NBER WORKING PAPER SERIES

HIGH-FREQUENCY, ALGORITHMIC SPILLOVERS BETWEEN NASDAQ AND FOREX

Takatoshi Ito
Masahiro Yamada

Working Paper 21122
http://www.nber.org/papers/w21122

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2015

Ito acknowledges financial support Grant-in-Aid for Scientific Research A-25245044. We thank Frank Hatheway (Nasdaq OMX) for providing Nasdaq OMX data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

High-frequency, Algorithmic Spillovers Between NASDAQ and Forex
Takatoshi Ito and Masahiro Yamada
NBER Working Paper No. 21122
April 2015
JEL No. F31,G12,G14,G15,G23

ABSTRACT

We empirically examine the order flows spillovers between Nasdaq and the Forex markets in 2008 and 2009. With emphasis on a role of high-frequency traders (HFTs) who aggregate information between the two markets as well as within each market, our results show that HFTs in Nasdaq trade intensively on the market-wide information more rapidly than other market participants, and that their order flows contain more information about the Forex rates than those of the Forex themselves. As a result, order flows by HFTs in Nasdaq significantly lead those in the Forex activities. Reflecting each market's exposures to the common shocks during the Global Financial crisis, these spillovers vary over time, and HFTs have increased their influences. These empirical results are consistent with theoretical predictions of the rational expectations model of multi-asset trading.

Takatoshi Ito
Columbia University
School of International and Public Affairs
International Affairs Building
Room 927, (MC 3333)
420 West 118th Street
New York, NY 10027
and NBER
ti2164@columbia.edu

Masahiro Yamada
Hitotsubashi University
2-1 Naka
Kunitachi, Tokyo, 186-8601
Japan
m.yamada@r.hit-u.ac.jp

1 Introduction

Recently, behavior of high-frequency traders (HFTs) becomes a focus of intensive examination among practitioners, regulators, as well as academics. Although there is a growing literature of HFTs in individual assets, their spillovers across different markets have not been investigated in the literature. In this study, using a dataset that identifies transactions by HFTs in Nasdaq OMX, we explore the spillovers between Nasdaq stock market and Foreign exchange (Forex) markets. We find that HFTs trade on the common information more aggressively and rapidly than non-HFTs and their order flows contain more information about the Forex rates than those of the Forex themselves.

Stock prices and Forex rates are correlated. Figure 1 plots three Forex rates (EUR/JPY, EUR/USD, USD/JPY), Nasdaq index, and the volume share of Nasdaq HFTs from January 2008 to August 2009. The plots shows that the correlation between Forex rates and the stock index are time-varying. After September 2008, the correlation seems increased as the volume share of HFTs rose. Then, what is the role of HFT in the determination of the correlation between stocks and Forex? Recent studies on HFTs show that HFTs trade in the same direction of permanent price changes and contribute to the price discovery as a market maker (Hirschey (2013); Menkveld (2013); Hendershott (2011); Carrion (2013); Brogaard, Hendershott and Riordan (2012)), but they do not investigate the spillovers of HFTs to other markets or the source of better price discovery by HFTs.

We aim to answer HFT's role in influencing cross-market correlation. To do so, we combine Nasdaq OMX high-frequency trader dataset and EBS Forex dataset. Nasdaq dataset identifies transactions by HFTs, allowing an analysis on HFTs as a group. By using the dataset, we compare market-wide order flows in Nasdaq and Forex activities. A standard vector auto-regression (VAR) is used to investigate the lead-lag, cross-price impacts, and the information share of order flows among order flows in the Forex, Nasdaq HFT, non-HFT, and the Forex returns.

Since we can interpret these empirical results in many ways, a theoretical framework which provides a consistent explanation of the results is presented. In the Kodres and Pritsker (2002) framework, where traders can trade multiple-assets in a one-shot game, since uninformed traders make signal extraction from prices, noise trading in one market can influence the other markets. We numerically examine the estimated cross-price impacts and price correlations and judge whether the model can replicate empirical results.

Our main empirical findings are as follows. First, we find strong evidence that HFT order flows are very influential to the Forex market. In particular, an empirical results show that order flows by Nasdaq HFTs proceed to the Forex activities but those by non-HFT do not have such property. Forex returns and order flows lag to HFT's order flows by 500 milliseconds. However, this lead-lag can occur due to not only some economic reasons but also mechanical reasons such as a difference of trading architectures. Theoretical analysis suggests that a response in a market to noise tradings in other markets can be different depending on the price informativeness in each market. This asymmetry can be interpreted as a short-run lead-lag relationship. To examine the informativeness, we estimate following Hasbrouck (1991), each order flow's contribution to the permanent changes of Forex rates. As a result, we find that the contribution by Nasdaq HFT order flows dominate the contribution from non-HFT order flows and Forex order flows. This result means that the HFT order flows contain more market-wide information that affects both stocks and Forex rates.

Second, we find that HFTs and non-HFTs differs in the weight and reaction speed for market-wide information. HFT order flows in individual stocks are more correlated with market-wide order flows than non-HFT order flows.[1] Since idiosyncratic order flows in individual stocks do not affect the Forex market at all, influences to Forex by Nasdaq traders can be different between HFTs and non-HFTs. In addition, the spillovers from the Forex are different between HFTs and non-HFTs. After the changes in Forex return, HFTs order more aggressively than non-HFTs for less than around two seconds. After that, orders from non-HFTs become larger than those from HFTs.

[1]Brogaard, Hendershott and Riordan (2012) also report this property.

These results suggest that HFTs are equipped to aggregate market-wide information and reflect them on prices more rapidly than non-HFTs. This makes their order flows informative for market-wide information, providing the impacts on Forex activities.

Lastly, we report that the correlation between stock and Forex is increased during September 2008, when the Global financial crisis emerged. Based on the theoretical model, we can interpret this transition as changes in currency's exposure to the common risk factors. During this time, HFTs raised the share of trading volume at the Nasdaq. Since HFTs trade more aggressively for the common information in stocks, the spillovers from HFT's activities to Forex are particularly pronounced during September 2008.

2 Related Literature

Theoretical background of inter-market relationships The theoretical market microstructure mechanism of multi-asset trading is studied by Admati (1985); Caballe and Krishnan (1994); Kodres and Pritsker (2002); Bernhardt and Taub (2008). Among all, Kodres and Pritsker (2002) extend the framework of Grossman and Stiglitz (1980) and study the case when information about fundamentals are nested and traders are competitive. They show that correlated macro shock affects prices through the information channel and portfolio rebalancing channel. Through these channels, even a noise trading can cause the correlation of price changes, because uninformed traders cannot tell whether the shock is from informed traders or noise.

Price discovery by High-frequency traders and Algorithmic traders There is a growing literature on behavior of high-frequency traders (HFTs) and Algorithmic traders (ATs). Many studies show empirical evidence that HFTs (or ATs) improve liquidity and price efficiency, which are represented as narrower spreads, reduced adverse selection, and reduced trade-related price discovery (e.g., Hendershott (2011); Hendershott and Menkveld (2014)).

Using unique dataset at Nasdaq OMX, Hirschey (2013) shows that HFT's liquidity demand predicts future returns and non-HFT's liquidity demand. Brogaard, Hendershott and Riordan (2012) use the same dataset to ours, and show that HFTs trade against transitory pricing error and improve price efficiency. Carrion (2013) also use the same dataset and shows that HFTs engage in successful intra-day market timing. Hagströmer and Nordén (2013) use data from Nasdaq OMX Stockholm which identify individual HFT firms, and report there are two-types of HFT strategies: market-making and momentum. Taking tick size changes as a natural experiment, they find that both HFTs mitigate intraday price volatility.

As for the correlation between HFT and market-wide trading activities, Brogaard, Hendershott and Riordan (2012) point that HFT is more correlated with market-return, and Carrion (2013) finds that price reflect more of the information in mast market return HFT's participation is high. Biais, Foucault and Moinas (2013) and Menkveld and Jovanovic (2012) point that HFTs can react to public news such as index returns more rapidly than non-HFTs because of their speed advantage. Public news are also refereed to as "hard" quantitative information. Zhang (2012) reports that HFT react to hard information faster and stronger than soft qualitative information.

HFT in the Foregin exchange (Forex) market is studied by Chaboud et al. (2014). Different from us, they use special data with identification of HFTs. Taking advantage of the identification of HFTs, they show that HFTs take the opportunities of triangular arbitrage. Ito et al. (2012) also study second-by-second triangular arbitrage opportunities and find that the triangular arbitrage opportunities have become less likely and disappear more quickly as the trading become more frequent.

3 Market Structure: Data

3.1 EBS Forex dataset

We use the EBS level five data set. The data set contains transaction prices and trade volume at a frequency of quarter-seconds from January 7, 2008 to August 30, 2009.[2] The EBS price history shows whether a deal is done on the bid side or the ask side. The EBS global system consists of three regional computer sites, based in Tokyo, London, and New York, and it matches orders either within the site or across different sites. The system offer a colocation platform from 2005, which facilitates the high-speed and low-latency trading for HFTs.

Among 24 currency pairs contained in the dataset, we use EUR/JPY, EUR/USD, and USD/JPY. The original time-stamp of the data is based on GMT. In order to merge with the Nasdaq data set, the time-stamp is adjusted to EST. Detailed data cleaning is found in Appendix.

Below the summary of the data and its cleaning process is explained. In this dataset, trades (or deals) at time t occurs between time $t-1$ and t, or $(t-1, t]$. The order book information is a snapshot at the time-stamp. Price is defined as a mid-price that is a equally weighted average of the best bid and best offer prices.[3] The FX-return, r_t, is defined as the the log before-the-deal-price changes from t to $t+1$. Price changes occur when one or both of the best bid or the best offer changes, which may be a result of quote addition, cancellation or taken by transactions.

[2]The EBS level five dataset contains the following data fields: (1) Currency pair, (2) Date, (3) Time in quarter-seconds, (4) Event type (quote or done), (5) Buy Sell indicator, (6) Distance of quote, (7) Price, (8) Amount, (9) Quote count or Number of counterparty. In precise, the EBS dataset allows quarter-seconds time stamps from January 24, 2008. EBS also started providing the 100 millisecond time stamps from August 31, 2009. Our dataset covers the dates after August 31, 2009, and we analyze them separately.

[3]The definition of mid-price can alternatively be defined by depth-weighted mid-prices. In addition to the standard definitions, we examined these alternative definitions of return as well. We obtained the consistent empirical results but we do not report them in this article for the interest of brevity.

3.2 Nasdaq dataset

Nasdaq OMX also offer the colocation service for high frequency traders (or HFTs), and is a one of the most liquid electric trading venue. For Nasdaq transaction data, we use data provided by NASDAQ OMX for academic study of HFTs.[4] The data is a sample of 120 randomly selected stocks listed on Nasdaq and the New York Stock Exchange. From three market capitalization groups (high, medium, and low), each 40 firms are picked up. We use the sample of trading date from January 7, 2008 to December 31, 2009.

Transaction records provide millisecond time stamps and identify the liquidity demander and supplier as a HFT or non-high-frequency trader (or non-HFT).[5] The transaction type takes one of four values, HH, HN, NH or NN. The first letter represents the liquidity demander group, and the second letter represents the liquidity provider group. Regardless of this identification, we cannot identify all HFT because HFT firms that also engage in lower frequency trading strategies are excluded (see Brogaard, Hendershott and Riordan (2012) for the detail). HFTs are known to square positions by the end of trading days, but data do not necessarily suggest the squaring the position. Each HFT firms can trade the same stocks at different markets, their position cannot become zero at each end of day.

In order to match the Nasdaq data to the Forex data which differs from its fineness of the time stamps, we consolidate the transactions of Nasdaq for every quarter second. To capture the components of HFT behavior that affect the Forex market, we calculate the sum of quoted order imbalances (OIBs) across stocks, and define aggregate order flows.[6] To take the advantage of HFT identification, we primary focus on order flows rather than prices.

We match the Nasdaq data to the Forex data by outer join. That is, we hold both transactions at each clock-time, but drop the clock-time when there is no transaction at

[4]The HFT dataset provided by Nasdaq contains the following data fields: (1) Symbole, (2) Date, (3) Time in miliseconds, (4) Shares, (5) Price, (6) Buy Sell indicator, (7) Type (HH,HN,NH,NN).

[5]26 Trading firms are categorized as HFT based on Nasdaq's knowledge of their customers and analysis of firm's trading (Brogaard, Hendershott and Riordan (2012)).

[6]We denote Nasdaq HFT's OIB and non-HFT's OIB as HFT-OIB and non-HFT-OIB. When we direct both, we simply say Nasdaq OIB.

both markets. This keeps 64% of clock-time. Because the number of transaction at whole-Nasdaq stocks is far larger than that of the Forex, our estimation results could be affected. To address this issue, we also construct matched data by inner join, which keeps transactions at times when both markets have transactions. Since the qualitative results do not change much, we present only the former results in this paper.

3.3 Descriptive statistics of data

Descriptive statistics are shown in Table 1. Panel A reports statistics about the Forex market, and Panel B reports statistics about Nasdaq. Each panel reports the statistics by each quarter sub-sample.

Panel A of Table 1 calculates volatility of quarter-second return ($Std(r_t)$), average of instantaneous price impact per one unit of deal ($mean(\frac{r_t}{oib})$), average trade size, average trade volume per day, standard deviation of volume and quoted-OIBs per quarter-second, and probability of quote change and deals, and number of trade deals per quarter-second.

Panel B summarizes the trades ignited by either HFT or non-HFT (as for the trade type, 'HH' and 'HN' are of HFTs, 'NN' and 'NH' are of non-HFTs). Each trade is aggregated over stocks for every quarter-second. The panel calculates the average quoted trade size per trade, quoted trading volume per day, standard deviation of quoted trading volume and quoted-OIB per quarter-second, trading frequency, and average number of trades per quarter-second.

Roughly, 3.41/(3.41+4.7) = 42% of the trading volume is HFT-originated trades, and the proportion rises during 2008 (see Figure 1 as well). While Forex currency pairs have a trade probability less than 10% over time, Nasdaq transactions are more frequent and simultaneous transactions within a quarter-second is more likely.

Last column of Panel B measures the correlation between order flows of individual stocks and aggregate order flows.[7] Notably, HFT order flows is more correlated with the aggregate order flows than non-HFT order flows, even though the proportion of volume is less for HFT.

[7] In fact, the values are the adjusted R-squared obtained by the regression of individual stock's HFT-OIB (or non-HFT-OIB) on the cross-sectional sum of OIBs.

This figure provides simple evidence that HFTs trade on common stock fundamentals rather than idiosyncratic ones. Later, we present a result that HFT leads Forex but non-HFT does not. Why is this true even though market participants cannot tell HFT from non-HFT orders? One answer can be the speed advantage of HFTs. Another answer can be that HFTs put more weights on common component of stock values which also affect Forex markets. To support this reasoning, we also show that idiosyncratic component of individual OIB does not predict the Forex at all.

3.4 Time-series of Prices

The top panel of Figure 1 plots the time series of asset prices (Nasdaq Composite Index (IXIC), EUR/JPY, EUR/USD, USD/JPY) and the proportion of HFT trading volume over time. All the variables are normalized by the first observations. The middle panel plots a centered rolling correlation of IXIC and Forex (i.e., $corr(Return\ of\ Nasdaq\ index, Return\ of\ Forex\ rate_i$, i =EUR/JPY, EUR/USD, USD/JPY).[8] The bottom panel plots rolling correlation of the proportion of HFT trading volume and the stock-Forex correlation.

The top panel shows that proportion of the HFT trading volume was particularly high during the financial crisis in 2008. In fact, HFTs tend to trade more when the market is volatile. We can describe this fact by a following simple regression.

$$HFT\ share_t = \underset{(65.8)}{0.977} - \underset{(-0.763)}{0.281} \times Ret_{IXCI,t} + \underset{(13.1)}{0.00518} \times VIX_t + \underset{(1.53)}{0.0106} \times Ret_{IXCI,t} \times VIX_t.$$

The adjusted R-squared is 0.541. The regression is day-by-day with 402 observations, and the values in parentheses are t-statistics. The standard errors are corrected by Newey-West method. The volatility index, VIX, is of S&P 500. In this estimation, the most significant variable is VIX, which has t-statistics of 13.1 The share of HFTs tend to become large in a volatile market.

[8] The rolling sample window is 25 days.

As observed in the middle panel, the stock-Forex correlation is not steady and differs among currency pairs. Before the 2008 financial crisis (represented by the bankruptcy of Lehman-brothers at Sep 15, 2008), IXIC is positively correlated with EUR/JPY and USD/JPY but is negatively correlated with EUR/USD. After October 2008, each three currency pair tends to behave in the same way. Since HFTs trade more on the common component of stock values, high stock-Forex correlation can result in strong spillovers of HFTs at the Forex market.

Thus, HFTs trade more when the market is volatile, and they influence the Forex market more when the stock-Forex correlation is high. Findings are consistent with a view that during a crisis (high VIX) period, HFT activities tend to become higher (comparing to non-HFT), which contributes to increasing correlation between Nasdaq and currency markets. Related to this, in section 5.6, we implement a regression analysis to examine when the HFTs can predict the Forex markets.

3.5 Conditional frequency of the transactions

In this analysis, we use the return defined as r_{t3}, i.e., the log-difference between after-the-trade prices at t and before-the trade prices at t. Different from other analysis, we use dataset ranges from 2009/09/09 to 2009/12/31. [9] Although we can take this advantage of finer grids, the dataset is defined as event-time and still has gaps in real-time basis.[10] We construct the dataset by filling the no-entry (event with zero returns and zero transaction volume.)

In order to overview the spillovers of trading, we first implement a simple event-study analysis. We calculate the frequency of the transactions that both Nasdaq and Forex trades occurs in the same direction around a designated event. The event is designed as a two-standard deviation shock in market i, and calculate the frequency of the market j's transac-

[9] After September 2009, EBS provides 100 millisecond time-stamps.

[10] Event-time means that the observation occurs only when quote changes or deals occur.

tions in the same direction. For example, we compute the following.

$$Prob(OIB_{FX,-k} > 0 | OIB_{HF,0} > std(OIB_{HF,0})) = \frac{count(OIB_{FX,t-k} > 0 \text{ and } OIB_{HF,t} > 2std(OIB_{HF,t}) \text{ over } t)}{count(OIB_{HF,t} > std(OIB_{HF,t}) \text{ over } t)},$$

where k is from -10 to +10 and t is from the beginning to the end of trading days. $count(\cdot)$ simply counts the number of observations in the data that matches the statement. This example describes the reaction of Forex order flows to the imbalances in HFT order flows; we can calculate reaction of Forex returns, reaction of HFT (and non-HFT) to the Forex events, vice versa. For normalization, we present the difference of the probabilities between the directions, i.e.,

$$Prob(OIB_{FX,0-k} > 0 | OIB_{HF,0} > std(OIB_{HF,0})) - Prob(OIB_{FX,0-k} < 0 | OIB_{HF,0} > std(OIB_{HF,0})).$$

Since buying and selling is equally likely on average, the unconditional spreads should be close to zero. We compute the conditional frequencies for each three currency pairs for each day, and take their means.

Figure 3 show the results. The first column panels show the reaction of the Forex market conditional on the HFT's (or non-HFT's) events. The second panels do the reverse. The results for EUR/JPY, EUR/USD, and USD/JPY are presented from the top to the bottom. Events happen at $t = 0$, and the frequency is calculated by every 100 milliseconds. We focus more on the shapes of the graphs than their values.

The first-column panels show asymmetric hump-shapes. The hump-shapes indicate that such currencies tend to move in the same direction to U.S. stocks.[11] This hump-shape is slightly skewed toward right and the peak lags by 100 to 200 milliseconds. This indicates that Forex tends to lag to stocks if the both Forex and stock react to the same public information. To contrast, the second-column panels, showing the reaction of stocks to Forex events, have hump-shapes skewed toward left. This figure indicates the same thing: stocks tend to lead

[11]This is not always the case. Reflecting economic situations, the relationships can change over-time (see Figure 1 and 5).

Forex. In addition to this finding, we can compare the HFT and non-HFT. In these plots, the height of the graphs is higher for HFT-related shapes than those of non-HFT.

4 Structural Change of spillovers: A Model

In the literature of market microstructure, there are two possible mechanisms by which order flows and asset returns in one market may spill over other markets. The first mechanism is the information effect (King and Wadhwani (1990)). Since rational agents attempt to draw information from prices at other markets, market prices correlates even though any two markets do not have shared exposure to common risks. The second mechanism is the portfolio balance effect which is stressed by Kodres and Pritsker (2002). Since order flows through portfolio rebalance are misconstrued as being information-based, resulting price movements are exaggerated. In both channels, comovements occur not only from information shocks but also from noisy liquidity trading, because there are asymmetry of information and uninformed traders cannot differentiate information from noise.

4.1 The trading model by Kodres and Pritsker (2002)

To address the theoretical background for our empirical results, we briefly review the model of Kodres and Pritsker (2002) and why Nasdaq stock prices and Forex rates correlate.

We assume the economy with N risk assets with zero net supply, and a riskless asset as the numeraire which is inelastically supplied. The risky assets are liquidated at the end of trading game and generate the random vector of value $\tilde{v}$. The riskless assets are assumed to have zero net return. Investors trade the assets in the first period, and consume the liquidation value of all assets in the second period, or the end of this game.

As in the model of Grossman and Stiglitz (1980), there are three types of agents: informed traders, uninformed traders, and noise traders. The number of traders are denoted as μ_I, and μ_U. Noise traders submit elastic supply of assets by following the probability distribution

$\tilde{\boldsymbol{e}} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_e)$. We assume $\boldsymbol{\Sigma}_e$ is diagonal. This represents independent liquidity shocks[12]. In order to derive a meaningful equilibrium, informed traders and uninformed traders are assumed to be risk-averse and have negative exponential utility function. The rate of relative risk aversion is set to $A > 0$.

Informed traders receive information on the liquidation value before the first period, and form the conditional expectation about $\tilde{\boldsymbol{v}}$. Namely,

$$\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{u}}.$$

Here, $\tilde{\boldsymbol{\theta}}$ is the informed trader's conditional expectation $E[\tilde{\boldsymbol{v}}|\text{Information}]$ and $\tilde{\boldsymbol{u}}$ is the regression residual. The unconditional joint distribution of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{u}}$ is assumed to be normal:

$$\begin{pmatrix} \tilde{\boldsymbol{\theta}} \\ \tilde{\boldsymbol{u}} \end{pmatrix} \sim N \left[\begin{pmatrix} \bar{\boldsymbol{\theta}} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_\theta & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_u \end{pmatrix} \right]. \tag{1}$$

We also assume that these random variables are jointly distributed with $\tilde{\boldsymbol{e}}$. Note that the information structure is hierarchical. Uninformed trader's information set is nested by that of informed traders. Also, informed traders obtain the "signal" to infer the liquidation value for all markets at the same time. On the other hand, uninformed traders attempt to infer information from prices. The presence of noise traders prevents the informed trader's information to be fully revealing.

The market makers receive the demand schedule from traders, and pins down a market clearing prices. Denoting each demand function as $\boldsymbol{X}_I(\tilde{\boldsymbol{P}}, \tilde{\boldsymbol{\theta}})$ and $\boldsymbol{X}_U(\tilde{\boldsymbol{P}})$, the market clearing condition is

$$\mu_I \boldsymbol{X}_I(\tilde{\boldsymbol{P}}, \tilde{\boldsymbol{\theta}}) + \mu_U \boldsymbol{X}_U(\tilde{\boldsymbol{P}}) + \tilde{\boldsymbol{e}} = 0. \tag{2}$$

The assumptions of negative exponential utility function and their terminal wealth following

[12] Bernhardt and Taub (2008) proposes that $\boldsymbol{\Sigma}_e$ can be a only a source of contagion when investors are risk neutral and internalize price impacts.

normal distribution, both traders have linear demand functions:

$$\boldsymbol{X}_I(\tilde{\boldsymbol{P}}, \tilde{\boldsymbol{\theta}}) = A^{-1} Var^{-1}(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}}) \left(E(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}}) - \tilde{\boldsymbol{P}} \right),$$
$$\boldsymbol{X}_U(\tilde{\boldsymbol{P}}) = A^{-1} Var^{-1}(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{P}}) \left(E(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{P}}) - \tilde{\boldsymbol{P}} \right). \tag{3}$$

As we assumed, informed traders' conditional expectation and conditional variance of the liquidation values are $E(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}}) = \tilde{\boldsymbol{\theta}}$ and $E[Var(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}})] = Var(\tilde{\boldsymbol{v}}) - Var(E(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}})) = \Sigma_u = Var(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}})$. Plugging investor's demand functions (3) into the market clearing condition (2), we can construct uninformed's "signal" inferred by prices as follows.

$$\tilde{\boldsymbol{\theta}}_P \equiv \tilde{\boldsymbol{\theta}} + \frac{A}{\mu_I} \boldsymbol{\Sigma}_u \tilde{\boldsymbol{e}}. \tag{4}$$

The expected liquidation value conditional on prices coincides with that of conditional on this signal:

$$E(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{P}}) = E(\tilde{\boldsymbol{v}}|\tilde{\boldsymbol{\theta}}_P). \tag{5}$$

Since $\tilde{\boldsymbol{v}}$, $\tilde{\boldsymbol{\theta}}$, and $\tilde{\boldsymbol{e}}$ are jointly normally distributed, we can calculate this conditional expectation and the conditional variance by projection theorem. Thus we can obtain the expression for the demand functions.

Equation (5) means that uninformed traders cannot differentiate the informed traders' expectation from noise. Different from the model of Grossman and Stiglitz (1980), $\boldsymbol{\Sigma}_u$ takes a matrix form, which defines the allocation of risky assets for informed traders. Thus, when $\boldsymbol{\Sigma}_u$ is not diagonal, uninformed trader can confuse the shocks in the excess demand from portfolio rebalancing with information shocks.

The rational expectations equilibrium is defined by the demand functions which maximize each agent's utility (or Equation 3), the prices that clear the market (or Equation 2), and agent's beliefs being consistent with information revealed by prices (or Equation 5).

Kodres and Pritsker (2002) show that there is no contagion when $\boldsymbol{\Sigma}_\theta$, $\boldsymbol{\Sigma}_u$, and $\boldsymbol{\Sigma}_e$ are all diagonal (their Proposition 2). Namely, liquidity shocks on $\tilde{e}_k$, information shock on $\tilde{\theta}_k$, changes in order flows, and changes in prices in k th market do not affect other $i \neq k$ markets' variables. In next section, we specify these matrices.

4.2 Information structure between Equity and Forex markets

Based on this model, we specify following information structure with one stock and three currency fundamentals.

$$\tilde{\boldsymbol{v}} = \boldsymbol{\Phi}\tilde{\boldsymbol{f}} + \tilde{\boldsymbol{\eta}},\ \boldsymbol{\Phi} \equiv \begin{bmatrix} \alpha_1 & 0 & 0 & 0 & \beta_1 \\ 0 & \alpha_E & 0 & 0 & \beta_E \\ 0 & 0 & \alpha_U & 0 & \beta_U \\ 0 & 0 & 0 & \alpha_J & \beta_J \end{bmatrix},\ \tilde{\boldsymbol{f}} = \begin{pmatrix} \tilde{f}_1 \\ \tilde{f}_E \\ \tilde{f}_U \\ \tilde{f}_J \\ \tilde{f}_c \end{pmatrix}.$$

$\tilde{\boldsymbol{f}}$ are the risk factors where $\tilde{f}_c$ is a common factor and the others are the country (or $\tilde{f}_E, \tilde{f}_U, \tilde{f}_J$) or stock market (or $\tilde{f}_1$) specific risk factors. They are independent each other. $\boldsymbol{\Phi}$ specify the exposure to the risk. $\tilde{v}_1$ is the fundamental of the stock market, and the others are for country specific fundamentals. We assume that fundamentals take a log form, and the difference of fundamentals represents the exchange rate. Since there are three Forex rates for three currency fundamentals, the system loses a degree of freedom. Thus we have,

$$\boldsymbol{T}\tilde{\boldsymbol{v}} = \begin{pmatrix} \tilde{v}_1 \\ \tilde{v}_E - \tilde{v}_J \\ \tilde{v}_E - \tilde{v}_U \end{pmatrix} = \boldsymbol{T}\boldsymbol{\Phi}\tilde{\boldsymbol{f}} + \boldsymbol{T}\tilde{\boldsymbol{\eta}},\ \boldsymbol{T} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix}.$$

Because we examine the correlation within quarter-seconds and look for the primary cause of this rapid co-price-discovery process, we focus on the information channel. Along with the notation of the general model, we define the expected fundamentals and the residual for informed traders.

$$\tilde{\boldsymbol{\theta}} \equiv \boldsymbol{T}\boldsymbol{\Phi}\tilde{\boldsymbol{f}}, \;\; \tilde{\boldsymbol{u}} \equiv \boldsymbol{T}\tilde{\boldsymbol{\eta}}.$$

The variance-covariance matrices of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\eta}}$ are not diagonal. Especially, the exposure to the common risk factor determines the comovement between the stock market and the Forex market. When the relative exposures $\beta_2 \equiv \beta_E - \beta_J$ and $\beta_3 \equiv \beta_E - \beta_U$ are both equal to zero and there are no correlated noise trading, there is no comovement between stock and Forex.

Kodres and Pritsker (2002) mainly assumes that information of common shocks are not available for informed traders. In contrast, we assume that the common shocks are included in the informed trader's information set. In this model, we think that for relatively short time length within one second, common information such as index future returns and macroeconomic announcement are observable for the informed but not for the others.

4.3 Theoretical implications

We can interpret the correlation of prices in Figure 1 in line with this model. When Nasdaq market is positively exposed to the common risk, or $\beta_1 > 0$, Nasdaq's positive comovement with EUR/JPY and USD/JPY indicates the exposure to the common risk is stronger for Euro and U.S. than Japan, i.e., $\beta_E - \beta_J > 0$ and $\beta_U - \beta_J > 0$.

In contrast to EUR/JPY and USD/JPY, changes of correlation between Nasdaq and EUR/USD indicates the exposure of Euro and U.S. to the common risk is time varying. Before the beginning of September 2008, U.S. and Euro were equally exposed to the risk, but after Lehman Brothers collapsed and TARP was introduced, the risk for Euro had been more emphasized. The magnitude of exposure was $\beta_E > \beta_U > \beta_J$ and Nasdaq stock market was positively correlated with all the currency pairs.

From next section, we examine the high-frequency correlation between Nasdaq HFT trading and Forex markets. Based on the model, we can draw a way to understanding the results of the empirical analysis. Since this multiple-asset model does not produce a simple analytic expression, we rely on numerical comparative statics.

Figure 2 shows the numerical comparative statics for how correlation of prices and responses to liquidity shocks are influenced by the exposure to the common risk factor, the aggressiveness of informed trading, and the magnitude of noise trading.[13]

First, when the price change at market k is less related to the common risk factor (or small $|\beta_k|$), uninformed trading at market k less relays on the other market. Thus, correlation of prices between market k and other markets is in proportion to β_k (top-left panel of Figure 2). The correlation becomes zero when $\beta_k = 0$.

We are also interested in the relation between the volatility of market k and the correlation with other markets. When the volatility of prices are induced by the changes in idiosyncratic factor α_k, the volatility of market k and the correlation have negative proportional relationships (top-right panel of Figure 2 where $k = 2$).

Second, we investigate the case when the informed traders trade more aggressively in market k than other markets. So far, we assume that μ_I and μ_U are a scalar. Here, we can rewrite them as a diagonal matrix. That is, we assume that the proportion of informed and uninformed traders differ across markets. Then we can interpret why trading aggressiveness or informed traders differs in the two market. The numerical example of this case is illustrated by the middle panel in Figure 2, which show how the price change of market i in response to normalized liquidity shocks ϵ_k varies depending on the ratio of informed trading in market 1 (or stock market).[14] The panels show that $\frac{dP_2}{d\epsilon_1} > \frac{dP_1}{d\epsilon_2}$ and $\frac{dX_2}{d\epsilon_1} > \frac{dX_1}{d\epsilon_2}$ when the ratio of informed trading in market 1 is large enough.[15]

Third, when the noise trading is larger at market k, or $\mathbf{\Sigma}_{e(k,k)}$ is larger than other elements, shocks in market k affect little to the other markets. This is because uninformed traders put smaller weight for the noisy market k to predicting other markets. Thus, even though shocks

[13]For the numerical experiments, each parameters are set as follows: $[\beta_1, \beta_E, \beta_U, \beta_J] = [1, 1, 0.5, 0.5]/\sqrt{10}$; For $j = 1, E, U, J$, $\alpha_j = 1/\sqrt{10}$; $Var(\tilde{e}_j) = 10$; $Var(\tilde{u}_j) = 10$. $\mu_I = 0.1$ and $\mu_U = 0.9$. The risk factors are assumed to follow the standard normal distribution.

[14]In this example, we assume that aggressiveness of informed trading in market 2 and 3 (or Forex markets) are the same and $\mu_I = 0.1$ and $\mu_U = 0.9$.

[15]Here, X_k denotes the uninformed trading at market k.

in Nasdaq market can affect the Forex market, the reverse is not necessarily true if trading in the Forex market is noisier and less informative. This case is illustrated by the bottom panels in Figure 2. The panels show that $\frac{dP_2}{d\epsilon_1} > \frac{dP_1}{d\epsilon_2}$ when the variance of noise trading is sufficiently large.

The first results can explain the time-variation of the spillover is governed by the variation of the exposure to risks, or α and β. Based on this result, we can interpret the empirical results in section 5.3 and 5.6. The second and third results explain the asymmetric lead-lag between HFT and the Forex. If HFT is better at representing the common risk factors than Forex traders, the Forex market can be influenced by HFTs while the reverse may not be true (this intuition is related to the analysis in section 5.8). This result is already observed in section 3.5, and are examined in section 5.1 and 5.2 by an alternate methodology.

The limitation of this model is that we do not distinguish the behavior of HFTs and non-HFTs in stock market, and this is the reason that our theoretical model does not exactly match our empirical analysis. There are a growing theoretical literature on this issue (e.g., Biais, Foucault and Moinas (2013); Menkveld and Jovanovic (2012)).

5 High-frequency VAR analysis

5.1 Base formulation

To measure the spillovers between Forex-OIB, HFT-OIB, non-HFT-OIB, and Forex-return, we estimate the following system of vector autoregression (VAR):

$$OIB_{fx,t} = \sum_{k=1}^{K} \alpha_{fx,k} OIB_{fx,t-k} + \sum_{k=1}^{K} \beta_{fx,k} OIB_{HFT,t-k} + \sum_{k=1}^{K} \gamma_{fx,k} OIB_{nHFT,t-k} + \sum_{k=1}^{K} \rho_{fx,k} r_{t-k} + \epsilon_{fx,t} \quad (6)$$

$$OIB_{HFT,t} = \sum_{k=1}^{K} \alpha_{hft,k} OIB_{fx,t-k} + \sum_{k=1}^{K} \beta_{hft,k} OIB_{HFT,t-k} + \sum_{k=1}^{K} \gamma_{hft,k} OIB_{nHFT,t-k} + \sum_{k=1}^{K} \rho_{hft,k} r_{t-k} + \epsilon_{hft,t} \quad (7)$$

$$OIB_{nHFT,t} = \sum_{k=1}^{K} \alpha_{nhf,k} OIB_{fx,t-k} + \sum_{k=1}^{K} \beta_{nhf,k} OIB_{HFT,t-k} + \sum_{k=1}^{K} \gamma_{nhf,k} OIB_{nHFT,t-k} + \sum_{k=1}^{K} \rho_{nhf,k} r_{t-k} + \epsilon_{nhf,t} \quad (8)$$

$$r_t = \sum_{k=1}^{K} \alpha_{ret,k} OIB_{fx,t-k} + \sum_{k=1}^{K} \beta_{ret,k} OIB_{HFT,t-k} + \sum_{k=1}^{K} \gamma_{ret,k} OIB_{nHFT,t-k} + \sum_{k=1}^{K} \rho_{ret,k} r_{t-k} + \epsilon_{ret,t}. \quad (9)$$

In equations, $OIB_{fx,t}$, $OIB_{HFT,t}$, $OIB_{nHFT,t}$ are quoted signed orders, or quoted order imbalances (quoted OIBs), of a Forex currency pair (EUR/JPY, EUR/USD, USD/JPY), HFT, and non-HFT.[16] r_t is the Forex return of the currency pair. The time interval is defined as an event time, and the minimum clock-time between t and $t-1$ is a quarter second. The return r_t is defined based on the mid-quote-price between before-the-first trade at t and before-the-first trade at $t-1$.[17]

A concern about this formulation is the proper ordering of trades across different market: we cannot identify the order of trades in a quarter second. This problem affects the variance decomposition. We take care of this problem by deriving the lower and the upper limit of the contribution to the variance.

We estimate Equations (6)-(9) with ten ($K = 10$) and 30 ($K = 30$) lags. $K = 10$ lag is determined by the Bayesian Information Criterion in a preliminary analysis for the sample day of Jan 8, 2008. Using 30 lags, we can capture the reaction to a shock over $10 \times 0.25 = 2.5$ seconds. We estimate the system for day-by-day samples and present the average of estimated coefficients and t-values. The standard errors are calculated using Fama-MacBeth method. In Table 2, the t-values are from simple cross-sectional means. Figure 4 presents the Fama-MacBeth t-statistics.

Estimation results

Table 2 presents the VAR(10) estimation results for EUR/JPY. In addition to the estimated coefficients and the Fama-MacBeth t-statistics, we also present the proportion of samples with less than one-percent p-values. For results, the coefficients of $\beta_{fx,k}$ and $\beta_{ret,k}$ are positive and statistically significant (at roughly 10 % significance level) for up to lag $k = 5$, providing evidence that HFTs can predict Forex activities. To contrast, $\rho_{hft,k}$ is significantly positive

[16]Forex OIB is divided by one million.

[17]Barclay and Hendershott (2003) defines this return as r_{t-1} and include contemporaneous terms of OIB in VAR.

up to three lags, meaning that HFT-OIB reacts to the Forex return and transmits the information produced in the Forex market. To contrast, the t-statistics of $\gamma_{fx,k}$ and $\gamma_{ret,k}$ are barely significant up to lag $k = 2$. Thus, non-HFT have only limited predictive power for the Forex activities. non-HFT-OIB also does not react to the Forex return. Therefore, in the high-frequency world, common-information between the Forex and Equities are transmitted mostly by the HFTs rather than non-HFTs. In later section, we confirm this statement by looking into Granger-causality and variance decomposition.

Moreover, the economic magnitude of the spillover is not small. In Table 1, one standard deviation of quoted OIB of Forex (EUR/JPY) and HFT are 51.9 million yen and 88.5 thousands dollar. With $\alpha_{ret,1} = 1.33 \times 10^{-8}$ and $\beta_{ret,1} = 1.37 \times 10^{-11}$, each deviations of OIBs leads to $51.9 \times 1.33 \times 10^{-8} = 0.0069$bp and $88.5 \times 10^3 \times 1.37 \times 10^{-11} = 0.012$bp increases in the EUR/JPY rate. Thus, HFT's order flows also have a significant impact on the Forex returns. The calculation measures the response of returns in a quarter-second after the order. The contemporaneous price impact is measured by $mean(\frac{r_t}{OIB|t})$ in the summary statistics table. For the long-run response, we implement the variance decomposition in section 5.5 following the method of Hasbrouck (1991).

While Nasdaq OIB can affect the Forex market, the reverse causality seems weaker. Since the t-statistics of $\alpha_{HFT,k}$ and $\alpha_{nHFT,k}$ are both close to zero, Forex-OIB does not predict both HFT- and non-HFT-OIB. Since the t-statistics of $\rho_{HFT,k}$ and $\rho_{nHFT,k}$ are smaller than $\beta_{HFT,k}$ or $\beta_{nHFT,k}$, the Forex return also has limited predictive power to future Nasdaq OIBs. As we explained in section 4.3, this finding is consistent with the case when the Forex market is less informative because of less accurate information signals and greater noise trading. Informativeness of order flows are estimated in section 5.5, supporting this argument.

For other currency pairs, Figure 4 shows the results. It shows the Fama-MacBeth t-statistics of the day-by-day estimation of VAR(30) for three currency pairs. For HFT-OIB and Forex variables, we observe the similar lead-lag relationships analyzed above. The pattern of coefficients are also very similar among the currencies. The estimated coefficients in sub-sample, however, show that the spillovers are time-varying. In section 5.3, we examine the time-series plot of the predictability of FX-returns by Nasdaq OIBs.

5.2 Granger-causality tests

Table 3 reports pairwise Granger-causality tests. For the null hypothesis that variable j does not Granger-cause variable i, we test whether the lagged coefficients of j are jointly zero when i is the dependent variable in the VAR.

In Table 3, (i, j) cell shows the p-value of the F-test (the causality of $j \rightarrow i$). A day-by-day VAR are conducted the means of p-value over time are presented. For results, HFT-OIB Granger-causes Forex-OIB, non-HFT-OIB, and Forex returns. For the reverse causality, we only find the significant spillover from the Forex returns to Nasdaq HFT-OIB.[18] We also find that non-HFT-OIB does not Granger-cause Forex variables.

Thus, Granger-causesality tests also confirm that the causality from Nasdaq-HFT to Forex exists. The reverse causality result implies HFTs are common information aggregators.

5.3 Time-series plot

In an earlier section, it was shown that the correlation between Nasdaq Index and Forex rates are time-varying. By plotting the spillover coefficients over time, the interdependence of markets and the role of HFTs related to this phenomenon can be shown clearly.

Figure 5 plots the time series of the sum of t-statistics for $y = r_{fx}$ over time.[19] This plots shows the intensity of spillover between Nasdaq and Forex in terms of their order flows. We focus on the sum of t-statistics of HFT-OIB and non-HFT-OIB (i.e, $\Sigma_{k=1}^{10} \beta_{ret,k}$ and $\Sigma_{k=1}^{10} \gamma_{ret,k}$). The top panel shows the t-statistics of HFT-OIB, and the bottom one shows that of non-HFT-OIB.

[18] When we keep the data only when both Nasdaq and Forex have observations, the Granger-causesality (FX-return $\rightarrow$ HFT-OIB) becomes stronger (estimation results are omitted from this paper). Thus this result can be partly driven by the scarcity of the Forex transactions. But still, the Nasdaq HFT have significant spillovers to the Forex market.

[19] We present these values because the level of coefficients have different units and are less comparable.

The spillovers between EUR/JPY and USD/JPY are mostly stable than EUR/USD, but they become relatively high between November 2008 and January 2009. To contrast, the spillover with EUR/USD is varying over time. It is close to zero before the end of May and goes downward between June and September in 2008. After September, it goes upward and the sign becomes positive.

The above results are consistent with the conventional view on the financial crisis. At the midst of the financial crisis, Japanese yen had become a safe haven currency (or exposure to the common shock is small). Thus, any good (bad) sentiment about the financial crisis push up (down) Nasdaq market and Japanese yen depreciates (appreciates) relative to other currencies. During 2008, it had been ambiguous which of U.S. or EU is hit by the financial crisis harder. Since many European banks had invested in "toxic" assets that were earlier created in the U.S., and dollar liquidity shortage was severe in Europe. This ambiguity is reflected in the time-varying sign of the reaction in EUR/USD. Before September 2008, U.S. economy was thought to be more sensitive to the crisis, leading the negative correlation between Nasdaq order flows and EUR/USD. After the collapse of the Lehman Brothers in September 2008, the financial crisis propagated to the Eurozone crisis, and Nasdaq order flows began to be positively correlated with EUR/USD.

During 2008-09, non-HFT exhibit the smaller variation of the spillover over time. The time-series of HFT, however, exhibits the clearer patterns. For example, during October in 2008, the spillovers by HFTs rose for every currency pairs. This result may provide evidence that HFTs have superior ability to trading on market-wide information, which is captured as the clearer spillovers on the Forex market.

5.4 Intraday Pattern

Intraday patterns of Forex and Stock market are different. Ito and Hashimoto (2006) showed that the Forex activity has three peaks at the beginning of each of London, New York, and Tokyo stock market, but it does not have peaks at the closing of each stock market. To contrast, stock markets have so called U-shape patterns for the activity. Also, the Forex market is the most liquid at the opening in each time-zone, but stock markets are illiquid at the opening. Do these different intraday patterns affect the spillovers between the two markets?

To examine the changes of spillovers along the intraday time of day, we separate a day into thirteen 30-minutes-windows and estimated the day-by-day VAR for each time window. The number of each sub-sample is as many as 5000, and we can estimate the same VAR system.

Figure 6 shows the Fama-MacBeth t-statistics o for $y = OIB_{fx}$ and $y = r_{fx}$ equations. Here, we only presents EUR/JPY results but the other currency pairs also show similar patterns. We estimated the VAR with 30 lags, but show only up to 10 lags for presentation purpose.

For both $y = OIB_{fx}$ and $y = r_{fx}$, the Fama-MacBeth t-statistics for HFT-OIB are larger in the morning than in the end of the trading day. Not only Nasdaq OIBs, but also Forex-OIB and returns show smaller significance and delayed responses in the afternoon session.

Related to this finding, Hirschey (2013) reports that HFT-OIB is more strongly correlated with future non-HFT-OIB in the morning than in the middle of the trading day. He explains this fact due to the impatient non-HFTs in the morning when there are much private information accumulated during overnight. In our context, the variable intensity of the spillovers can be caused by the informativeness, or amount of private information, of the equity market order flows.

5.5 Variance decomposition (Hasbrouck 1991)

To address the importance of Nasdaq transactions to the price formation process in the Forex market, we follow Hasbrouck (1991, 1995) to decompose the variance of the efficient price into the portion of total price discovery that is correlated with Forex, HFT, and Nasdaq nonon-HFT trades.

Following Hasbrouck (1991), the VAR system is modified to incorporating a contemporaneous term of Forex-OIB for $y = r_{fx}$ regression:

$$r_t - \alpha_{ret,0} OIB_{fx,t} = \sum_{k=1}^{K} \alpha_{ret,k} OIB_{fx,t-k} + \sum_{k=1}^{K} \beta_{ret,k} OIB_{HFT,t-k} + \sum_{k=1}^{K} \gamma_{ret,k} OIB_{nHFT,t-k} + \sum_{k=1}^{K} \rho_k r_{t-k} + \epsilon_{ret,t}. \tag{10}$$

The other equations are unchanged. After estimating Equation (10), we transform this little SVAR model to VAR. Then we invert the system to generate a vector moving average model. We obtain

$$\boldsymbol{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\epsilon}_t,$$

where $\boldsymbol{y}_t$ and $\boldsymbol{\epsilon}_t$ are the (4×1) column vectors of independent and error variables. The variance decomposition is done through calculating the fraction of total price discovery from private information revealed through trades. The bid-ask midpoint p_t can be decomposed into a random-walk component m_t and a stationary component s_t: $p_t = m_t + s_t$, where $m_t = m_{t-1} + v_t$ and $v_t \sim N(0, \sigma_v^2)$ with $E[v_t v_s] = 0$ for $t \neq s$. m_t is referred to the permanent component of price, and s_t is referred to the transitory component of price. Thus, $r_t = p_t - p_{t-1} = v_t + \Delta s_t$. Matching the autocovariance generating function of this expression and that of VMA representation of equations (6) - (8) and (10) , we can decompose the variance of the permanent component of the quote changes σ_v^2 into price changes triggered by the Forex trades, HFT and nonon-HFT trades, and the arrivals of public information[20]. Assuming that the covariance matrix of the residuals $var(\boldsymbol{\epsilon}_t)$ is diagonal, we have

$$\sigma_v^2 = \boldsymbol{\psi} diag([\sigma_{fx}^2, \sigma_{hft}^2, \sigma_{nhf}^2, \sigma_{ret}^2])\boldsymbol{\psi}', \;\; \boldsymbol{\psi} \equiv \left[\psi^{fx}, \psi^{hft}, \psi^{nhf}, \psi^{ret}\right] = \boldsymbol{\Psi}(1)_{(4,:)}$$

where $\boldsymbol{\Psi}(1) = \boldsymbol{I} + \boldsymbol{\Psi}_1 + \boldsymbol{\Psi}_2 + \cdots$. Intuitively, $\boldsymbol{\Psi}(1)\boldsymbol{\epsilon}_t$ means the long-run impact of a disturbance on each dependent variables. $\boldsymbol{\psi}$ is the vector of the VMA coefficients for $y = r_t$, or the fourth row vector of $\boldsymbol{\Psi}(1)$. σ_j^2 is the variance of the residuals for j's regression

[20] See Hasbrouck (1991) for the detailed proof.

equation. Empirically, twenty lags of the VMA coefficients are used. The information share of j's variable is defined as

$$IS_j \equiv \frac{\left(\psi^j \sigma_j\right)^2}{\sigma_v^2}.$$

We interpret the measure as the component of price discovery attributable to public information and the private information revealed by some party of traders. For example, the ratio of private information conveyed by HFT trades is IS_{hft}.

Following Hasbrouck (1995), we can also calculate the lower and the upper bound of the information share when we consider the structure of $Var(\boldsymbol{\epsilon}_t)$. To do so, we first make a Cholesky factorization of $var(\boldsymbol{\epsilon}_t)$, and denote the lower triangular matrix as $\boldsymbol{F}$. Thus we have $\boldsymbol{\epsilon}_t = \boldsymbol{F}\boldsymbol{z}_t$ with $E[\boldsymbol{z}_t] = \boldsymbol{0}$ and $var(\boldsymbol{z}_t) = I$. The information share is now defined as

$$IS_j \equiv \frac{\left([\boldsymbol{\psi}\boldsymbol{F}]_j\right)^2}{\boldsymbol{\psi} var(\boldsymbol{\epsilon}_t)\boldsymbol{\psi}'}$$

where $[\boldsymbol{\psi}\boldsymbol{F}]_j$ is the jth element of the row matrix $\boldsymbol{\psi}\boldsymbol{F}$. The calculation of the information shares now depends on the order of the VAR system. The factorization imposes a hierarchy that maximizes the information share on the first variable and minimizes the information share on the last variable. [21]

Table 4 reports the variance decomposition results, including the average of the decomposition results for day-by-day estimations, associated with the standard deviation; the lower and the upper bounds of the information shares ($\underline{IS}_j$ and $\overline{IS}_j$) by taking a permutation of the orders of the VAR system. Note that the IS_j under the independence assumption does not necessarily lie between the lower and the upper bounds, because of negative correlations between the residuals. In addition to the statistics using all samples, sub sample statistics for every quarter are presented.

The result confirms that HFTs have significant spillovers on the price discovery in the Forex market. The fraction of total price discovery due to HFTs is 4.27% for EUR/JPY,

[21] In fact, as Hasbrouck (2002) points, this ordering does not ensure the correct lower and upper bound. We check all the permutation to provide the lower and upper bounds.

1.87% for EUR/USD, and 3.23% for USD/JPY. These figures are high relative to the Forex OIB.[22] For EUR/USD, the spillovers by HFT-OIB is relatively smaller. This negative result can be caused by the time-variation of the spillovers between the currency and U.S. stocks: while the price discovery by HFT-OIB is smaller in the first and second quarters, it becomes higher in the third and fourth quarter.

Although HFTs take a significant role in the price discovery process of the Forex market, non-HFTs do not. For all currency pairs, the fraction of price discovery due to non-HFT-OIB is the lowest. Price discovery by the quote changes dominate. Probably this is because the frequency of quotes change is 10 times as much as that of trades (see summary statics table).

In the sub-sample analysis, we can observe that the contribution by HFT-OIB varies over time and is not monotonically increasing. The pattern across currency pairs also differs. Since the contribution varies in time-series as much as cross-sectional, it is likely that these variation can be due to the correlation between the currency rate and U.S. stock market rather than the changes of HFT's activities.

These results about price discovery confirms that Nasdaq order flows (especially those by HFTs) are informative for Forex rate. As our theoretical prediction tells, this leads the asymmetric response to the noise trading in markets, which can be viewed as a lead-lag between the markets.[23]

5.6 Determinants of predictability

The theoretical model predicts that the correlation of price between markets i and k, or $corr(P_i, P_k)$, depends on the covariance structure of the fundamentals defined by the idiosyncratic risk exposures $\boldsymbol{\alpha}$ and common risk exposures $\boldsymbol{\beta}$ (top panels in Figure 2). When the asset k's price volatility is ascribed to the increase in α_k (or β_k), the volatility is neg-

[22]The estimated values depend on the formation of data, especially returns. We repeated this analysis for different formation of data and confirm that the qualitative results are similar.

[23]For stating this argument, we have to show that Forex variables are not very informative for the price discovery in the Nasdaq market. We do not implement this analysis here, but this is highly likely because Forex variables are mostly insignificant to determine the Nasdaq OIBs.

atively (or positively) correlated with $corr(P_i, P_k)$. Both α_k and β_k refer to the degree of asymmetric information and contribute to the price volatility, but their effect on $corr(P_i, P_k)$ are opposite.

Based on these observation, we examine the determinants of predictability of Forex returns. Using the price volatility and the measure of asymmetric information $\frac{|OIB|}{VOL}$, we estimate the following equation.[24]

$$logit\left(1 - \frac{R^2_{-hf}}{R^2}\right) = b_1\sigma_{p,fx} + b_2\frac{|OIB_{fx}|}{VOL_{fx}} + b_3 r_{fx}$$
$$b_4\sigma_{p,nas} + b_5\frac{|OIB_{nas}|}{VOL_{nas}} + b_6 r_{nas} + b_7\frac{VOL_{hft}}{VOL_{all}} + \sum_{k=1:4} b_{7+k} Dummy.Q_k$$

Time subscripts are omitted. All the variables are at daily frequency. The dependent variable is the logit transformation of the percentage loss of the R-squared by omitting HFT-OIBs from the VAR equation (9). This gives a measure of predictive power of HFT to the Forex returns. Independent variables include the volatility, absolute OIB to volume ratio, and returns of both market. Here, the Nasdaq variables are the sum of HFT's and non-HFT's. The price volatility is defined by using the daily high and low prices[25]. The regression also includes HFT-volume to All-volume ratio and seasonal dummies for each quarter periods. All independent variables but dummies are standardized. The estimation is done through the OLS with the Newey-West standard errors.

Table 5 presents the regression results for each currency pair. The daily volatility σ_{fx} and absolute OIB to volume ratio $\frac{|OIB_{fx}|}{VOL_{fx}}$ tend to have negative coefficients. This implies that the Forex volatility and information asymmetry is mainly induced by idiosyncratic factors, reducing the predictability of Forex-returns by Nasdaq variables. To contrast, absolute OIB to volume ratio in Nasdaq is more like insignificant and the volatility tend to have mixed results. Thus, the volatility in Nasdaq is more likely to capture the varying exposure to

[24] The rationale for using $\frac{|OIB|}{VOL}$ as a measure of asymmetric information, please refer to Kaul, Lei and Stoffman (2008); Easley et al. (2008); Easley, O'Hara and Marcos (2012).

[25] i.e., $\sigma^2_{nas} = \frac{(P_{high} - P_{low})^2}{4log(2)}$.

common risk factors.

The HFT-volume ratio has significantly positive spillovers on the predictability measure. The more HFTs trade, the predictability of Forex by HFTs increases more. This provides evidence that HFT can improve the efficiency not only of the stock market, which is discussed in past literature, but also of the Forex market.

5.7 HFT v.s. non-HFT

In previous sections, a focus of investigation was on the reaction of Forex to the HFT and non-HFT order flows. In this section, the reverse is examined: the reaction of HFT and non-HFT to the fluctuation at Forex, with emphasis on the difference between the HFT and non-HFT.

To this aim, the difference of turnover captured by $OIB_{HFT} - OIB_{nHFT}$ is used. When this measure is positive (negative), HFT accumulate long (short) position more aggressively than non-HFT. We estimate the following equation:

$$OIB_{HFT,t} - OIB_{nHFT,t} = \sum_{k=1}^{K} \alpha_{fx,k} OIB_{fx,t-k} + \sum_{k=1}^{K} \rho_{fx,k} r_{t-k} + \epsilon_{fx,t}.$$

We estimate this equation for day-by-day samples for each three currency pairs.The Fama-MacBeth t-statistics are reported in Figure 8.

The result shows that the coefficients on $OIB_{fx}$are negative and those on returns move from positive to negative as the number of lag increases. This result indicates that HFTs trade more aggressively than non-HFT for around 2.5 seconds after a shock in Forex returns. After 2.5 seconds, non-HFTs become more aggressive. Potentially, this difference of reaction can explain why HFT-OIB predicts non-HFT OIB. HFTs aggregate market-wide information from real-time financial information including Forex markets, and move faster than non-HFTs.

5.8 Idiosyncratic stock trading matters for Forex?

So far aggregate OIB was used to represent the activities in Nasdaq. Does idiosyncratic fluctuation in each stock have predictive power to the Forex market? In this section, we use individual stock transaction data and examine if idiosyncratic component matters or not. The same VAR formula is used but replace the aggregate OIB with idiosyncratic component of stock i's OIB, and repeat stock-day by stock-day regressions.

To extract the idiosyncratic component, we regress the following equation for each stock i:

$$OIB_{i,t} = constant + \beta \sum_{i=1}^{120} OIB_{i,t} + \epsilon_t.$$

OIB_i is quoted order imbalances of stock i and can be either HFT's or non-HFT's. The predicted values and residuals are denoted as $\hat{O}IB_{i,t}$ and $\tilde{O}IB_{i,t}$. $\hat{O}IB_{i,t}$ is the common component of stock i and $\tilde{O}IB_{i,t}$ is the idiosyncratic component. The revised regression equation for Forex-return on Forex-OIB, stock i's HFT and non-HFT-OIB is

$$\begin{aligned} r_t = &\sum_{k=1}^{K} \alpha_{ret,k} OIB_{fx,t-k} + \sum_{k=1}^{K} (\beta_{c,ret,k} \hat{O}IB_{HFT,i,t-k} + \beta_{id,ret,k} \tilde{O}IB_{HFT,i,t-k}) \\ &+ \sum_{k=1}^{K} (\gamma_{c,ret,k} \hat{O}IB_{nHFT,i,t-k} + \gamma_{id,ret,k} \tilde{O}IB_{nHFT,i,t-k}) + \sum_{k=1}^{K} \rho_k r_{t-k} + \epsilon_{ret,t}. \end{aligned} \quad (11)$$

This equation is estimated stock-day by stock-day, and take the average of estimated coefficients across time. The estimation results of Equation (11) are summarized in Figure 7. Here we present the sum of stock i's Fama-MacBeth t-statistics up to ten lags for HFT-OIB and non-HFT-OIB coefficients, or $\sum_{k=1}^{10} FM_i(\beta_{id,ret,i,k})$, $\sum_{k=1}^{10} FM_i(\beta_{c,ret,i,k})$, $\sum_{k=1}^{10} FM_i(\gamma_{id,ret,i,k})$, and $\sum_{k=1}^{10} FM_i(\gamma_{c,ret,i,k})$.

The results show that an idiosyncratic component of OIB has almost zero predictive power to the future Forex-returns. Most of the predictive power comes from the common components. This findings holds for both HFT-OIB and non-HFT-OIB coefficients.

The result also is indirect evidence of why only HFT-OIB predict Forex: To predict Forex, only common component of Nasdaq OIB matters and HFT-OIB better represents the common component. Forex traders do not differentiate HFT-OIB from non-HFT-OIB but they just respond to the common component of OIB which is more likely to be driven by HFTs.

6 Discussion

Lead-lag relationships Our dataset has the identification of HFTs for Nasdaq but not for Forex. This limitation of data, in fact, makes it difficult to compare HFTs in Nasdaq and HFTs in Forex. What we compare is HFTs in Nasdaq with Forex overall. Therefore there is a possibility that HFTs in Forex are as faster as HFTs in Nasdaq and make the Forex rates more efficient. Since our VAR analysis shows that OIBs in Forex do not predict Nasdaq HFT-OIB but returns in Forex do, market makers in Forex are less likely to lag to Nasdaq HFTs. The response should be less than 200 milliseconds, which is hard for human traders to catch up and is likely to be an algorithm trader. We leave this possibility to future works.

Another concern over our lead-lag analysis is that Nasdaq system may simply be quicker than that of EBS. In this situation, although public news is received by traders in the both markets simultaneously, the difference in technological architecture may generate an apparent comovement with lead-lag relationship. Our rational expectations model, however, proposes that the comovement also occurs from liquidity trading shocks because uninformed traders try to infer information from prices and order flows. The model also implies that the reactions can be asymmetric between markets when one market is noiser. In fact, our empirical evidence suggests such an asymmetry, supporting the second view.

Hard information and HFTs The reason why only HFTs can lead Forex activities is that HFT-OIBs are more correlated with the market-wide activities than non-HFT-OIBs. Since the Forex market only reacts to the common component of Nasdaq OIBs, HFTs tend to predict the Forex activities better. Then, why are HFT-OIBs correlated with market-wide activities? Unfortunately, this paper does not answer this question. Potentially, HFTs engage in aggregating "hard" quantitative information in different markets such as Index futures, Government bond futures, and so on.

In fact, as suggested in the significant estimation of ρ_{hft} in the VAR system, HFTs put the Forex information together and reflect it in the stock prices. When we examine stock-by-stock trading, HFTs respond to the Forex rates more rapidly than non-HFTs and quickly reverse their trades. If HFTs respond to other financial assets in the same fashion, their trades necessarily predict non-HFT's behavior as literature claims.

7 Conclusion

In this paper, cross-market spillovers of returns, order flows and transactions between NASDAQ and the foreign exchange markets, in a fraction of one second, literally, were examined, exploiting high-frequency data sets in the two markets. In our research, a distinction between high-frequency traders (HFTs) and non-HFTs in the NASDAQ is important. It is expected that HFTs react to new arrival of news faster than non-HFTs, and we quantify the difference. The lead-lag relationship between order flows in Nasdaq and foreign exchange markets is examined, distinguishing HFTs and non-HFTs.

Major findings are summarized as follows. First, HFTs in the Nasdaq market are found to use market information of the foreign exchange market as well as the Nasdaq market. There is a strong evidence that HFTs in the Nasdaq aggregate information in the two market in order to place orders.

Second, HFTs react to the market conditions of the foreign exchange market faster than non-HFTs, by roughly 250 millisecond. The degree of the lag by non-HFTs may be surprisingly small, but it would make a large difference in return performance over the time.

Third, order flow imbalances and returns in the Nasdaq leads those in the foreign exchange market. Possible reasons were discussed in Section 6.

Four, when the share of HFT trades in total trade increases, the correlation between

Nasdaq returns and foreign exchange returns becomes higher.

Five, during a crisis period, the degree of the above findings—such as the correlation of HFTs trading share and return co-movement between Nasdaq and foreign exchange—is intensified. The crisis period increases risk and opportunities of trading assets and advantages of HFTs over non-HFTs become more distinctive.

In sum, the strong presence of HFTs in the Nasdaq market contributes to faster price discovery not only in Nasdaq market but in the foreign exchange market with lag of a fraction of one second. The novelty of this paper is to match the high-frequency datasets in the two market and found the above-mentioned empirical findings. Theoretical framework and prediction of the Kodres and Pritsker (2002) is confirmed by our empirical examination.

A Data construction

- Nasdaq OMX High-frequency trading data

  1. Making second-by-second data

     (a) For each HFT or non-HFT initiated trades, trades are summed up within a quarter-second. We calculate total quoted-volume and quoted order imbalance (OIB). OIB is defined as buyer-initiated quoted trade volumes less seller-initiated quoted trade volumes. Thus we have OIB of individual stocks for each HFT and non-HFT. This individual data is used in section 5.8.

     (b) The volume and OIB are aggregated across individual stocks for each quarter-second, generating HFT-aggregate data and non-HFT-aggregate data.

     (c) There are gaps between transactions when there is no trade within some quarter-second. We fill in the gap and keep the time-stamp as real clock time. This interpolated data is used to calculate the descriptive statistics and conditional frequencies.

  2. Making daily frequency data

     (a) We calculate the daily return, volume, and OIB. Daily frequency data are used to analyze the determinant of the magnitude of the spillovers (in section 5.6).

- ICAP EBS Foreign Exchange data (refer Table 6 to the dataset after data-cleaning below)

  1. Making quarter-second by quarter-second data

     (a) Original level-five data have quarter-second by quarter-second time stamps.

        i. When there are multiple trades within a quarter-second, the orders have the same time stamp. We calculate the total volume and OIB by summing up within the quarter-second.

ii. When there are both quotes and prices within a quarter-second, the records do not specify the correct ordering of the transactions. In this case, we identify the order for matching the direction of deals and the prices on the LOB. For example, buying orders are followed by the increase of the LOB prices while selling orders are followed by the drop of prices. When the prices on the LOB do not change before and after the deals, we can safely ignore the orders: we randomly determine the orders.

iii. When there are no quotes update between subsequent trades, we consolidate the trades and give the youngest time-stamp.

(b) The price is defined as a mid-price calculated form the order book data .[26] We define two prices during a time segment t: the log-price before the first trade within a quarter-second (denoted as $p_{b,t}$) and the log-price after the last trade within a quarter-second (denoted as $p_{a,t}$). When there is only one quote change between trades, another quote is interpolated. Forex returns are defined in three ways. The first one, or r_{t1}, is defined by the log-price changes between the before-the-trade-prices at $t+1$ and the before-the-trade-prices at t, or $p_{b,t+1} - p_{a,t}$. Thus, the return r_{t1} represents the price changes involving trades. The second one, r_{t2}, is defined by the log-price changes between the before-the-trade prices at $t + 1$ and the after-the-trade prices at t. r_{t2} does not include the price changes that is caused by trades. The last definition, r_{t3}, is the log-difference between the after-the-trade price at t and the before-the-trade price at t. The definition r_{t1} and r_{t2} have a bias in the time-stamp. The time-stamp shows the occurrence of the price change too early especially when price changes are less frequent. This is because the returns at t use future prices at $t + 1$. This problem pronounces when we focus on real-time rather than event-time. To contrast, the disadvantage of using r_{t3} is that

[26]The price can also be defined as a depth weight mid-price. In this article, we take a equally weighted mid-price.

it misses inter-seconds price changes. Since inter-second price changes occur when a trade is not associated. In this article, we employ r_{t1} for regression analysis and use r_{t3} for calculating conditional probability of order directions .

(c) The Forex data are analyzed currency pair by currency pair. We do not aggregate the data like Nasdaq data.

(d) We fill in the gap of the time to calculate the descriptive statistics and conditional frequencies.

2. Making daily frequency data

   (a) We make daily frequency data in a similar manner of the Nasdaq case.

We merge the two data-set for estimations.

- Merge the Nasdaq and Forex data

  1. We convert the time-stamp of Forex from GMT to EST.
  2. We discard the observation before 9:30 and after 16:00 EST for both the Nasdaq and Forex data-set.
  3. We merge each data by outer join, meaning that we keep all the transactions. Thus we have event-time data. We use the combined data for estimating the VAR.
  4. For robustness check, we also combined the data by keeping calender time, which allows no transactions in both markets. We use the data in our unreported manuscript.

References

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

Barclay, Michael J., and Terrence Hendershott. 2003. "Price Discovery and Trading After Hours." *Review of Financial Studies*, 16(4): 1041–1073.

Bernhardt, Dan, and Bart Taub. 2008. "Cross Asset Speculation in Stock Markets." *The Journal of Finance*, LXIII(5).

Biais, Bruno, Thierry Foucault, and Sophie Moinas. 2013. "Equilibrium high-frequency trading." *Working Paper.*

Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2012. "High frequency trading and price discovery." *SSRN eLibrary.*

Caballe, Jordi, and Murugappa Krishnan. 1994. "Imperfect competition in a multi-security market with risk neutrality." *Econometrica*, 62(3): 695–704.

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

Chaboud, Alain, Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega. 2014. "Rise of the machines: Algorithmic trading in the foreign exchange market." *The Journal of Finance*, 69(5): 2045–2084.

Easley, David, Maureen O'Hara, and M Marcos. 2012. "Flow toxicity and liquidity in a high-frequency world." *Review of Financial Studies*, 25(2): 1457–1493.

Easley, David, Robert Engle, Maureen O'Hara, and Liuren Wu. 2008. "Time-varying arrival rates of informed and uninformed trades." *Journal of Financial Econometrics*, 6(2): 171–207.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the impossibility of informationally efficient markets." *The American Economic Review*, 70(3): 393–408.

Hagströmer, Björn, and Lars Nordén. 2013. "The diversity of high-frequency traders." *Journal of Financial Markets*, 16(4): 741–770.

Hasbrouck, Joel. 1991. "The summary informativeness of stock trades: An econometric analysis." *Review of Financial Studies*, 4(3): 571–595.

Hasbrouck, Joel. 1995. "One security, many markets: Determining the contributions to price discovery." *The Journal of Finance*, L(4): 1175–1199.

Hasbrouck, Joel. 2002. "Stalking the "efficient price" in market microstructure specifications: an overview." *Journal of Financial Markets*, 5(3): 329–339.

Hendershott, Terrence. 2011. "Does algorithmic trading improve liquidity?" *The Journal of Finance*, LXVI(1): 1–33.

Hendershott, Terrence, and Albert J. Menkveld. 2014. "Price pressures." *Journal of Financial Economics*, 114(3): 405–423.

Hirschey, Nicholas. 2013. "Do High Frequency Traders Anticipate Buying and Selling Pressure." *Working Paper.*

Ito, Takatoshi, and Yuko Hashimoto. 2006. "Intra-day Seasonality in Activities of the Foreign Exchange Markets: Evidence from the Electronic Broking System." *Journal of the Japanese and International Economies*, 20(4): 637–664.

Ito, Takatoshi, Kenta Yamada, Misako Takayasu, and Hideki Takayasu. 2012. "Free Lunch! Arbitrage Opportunities in the Foreign Exchange Markets." *NBER working paper 18541.*

Kaul, Gautam, Qin Lei, and Noah Stoffman. 2008. "Aiming at Pin: Order Flow, Information, and Liquidity." *Working paper.*

King, Mervyn A., and Sushil Wadhwani. 1990. "Transmission of volatility between stock markets." *Review of Financial studies*, 3(1): 5–33.

Kodres, Laura E., and Matthew Pritsker. 2002. "A rational expectations model of financial contagion." *The Journal of Finance*, LVII(2).

Menkveld, Albert J. 2013. "High frequency trading and the new market makers." *Journal of Financial Markets*, 16(4): 712–740.

Menkveld, Albert J., and Boyan Jovanovic. 2012. "Middlemen in Limit Order Markets." *Working Paper.*

Zhang, S. Sarah. 2012. "Need for Speed: An Empirical Analysis of Hard and Soft Information in a High Frequency World." *SSRN Electronic Journal*.

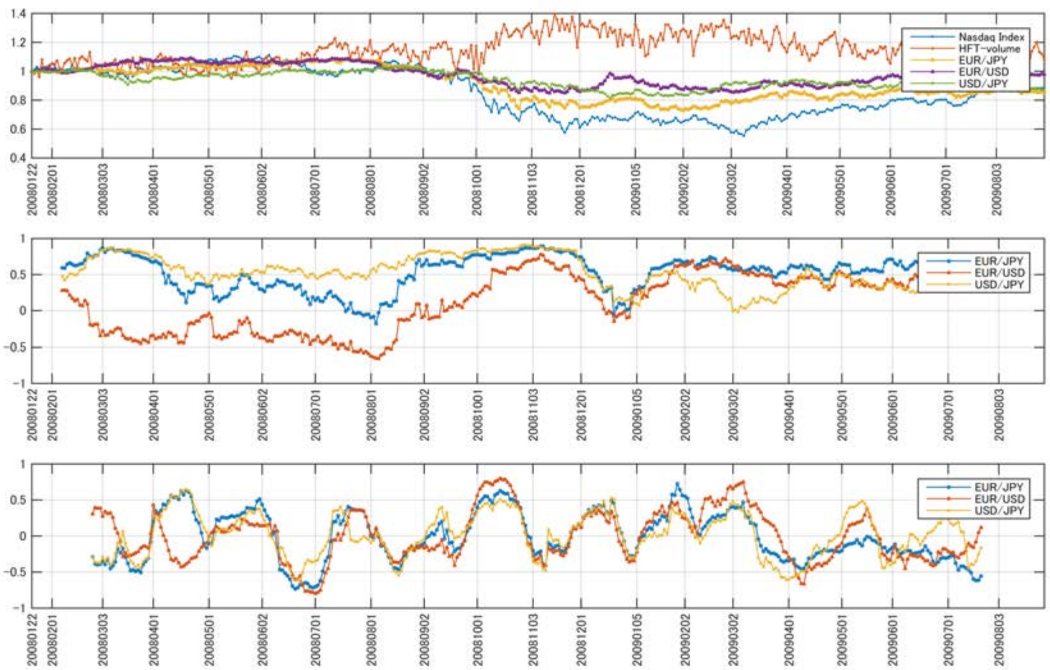

Figure 1: (Top) Time series plots of Nasdaq Composite Index and Forex rates. This figure plots the daily assets price at 16:00 (EST) over time. The prices are normalized by the prices at Jan 7, 2008.
(Middle) The centered rolling correlation between the Nasdaq Composite Index return and Forex returns. The rolling window is 25 days.
(Bottom) The centered rolling correlation between the level of the proportion of HFTs trading volume and the stock-Forex correlations. The rolling window is 25 days.

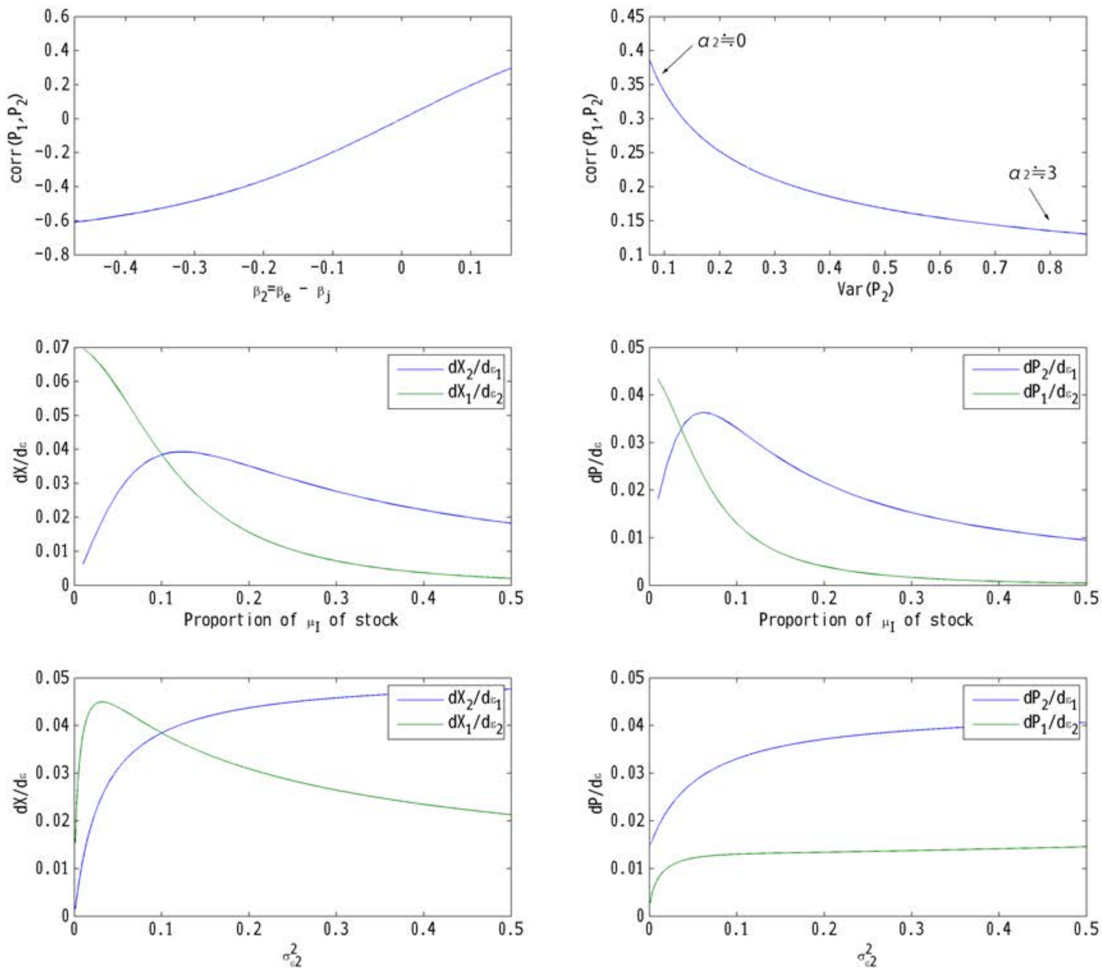

Figure 2: Simulation results of the model. This simulation describes how the price and order flows can be influenced by the changes in underlying parameters such as the exposure to the risk, number of informed traders, quantities of noise trading.
(1) Top-left panel: The correlation of prices P_1 and P_2 as a function of common exposure to the fundamentals $\beta_2 = \beta_e - \beta_j$.
(2) Top-right panel: The relationships between $corr(P_1, P_2)$ and $Var(P_2)$ as a function of idiosyncratic risk exposure α_2.
(3) Middle panels: The impacts on noise order flows X_1 and X_2 (left) and prices P_1 and P_2 (right) as a function of the proportion of informed traders μ_I of asset 1.
(4) Bottom panels: The impacts on noise order flows (left) and prices (right) as a function of the variance of noise trading of asset 2 $\sigma^2_{e,2}$.
In panels (3) and (4), the gaps between blue (response of asset 2 to shocks in 1) and green (response of asset 1 to shocks in 2) show the asymmetric spillovers.

Note: Baseline parameters are set as follows: $[\beta_1, \beta_E, \beta_U, \beta_J] = [1, 1, 0.5, 0.5]/\sqrt{10}$; For $j = 1, E, U, J$, $\alpha_j = 1/\sqrt{10}$; $Var(\tilde{e}_j) = 10$; $Var(\tilde{u}_j) = 10$. $\mu_I = 0.1$ and $\mu_U = 0.9$.

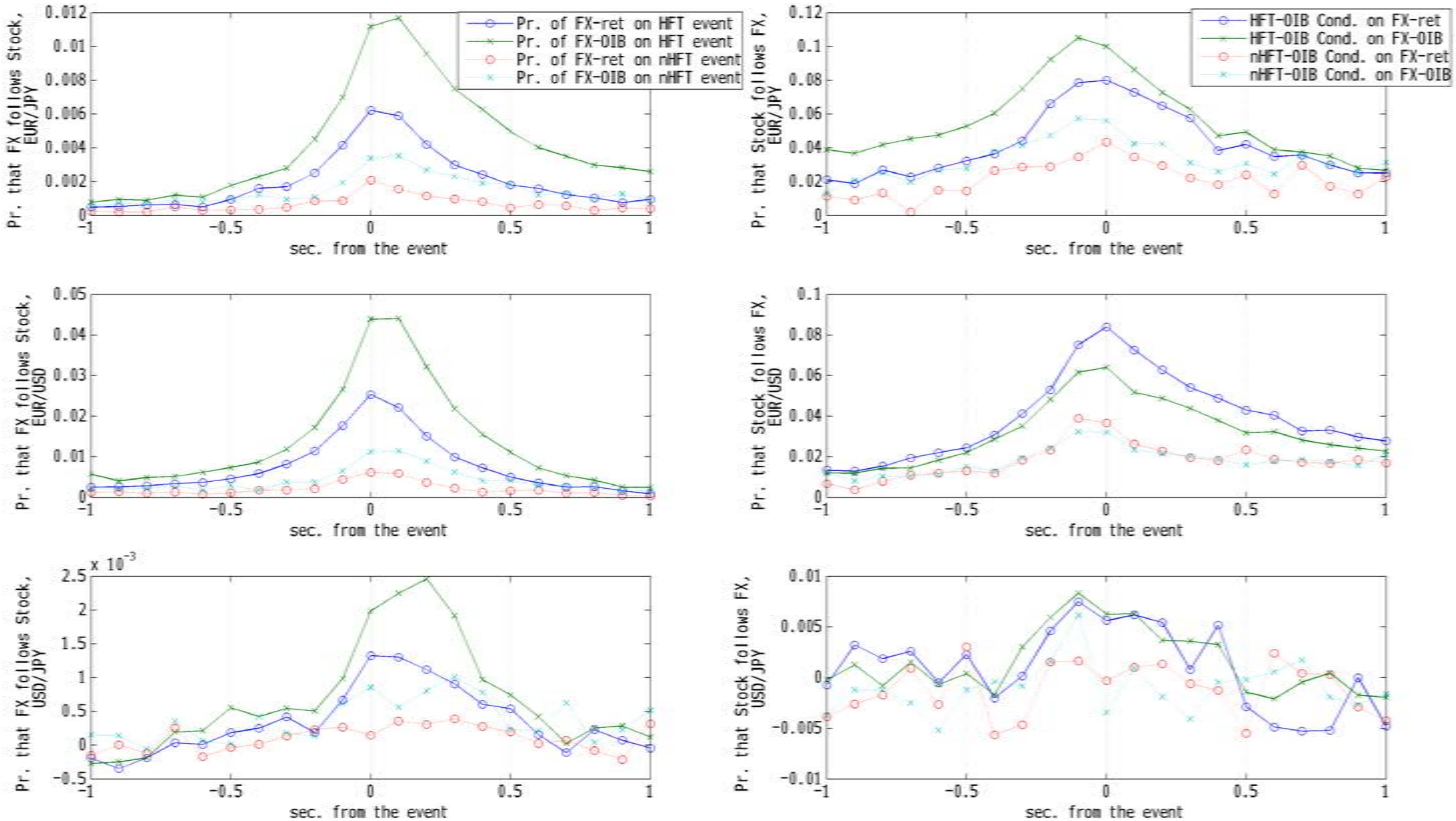

Figure 3: Conditional probability plots. We present the differences for normalizing the unconditional probability equal to zero: $Prob(OIB_{FX,0-k} > 0|OIB_{HF,0} > std(OIB_{HF,0})) - Prob(OIB_{FX,0-k} < 0|OIB_{HF,0} > std(OIB_{HF,0}))$.

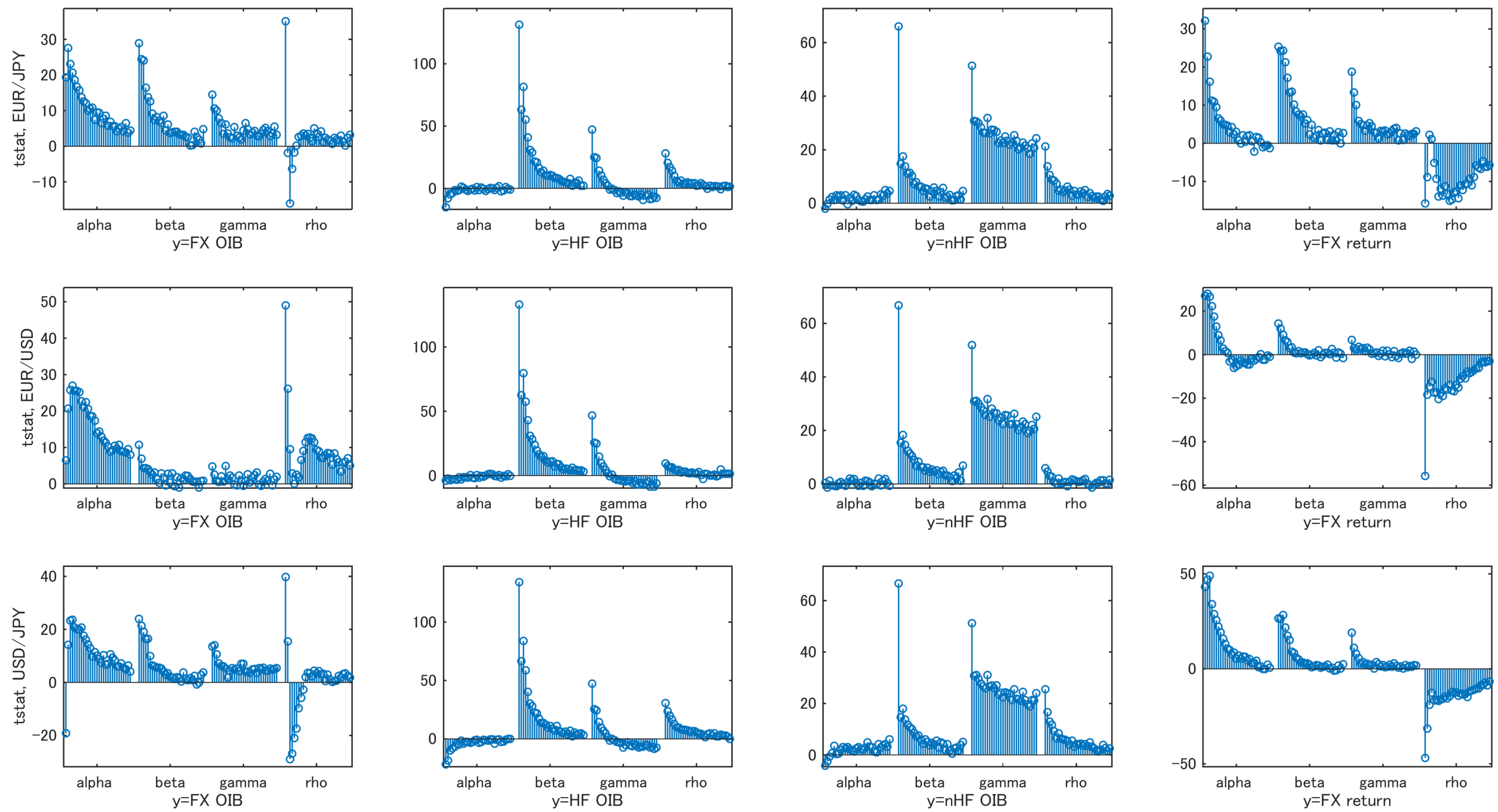

Figure 4: VAR(30) result for the equations of (6) to (9). The estimation is day-by-day, and each estimation contains roughly 65000 observations. The panels plot Fama-MacBeth t-statistics of the coefficients. The horizontal axis shows the independent variables. (1) FX-OIB (α), (2) HFT-OIB (β) (3) non-HFT-OIB (γ), and (4) FX-return (ρ). Each independent variable has coefficients (and t-statistics) from one to 30 lags which are plotted from left to right.

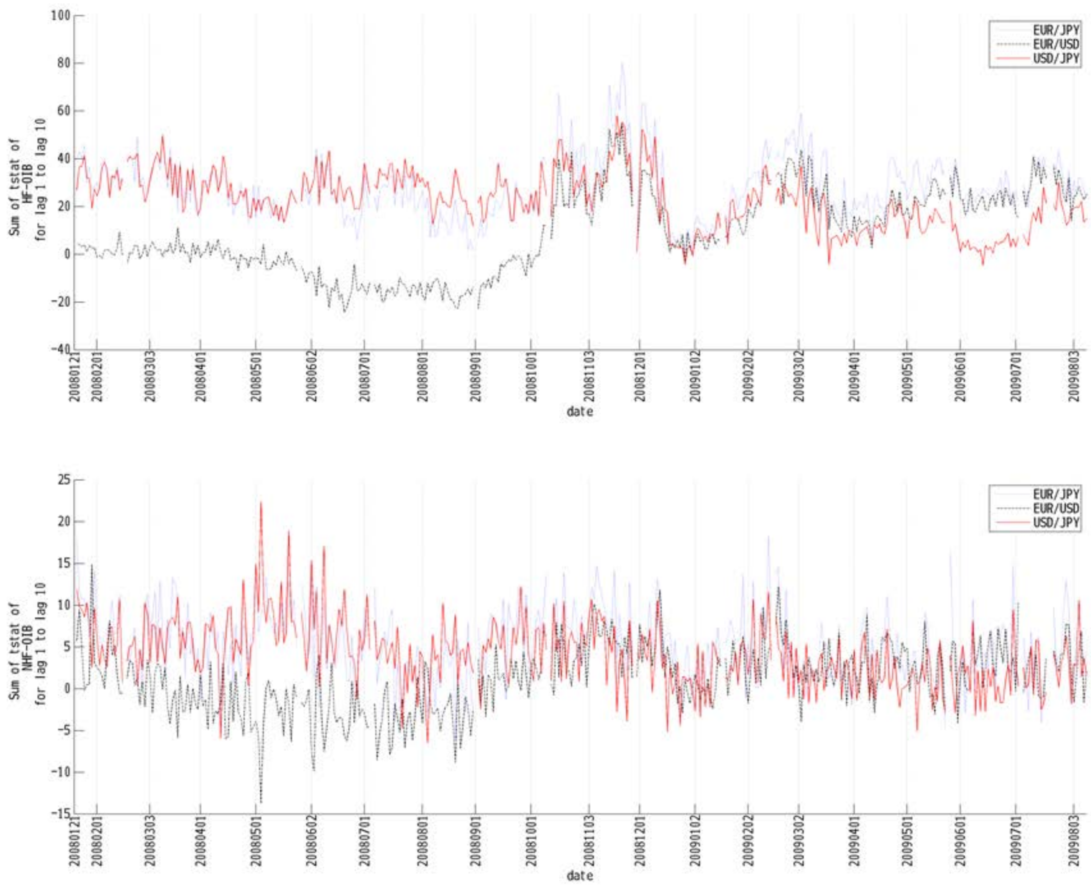

Figure 5: Time series plots of the sums of t-statistics in the equation of $y = OIB_{fx}$. For each day-estimation, the sum of t-statistics are calculated for the coefficients of $OIB_{HFT,t-k}$ (top) and $OIB_{nHFT,t-k}$ (bottom) up to 10 lags, i.e., $\sum_{k=1,\ldots,10} t_{t-k}$.

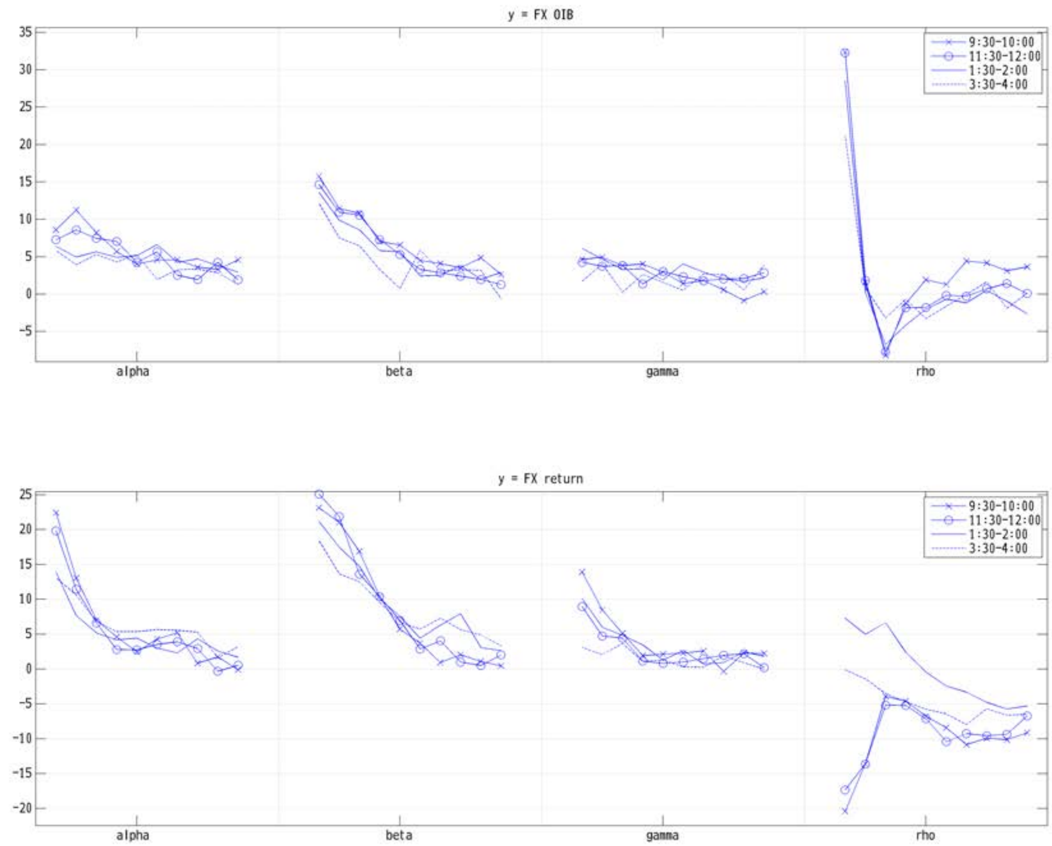

Figure 6: The VAR results for time-of-day regression. The currency pair is EUR/JPY. The dependent variable is either FX OIB or FX return. The way of presentation is the same as Figure 4: the horizontal axis shows the independent variables. (1) FX-OIB (α), (2) HFT-OIB (β) (3) non-HFT-OIB (γ), and (4) FX-return (ρ). Each independent variable has coefficients (and t-statistics) from one to 10 lags which are plotted from left to right.

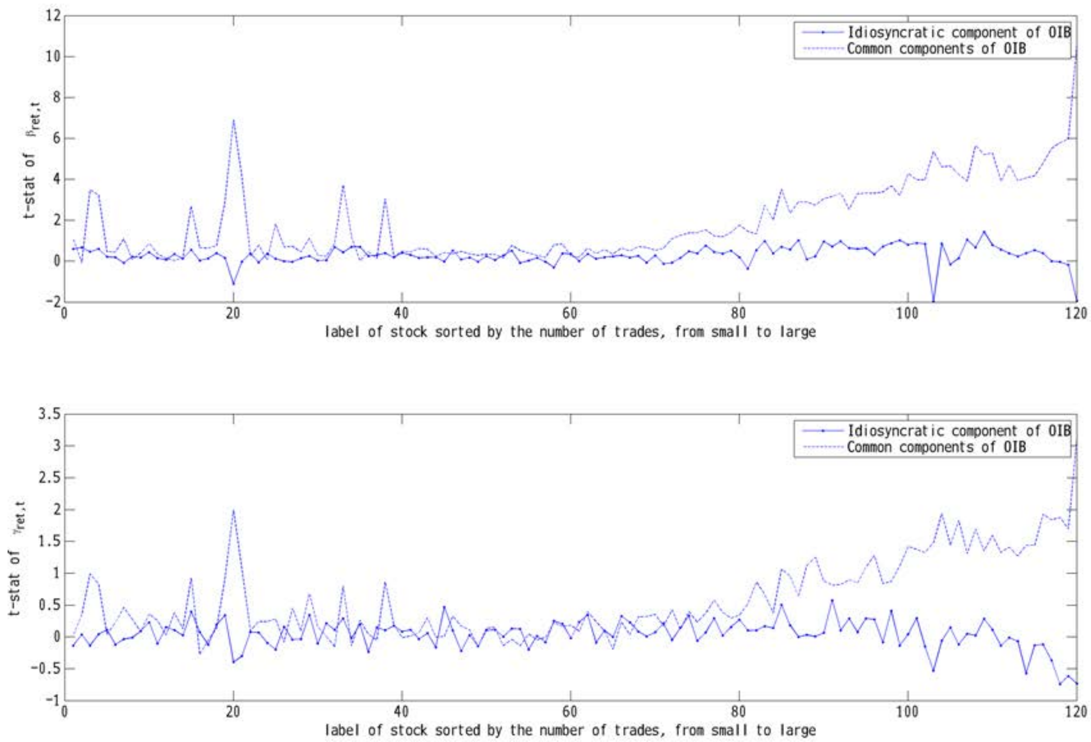

Figure 7: Plot of $\sum_{k=1,\dots,10} t_{\beta_{ret},k}$ (top) and $\sum_{k=1,\dots,10} t_{\gamma_{ret},k}$ (bottom) for idiosyncratic residual component and the common expected component, sorted by the number of trades from the smallest (x=1) to largest (x=120).

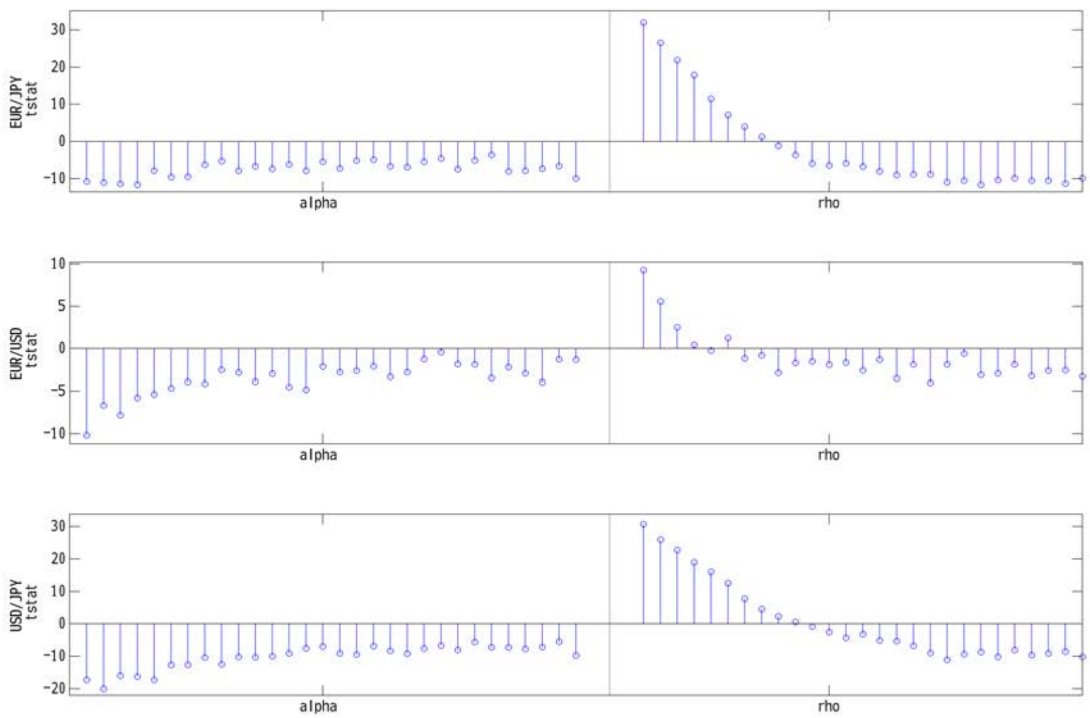

Figure 8: A Plot of t-satistics for the regression coefficients:

$OIB_{hf,t} - OIB_{nhf,t} = \sum_{k=1}^{30} \alpha_{fx,k} OIB_{fx,t-k} + \sum_{k=1}^{30} \rho_{fx,k} r_{t-k} + \epsilon_{fx,t}$.

The currency pairs are EUR/JPY, EUR/USD, and USD/JPY (from top to bottom). This plot describes who becomes a net-aggressive-buyer after the shocks for Forex OIBs and Forex returns. When $OIB_{hf,t} - OIB_{nhf,t}$ is positive (negative), HFT accumulate long (short) position more aggressively than non-HFT. Therefore the positive coefficients imply future relative intensity of aggressive buying for HFTs to non-HFTs.

Panel A: FX		Std(r_t)	mean($\frac{r_t}{oib}$)	Trade size	Vol./day	Std(Vol./qsec)	Std(OIB./qsec)	Pr(Quote)	Pr(Trade)	Trade
unit		bp	bp/unit	¥ or $ (mil.)	¥ or $ (bil.)	Yen or $ (mil.)	Yen or $ (mil.)			#
EUR/JPY	Q1-Q4	0.149	0.182	216	638	55.6	51.9	0.278	0.0282	1.05
EUR/JPY	Q1	0.112	0.117	267	939	74.8	70.2	0.284	0.0337	1.04
EUR/JPY	Q2	0.15	0.186	223	708	58.6	55.3	0.285	0.0305	1.04
EUR/JPY	Q3	0.2	0.282	172	486	41.3	38.2	0.309	0.0272	1.04
EUR/JPY	Q4	0.138	0.16	187	417	40.4	36.7	0.234	0.0212	1.05
EUR/USD	Q1-Q4	0.121	0.0787	3.14	31.7	1.75	1.63	0.293	0.0908	1.08
EUR/USD	Q1	0.078	0.0501	4	35	2.08	1.96	0.216	0.0789	1.08
EUR/USD	Q2	0.129	0.0716	3.32	39.5	2.02	1.9	0.3	0.107	1.08
EUR/USD	Q3	0.162	0.108	2.51	26.8	1.35	1.24	0.326	0.0969	1.09
EUR/USD	Q4	0.11	0.0828	2.86	25.4	1.4	1.27	0.33	0.0803	1.09
USD/JPY	Q1-Q4	0.128	0.118	213	1.3e+03	91.6	84.8	0.264	0.0555	1.07
USD/JPY	Q1	0.124	0.0914	255	1.9e+03	119	111	0.309	0.0685	1.07
USD/JPY	Q2	0.153	0.114	223	1.75e+03	108	101	0.326	0.0718	1.07
USD/JPY	Q3	0.15	0.16	168	876	65.1	60	0.239	0.0477	1.07
USD/JPY	Q4	0.102	0.125	179	673	58.4	53.6	0.185	0.0343	1.07

Panel B: Nasdaq		Trade size	Vol./day	Std(Vol./qsec)	Pr(Trade)	Trade	Std(OIB./qsec)	$corr^2(X, OIB_{all})$
unit		$	$ (bil.)	$		#	$	
HFT	Q1-Q4	7.21e+03	3.41	9.21e+04	0.604	4.55	8.85e+04	0.0124
HFT	Q1	1.07e+04	4.18	1.05e+05	0.615	3.72	1.02e+05	0.0102
HFT	Q2	7.67e+03	4.28	1.02e+05	0.665	5.56	9.9e+04	0.0105
HFT	Q3	5e+03	3.05	8.04e+04	0.627	5.58	7.8e+04	0.0139
HFT	Q4	5.39e+03	2.13	7.4e+04	0.508	3.42	7.03e+04	0.0152
non-HFT	Q1-Q4	6.69e+03	4.7	1.1e+05	0.793	6.29	9.47e+04	0.00834
non-HFT	Q1	9.02e+03	6.74	1.37e+05	0.846	6.74	1.24e+05	0.00817
non-HFT	Q2	7.05e+03	5.75	1.17e+05	0.849	7.61	1.03e+05	0.00909
non-HFT	Q3	4.87e+03	3.45	8.38e+04	0.779	6.44	7.05e+04	0.0081
non-HFT	Q4	5.59e+03	2.86	8.94e+04	0.698	4.36	7.14e+04	0.00809

Table 1: Descriptive summary statistics.

	y=FX-OIB			y=HFT-OIB			y=non-HFT-OIB			y=FX-return		
lag	coef.	t-stat	1% sig	coef.	t-stat	1% sig	coef.	t-stat	1% sig	coef.	t-stat	1% sig
α(FX-OIB)												
1	0.0212	5.37	0.714	-6.87	-1.01	0.163	0.155	0.044	0.0213	1.33e-08	7.18	0.898
2	0.0163	4.18	0.693	-2.81	-0.432	0.0402	0.814	0.131	0.0284	5.82e-09	3.21	0.619
3	0.0103	2.63	0.504	-1.68	-0.267	0.0402	1.55	0.221	0.0213	3.32e-09	1.71	0.324
4	0.009	2.28	0.485	-1.13	-0.189	0.0213	1.99	0.318	0.0331	1.9e-09	1.06	0.229
5	0.00738	1.86	0.397	-0.353	-0.0755	0.0284	1.52	0.243	0.0189	1.71e-09	0.999	0.187
7	0.0061	1.53	0.312	-0.165	-0.0181	0.0142	2.22	0.351	0.0378	1.06e-09	0.61	0.168
10	0.00539	1.37	0.366	0.315	0.0358	0.0213	3.36	0.52	0.0544	4.2e-10	0.287	0.113
β(HFT)												
1	7.37e-06	3.22	0.567	0.235	58.6	0.953	0.0505	12.6	0.936	1.37e-11	11.9	0.936
2	4.96e-06	2.07	0.369	0.0577	14.1	0.934	0.0084	1.91	0.57	6.89e-12	5.95	0.846
3	3.77e-06	1.57	0.262	0.0348	8.44	0.946	0.00804	1.89	0.508	4.04e-12	3.49	0.681
4	2.34e-06	0.979	0.137	0.0204	4.94	0.851	0.00539	1.25	0.371	2.29e-12	2	0.366
5	1.71e-06	0.717	0.0662	0.013	3.11	0.671	0.00343	0.803	0.241	1.42e-12	1.27	0.196
7	1.07e-06	0.457	0.0449	0.00766	1.83	0.324	0.00243	0.584	0.158	8.71e-13	0.762	0.104
10	1.48e-06	0.627	0.0686	0.00583	1.43	0.26	0.00448	1.09	0.251	5.59e-13	0.444	0.0567
γ(non-HFT)												
1	2.3e-06	0.93	0.118	0.04	10.1	0.924	0.124	30.5	0.953	2.17e-12	2	0.381
2	1.91e-06	0.773	0.125	0.0104	2.63	0.527	0.0565	13.6	0.948	1.15e-12	0.948	0.147
3	1.69e-06	0.7	0.0638	0.0078	1.92	0.359	0.0415	9.89	0.943	7.63e-13	0.576	0.0567
4	1.47e-06	0.574	0.0662	0.00425	1.02	0.194	0.0364	8.69	0.939	5.24e-13	0.396	0.0449
5	1.21e-06	0.495	0.0473	0.00259	0.627	0.109	0.0265	6.33	0.884	4.08e-13	0.309	0.0402
7	1.29e-06	0.507	0.052	0.00105	0.237	0.0851	0.0198	4.8	0.818	3.38e-13	0.257	0.0378
10	8.83e-07	0.363	0.0331	-0.000124	-0.0585	0.0851	0.0183	4.44	0.806	3.85e-13	0.197	0.0426
ρ(FX return)												
1	3.2e+05	29.7	0.953	1.82e+08	9.78	0.898	6.78e+07	3.35	0.603	-0.0266	-6.53	0.761
2	-3.13e+03	-1.42	0.567	6.46e+07	3.13	0.527	2.01e+07	0.917	0.118	-0.00976	-2.36	0.624
3	-2.59e+04	-2.01	0.444	3.28e+07	1.55	0.246	1.28e+07	0.61	0.0733	0.00136	0.403	0.444
4	-6.86e+03	-0.255	0.189	2e+07	0.975	0.132	8.93e+06	0.519	0.052	0.000637	0.23	0.331
5	-2.1e+03	0.031	0.0922	1.24e+07	0.583	0.0733	8.42e+06	0.475	0.0449	-0.00245	-0.551	0.274
7	1.23e+03	0.206	0.078	5.61e+06	0.323	0.0426	4.11e+06	0.316	0.0331	-0.00508	-1.24	0.284
10	3.21e+03	0.395	0.0922	5.41e+06	0.33	0.0378	7.99e+06	0.468	0.052	-0.0032	-0.792	0.182

Table 2: VAR(10) result for the equations of (6) to (9). The estimation is day-by-day, and each estimation contains roughly 65000 observations. We present only the estimation of EUR/JPY for the interest of brevity. We show the means of estimated coefficients and the t-statistics. We also show the proportions of the coefficients that is significant at the 1% level.

EUR/JPY		FX-OIB	NS HFT-OIB	NS non-HFT-OIB	FX-Return
	FX-OIB	0	0.053	0.294	0
	NS HFT-OIB	0.414	0	0	0.0128
	NS non-HFT-OIB	0.467	0	0	0.151
	FX-Return	0	0	0.179	0
EUR/USD					
	FX-OIB	0	0.202	0.405	0
	NS HFT-OIB	0.498	0	0	0.178
	NS non-HFT-OIB	0.534	0	0	0.364
	FX-Return	0	0.108	0.366	0
USD/JPY					
	FX-OIB	0	0.166	0.289	0
	NS HFT-OIB	0.383	0	0	0.0205
	NS non-HFT-OIB	0.533	0	0	0.175
	FX-Return	0	0.00967	0.275	0

Table 3: Granger Causality Tests (Null hypothesis: for (i, j) element, column variable j does not Granger-cause row variable i) for the data using inner join. The figures are the p-value of the test. The p-values less than 0.001 are rounded to zero.

unit = %		ALL				Q1	Q2	Q3	Q4
		mean	std	$\underline{IS}_j$	$\overline{IS}_j$				
EUR/JPY	FX-OIB	0.604	0.529	0.584	0.614	0.838	0.79	0.441	0.424
	HFT-OIB	4.27	3.37	3.88	7.36	4.46	3.16	5.72	3.93
	non-HFT-OIB	0.378	0.464	0.345	1.43	0.516	0.241	0.424	0.282
	FX-ret	91.6	16.8	88.5	91.4	91.2	92.8	90.4	91.4
EUR/USD	FX-OIB	4.41	1.98	4.37	4.41	5.66	5.37	3.59	3.42
	HFT-OIB	1.87	2.2	1.72	3.08	0.125	1.48	3.19	2.63
	non-HFT-OIB	0.16	0.259	0.149	0.586	0.0672	0.111	0.242	0.186
	FX-ret	90.4	16.4	89.2	90.3	91.1	90	90	89.8
USD/JPY	FX-OIB	2.83	1.29	2.76	2.82	3.76	2.82	2.23	2.65
	HFT-OIB	3.23	2.8	2.97	5.56	4.61	4.2	3.37	0.925
	non-HFT-OIB	0.311	0.446	0.287	1.11	0.559	0.262	0.274	0.108
	FX-ret	90.5	16.6	88.2	90.3	88.1	89.7	91.1	92.3

Table 4: Long-run variance decomposition of the efficient prices. The figures are denoted as percentage to the total variance. The higher figure indicates stronger spillovers on the efficient price discovery.

$y = logit\left(1 - \frac{R^2_{-hft}}{R^2}\right)$	EUR/JPY	EUR/USD	USD/JPY
volatility of P_{fx}	-0.0166	0.0967 ***	-0.0236
	(-0.32)	(5.82)	(-0.49)
$\frac{\lvert OIB_{fx}\rvert}{VOL_{fx}}$	-0.172 *	-0.0473	-0.163 **
	(-1.47)	(-0.668)	(-1.92)
FX-return	0.0651	0.0873 *	0.036
	(1.1)	(1.49)	(0.573)
volatility of P_{nas}	-0.118 **	-0.0444	0.0867 *
	(-1.92)	(-0.63)	(1.51)
$\frac{\lvert OIB_{nas}\rvert}{VOL_{nas}}$	0.00235	0.0821 **	-0.0148
	(0.0497)	(1.75)	(-0.337)
Nasdaq return	-0.0847 *	0.00375	-0.0229
	(-1.52)	(0.077)	(-0.4)
VOL.HFT/VOL.ALL	0.296 ***	0.453 ***	0.474 ***
	(3.11)	(3.77)	(4.52)
Dummy Q1	0.0291	-3.67 ***	-0.723 ***
	(0.204)	(-17.7)	(-4.52)
Dummy Q2	-0.74 ***	-2.49 ***	-0.729 ***
	(-4.51)	(-15.1)	(-5.8)
Dummy Q3	-0.717 ***	-2.14 ***	-1.91 ***
	(-3.21)	(-7.87)	(-8.3)
Dummy Q4	-0.0779	-1.01 ***	-2.04 ***
	(-0.521)	(-5.81)	(-7.44)
# of obs	402	402	402
R^2	0.1117	0.6282	0.333

Table 5: Determinants of predictability. All the variables are at daily frequency.
The dependent variable is the logit transformation of the percentage loss of the R-squared by omitting HFT-OIBs for the regressor from the VAR.
For independent variables, we have the volatility, absolute OIB to volume ratio, and return for both market. We also have HFT-volume to All-volume ratio and seasonal dummies. Dummies are for each quarter periods. The Nasdaq variables are the sum of HFT's and non-HFT's. The volatility of Nasdaq is calculated from the prices of daily high and low. All independent variables but dummies are standardized.
The estimation is done through the OLS with Newey-West standard errors.

Date	Time	Deal/Quote	Price before deal	Price after deal	Buy #	Sell #	OIB	r_{t1}	r_{t2}	r_{t3}
20080121	136864	Q	153.715	153.715			0	6.5e-05	6.50e-05	0
20080121	136868	Q	153.725	153.725			0	0	0	0
20080121	136872	D	153.725	153.715	0	1	-1	-6.5e-05	0	-6.5e-05
20080121	136876	Q	153.715	153.715			0	0	0	0
20080121	136880	Q	153.715	153.715			0	-3.25e-05	-3.25e-05	0
20080121	136884	Q	153.710	153.710			0	0	0	0

Table 6: Example of the Forex dataset (EUR/JPY). Time-stamp is a quarter-second from GMT00:00. When there is no deals at the interval, the price after deals is equal to the price before deals. Each price is a mid-price and is obtained from the state of the limit order book.

Return r_{t1} is defined as $log(Price\ before\ deals\ at\ t+1) - log(Price\ before\ deals\ at\ t)$. Return r_{t2} is defined as $log(Price\ before\ deals\ at\ t+1) - log(Price\ after\ deals\ at\ t)$. Thus, Return r_{t1} includes the price change associated with trades, while Return r_{t2} does not. The deals at t is defined as a deals occurring in $(t-1, t]$. Return r_{t3} is defined as $log(Price\ after\ deals\ at\ t) - log(Price\ before\ deals\ at\ t)$.

In this example, the deal at 136872 occurs in (136871,136872]. Nasdaq orders are matched with these intervals.