented trading restrictions.

[27] consider that this assumption only works if the market is truly partitioned, i.e. an informed trader is restricted to the A market; and this will not hold if the traders have free access to alternate assets within the market.

the second order constraint is $\tilde{\mathbb{E}}[\tilde{a}_i|\mathcal{F}_n] = \tilde{B}_0 + \tilde{B}_1\tilde{\mathbb{E}}[\tilde{\xi}_i|\mathcal{F}_n]$, where $\tilde{\mathbb{E}}[\tilde{\xi}_j|\mathcal{F}_n]$ is the expectation of another trader $j \in N_i$ estimate of $\tilde{\delta}(t)$. The conditional variance and covariance of the system in Equation (A.1), for any $j \notin n$ is therefore

$$\tilde{\Xi} \equiv \text{var}(\tilde{\xi}_n) = \tilde{\Sigma}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})^{-1}\tilde{\Sigma} \qquad \text{Covariance} \tag{A.3}$$

$$\tilde{\Xi}_c \equiv \text{cov}(\tilde{\xi}_i, \tilde{\xi}'_n) = \tilde{\Sigma}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})^{-1}(\tilde{\Sigma} + \tilde{\Gamma})(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})^{-1}\tilde{\Sigma} \qquad \text{Cross-covariance} \tag{A.4}$$

Then, setting the prior mean $\tilde{\mathbb{E}}(\tilde{\xi}_i) = 0$, the conditional update simplifies to $\tilde{\mathbb{E}}[\tilde{\xi}_i|\mathcal{F}_n] = \tilde{\Xi}_c\tilde{\Xi}^{-1}\tilde{\xi}_n$. We now diverge somewhat from Watanabe (2008) as the system consistent first order condition with informed traders will be

$$0 = \tilde{\delta} + \tilde{\xi}_n - \tilde{\Lambda}_0 - 2\tilde{\Lambda}_1(\tilde{B}_0 + \tilde{B}_1\tilde{\xi}_n) - (\tilde{N} - 1)\tilde{\Lambda}_1[\tilde{B}_0 + \tilde{B}_1\tilde{\Xi}_c\tilde{\Xi}^{-1}\tilde{\xi}_n] \tag{A.5}$$

note that the inclusion of $(\tilde{N} - 1)\tilde{\Lambda}_1[\tilde{B}_0 + \tilde{B}_1\tilde{\Xi}_c\tilde{\Xi}^{-1}\tilde{\xi}_n]$, holds only if $tr\tilde{N} = \tilde{N}$ under the Watanabe (2008) approach. This now solves for our system of equations (4) to (6) of Theorem 1 in the paper, first set

$$0 = \tilde{I} - 2\tilde{\Lambda}_1\tilde{B}_1 - (\tilde{N} - 1)\tilde{\Lambda}_1\tilde{B}_1\tilde{\Xi}_c\tilde{\Xi}^{-1}, \quad or \tag{A.6}$$

$$\tilde{\Lambda}_1\tilde{B}_1 = \tilde{J}^{-1}, \quad with \quad \tilde{J} \equiv 2\tilde{I} + (\tilde{N} - 1)\tilde{\Xi}_c\tilde{\Xi}^{-1} \tag{A.7}$$

back substitution equations from (A.3) and (A.4), reproduces the expression suggested for $\tilde{J}$ in Theorem 1 Equation (6). Again we see the Hadamard product enter the system, $\tilde{\Lambda}_0 = \tilde{\delta} - (\tilde{N} - 1)\tilde{\Lambda}_1\tilde{B}_0$ following from this, the market maker efficiency condition will be

$$\tilde{f}(t) = \tilde{\delta} + \tilde{\mathbb{E}}[\tilde{\delta}|\mathcal{F}\theta_m] = \tilde{\delta} + \text{cov}(\tilde{\delta}, \tilde{e}')\text{var}^{-1}(\tilde{e})(\tilde{e} - \tilde{N}\tilde{B}_0) \tag{A.8}$$

proceeding in the standard fashion reveals the market maker variance covariance expression

$$\text{cov}(\tilde{\delta}, \tilde{e}') = \sum_{i=1}^{\tilde{N}_i} \text{cov}(\tilde{\delta}, \tilde{\delta}'_i)\tilde{B}'_1 = \tilde{N}\tilde{\Xi}\tilde{B}'_1 \tag{A.9}$$

by extension[28] this gives $\mathrm{var}(\tilde{e}) = \tilde{B}_1 \mathrm{var}(\sum_{i=1}^{\tilde{N}_i} \tilde{\xi}_n) \tilde{B}_1' + \tilde{\Psi}$ which can then be explicitly determined by $\mathrm{var}(\tilde{e}) = \tilde{N}\tilde{B}_1\{\tilde{\Xi} + (\tilde{N}-1)\tilde{\Sigma}_{\tilde{c}}\}\tilde{B}_1' + \tilde{\Psi} \equiv \tilde{\Sigma}_{\tilde{e},\tilde{c}}$ this yields the liquidity adjusted pricing formula to be

$$\begin{aligned} \tilde{f}(t) &= \tilde{\delta} + \tilde{N}\tilde{\Xi}\tilde{B}_1'[\tilde{N}\tilde{B}_1\{\tilde{\Xi} + (\tilde{N}-1)\tilde{\Xi}_c\}\tilde{B}_1' + \tilde{\Psi}]^{-1}(\tilde{e} - \tilde{N}\tilde{B}_0) \\ &= \tilde{\delta} + [\tilde{B}_1\{\tilde{I} + (\tilde{N}-1)\tilde{\Xi}_c\tilde{\Xi}^{-1}\} + \tilde{N}^{-1}\tilde{\Psi}\tilde{B}_1'^{-1}\tilde{\Xi}^{-1}]^{-1}(\tilde{e} - \tilde{N}\tilde{B}_0) \end{aligned}$$

again we proceed along the lines of Watanabe (2008), which, in turn, is generalized from Admati and Pfleiderer (1988) with the addition of our separated trader condition,

$$\tilde{\Lambda}_1 = [\tilde{B}_1(\tilde{I} + (\tilde{N}-1)\tilde{\Xi}_c\tilde{\Xi}^{-1}) + \tilde{N}^{-1}\tilde{\Psi}\tilde{B}_1'^{-1}\tilde{\Xi}^{-1}]^{-1}, \quad \textit{or} \tag{A.10}$$

$$\tilde{\Lambda}_1\tilde{B}_1 = [\tilde{J} - \tilde{I} + \tilde{N}^{-1}\tilde{B}_1^{-1}\tilde{\Psi}\tilde{B}_1'^{-1}\tilde{\Xi}^{-1}]^{-1}, \quad \textit{and} \quad \tilde{\Lambda}_0 = \tilde{\delta} - \tilde{N}\tilde{\Lambda}_1\tilde{B}_0 \tag{A.11}$$

eliminating $(\tilde{N}-1)\tilde{\Xi}_c\tilde{\Xi}^{-1}$ leaves a system of four equations for the four unknowns $\tilde{\Lambda}_1\tilde{e}$ and $\tilde{B}_1$ and the vectors $\tilde{\Lambda}_0\tilde{e}$. Cancelling for $\tilde{J}$ yields $\tilde{B}_1^{-1}\tilde{\Psi}\tilde{B}_1'^{-1} = \tilde{N}\tilde{\Xi}$ and then substituting for our standard functional form of $\tilde{B}$, whereby $\tilde{B}_1 = \tilde{\Lambda}_1^{-1}\tilde{J}^{-1}$, we find that $\tilde{\Lambda}_1\tilde{\Psi}\tilde{\Lambda}_1 = \tilde{N}\tilde{J}^{-1}\tilde{\Xi}\tilde{J}'^{-1}$ as such

$$(\tilde{\Psi}^{\frac{1}{2}}\tilde{\Lambda}_1\tilde{\Psi}^{\frac{1}{2}})^2 = \tilde{N}\tilde{\Psi}^{\frac{1}{2}}\tilde{J}^{-1}\tilde{\Xi}\tilde{J}'^{-1}\tilde{\Psi}^{\frac{1}{2}} \equiv \tilde{N}\tilde{M} \tag{A.12}$$

element by element dividing by $\tilde{N}$ yields an expression for $\tilde{M}$, which is of the form VUV' and therefore PD. Therefore a Cholesky factor $\tilde{M}^{\frac{1}{2}}$ exists. The resultant order flow is therefore

$$\tilde{a}_n = \tilde{B}_1\tilde{\xi}_n = \tilde{\Lambda}_1^{-1}\tilde{J}^{-1}\tilde{\Sigma}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})^{-1}\tilde{b}_n = \tilde{\Lambda}_1^{-1}\tilde{J}^{-1}\tilde{\Xi}\tilde{\Sigma}^{-1}\tilde{b}_n \tag{A.13}$$

The matrix $\tilde{\Lambda}_1^{-1}\tilde{J}^{-1}\tilde{\Sigma}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})^{-1}$ is the solution to the autoregressive terms of B_1, given the structure of the solution to $\tilde{\Lambda}_1 = \tilde{\Lambda}$, which is PD the multivariate equivalence of Kyle's lambda is a PD matrix autoregressive coefficient. ■

[28] Once again from standard matrix commutation rules this holds, only under the condition that $tr\tilde{N} = \tilde{N}$.

Appendix A.2. Proof of Proposition 1

From Theorem 1, the variance of the price process is mechanistically specified from the multivariate autoregression as

$$\Delta t\tilde{H} \equiv \mathrm{var}(\tilde{f}(t+\Delta t)-\tilde{f}(t)) = \mathrm{var}(\tilde{\Lambda}\tilde{e}) + \mathrm{var}(\tilde{\delta}-\tilde{\Lambda}\tilde{e}) \tag{A.14}$$

by induction the variance follows from the noise of the submission process and the multivariate extension of Kyle's lambda, therefore the variance iteration will be

$$\mathrm{var}(\tilde{\Lambda}\tilde{e}) = \tilde{\mathbb{E}}[\tilde{\Lambda}\tilde{e}\tilde{e}'\tilde{\Lambda}] = \tilde{\mathbb{E}}[\tilde{\Lambda}\tilde{\mathbb{E}}[\tilde{e}\tilde{e}']\tilde{\Lambda}] = \tilde{\mathbb{E}}[\tilde{\Lambda}\mathrm{var}(\tilde{e})\tilde{\Lambda}] \tag{A.15}$$

Simple rearrangement and substitution from the definitions in Theorem 1 yields

$$\mathrm{var}(\tilde{e}) = \tilde{\Lambda}^{-1}\mathrm{cov}(\tilde{\delta},\tilde{e}') = \tilde{N}\tilde{\Lambda}^{-1}\tilde{\Xi}\tilde{B}_1' = \tilde{N}\tilde{\Lambda}^{-1}\tilde{\Xi}\tilde{J'}^{-1}\tilde{\Lambda}^{-1} \tag{A.16}$$

and therefore the expectation collapses to $\mathrm{var}(\tilde{\Lambda}\tilde{e}) = \tilde{\mathbb{E}}[\tilde{N}\tilde{\Xi}\tilde{J'}^{-1}]$. We now show that the imposition of $\tilde{N} = [N_i]$ runs through the derivation without loss of generality, therefore

$$\tilde{N}\tilde{\Xi}\tilde{J'}^{-1} = \tilde{N}(\tilde{J}'\tilde{\Xi}^{-1})^{-1} = \tilde{N}[2\tilde{\Xi}^{-1} + (\tilde{N}-1)\tilde{\Xi}^{-1}\tilde{\Sigma}_{\tilde{c}}\tilde{\Xi}^{-1}]^{-1} \tag{A.17}$$

The diagonal condition imposed on $\tilde{N}$ reduces the noise to the following diagonal matrix

$$\tilde{N}\tilde{\Xi}\tilde{J'}^{-1} = \tilde{N}[2\tilde{\Xi}^{-1} + (\tilde{N}-1)\tilde{\Sigma}^{-1}(\tilde{\Sigma}+\tilde{\Gamma})\tilde{\Sigma}^{-1}]^{-1} \tag{A.18}$$

Substitution from Theorem 1 definitions yields

$$\tilde{N}\tilde{\Xi}\tilde{J'}^{-1} = \tilde{N}\tilde{\Sigma}[(\tilde{N}+1)(\tilde{\Sigma}+\tilde{\Gamma}) + 2\tilde{\Phi}]^{-1}\tilde{\Sigma} \equiv \tilde{\Sigma}_{\tilde{\Lambda}\tilde{e}} \tag{A.19}$$

Add up all of the variances and impose diagonal restrictions on the covariances to derive the variance condition,

$$\mathrm{var}(\tilde{\delta} - \tilde{\Lambda}\tilde{e}) \ = \tilde{\Sigma} - \mathrm{cov}(\tilde{\delta}, \tilde{e}')\tilde{\Lambda} - [\mathrm{cov}(\tilde{\delta}, \tilde{e}')\tilde{\Lambda}]' + \mathrm{var}(\tilde{\Lambda}\tilde{e}) = \tilde{\Sigma} - \tilde{\Sigma}_{\tilde{\Lambda}\tilde{e}} \tag{A.20}$$

therefore yielding the volatility expectation. ■

Appendix A.3. Proof of Proposition 2

We start with Lemma 2 from Watanabe (2008), but replace and re-derive $(\tilde{B}\sum_{n=1}^{\tilde{N}_i}\tilde{b}_n)$ with $(\tilde{B}\sum_{n=1}^{tr\tilde{N}_i}\tilde{b}_n)$. From this the net order flow in our notation will be,

$$\tilde{\Sigma}_{\tilde{e}} = \mathrm{var}(\tilde{e}) = \mathrm{var}(\tilde{B}\sum_{n=1}^{tr\tilde{N}_i}\tilde{b}_n) + \mathrm{var}(\tilde{d}) = \tilde{N}\tilde{B}[\tilde{N}(\tilde{\Sigma} + \tilde{\Gamma}) + \tilde{\Phi}]\tilde{B}' + \tilde{\Psi} \tag{A.21}$$

therefore in expectations the volume will be $\tilde{\mathbb{E}}_0[\tilde{V}] = \frac{1}{2}\tilde{\mathbb{E}}_0[\tilde{\mathbb{E}}|\sum_{n=1}^{tr\tilde{N}_i}\tilde{a}_n| + \tilde{\mathbb{E}}|\tilde{d}| + \tilde{\mathbb{E}}|\tilde{e}|]$. The last part is actually identical to Lemma 2 of Watanabe (2008), except with the inclusion of the Hadamard product expression, under our diagonal specification, it is easy to see that this is in effect two separate Watanabe (2008) models. This is sequentially simplified to

$$\begin{aligned} \mathrm{var}(\tilde{a}_n) \quad &= \quad \tilde{B}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})\tilde{B}' = \tilde{\Lambda}^{-1}\tilde{J}^{-1}\tilde{\Xi}\tilde{\Sigma}^{-1}(\tilde{\Sigma} + \tilde{\Gamma} + \tilde{\Phi})\tilde{\Sigma}^{-1}\tilde{\Xi}^{-1}\tilde{\Xi}\tilde{J}'^{-1}\tilde{\Lambda}^{-1} \\ &= \quad \tilde{\Lambda}^{-1}\tilde{J}^{-1}\tilde{\Xi}\tilde{J}'^{-1}\tilde{\Lambda}^{-1} = \tilde{N}^{-1}\tilde{\Psi} \end{aligned} \tag{A.22}$$

recalling the definitions from Proposition 1 and 2 and hence the cross market volume. ■

Appendix B. Our Adapted IV Robinson Estimator

We adjust the standard instrumental variable semi-parametric partially linear for our purposes. The estimator runs over both panel and time series data and we have tested a variety of different Kernels with similar results. One of the major issues with the standard implementations of the IV Robinson Estimator is speed, this approach is usually designed for smaller panel datasets than the one we have deployed here. Furthermore, we need to be able to compare instrumental variables across a number of equations describing the market quality and execution risk and our approach to dealing with this problem is below.

The conditional expectation is unknown as such we utilize the Robinson (1988) double residual methodology by applying a nonparametric conditional mean estimator $\mathbb{E}[Y_i|Z_i]$, $\mathbb{E}[X_i|Z_i]$ and $\mathbb{E}[\varepsilon_i|Z_i]$ on Equation (20) in the paper. We will follow Newey and Powell (2003) and adhere to the exponential family of kernels and more specifically a Gaussian kernel. Our major departure from Robinson (1988) and Florens et al. (2012) will be in the bootstrapping procedure we implement to generate the confidence bounds of the model. To begin with we will outline the family of models, using an adaptation of notation in Florens et al. (2012) and then introduce our bootstrap procedure. Let the general instrumented semi-parametric regression be denoted in expectations as:

$$\mathbb{E}[Y_i|Z_i] = \mathbb{E}[X_i|Z_i]\beta + \mathscr{G}[Z_i] + \mathbb{E}[\varepsilon_i|Z_i] \tag{B.1}$$

By subtracting Equation (B.1) from Equation (20), we can build an ordinary least squares (OLS) estimation of model as follows,

$$Y_i - \mathbb{E}[Y_i|Z_i] = (X_i - \mathbb{E}[X_i|Z_i])\beta + (\varepsilon_i - \mathbb{E}[\varepsilon_i|Z_i]) \tag{B.2}$$

where we define $\tilde{Y}_i = Y_i - \mathbb{E}[Y_i|Z_i]$, $\tilde{X}_i = X_i - \mathbb{E}[X_i|Z_i]$ and $\tilde{\varepsilon}_i = \varepsilon_i - \mathbb{E}[\varepsilon_i|Z_i]$. We can recover $\tilde{Y}_i$, $\tilde{X}_i$ from the nonparametric regressions of Y_i and X_i on $\hat{Z}_i$. Then the estimated coefficients vector β can be estimated by Equation (B.2) in the absence of the nonparametric function $\mathscr{G}[Z_i]$. Let β be the standard OLS estimates:

$$\beta = (\textstyle\sum_i \tilde{X}_i \tilde{X}_i'^{-1})(\sum_i \tilde{X}_i \tilde{Y}_i') \tag{B.3}$$

we then regress Z_i onto the filtered values of Y_i to provide an estimate of the non-parametric function $\mathscr{G}[Z_i]$,

$$Y_i - X_i\beta = \mathscr{G}[Z_i] \tag{B.4}$$

However, prior to this estimation stage we need to deal with the problem of the variables X_i and Z_i being endogenous with respect to the disturbance term, as such $\mathbb{E}[\varepsilon_i|X_i, Z_i] \neq 0$, $\mathbb{E}[Y_i|Z_i]$ and $\mathbb{E}[X_i|Z_i]$ cannot be estimated consistently via the non-parametric approach proposed herein.

Our approach is in the spirit of Florens et al. (2012) in that we estimate specifically a time series/panel version of the instrumented semi-parametric model with linear covariates. The target function $\mathscr{G}(\cdot)$ and parameter β are the solution of the functional equation

$$\mathbb{E}[Y_i|W_i] = \mathbb{E}[\mathscr{G}[Z_i]|W_i] + \mathbb{E}[X_i\beta|W_i] \tag{B.5}$$

therefore the optimal non-parametric function should be such that the model disturbances are uncorrelated with the time to maturity and/or the total messages per day. As such, following Florens et al. (2012) who in turn derived their approach from Robinson (1988) therefore Equation (B.5) can be rewritten as the following indefinite integrals:

$$\int \mathrm{d}y\; y\, \frac{\mathscr{F}_{Y|W}(y,\cdot)}{\mathscr{F}_W(\cdot)} = \int \mathrm{d}z\; \mathscr{G}[z]\, \frac{\mathscr{F}_{Z|W}(z,\cdot)}{\mathscr{F}_W(\cdot)} + \int \mathrm{d}x\; x\beta\, \frac{\mathscr{F}_{X|W}(x,\cdot)}{\mathscr{F}_W(\cdot)} \tag{B.6}$$

$$\int \mathrm{d}y\; y\, \mathscr{K}_{Y|W}(y) = \int \mathrm{d}z\; \mathscr{G}[z]\, \mathscr{K}_{Z|W}(z) + \int \mathrm{d}x\; x\beta\, \mathscr{K}_{X|W}(x) \tag{B.7}$$

where $\mathscr{F}_{Y|W}$ denotes the joint density of Y and W, and similarly for $\mathscr{F}_{Z|W}$ and $\mathscr{F}_{X|W}$, and $\mathscr{K}_{Y|W}$, $\mathscr{K}_{Z|W}$ and $\mathscr{K}_{X|W}$ indicate the conditional densities of Y_i, X_i and Z_i given W_i respectively. Computing the instrumental version of the partially linear semi-parametric model is non-trivial as both the first and second steps much account for the non-parametric element recalling the objective of othogonalizing the disturbances with respect to the instruments.

We define $L^2_{\mathscr{V}}(\mathbb{R}^f)$ and $L^2_{\mathscr{U}}(\mathbb{R}^g)$ as the Hilbert spaces of square integrable functions with respect to the two densities $\mathscr{V}$ and $\mathscr{U}$. Florens et al. (2012) indicate that $\mathscr{V}$ and $\mathscr{U}$ are the functions $\mathscr{F}_Z$ and $\mathscr{F}_W$ respectively then the approach is identical to Darolles et al. (2011), or if the functions $\mathscr{V}$ and $\mathscr{U}$ are indicator functions then the approach tends to that of Hall et al. (2005). Indeed, these two cases can be seen as points at the end of a continuum of model choices the econometrician faces. If $\mathscr{V}$ and $\mathscr{U}$ are identified in the continuous domain $[0, 1]$ we can also recover the approach of Newey and Powell (2003), which is a fully non-parametric approach across the model space.

By multiplying with functions of W, integral Equation (B.7) can be formalized as equation of operators as follows,

$$\mathscr{R} = \mathscr{O}_Z\mathscr{G} + \mathscr{O}_X\beta, \tag{B.8}$$

with

$$\mathscr{R} = \mathbb{E}[Y|W]\mathscr{F}_W/\mathscr{U} \tag{B.9}$$

$$\mathscr{O}_X : \mathbb{R}^e \to L^2_{\mathscr{U}}(\mathbb{R}^g) : \tilde{\beta} \mapsto \mathbb{E}[X'\tilde{\beta}|W]\mathscr{F}_W/\mathscr{U} \tag{B.10}$$

$$\mathscr{O}_Z : L^2_{\mathscr{V}}(\mathbb{R}^f) \to L^2_{\mathscr{U}}(\mathbb{R}^g) : \tilde{\mathscr{G}} \mapsto \mathbb{E}[\tilde{\mathscr{G}}[Z]|W]\mathscr{F}_W/\mathscr{U} \tag{B.11}$$

where $\mathscr{R} \in L^2_{\mathscr{U}}(\mathbb{R}^g)$, $\mathscr{G} \in L^2_{\mathscr{V}}(\mathbb{R}^f)$, $\beta \in \mathbb{R}^e$, and $\mathscr{R} \in \mathcal{R}(\mathscr{O}_X) + \mathcal{R}(\mathscr{O}_Z)$ where $\mathcal{R}(\mathscr{O})$ is the range or co-domain of the operator $\mathscr{O}$. Let $\mathscr{O}^*_X$ and $\mathscr{O}^*_Z$ for $\mathscr{O}_X$ and $\mathscr{O}_Z$, be the following corresponding adjoint operators

$$\mathscr{O}^*_X : L^2_{\mathscr{U}}(\mathbb{R}^g) \to \mathbb{R}^e : \mathscr{J} \mapsto \mathbb{E}[X\mathscr{J}(W)] \tag{B.12}$$

$$\mathscr{O}^*_Z : L^2_{\mathscr{U}}(\mathbb{R}^g) \to L^2_{\mathscr{V}}(\mathbb{R}^f) : \mathscr{J} \mapsto \mathbb{E}[\mathscr{J}(W)|Z]\mathscr{F}_Z/\mathscr{V} \tag{B.13}$$

We can identify the operators $\mathscr{O}_X, \mathscr{O}^*_X, \mathscr{O}_Z$ and $\mathscr{O}^*_Z$ using the iid vectors (Y_i, X_i, Z_i, W_i) from the partially linear model (20) in the first step. Recall from Equation (B.8), the normal equations can be expressed as below,

$$\mathscr{O}^*_Z\mathscr{R} = \mathscr{O}^*_Z\mathscr{O}_Z\mathscr{G} + \mathscr{O}^*_Z\mathscr{O}_X\beta, \quad \mathscr{O}^*_X\mathscr{R} = \mathscr{O}^*_X\mathscr{O}_Z\mathscr{G} + \mathscr{O}^*_X\mathscr{O}_X\beta \tag{B.14}$$

Florens et al. (2012) demonstrate that if $\mathscr{O}_X$ is orthogonal to $\mathscr{O}_Z$ (namely $\mathscr{O}^*_Z\mathscr{O}_X = 0$ with $\mathcal{R}(\mathscr{O}_X) \perp \mathcal{R}(\mathscr{O}_Z)$, where $\mathscr{O}^*_Z\mathscr{O}_X$ is a dot product), then we can avoid having to try an iterate estimating $\mathscr{G}(\cdot)$ whilst simultaneously estimating β. So the normal Equation (B.14) is equivalent to

$$\mathscr{O}^*_Z\mathscr{R} = \mathscr{O}^*_Z\mathscr{O}_Z\mathscr{G}, \qquad \mathscr{O}^*_X\mathscr{R} = \mathscr{O}^*_X\mathscr{O}_X\beta \tag{B.15}$$

Using kernel estimators to replace the operators $\mathscr{O}_X^*$, $\mathscr{O}_X$and $\mathscr{O}_X^*\mathscr{R}$, yields

$$\begin{aligned}\sum_{i,j} Y_i X_j \frac{\mathscr{K}_h(W_i - W_j)}{\mathscr{U}(W_i)} &= \sum_{i,j} X_i X_j^t \frac{\mathscr{K}_h(W_i - W_j)}{\mathscr{U}(W_i)} \beta \\ \beta &= (\sum_{i,j} X_i X_j^t \frac{\mathscr{K}_h(W_i - W_j)}{\mathscr{U}(W_i)})^{-1} (\sum_{i,j} Y_i X_j \frac{\mathscr{K}_h(W_i - W_j)}{\mathscr{U}(W_i)})\end{aligned} \tag{B.16}$$

where the kernel $\mathscr{K}$ is a Gaussian kernel function with bandwidth parameter h $(h > 0)$ and where the scaled kernel $\mathscr{K}_h(w) = h^{-g}\mathscr{K}(w/h)$. After identifying the parameter β, estimation of $\mathscr{G}$ can be obtain by purely nonparametric regression (Darolles et al., 2011; Hall et al., 2005).

We can now presume that $\mathcal{R}(\mathscr{O}_X) \not\subseteq \mathcal{R}(\mathscr{O}_Z)$ with $\mathscr{O}_Z^*\mathscr{O}_X \neq 0$, then so the normal Equation (B.17) can be expressed as

$$\mathscr{O}_Z^*(I - \mathscr{P}_X)\mathscr{R} = \mathscr{O}_Z^*(I - \mathscr{P}_X)\mathscr{O}_Z\mathscr{G}, \quad \mathscr{O}_X^*(I - \mathscr{P}_Z)\mathscr{R} = \mathscr{O}_X^*(I - \mathscr{P}_Z)\mathscr{O}_X\beta \tag{B.17}$$

where $\mathscr{P}_X$ and $\mathscr{P}_Z$ are the orthogonal projections for the $\mathscr{O}_X$ and $\mathscr{O}_Z$, respectively. Hence, $\mathscr{P}_X = \mathscr{O}_X(\mathscr{O}_X^*\mathscr{O}_X)^{-1}\mathscr{O}_X^*$ and $\mathscr{P}_Z = \mathscr{O}_Z(\mathscr{O}_Z^*\mathscr{O}_Z)^{-1}\mathscr{O}_Z^*$. We define the parameter estimators β based on Equation (B.17),

$$\beta = \frac{\mathscr{O}_X^*(I - \mathscr{P}_Z)\mathscr{R}}{\mathscr{O}_X^*(I - \mathscr{P}_Z)\mathscr{O}_X} = \frac{\mathscr{O}_X^*(I - \mathscr{O}_Z(\alpha I + \mathscr{O}_Z^*\mathscr{O}_Z)^{-1}\mathscr{O}_Z^*)\mathscr{R}}{\mathscr{O}_X^*(I - \mathscr{O}_Z(\alpha I + \mathscr{O}_Z^*\mathscr{O}_Z)^{-1}\mathscr{O}_Z^*)\mathscr{O}_X} \tag{B.18}$$

where α is the positive regularization parameter, which relates to the value of n. We follow the standard approach and assume that the operator $\mathscr{R}$ belongs to $L^2_{\mathscr{U}}(\mathbb{R}^g)$, functions $\mathbb{E}(\mathscr{P}(Z)|W = \cdot)\mathscr{F}_W(\cdot)/\mathscr{U}(\cdot)$ and $\mathbb{E}(X_i|W = \cdot)\mathscr{F}_W(\cdot)/\mathscr{U}(\cdot)$ belong to $L^2_{\mathscr{U}}(\mathbb{R}^g)$ for all $\mathscr{P} \in L^2_{\mathscr{V}}(\mathbb{R}^f)$, and $i \in \{1, ..., e\}$. The unknown nonparametric densities estimators (i.e. $\mathscr{O}_X, \mathscr{O}_X^*, \mathscr{O}_Z$ and $\mathscr{O}_Z^*$) can be identified by kernel estimators from the data vectors (Y_i, X_i, Z_i, W_i) from the partially linear model as follows:

$$\mathscr{O}_X\beta = \frac{1}{n}\sum_{i=1}^{n} X_i\beta\frac{\mathscr{K}_{h_W}(W_i)}{\mathscr{U}} \tag{B.19}$$

$$\mathscr{O}^*_X\mathscr{W} = \frac{1}{n}\sum_{i=1}^{n} X_i\int\mathscr{K}_{h_W}(W_i-w)\mathscr{W}(w)\mathrm{d}w \tag{B.20}$$

$$\mathscr{O}^*_Z\mathscr{G} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathscr{K}_{h_W}(W_i)}{\mathscr{U}}\int\mathscr{K}_{h_Z}(Z_i-z)g(z)\mathrm{d}z \tag{B.21}$$

$$\mathscr{O}^*_Z\mathscr{W} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathscr{K}_{h_Z}(Z_i)}{\mathscr{V}}\int\mathscr{K}_{h_W}(W_i-w)\mathscr{W}(w)\mathrm{d}w \tag{B.22}$$

$$\mathscr{R} = \frac{1}{n}\sum_{i=1}^{n} Y_i\frac{\mathscr{K}_{h_W}(W_i)}{\mathscr{U}} \tag{B.23}$$

where $\mathscr{W}\in L^2_{\mathscr{U}}(\mathbb{R}^g)$, $\mathscr{G}\in L^2_{\mathscr{V}}(\mathbb{R}^f)$ and the bandwidth parameters h_W, h_Z are dependent on sample size N, and the kernel $\mathscr{K}$. The kernels can be any of the standard kernels: Epanechnikov, Gaussian or fractional polynomials. For the results presented in this paper, we implemented a gaussian kernel weighted local polynomial fit.

To provide easy to compare results across different models we report the marginal effects of changes in z on y, hence we set $\mathbf{M} = \mathrm{d}y/\mathrm{d}z$ as a function of the dependent variable. If the bootstrapped confidence interval straddles the abscissa axis across its domain then the independent variable has no marginal impact on the dependent variable.

We compute confidence bounds for our partially linear semi-parametric regressions via an iid bootstrap with 99 resamples. Monte-Carlo studies for this estimator utilizing the function $\tilde{\mathscr{G}} = \mathcal{B}(Z,\mathscr{V},2\mathscr{V})\times\sin(H\mathscr{V}Z)$, where $\mathcal{B}(z,a,b)$ is the probability density function of the beta distribution with shape parameters a, b, the dependent variable is $Z\in[0,1]$ and the frequency scaling parameter set to $H = 10$.

The bootstrap consistency theory for these types of models is somewhat sparse; however, most implementations use bootstrap to determine the confidence intervals, for single index models a consistency proof for i.i.d. bootstrap is available from Yu and Ruppert (2002).[29] Whilst the choice

[29]For instance the semipar implementation in Stata uses this approach, we have included some models estimated via a Gaussian Kernel in Stata in the Online Supplement to illustrate that the pattern of the marginal effects is very

that we make is the 'third best' approach for providing evidence for the consistency of the bootstrap, it does provide a useful guide across the sets of choices available for our customized implementation versus those in other software packages.

We compute the iid bootstrap in the following steps. *Step 1*: Estimate the semiparametric regression using the original sample (Y_i, X_i, Z_i, W_i), and compute the fitted value $\hat{Y}_i$ and the residual r_i for each observation.

$$\hat{Y}_i = X_i\beta + \mathscr{G}[Z_i] + \varepsilon_i, \quad r_i = Y_i - \hat{Y}_i \tag{B.24}$$

Step 2: For each fitted value $\hat{Y}_i$, randomly choose a residual r_{ci} to build the bootstrapped Y^* value,

$$Y^* = [\hat{Y}_1 + r_{c1}, \hat{Y}_2 + r_{c2}, ..., \hat{Y}_i + r_{ci}, ..., \hat{Y}_n + r_{cn}] \tag{B.25}$$

where random number $c \in \{1, 2, ..., N\}$. Hence, the bootstrap data sample includes the bootstrapped Y^* value with fitted dependent variables. Refit the model using the bootstrap data sample, (Y_i^*, X_i, Z_i, W_i), and compute the marginal effects $\mathbf{M}$. Repeat the resampling procedure 99 times. The standard deviation of marginal effects $\sigma^{\mathbf{M}}$ can be computed as: $\sigma^{\mathbf{M}} = \sqrt{\frac{\sum(\mathbf{M}-\overline{\mathbf{M}})^2}{N-1}}$.

Several prior studies, see Florens et al. (2012) for overview have indicated that the distribution of the marginal effects is normal. Therefore, the 95% lower and upper confidence bounds, C^L and C^H, can be calculated by the 2.5 and 97.5 percentiles of the distribution of $\mathbf{M}$. For each plot we overlay on a second set of abscissa and ordinate axes the variation of our HFT proxy with respect to time to maturity from 10 to 0 years. Inference via confidence bounds is inherently more difficult than via a specific test, such as a standard Wald test, Lagrange multiplier or likelihood ratio. However, our regression framework is relatively simple in construction, with a one contemporaneous non-parametric and one lagged independent variable. For optimal specification of the non-parametric function we use the Akiake Information Criteria (this is used within each bootstrap too hence the variation in the pattern of the confidence bounds).

similar.