



Cross-impact of order flow imbalance in equity markets

Rama Cont, Mihai Cucuringu & Chao Zhang

To cite this article: Rama Cont, Mihai Cucuringu & Chao Zhang (2023) Cross-impact of order flow imbalance in equity markets, Quantitative Finance, 23:10, 1373-1393, DOI: [10.1080/14697688.2023.2236159](https://doi.org/10.1080/14697688.2023.2236159)

To link to this article: <https://doi.org/10.1080/14697688.2023.2236159>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 17 Aug 2023.



Submit your article to this journal



Article views: 7204



View related articles



View Crossmark data



Citing articles: 11 View citing articles



© 2023 iStockphoto LP

Cross-impact of order flow imbalance in equity markets

RAMA CONT  , MIHAI CUCURINGU   and CHAO ZHANG  

[†]Mathematical Institute, University of Oxford, Oxford, UK

[‡]Oxford Suzhou Centre for Advanced Research, Suzhou, People's Republic of China

[§]Department of Statistics, University of Oxford, Oxford, UK

[¶]Oxford-Man Institute of Quantitative Finance, University of Oxford, Oxford, UK

^{||}The Alan Turing Institute, London, UK

(Received 24 September 2022; accepted 7 July 2023; published online 17 August 2023)

We investigate the impact of order flow imbalance (OFI) on price movements in equity markets in a multi-asset setting. First, we propose a systematic approach for combining OFIs at the top levels of the limit order book into an integrated OFI variable which better explains price impact, compared to the best-level OFI. We show that once the information from multiple levels is integrated into OFI, multi-asset models with cross-impact do not provide additional explanatory power for contemporaneous impact compared to a sparse model without cross-impact terms. On the other hand, we show that lagged cross-asset OFIs do improve the forecasting of future returns. We also establish that this lagged cross-impact mainly manifests at short-term horizons and decays rapidly in time.

Keywords: Market impact; Cross-impact; Order flow imbalance; Return prediction

JEL Codes: C31, C53, G14

1. Introduction

Accurately estimating and forecasting the impact of trading behavior of market participants on the price movements of assets carries practical implications for both practitioners and academics, such as trading cost analysis and optimal execution of trades. The impact of trades on asset prices, known as *price impact*, has been the focus of many studies and modeling efforts (Lillo *et al.* 2003, Cont *et al.* 2014). In a multi-asset setting, several studies have focused on the concept of **cross-impact**, which attempts to describe the impact of trading a given asset on the price of *other* assets (see Pasquariello and Vega 2015, Benzaquen *et al.* 2017, Capponi and Cont 2020).

Several studies have investigated *contemporaneous cross-impact* of order flow on returns by examining their cross-correlation structure. For example, Hasbrouck and Seppi (2001) revealed that commonality in returns among Dow 30 stocks is mostly attributed to order flow commonality. Tomas *et al.* (2022) built a principled approach to choosing a cross-impact model for various markets. Capponi and Cont (2020) showed that the positive covariance between returns of a specific stock and order flow imbalances of other stocks does not necessarily constitute evidence of cross-impact. They further demonstrated that, as long as the common factor in order flow imbalances is taken into account, adding cross-impact terms only marginally improves model performance, and thus may be disregarded. Our study complements Capponi and Cont (2020) in several ways: unlike Capponi and Cont (2020) which focuses on in-sample performance, we also consider the forecasting power of cross-order

*Corresponding author. Email: chao.zhang@stats.ox.ac.uk, chao.zhang@queens.ox.ac.uk

flow using both single and multi-level OFIs. To the best of our knowledge, there have been no studies that examine the influence of order flows on price movements in a multi-asset setting, while also taking into account the deeper levels in the limit order book (LOB).[†]

A more challenging problem than explaining contemporaneous returns is to examine the impact of trade orders on prices over future horizons, which has received a lot less attention in the literature, despite its important economic implications. Some studies have examined the relationship between order imbalances and *future daily* returns.[‡] Chordia *et al.* (2002) revealed that daily stock market returns are strongly related to contemporaneous and lagged order imbalances. Chordia and Subrahmanyam (2004) further found that there exists a positive relation between lagged order imbalances and daily individual stock returns. The authors also showed that imbalance-based trading strategies, i.e. buy if the previous day's imbalance is positive, and sell if the previous day's imbalance is negative, are able to yield statistically significant profits. Pasquariello and Vega (2015) provided empirical evidence of cross-asset informational effects in NYSE and NASDAQ stocks between 1993 and 2004, and demonstrated that the daily order flow imbalance in one stock, or across one industry, has a significant and persistent impact on daily returns of other stocks or industries. Rosenbaum and Tomas (2021) provided a characterization of the class of cross-impact kernels for a market that employs Hawkes processes to model trades and applied their method to two instruments from E-Mini Futures.

Given the recent progress in high-frequency trading (HFT), it is increasingly crucial to obtain accurate estimations of the cross-impact on *future intraday* returns. Benzaquen *et al.* (2017) introduced a multivariate linear model (see Kyle 1985) to describe the structure of cross-impact and found that a significant fraction of the covariance of stock returns can be accounted for by this model. Wang *et al.* (2016a, 2018) empirically analyzed and discussed the impact of trading a specific stock on the average price change of the whole market or of individual sectors. Schneider and Lillo (2019) derived theoretical limits for the size and form of cross-impact and verified them on sovereign bonds data. However, when modeling cross-impact, these methods do not consider the possibility of high correlations between cross-asset order flows, which may result in overfitting issues. This is also evidenced by studies such as Benzaquen *et al.* (2017) and Tomas *et al.* (2022). Moreover, these studies mainly investigated the cross-impact coefficients for a fixed time period (i.e. in a static setting), ignoring the temporal dynamics of cross-impact.

In recent years, machine learning models including deep neural networks, have achieved substantial developments, leading to their applications in financial markets, especially for the task of modeling stock returns. For example,

[†] Xu *et al.* (2018) studied the contemporaneous price impact (not cross-impact) model by extending the model of Cont *et al.* (2014) to multi-level order flow imbalance.

[‡] Several studies, such as Hou (2007), Menzly and Ozbas (2010), Chincio *et al.* (2019), and Buccheri *et al.* (2021), investigated the lead-lag effect in equity returns across various assets, but did not take into account order flows.

Huck (2019) utilized state-of-the-art techniques, such as random forests, to construct a portfolio over a period of 22 years, and the results demonstrated the power of machine learning models to produce profitable trading signals. Krauss *et al.* (2017) applied a series of machine learning methods to forecast the probability of a stock outperforming the market index, and then constructed long-short portfolios from the predicted one-day-ahead trading signals. Gu *et al.* (2020) employed a set of machine learning methods to make one-month-ahead return forecasts, and demonstrated the potential of machine learning approaches in empirical asset pricing, due to their ability to handle nonlinear interactions. Ait-Sahalia *et al.* (2022) investigated the predictability of high-frequency stock returns and durations using LASSO and tree methods via many relevant predictors derived from returns and order flows. Tashiro *et al.* (2019) and Kolm *et al.* (2023) applied deep neural networks with LOB-based features to predict high-frequency returns. Nonetheless, to the best of our knowledge, cross-asset order flow imbalances have not been considered as predictors for forecasting future high-frequency returns in the literature, which is one of the main directions we explore in the second half of this paper.

1.1. Main contributions

The present study makes two main contributions to the literature regarding the *contemporaneous* and *predictive* cross-impact of order flow imbalances on price returns.

First, we revisit the significance of contemporaneous cross-impact by considering various definitions of order flow imbalance (OFI). Instead of only looking at the best-level orders, we systematically examine the impact of multi-level order flows in a cross-asset setting. Our results show that, once information from multi-level order flow is incorporated in the definition of order flow imbalance, cross-impact terms do not provide additional explanatory power for contemporaneous impact, compared to a parsimonious model without cross-impact. To the best of our knowledge, this is the first study to comprehensively analyze the relations between contemporaneous individual returns and multi-level orders in both single-asset and multi-asset settings.

Furthermore, we consider the associated forecasting problem and investigate the predictive power of the cross-asset order flows on future price returns. Our results suggest that cross-impact terms do provide significant information content for intraday forecasting of future returns over a short horizon of up to several minutes, but their predictability decays quickly through time.

1.2. Outline

Section 2 describes our dataset and defines the variables of interest. Section 3 discusses modeling of contemporaneous cross-impact. In Section 4, we first discuss the out-of-sample forecasting performance of cross-impact models over one-minute-ahead horizon from two perspectives: R^2 values and economic gains, and then examine the predictability over longer horizons. Finally, we conclude the analysis in Section 5 and highlight potential future research directions.

2. Data and variables

2.1. Data

We use the Nasdaq ITCH data from LOBSTER to compute the independent and dependent variables. Our data includes the top 100 components of S&P 500 index, existing from 2017-01-01 to 2019-12-31.[†]

Cont *et al.* (2014) found that over short time intervals, price changes are mainly driven by the Order Flow Imbalance (henceforth denoted as OFI). Kolm *et al.* (2023) also demonstrated that forecasting deep learning models trained on OFIs significantly outperform most models trained directly on order books or returns. Therefore, we adopt the OFIs as features in our below analysis.

During the interval $(t-h, t]$, we enumerate the observations of all order book updates by n . Given two consecutive order book states for a given stock i at $n-1$ and n , we compute the bid order flows ($\text{OF}_{i,n}^{m,b}$) and ask order flows ($\text{OF}_{i,n}^{m,a}$) of stock i at level m at time n as

$$\begin{aligned}\text{OF}_{i,n}^{m,b} &:= \begin{cases} q_{i,n}^{m,b}, & \text{if } P_{i,n}^{m,b} > P_{i,n-1}^{m,b}, \\ q_{i,n}^{m,b} - q_{i,n-1}^{m,b}, & \text{if } P_{i,n}^{m,b} = P_{i,n-1}^{m,b}, \\ -q_{i,n}^{m,b}, & \text{if } P_{i,n}^{m,b} < P_{i,n-1}^{m,b}, \end{cases} \\ \text{OF}_{i,n}^{m,a} &:= \begin{cases} -q_{i,n}^{m,a}, & \text{if } P_{i,n}^{m,a} > P_{i,n-1}^{m,a}, \\ q_{i,n}^{m,a} - q_{i,n-1}^{m,a}, & \text{if } P_{i,n}^{m,a} = P_{i,n-1}^{m,a}, \\ q_{i,n}^{m,a}, & \text{if } P_{i,n}^{m,a} < P_{i,n-1}^{m,a}, \end{cases}\end{aligned}$$

where, $P_{i,n}^{m,b}$ and $q_{i,n}^{m,b}$ denote the bid price and size (in number of shares) of stock i at level m , respectively. Similarly, $P_{i,n}^{m,a}$ and $q_{i,n}^{m,a}$ denote the ask price and ask size at level m , respectively. Note that the variable $\text{OF}_{i,t}^{m,b}$ is positive when (i) the bid price increase; (ii) the bid price remains the same and the bid size increases. $\text{OF}_{i,t}^{m,b}$ is negative when (i) the bid price decreases; (ii) the bid price remains the same and the bid size decreases. One can perform an analogous analysis and interpretation for the ask order flows $\text{OF}_{i,t}^{m,a}$.

2.1.1. Best-level OFI. It calculates the accumulative OFIs at the best bid/ask side during a given time interval (see Cont *et al.* 2014, Kolm *et al.* 2023), and is defined as[‡]

[†] We select the top 100 components based on their market capitalization as of the most recent market close at the time of our analysis, i.e. 2019-12-31.

[‡] In Cont *et al.* (2014), OFI was mathematically defined as $\text{OFI}_{i,t}^{1,h} = L_{i,h}^{1,b} - C_{i,h}^{1,b} - M_{i,h}^{1,b} - L_{i,h}^{1,a} + C_{i,h}^{1,a} - M_{i,h}^{1,a}$, where $L_{i,h}^{1,b}$ denotes the total size of buy orders that arrived to the current best bid during the time interval $(t-h, t]$; $C_{i,h}^{1,b}$ denotes the total size of buy orders that canceled from the current best bid during the time interval $(t-h, t]$; $M_{i,h}^{1,b}$ denotes the total size of marketable buy orders that arrived to current best ask during the time interval $(t-h, t]$. The quantities $L_{i,h}^{1,a}, C_{i,h}^{1,a}, M_{i,h}^{1,a}$ for sell orders are defined analogously. However, in the empirical study of Cont *et al.* (2014), the OFI was computed from fluctuations in best bid/ask prices and their sizes according to equation (1). The reason is that information about individual orders is not available in the data set. For better comparison, we employ the same formula, i.e. equation (1), to compute OFI.

$$\text{OFL}_{i,t}^{1,h} := \sum_{n=N(t-h)+1}^{N(t)} \text{OF}_{i,n}^{1,b} - \text{OF}_{i,n}^{1,a}, \quad (1)$$

where $N(t-h)+1$ and $N(t)$ are the indexes of the first and the last order book event in the interval $(t-h, t]$.

2.1.2. Deeper-level OFI. A natural extension of the best-level OFI defined in equation (1) is deeper-level OFI (see Xu *et al.* 2018, Kolm *et al.* 2023). We define OFI at level m ($m \geq 1$) as follows

$$\text{OFL}_{i,t}^{m,h} := \sum_{n=N(t-h)+1}^{N(t)} \text{OF}_{i,n}^{m,b} - \text{OF}_{i,n}^{m,a}. \quad (2)$$

Due to the intraday pattern in limit order depth, we use the average size to scale OFIs at the corresponding levels (consistent with Ahn *et al.* 2001, Harris and Panchapagesan 2005), and consider

$$\text{ofi}_{i,t}^{m,h} = \frac{\text{OFL}_{i,t}^{m,h}}{Q_{i,t}^{M,h}}, \quad (3)$$

where $Q_{i,t}^{M,h} = \frac{1}{M} \sum_{m=1}^M \frac{1}{2\Delta N(t)} \sum_{n=N(t-h)+1}^{N(t)} [q_{i,n}^{m,b} + q_{i,n}^{m,a}]$ is the average order book depth across the first M levels and $\Delta N(t) = N(t) - N(t-h)$ is the number of events during $(t-h, t]$. In this paper, we consider the top $M = 10$ levels of LOB and denote the multi-level OFI vector as $\text{ofi}_{i,t}^{(h)} = (\text{ofi}_{i,t}^{1,h}, \dots, \text{ofi}_{i,t}^{10,h})^T$.

2.1.3. Integrated OFI. Our following analysis in Section 2.2 will show that there exist strong correlations between multi-level OFIs, and that the first principal component can explain over 89% of the total variance among multi-level OFIs. In order to make use of the information embedded in multiple LOB levels and avoid overfitting, we propose an integrated version of OFIs via Principal Components Analysis (PCA) as shown in equation (4), which only preserves the first principal component.[§] We further normalize the first principal component by dividing by its l_1 norm so that the weights of multi-level OFIs in constructing integrated OFIs sum to 1, leading to

$$\text{ofi}_{i,t}^{I,h} = \frac{\mathbf{w}_1^T \text{ofi}_{i,t}^{(h)}}{\|\mathbf{w}_1\|_1}, \quad (4)$$

where \mathbf{w}_1 is the first principal vector computed from historical data. To the best of our knowledge, this is the first work to aggregate multi-level OFIs into a single variable.

2.1.4. Logarithmic returns. Our dependent variable is the logarithmic asset return. Specifically, we define the returns

[§] We would like to thank an anonymous reviewer for suggesting an analysis of various aggregations of multi-level OFIs, as detailed in Appendix 1.

Table 1. Summary statistics of OFIs and returns.

	Mean (bp)	Std (bp)	Skewness	Kurtosis	10% (bp)	25% (bp)	50% (bp)	75% (bp)	90% (bp)
ofi ^{1,(1m)}	-0.01	6.26	-0.04	1.89	-7.97	-3.45	0.03	3.47	7.90
ofi ^{2,(1m)}	0.01	6.86	-0.04	1.04	-8.86	-3.88	0.02	3.95	8.85
ofi ^{3,(1m)}	-0.01	7.05	-0.04	0.71	-9.26	-4.08	0.01	4.11	9.19
ofi ^{4,(1m)}	-0.02	7.22	-0.05	0.68	-9.50	-4.21	0.01	4.24	9.40
ofi ^{5,(1m)}	-0.03	7.14	-0.05	0.79	-9.38	-4.14	0.01	4.15	9.25
ofi ^{6,(1m)}	-0.03	6.87	-0.04	0.96	-8.98	-3.94	0.01	3.95	8.85
ofi ^{7,(1m)}	-0.03	6.39	-0.05	1.29	-8.31	-3.59	0.01	3.59	8.16
ofi ^{8,(1m)}	-0.03	6.03	-0.05	1.59	-7.80	-3.37	0.01	3.36	7.66
ofi ^{9,(1m)}	-0.05	5.71	-0.05	1.96	-7.38	-3.18	0.01	3.14	7.19
ofi ^{10,(1m)}	-0.05	5.38	-0.05	2.52	-6.92	-2.97	0.01	2.91	6.74
ofi ^{I,(1m)}	0.01	6.53	-0.05	0.76	-8.52	-3.81	0.05	3.89	8.47
$r^{(1m)}$	0.02	4.81	-0.04	1.85	-6.22	-2.71	0.00	2.79	6.23

Note: These statistics are computed at the minute level across each stock and the full sample period. 1bp = 0.0001 = 0.01%.

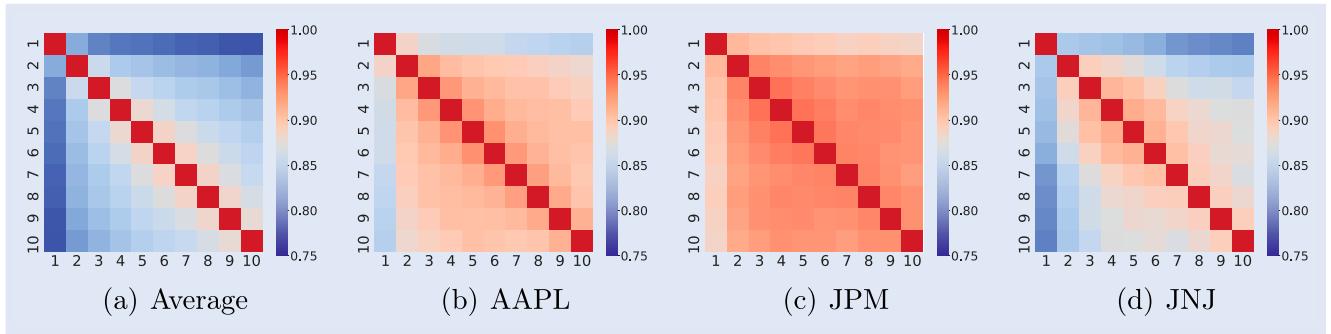


Figure 1. Correlation matrix of multi-level OFIs. (a) Average. (b) AAPL. (c) JPM and (d) JNJ.

Note: Plot (a) is averaged across each stock and each trading day, Plots (b)-(d): correlation matrix of Apple (AAPL), JPMorgan Chase (JPM), and Johnson & Johnson (JNJ) averaged across each trading day. The x-axis and y-axis represent different levels of OFIs.

over the interval $(t - h, t]$ as follows:

$$r_{i,t}^{(h)} = \log \left(\frac{P_{i,t}}{P_{i,t-h}} \right), \quad (5)$$

where $P_{i,t}$ is the mid-price at time t , i.e. $P_{i,t} = \frac{P_{i,t}^{1,b} + P_{i,t}^{1,a}}{2}$.

2.2. Summary statistics

Table 1 presents the summary statistics of multi-level OFIs, integrated OFIs, and returns for the top 100 components of S&P 500 index. These descriptive statistics (e.g. mean, std, etc) are computed at the minute level and aggregated across trading days and stocks.

Figure 1 reveals that even though the correlation structure of multi-level OFIs may vary across stocks, they all show strong relationships (above 75%). It is worth pointing out that the best-level OFI exhibits the smallest correlation with any of the remaining nine levels, a pattern that persists across different stocks. Table 2 further reveals that the first principal component explains more than 89% of the total variance.

In figure 2, we show statistics pertaining to the weights attributed to the top 10 levels in the first principal component. Plot figure 2(a) shows the average weights, and the one standard deviation bars, across all stocks in the universe. Plot figure 2(a) reveals that the best-level OFI has the smallest weight in the first principal component, but the

highest standard deviation, hinting that it fluctuates significantly across stocks. Plots (b-d) show various patterns for the first principal component of multi-level OFIs, for each quantile bucket of various stock characteristics, in particular, for volume, volatility and spread. For instance, in figure 2(b), the red curve shows the average weights in the first principal component for each of the 10 levels, where the average is taken over all the top 25% largest volume stocks. A striking pattern that emerges from this figure is that for *high-volume* (red line in figure 2(b)), and *low-volatility stocks* (blue line in figure 2(c)), OFIs deeper in the LOB receive more weights in the first component. However, for *low-volume* (blue line in figure 2(b)), and *large-spread stocks* (red line in figure 2(d)), the best-level OFIs account more than the deeper-level OFIs.

3. Contemporaneous cross-impact

In this section, we study the existence of contemporaneous cross-impact by comparing it with the price impact model studied in Cont *et al.* (2014).

3.1. Models

3.1.1. Price impact of best-level OFIs. We first pay attention to the price impact of best-level OFI ($ofi_{i,t}^{1,h}$) on contemporaneous returns ($r_{i,t}^{(h)}$) that materialize over the same

Table 2. Average percentage and the standard deviation (in parentheses) of variance attributed to each principal component.

Principal Component	1	2	3	4	5	6	7	8	9	10
Explained Variance Ratio	89.06 (6.12)	4.99 (3.52)	2.28 (1.26)	1.28 (0.74)	0.80 (0.48)	0.54 (0.34)	0.39 (0.25)	0.29 (0.19)	0.21 (0.15)	0.15 (0.11)

Note: The table reports the ratio (in percentage points) between the variance of each principal component and the total variance averaged across each stock and trading day.

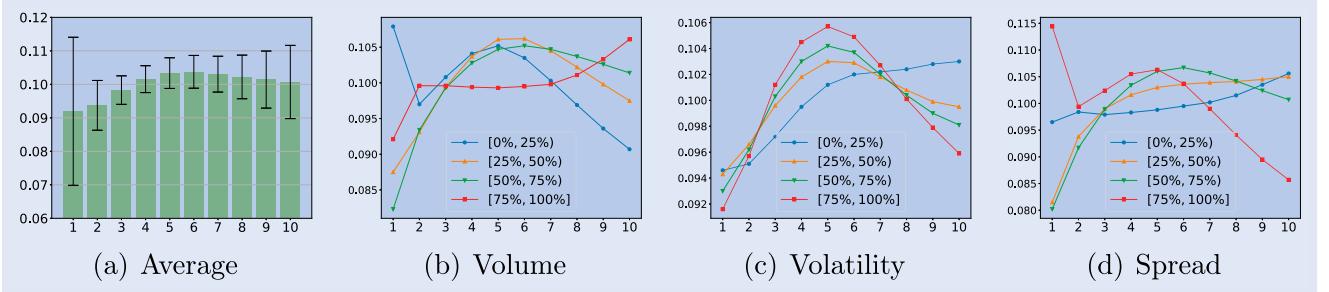


Figure 2. First principal component of multi-level OFIs, in quantile buckets for various stock characteristics. (a) Average. (b) Volume. (c) Volatility and (d) Spread.

Note: The x -axis indexes the top 10 levels of the OFIs. Volume: trading volume on the previous trading day. Volatility: volatility of 1-minute returns during the previous trading day. Spread: average bid-ask spread during the previous trading day. [0%, 25%], respectively [75%, 100%], denote the subset of stocks with the lowest, respectively highest, 25% values for a given stock characteristic.

time bucket as the OFI, via the model

$$\mathbf{PI}^{[1]} : r_{i,t}^{(h)} = \alpha_i^{[1]} + \beta_i^{[1]} \text{ofi}_{i,t}^{1,h} + \epsilon_{i,t}^{[1]}. \quad (6)$$

Here, $\alpha_i^{[1]}$ and $\beta_i^{[1]}$ are the intercept and slope coefficients, respectively. $\epsilon_{i,t}^{[1]}$ is a noise term summarizing the influences of other factors, such as the OFIs at even deeper levels, and potentially the trading behaviors of other stocks. For the sake of simplicity, we refer to the above regression model as $\mathbf{PI}^{[1]}$ and use ordinary least squares (OLS) to estimate it.

3.1.2. Price impact of integrated OFIs. The second model specification takes into account the impact of multi-level OFIs by leveraging the integrated OFIs, which we set up as follows and use OLS for estimation.

$$\mathbf{PI}^I : r_{i,t}^{(h)} = \alpha_i^I + \beta_i^I \text{ofi}_{i,t}^{I,h} + \epsilon_{i,t}^I. \quad (7)$$

3.1.3. Cross-impact of best-level OFIs. Assuming there are N stocks in the studied universe, we incorporate the multi-asset best-level OFIs, $\text{ofi}_{j,t}^{1,h}$ ($j = 1, \dots, N$), as candidate features to help fit the returns of the i th stock $r_{i,t}^{(h)}$. For simplicity, we denote the impact from itself (stock i) as *Self* and that from other stocks as *Cross*, as shown below,

$$\mathbf{CI}^{[1]} : r_{i,t}^{(h)} = \alpha_i^{[1]} + \underbrace{\beta_{i,i}^{[1]} \text{ofi}_{i,t}^{1,h}}_{\text{Self}} + \sum_{j \neq i} \underbrace{\beta_{i,j}^{[1]} \text{ofi}_{j,t}^{1,h}}_{\text{Cross}} + \eta_{i,t}^{[1]}. \quad (8)$$

Therefore, $\beta_{i,j}^{[1]}$ represents the influence of the j th stock's best-level OFIs on the returns of stock i .

3.1.4. Cross-impact of integrated OFIs. Finally, we incorporate the cross-asset integrated OFIs to explore the impact of

multi-level OFIs from other assets, resulting in the following \mathbf{CI}^I model,

$$\mathbf{CI}^I : r_{i,t}^{(h)} = \alpha_i^I + \underbrace{\beta_{i,i}^I \text{ofi}_{i,t}^{I,h}}_{\text{Self}} + \sum_{j \neq i} \underbrace{\beta_{i,j}^I \text{ofi}_{j,t}^{I,h}}_{\text{Cross}} + \eta_{i,t}^I. \quad (9)$$

3.1.5. Sparsity of cross-impact. As we are aware, OLS regression becomes ill-posed when there are fewer observations than parameters. Recall that we are now considering $N \approx 100$ independent variables in equations (8) and (9). Assuming the time interval is one minute and we are interested in estimating the intraday cross-impact models, e.g. relying on the 30-min estimation window and 1-min returns (as in Cont *et al.* 2014), then it seems inappropriate to estimate $\mathbf{CI}^{[1]}$ and \mathbf{CI}^I for intraday scenarios using the OLS regression with more variables than observations. Moreover, the multicollinearity issue of features contradicts the necessary condition for a well-posed OLS. As displayed in figure 3, a significant portion of the cross-asset correlations based on the best-level OFIs cannot be ignored. For example, approximately 10% of correlations are larger than 0.30. Last, Capponi and Cont (2020) found that a certain number of cross-impact coefficients from their OLS regressions are not statistically significant at the 1% significance level.

With all the above considerations in mind, we assume that there is a small number of assets having a significant impact on a specific stock i , as opposed to the entire universe, in $\mathbf{CI}^{[1]}$ and \mathbf{CI}^I . To this end, we apply the Least Absolute Shrinkage and Selection Operator (LASSO)[†] to solve equations (8)

[†]LASSO is a regression method that performs both variable selection and regularization, in order to enhance the prediction accuracy and interpretability of regression models (see more in Hastie *et al.* 2009). It can be formulated as a linear regression model and the objective function consists of two parts, i.e. the sum of squared residuals, and the l_1 constraint on the regression coefficients. In this

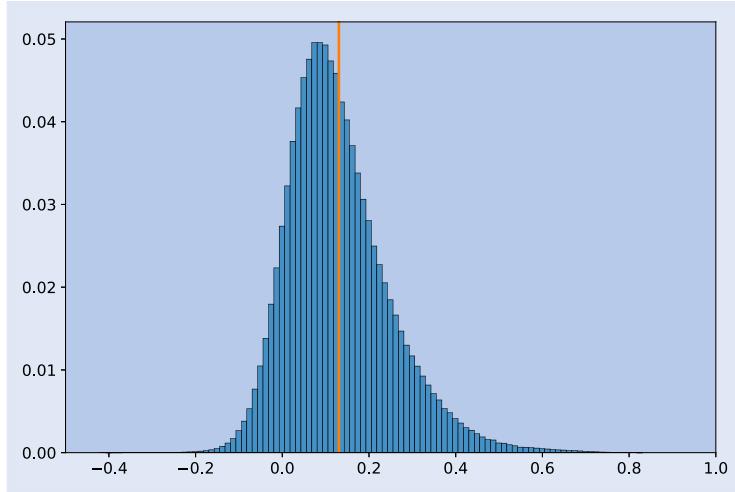


Figure 3. Distribution of correlations based on the best-level OFIs.

Note: The orange vertical line represents the average correlation.

and (9). The sparsity of cross-impact terms also facilitates the explanation of coefficients. Note that even though the sparsity of the cross-impact terms is not theoretically guaranteed, our empirical evidence confirms this modeling assumption.

3.2. Empirical results

For a more representative and fair comparison with previous studies, we apply a similar procedure described in Cont *et al.* (2014) to our experiments. We exclude the first and last 30 minutes of the trading day due to the increased volatility near the opening and closing sessions, in line with Hasbrouck and Saar (2002), Chordia *et al.* (2002) and Chordia and Subrahmanyam (2004), Cont *et al.* (2014), Capponi and Cont (2020). In particular, we use each non-overlapping 30-minute estimation window during the intraday time interval 10:00 am–3:30 pm to estimate the regressions, namely equations (6)–(9). Within each window, returns and OFIs are computed for every minute.

3.2.1. In-sample performance. We first measure the model performance via in-sample adjusted- R^2 , denoted as the **in-sample R^2** or **IS R^2** . From table 3, we first observe that $\mathbf{PI}^{[1]}$ can explain 71.16% of the in-sample variation of a stock's contemporaneous returns, consistent with the findings of Cont *et al.* (2014). Meanwhile, \mathbf{PI}^I displays higher and more consistent explanation power, with an average adjusted R^2 value of 87.14% and a standard deviation of 9.16%, indicating the effectiveness of our integrated OFIs.[†]

Table 3 also shows that the in-sample R^2 values increase as cross-asset OFIs are included as additional features, which is not surprising given that $\mathbf{PI}^{[1]}$ (respectively, \mathbf{PI}^I) is a nested model of $\mathbf{CI}^{[1]}$ (respectively, \mathbf{CI}^I). However, the increments of the in-sample R^2 are smaller when using integrated OFIs.

work, we employ the cross-validation to choose the l_1 penalty weight for each regression.

[†] We also investigate price impact with multi-level OFIs in Appendix 2. The results demonstrate that the price impact model using integrated OFIs outperforms those using multi-level OFIs in out-of-sample tests.

Table 3. In-sample performance for contemporaneous returns.

	Best-level OFIs		Integrated OFIs	
	$\mathbf{PI}^{[1]}$	$\mathbf{CI}^{[1]}$	\mathbf{PI}^I	\mathbf{CI}^I
IS R^2	71.16 (13.80)	73.87 (12.23)	87.14 (9.16)	87.85 (8.58)

Note: The table reports the mean values and standard deviations (in parentheses) of in-sample R^2 (in percentage points) of various models when modeling contemporaneous returns. The models include $\mathbf{PI}^{[1]}$ (equation (6)), $\mathbf{CI}^{[1]}$ (equation (8)), \mathbf{PI}^I (equation (7)), and \mathbf{CI}^I (equation (9)). These statistics are averaged across each stock and each regression window.

(87.85%–87.14% = 0.71%), compared to the counterpart using best-level OFIs (73.87%–71.16% = 2.71%). This indicates that cross-asset multi-level OFIs may not provide additional information on the variance in returns compared to the price impact model with integrated OFIs.

Next, we take a closer look at the cross-impact coefficients based on either the best-level or integrated OFIs, i.e. $\beta_{i,j}^{[1]}$ and $\beta_{i,j}^I$ ($i, j = 1, \dots, N$). Table 4 reveals the frequency of self-impact and cross-impact variables selected by LASSO, i.e. the frequency of $\beta_{i,j}^{[1]} \neq 0$ (respectively, $\beta_{i,j}^I \neq 0$). We observe that self-impact variables are consistently chosen in both $\mathbf{CI}^{[1]}$ and \mathbf{CI}^I , as found in Cont *et al.* (2014). However, another interesting observation is that the frequency of a cross-asset integrated OFI variable selected by \mathbf{CI}^I is around 1/2 of its counterpart in $\mathbf{CI}^{[1]}$. When we turn to the size of the average regression coefficients as shown in table 4, we obtain reasonably consistent results. The self-impact is much higher than the cross-impact in both the $\mathbf{CI}^{[1]}$ and \mathbf{CI}^I models, while the cross-impact coefficients in \mathbf{CI}^I are about 1/3 in scale of their counterparts in $\mathbf{CI}^{[1]}$. This difference in scale may suggest that the cross-impact terms are less important in the \mathbf{CI}^I model, however, it is worth noting that even small cross-term coefficients can have a non-negligible effect when aggregated at the portfolio level.

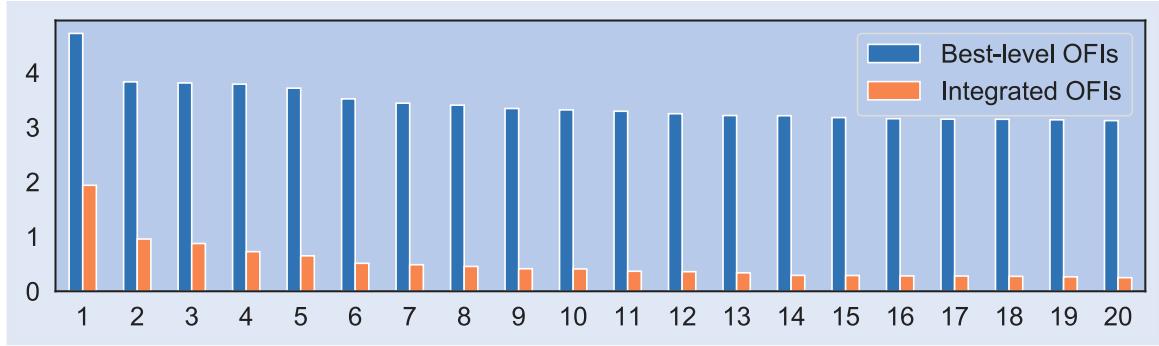


Figure 4. Barplot of singular values for the coefficient matrix in contemporaneous cross-impact models.

Note: We perform Singular Value Decomposition (SVD) on the coefficient matrix to obtain the singular values. Singular values are in descending order and the coefficients are averaged over each regression window between 2017–2019. The x-axis represents the singular value rank, and the y-axis represents the singular values.

Table 4. Summary statistics of coefficients in the cross-impact models $\mathbf{CI}^{[1]}$ and \mathbf{CI}^I .

	Frequency (%)		Magnitude	
	$\mathbf{CI}^{[1]}$	\mathbf{CI}^I	$\mathbf{CI}^{[1]}$	\mathbf{CI}^I
<i>Self</i>	99.85 (0.34)	99.96 (0.18)	1.02 (0.31)	1.24 (0.34)
<i>Cross</i>	17.34 (2.78)	8.29 (2.56)	$4.5e^{-3}$ ($1.3e^{-3}$)	$1.6e^{-3}$ ($0.7e^{-3}$)

Note: The table is calculated over each stock and each regression window. The first two columns describe the frequency of *Self* and *Cross* variables chosen by the corresponding model with a standard deviation (in parentheses); The last two columns describe the magnitude of *Self* and *Cross* coefficients in the corresponding model with a standard deviation (in parentheses).

In addition, cross-impact being large/small is a statement about a matrix, more related to its singular values and relative magnitudes, rather than the individual value of the coefficients. Figure 4 shows a comparison of the top 20 singular values of the coefficient matrices given by the best-level and integrated OFIs.[†] The relatively large singular values of the best-level OFI matrix are a consequence of the higher edge density, and thus average degree, of the network. Note that both networks exhibit a large top singular value of the adjacency matrix (akin to the usual *market mode* in Laloux *et al.* 2000), and the integrated OFI network has a faster decay of the spectrum, thus revealing its low-rank structure.

We visualize a network for each coefficient matrix, which only preserves the edges larger than a given threshold (following Kenett *et al.* 2010, Curme *et al.* 2015), as shown in figure 5. We color stocks according to the GICS sector division, and sort them by their market capitalization within each sector.[‡] As one can see from figure 5(a), the cross-impact coefficient matrix $[\beta_{i,j}^{[1]}]_{i \neq j}$ displays a sectorial

structure, in accordance with previous studies (e.g. Benzaquen *et al.* 2017). This behavior could be fueled by index arbitrage strategies, where traders may, for example, trade an entire basket of stocks coming from the same sector against an index.

Figure 5(b) presents the network of cross-impact coefficients based on integrated OFIs, i.e. $[\beta_{i,j}^I]_{i \neq j}$. Compared with figure 5(a), the connections in figure 5(b) are much weaker, implying that the cross-impact from stocks can be potentially explained by a stock's own multi-level OFIs, to a large extent. Note that there is only one connection from GOOGL to GOOG, as pointed out at the top of figure 5(b). This stems from the fact that both stock ticker symbols pertain to Alphabet (Google). Our study also reveals that OFIs of GOOGL have more influence on the returns of GOOG, not the other way around. The main reason might be that GOOGL shares have voting rights, while GOOG shares do not.

In figures 5(c,d), we set lower threshold values (75th, respectively, 25th percentile of coefficients) in order to promote more edges in the networks based on integrated OFIs. Interestingly, we observe only four connections in figure 5(c). Except from bidirectional links between GOOGL and GOOG, there exists a one-way link from Cigna (CI) to Anthem (ANTM), and another one-way link from Duke Energy (DUK) to NextEra Energy (NEE). Anthem announced to acquire Cigna in 2015. After a prolonged breakup, this merger finally failed in 2020. Therefore, it is unsurprising that the OFIs of Cigna can affect the price movements of Anthem. Conversely, Anthem's OFIs also have an impact on the price movements of Cigna, but to a lesser extent. Further research should be undertaken to investigate this phenomenon. In terms of the second pair, Duke Energy rebuffed NextEra's acquisition interest in 2020. Note that 2020 is not in our sample period. This finding hints that certain market participants may have noticed the special relationship between Duke Energy and NextEra Energy before this mega-merger was proposed.

[†] Here we only use the off-diagonal elements, i.e. $[\beta_{i,j}^{[1]}]_{i \neq j}$ and $[\beta_{i,j}^I]_{i \neq j}$.

[‡] The Global Industry Classification Standard (GICS) is an industry taxonomy developed in 1999 by MSCI and Standard & Poor's (S&P) for use by the global financial community.

3.2.2. Out-of-sample performance. Although the in-sample estimation yields interesting findings, practitioners are eventually concerned about the out-of-sample estimation. Therefore, we propose to perform the following out-of-sample tests.

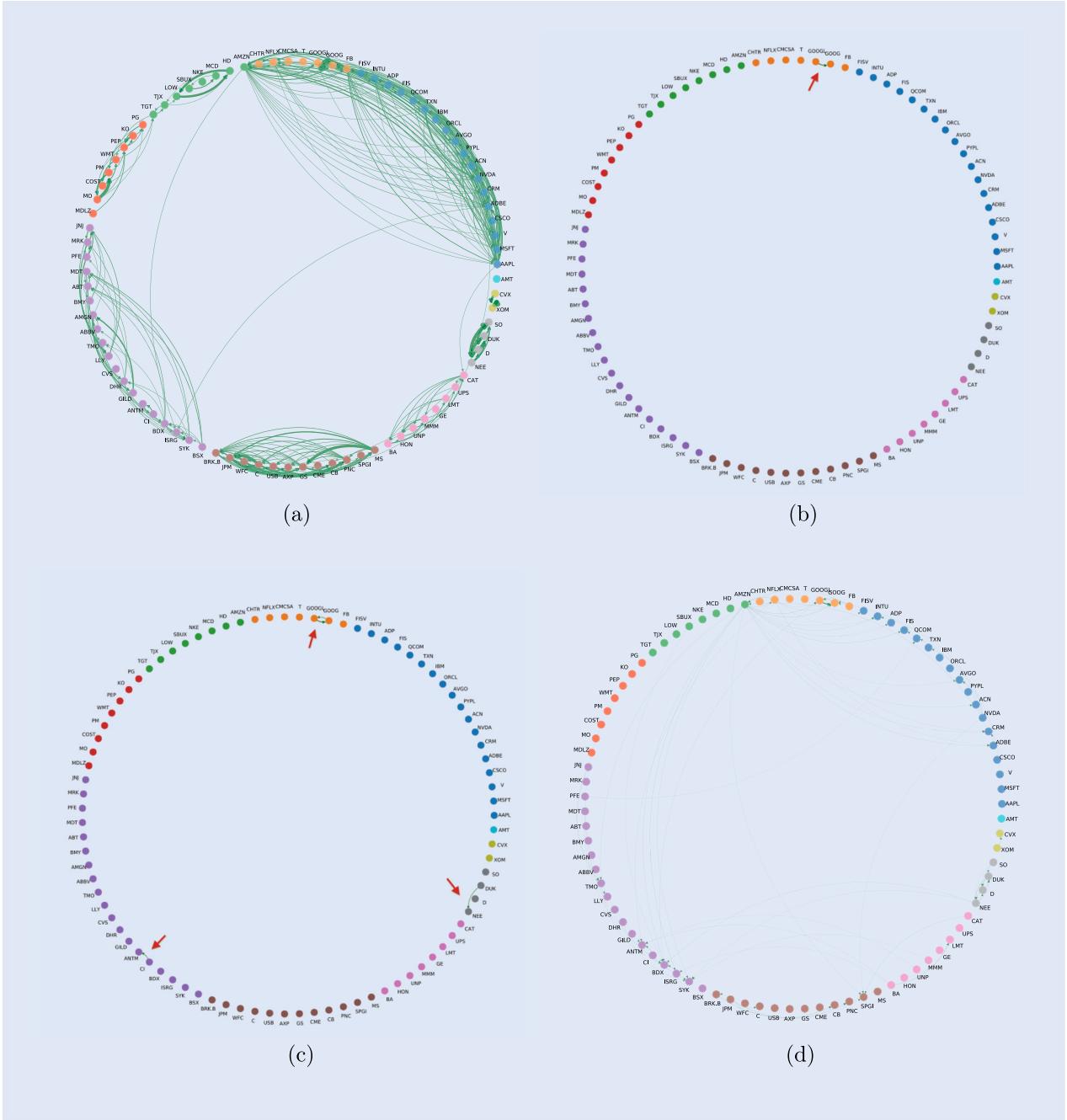


Figure 5. Illustrations of the coefficient networks constructed from contemporaneous cross-impact models. (a) Threshold = 95th percentile, based on best-level OFIs. (b) Threshold = 95th percentile, based on integrated OFIs. (c) Threshold = 75th percentile, based on integrated OFIs and (d) Threshold = 25th percentile, based on integrated OFIs.

Note: To render the networks more interpretable and for ease of visualization, we only plot the top 5% largest (a-b), or top 25% largest (c), or top 75% largest (d), in magnitude coefficients. The coefficients are averaged over each regression window between 2017–2019. Nodes are colored by the GICS structure and sorted by market capitalization. Green links represent positive values while black links represent negative values. The width of edges is proportional to the absolute values of their respective coefficients.

We use the above-fitted models to estimate returns on the following 30-minute data and compute the corresponding R^2 , denoted as **out-of-sample R^2** or **OS R^2** .[†]

[†] Previous studies either investigated the in-sample R^2 (including Cont *et al.* 2014, Capponi and Cont 2020), or adopted a cross-validation method (e.g. Xu *et al.* 2018). However, these works failed to consider the generalization error of their models or damaged the chronological order of the time-series data. In contrast, we obey the

Table 5 reports the average values and their standard deviations of out-of-sample R^2 of $\mathbf{PI}^{[1]}$, $\mathbf{CI}^{[1]}$, \mathbf{PI}^I , and \mathbf{CI}^I . We first focus on the models using best-level OFIs. It appears $\mathbf{CI}^{[1]}$ has a slight advantage compared with $\mathbf{PI}^{[1]}$ for out-of-sample tests

temporal ordering in our study. These matters are vital to practitioners, as only the historical data are accessible for the model fit in practice.

Table 5. Out-of-sample performance for contemporaneous returns.

	Best-level OFIs		Integrated OFIs	
	$\text{PI}^{[1]}$	$\text{CI}^{[1]}$	PI^I	CI^I
OS R^2	64.64 (21.82)	66.03 (19.51)	83.83 (16.90)	83.62 (14.53)

Note: The table reports the mean values and standard deviations (in parentheses) of out-of-sample R^2 (in percentage points) of various models when modeling contemporaneous returns. The models include $\text{PI}^{[1]}$ (equation (6)), $\text{CI}^{[1]}$ (equation (8)), PI^I (equation (7)), and CI^I (equation (9)). These statistics are averaged across each stock and each regression window.

with an improvement of 1.39% (=66.03%–64.64%). However, when involving multi-level or integrated OFIs, the performance of CI^I is slightly worse than PI^I , indicating that the cross-impact model with integrated OFIs cannot provide extra explanatory power to the price impact model with integrated OFIs. Overall, we observe that the models using integrated OFIs unveil significant and consistent improvements over those using only best-level OFIs.

In general, we observe strong evidence implying $\text{CI}^{[1]}$ provides a better out-of-sample estimate than $\text{PI}^{[1]}$, while for PI^I and CI^I , the evidence is opposite. However, it is important to note that these conclusions are based on a point estimate and do not necessarily indicate statistical significance. Therefore, we perform the following hypothesis test for each stock on the out-of-sample data to assess statistical significance,

$$\begin{aligned}\mathcal{H}_0 : \mathbb{E}[R_{\text{OS}}^2(\text{CI}^{[1]}) - R_{\text{OS}}^2(\text{PI}^{[1]})] &\leq 0 \text{ vs.} \\ \mathcal{H}_1 : \mathbb{E}[R_{\text{OS}}^2(\text{CI}^{[1]}) - R_{\text{OS}}^2(\text{PI}^{[1]})] &> 0.\end{aligned}$$

We employ the approach from Giacomini and White (2006) and Chinco *et al.* (2019) to assess statistical significance through a Wald-type test (see Ward and Ahlquist 2018). Theorem 1 in Giacomini and White (2006) implies that we can use a standard t -test to evaluate the statistical significance of changes in R^2 . A p -value less than a given significance level α rejects the null hypothesis in favor of the alternative at the $1 - \alpha$ confidence level, implying $\text{CI}^{[1]}$ has significantly better estimation than $\text{PI}^{[1]}$. We also implement this test for the comparison between PI^I and CI^I .

Figure 6 illustrates the main results from the above hypothesis tests. When using only the best-level OFIs, the cross-impact model is superior to the price impact model for 91.0% (94.4%) of stocks, at the 1% (5%) confidence level. However, when examining the models using integrated OFIs, we reject the null hypothesis (i.e. in favor of the cross-impact model) only for 28.1% (33.7%) of stocks at the 1% (5%) confidence level. As expected, cross-impact terms can significantly improve the explanatory power of the price impact model for GOOG and GOOGL.

Dynamics of limit order book may depend on the tick-to-price ratio, or alternatively, the fraction of time that the bid-ask spread is equal to one tick for a given stock (Curato and Lillo 2015). We examine whether this dependence also

Table 6. Out-of-sample R^2 of various contemporaneous models sorted by tick-to-price ratio.

	[0%, 25%)	[25%, 50%)	[50%, 75%)	[75%, 100%)
$\text{PI}^{[1]}$	44.38	62.51	77.55	70.70
$\text{CI}^{[1]}$	53.01	66.32	72.50	78.34
PI^I	68.14	84.58	88.14	89.86
CI^I	72.01	84.80	88.51	91.01

Note: [0%, 25%), respectively [75%, 100%], denote the subset of stocks with the lowest, respectively highest, 25% values according to the tick-to-price ratio.

extends to cross-asset OFIs.[†] Our findings, presented in table 6, suggest that cross-asset OFIs can better explain the price dynamics of stocks with a larger tick-to-price ratio.

3.3. Discussion about contemporaneous cross-impact

3.3.1. Impact on stocks. In summary, our previous results mainly show that when considering only the best-level OFI of a single stock, the addition of the best-level OFI from other stocks slightly increases the explanatory power. On the other hand, when the information from multiple levels is integrated into the OFI, the improvement is negligible. In the meantime, it is unsurprising that taking into account more levels in the LOB (PI^I) could better explain price changes, compared to only considering best-level orders ($\text{PI}^{[1]}$).

After observing these results, several natural questions may arise: How can the above facts be reconciled?[‡] How do the cross-asset best-level OFIs interact with the multi-level OFIs, when modeling contemporaneous returns?

To address these questions, we consider the following scenario, also depicted in figure 7. For simplicity, we denote the order from trading strategy A on stock i (respectively, j) as A_i (respectively, A_j). Analogously, we define orders from strategy B and S . Let us next consider the orders of stock i . There are three orders from different portfolios, given by A_i , B_i and S_i . A_i is at the third bid level of stock i and linked to an order at the best ask level of stock j , i.e. A_j . Also, B_i is at the best ask level of stock i and linked to an order at the best bid level of stock j , i.e. B_j . Finally, S_i is an individual bid order at the best level of stock i .

We observe that the best-level limit orders from stock j may be linked to price movements of stock i through paths $B_j \rightarrow B_i \rightarrow \text{ofi}_i^1 \rightarrow r_i$ and $A_j \rightarrow A_i \rightarrow \text{ofi}_i^3 \rightarrow r_i$. Thus the price impact model for stock i which only utilizes its own best-level orders will ignore the information of A_i , while the cross-impact model can partially collect it along the path $A_j \rightarrow A_i$. This might illustrate why the best-level OFIs of multiple

[†] We would like to thank an anonymous reviewer for suggesting this analysis.

[‡] One possible explanation for those facts is that the duration of the cross-impact terms might be shorter than the current time interval (30 minutes) used in our experiments, rendering the cross-impact terms vanish in out-of-sample tests. To verify this assertion, we implement additional experiments where models are updated more frequently. The results (deferred to Appendix 4) reveal that even under higher-frequency updates (1-min) of the models, there is no benefit from introducing cross-impact terms to the price impact model with integrated OFIs.

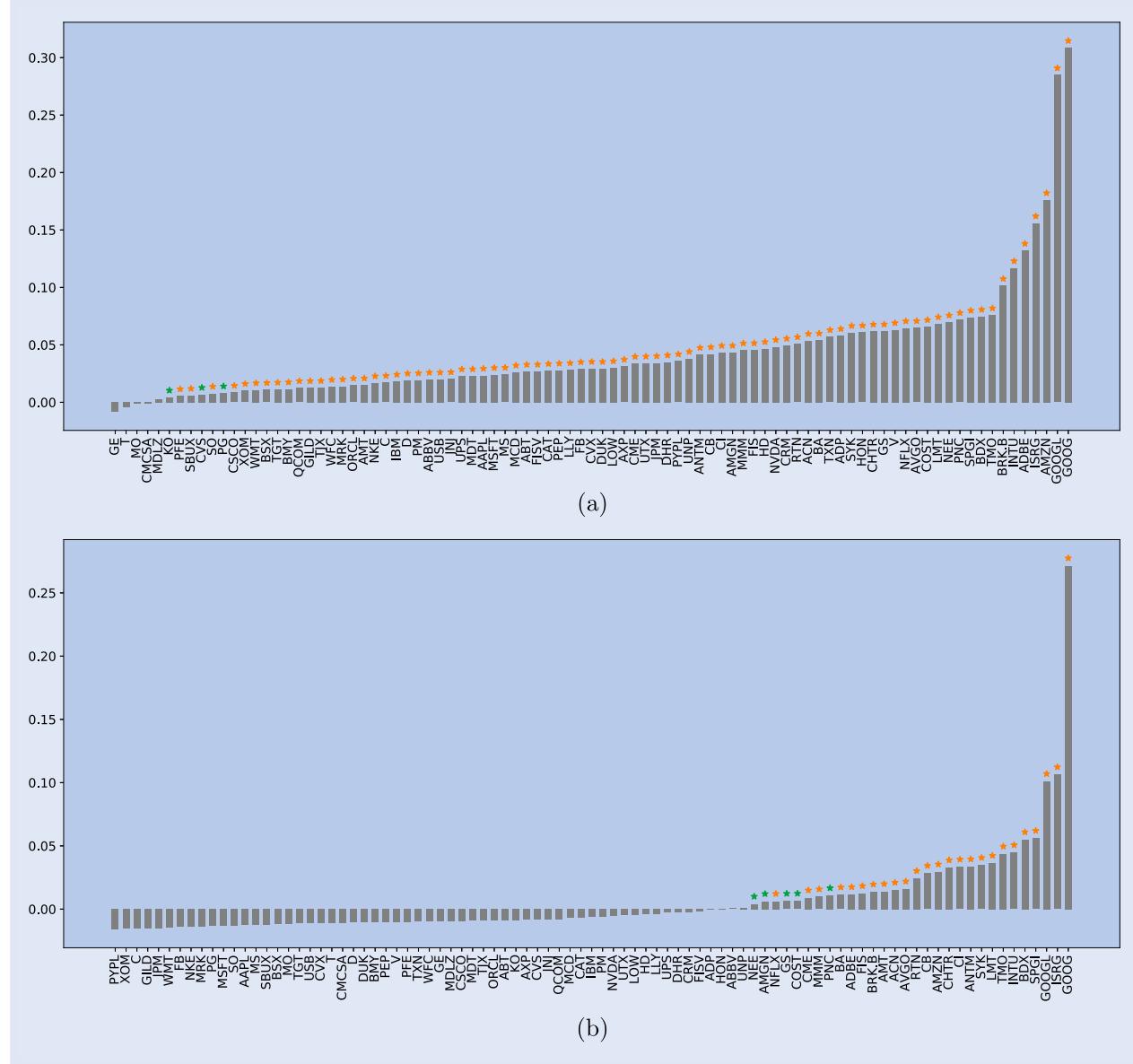


Figure 6. Mean differences of out-of-sample R^2 between CI and PI models. (a) $R^2_{OS}(CI^{[1]}) - R^2_{OS}(PI^{[1]})$ and (b) $R^2_{OS}(CI^I) - R^2_{OS}(PI^I)$.

Note: A positive (negative) number indicates superiority for the CI (PI) model. The y-axis represents the average difference of OS R^2 between CI and PI, while the x-axis lists the stock symbols. Stars indicate the p -values, with orange, green, and blue representing significance at the 1%, 5%, and 10% levels, respectively.

assets can provide slightly additional explanatory power to the price impact model using only the best-level OFIs.

Nonetheless, if we can integrate multi-level OFIs in an efficient way (in our example, aggregate order imbalances caused by orders A_i , B_i and S_i), then there is no need to consider OFIs from other stocks for modeling price dynamics. For example, information hidden in the path $A_j \rightarrow A_i \rightarrow \text{ofi}_i^3 \rightarrow r_i$ can be leveraged as long as A_i is well absorbed into new integrated OFIs. In this sense, for stock i , cross-asset best-level OFIs (including A_j) are surrogates of its own OFIs at different levels (here A_i), to a certain extent. The likelihood of this relationship is attributed to massive portfolio trades that submit or cancel limit orders across a variety of assets at different levels.[†] We put forward this mechanism which potentially

explains why the cross-impact model with integrated OFIs cannot provide additional explanatory power compared to the price impact model with integrated OFIs.[‡]

3.3.2. Impact on portfolios. A related question is about the aggregation of cross-impact at the portfolio level. Let us consider the OFI of portfolio p as $\text{ofi}_{p,t}^{1,h} := \sum_{i=1}^N w_i \text{ofi}_{i,t}^{1,h}$, where w_i is the weight of asset i in a portfolio. Then the price impact for portfolio p is

$$r_{p,t}^{(h)} = a_p^{[1]} + \beta_p^{[1]} \text{ofi}_{p,t}^{1,h} + e_{p,t}^{[1]}$$

[†]Cao *et al.* (2009), Hautsch and Huang (2012), Sirignano (2019), and Chakrabarty *et al.* (2022) showed that the depth of some deeper levels (such 2–3) is higher than the best level depth.

[‡]It would be a very interesting research direction to derive testable predictions for this mechanism in future work. However, so far it hinges on the availability of client-ID based LOB data, i.e. knowing that the same market participant is behind the orders for two related instruments, released at approximately the same time, such as (A_i, A_j) , (B_i, B_j) .

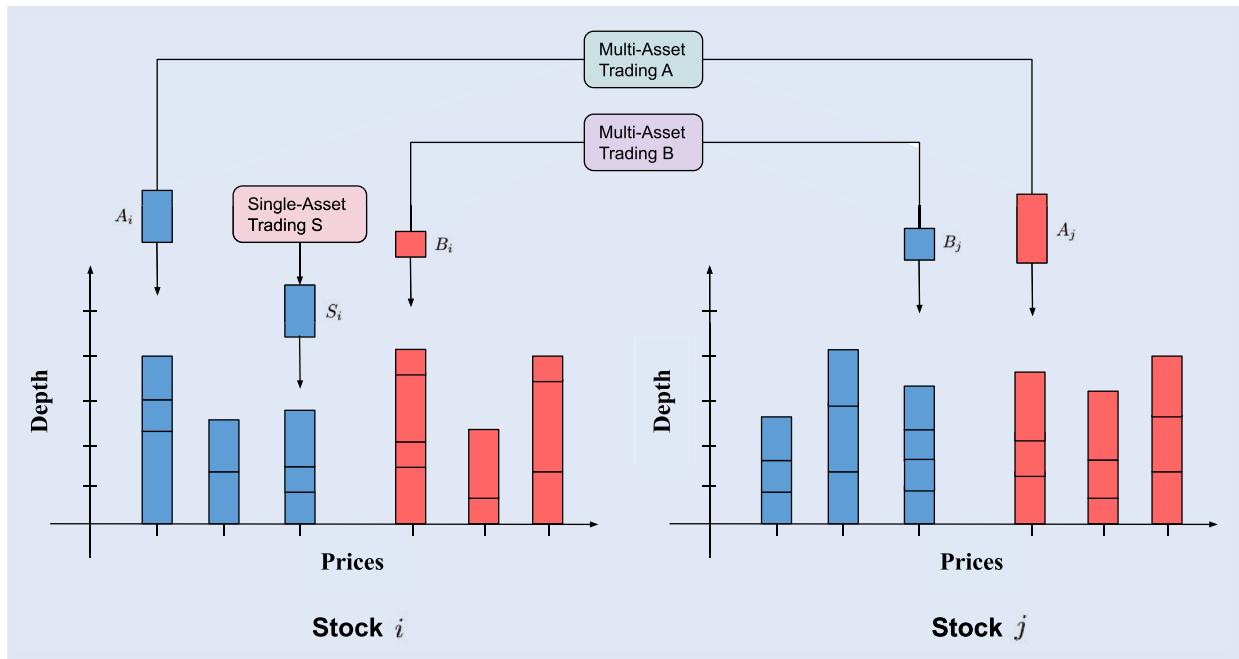


Figure 7. Illustration of the cross-impact model.

Note: The orders at different levels of each stock may come from single-asset and multi-asset trading strategies. The returns of stock i are potentially influenced by orders of stock j through the connections $B_j \rightarrow B_i \rightarrow \text{ofi}_i^1 \rightarrow r_i$ and $A_j \rightarrow A_i \rightarrow \text{ofi}_i^3 \rightarrow r_i$. Information along the path $A_j \rightarrow A_i \rightarrow \text{ofi}_i^3 \rightarrow r_i$ can be collected by the price impact model with integrated OFIs but not by the price impact model with only best-level OFIs.

$$= a_p^{[1]} + \beta_p^{[1]} \sum_{i=1}^N w_i \text{ofi}_{i,t}^{1,h} + e_{p,t}^{[1]}, \quad (10)$$

where $a_p^{[1]}$ is the intercept and $e_{p,t}^{[1]}$ is the noise term.

On the other hand, the cross-impact for portfolio p is

$$\begin{aligned} r_{p,t}^{(h)} &= \sum_{i=1}^N w_i r_{p,t}^{(h)} \\ &= \sum_{i=1}^N w_i (\alpha_i^{[1]} + \beta_i^{[1]} \text{ofi}_{i,t}^{1,h} + \epsilon_{i,t}^{[1]}) \\ &= \alpha_p^{[1]} + \sum_{i=1}^N \beta_i^{[1]} w_i \text{ofi}_{i,t}^{1,h} + e_{p,t}^{[1]}, \end{aligned} \quad (11)$$

where $\alpha_p^{[1]} = \sum_{i=1}^N w_i \alpha_i^{[1]}$ is the intercept.

Comparing (10) and (11) shows that cross-impact at portfolio level depends on the angles of $\vec{\beta} = (\beta_1^{[1]}, \dots, \beta_N^{[1]})$ and $\vec{w} = (w_1, \dots, w_N)$. On one hand, if individual assets exhibit a universal pattern of price dynamics, i.e. $\beta_i^{[1]} \approx \beta_j^{[1]}, \forall i \neq j$, we may conclude that the portfolio return is driven by its own order flows, to a large extent. On the other hand, if the portfolio places most of the weight on a specific stock or a set of stocks with similar patterns of price dynamics, then the portfolio returns are also driven by its own order flows, rather than by cross-impact. In other scenarios, it is necessary to consider the cross-impact. One can perform an analogous analysis for models with integrated OFIs.

In fact, the mechanism depicted in figure 7 aligns with this analysis. For example, orders placed on stock i represented

by A_i and S_i have the potential to affect the price movements of that stock, which subsequently affect the returns of portfolio B , even if A_i and S_i have no direct association with B . Additionally, A_j may have a different influence on the performance of B because of the (potentially) different price dynamics of stock j . Therefore, it may be necessary to take into account cross-asset OFIs when developing models for portfolio returns.

To examine the potential of cross-asset OFIs in explaining portfolio returns, we choose two widely-used portfolio construction methods: the equal-weighted portfolio (EW), and the eigenportfolio[†] using the 1st principal component (Eig1). Table 7 summarizes the out-of-sample R^2 of various models on these two portfolios. As table 7 shows, there is a significant difference between $\mathbf{PI}^{[1]}$ and $\mathbf{CI}^{[1]}$, \mathbf{PI}^I and \mathbf{CI}^I , indicating that cross-impact is a crucial factor when modeling portfolio returns. Again, the models using the integrated OFIs outperform their counterparts using the best-level OFIs.

4. Forecasting future returns

In the previous section, the definitions of price impact and cross-impact are based on contemporaneous OFIs and returns, meaning that both quantities pertain to the same bucket of time. In this section, we extend the above studies to future returns, and probe into the forward-looking price impact and cross-impact models.

[†] See Avellaneda and Lee (2010).

Table 7. Out-of-sample performance on portfolio returns.

	Best-level OFIs		Integrated OFIs	
	PI ^[1]	CI ^[1]	PI ^I	CI ^I
EW	79.29 (6.24)	81.03 (6.16)	85.26 (3.54)	87.97 (2.78)
Eig1	80.73 (6.75)	81.95 (6.36)	84.69 (3.70)	87.70 (2.84)

Note: EW: equal-weighted portfolio. Eig1: eigenportfolio using the 1st principal component of multi-asset returns.

4.1. Predictive models

We first propose the following forward-looking price impact and cross-impact models, denoted as **FPI**^[1], **FPI**^I, **FCI**^[1], and **FCI**^I, respectively. **FPI**^[1] (**FPI**^I) uses the lagged best-level (integrated) OFIs of stock i to predict its own future return $r_{i,t+f}^{(f)}$ during $(t, t+f]$, while **FCI**^[1] (**FCI**^I) involves the lagged multi-asset best-level (integrated) OFIs. We employ OLS to fit the forward-looking price impact models and LASSO to fit the cross-impact models.

$$\mathbf{FPI}^{[1]} : r_{i,t+f}^{(f)} = \alpha_i^{[1]} + \sum_{k \in L} \beta_i^{[1],k} \text{ofi}_{i,t}^{1,(kh)} + \epsilon_{i,t+f}^{[1]}, \quad (12)$$

$$\mathbf{FCI}^{[1]} : r_{i,t+f}^{(f)} = \alpha_i^{[1]} + \sum_{j=1}^N \sum_{k \in L} \beta_{i,j}^{[1],k} \text{ofi}_{i,t}^{1,(kh)} + \eta_{i,t+f}^{[1]}, \quad (13)$$

$$\mathbf{FPI}^I : r_{i,t+f}^{(f)} = \alpha_i^I + \sum_{k \in L} \beta_i^{I,k} \text{ofi}_{i,t}^{I,(kh)} + \epsilon_{i,t+f}^I, \quad (14)$$

$$\mathbf{FCI}^I : r_{i,t+f}^{(f)} = \alpha_i^I + \sum_{j=1}^N \sum_{k \in L} \beta_{i,j}^{I,k} \text{ofi}_{i,t}^{I,(kh)} + \eta_{i,t+f}^I, \quad (15)$$

where f is the forecasting horizon of future returns and $L = \{1, 2, 3, 5, 10, 20, 30\}$ represents the set of lags.

Furthermore, we compare OFI-based models with return-based models studied in previous works, where lagged returns are involved as predictors. **AR** (equation (16)) is an autoregressive (AR) model using various returns over different time horizons, inspired by Corsi (2009), Ait-Sahalia *et al.* (2022). **CAR** (Chinco *et al.* 2019) uses the entire cross-section lagged returns as candidate predictors, as detailed in equation (17). We employ OLS to fit the **ARs** and LASSO to fit the **CARs**.

$$\mathbf{AR} : r_{i,t+f}^{(f)} = \alpha_i + \sum_{k \in L} \beta_i^{r,k} r_{i,t}^{(kh)} + \epsilon_{i,t+f} \quad (16)$$

$$\mathbf{CAR} : r_{i,t+f}^{(f)} = \alpha_i + \sum_{j=1}^N \sum_{k \in L} \beta_{i,j}^{r,k} r_{i,t}^{(kh)} + \eta_{i,t+f} \quad (17)$$

4.2. Empirical results

In this experiment, observations associated with returns and OFIs are computed minutely, i.e. $h = 1$ minute.[†] Following Chinco *et al.* (2019), we use data from the previous 30

[†] Note that we choose to use the physical time as opposed to the trading time. This is because each stock has its own specific trading time, which is asynchronous with that of others. Thus it is difficult

Table 8. Out-of-sample performance for one-minute-ahead returns.

	Best-level OFIs		Integrated OFIs		Returns	
	FPI ^[1]	FCI ^[1]	FPI ^I	FCI ^I	AR	CAR
OS R^2	-0.37 (0.10)	-0.10 (0.05)	-0.36 (0.08)	-0.10 (0.05)	-0.36 (0.11)	-0.10 (0.05)

Note: The table reports the mean values and standard deviations (in parentheses) of out-of-sample R^2 of various models when modeling one-minute-ahead returns. The predictive models include **FPI**^[1] (equation (12)), **FCI**^[1] (equation (13)), **FPI**^I (equation (14)), **FCI**^I (equation (15)), **AR** (equation (16)) and **CAR** (equation (17)). These statistics are averaged across each stock and each regression window.

minutes to estimate the model parameters and apply the fitted model to forecast future f -minute returns. We then move one minute forward and repeat this procedure to compute the rolling f -minute-ahead return forecasts. For all models, we initially focus on the 1-minute forecasting horizon. In Section 4.3, we consider return forecasts over longer horizons, including $f \in \{2, 3, 5, 10, 20, 30\}$ minutes, to assess the strength and duration of price impact and cross-impact.

Following the analysis of Cartea *et al.* (2018) and Chinco *et al.* (2019), we demonstrate the effectiveness of the forward-looking price impact and cross-impact models from two perspectives: (1) statistical performance, and (2) economic gain.

4.2.1. Statistical performance. Table 8 summarizes the out-of-sample R^2 values of the aforementioned predictive models when predicting the subsequent 1-minute returns, i.e. $f = 1$. It appears the cross-impact models **FCI**^[1] (respectively, **FCI**^I, **CAR**) achieve higher out-of-sample R^2 statistics compared to the price impact models **FPI**^[1] (respectively, **FPI**^I, **AR**). We also implement the same hypothesis test described in Section 3 to investigate the statistical significance (unreported) of these results. We observe that the cross-impact models exhibit significantly superior performance than the price impact models across all stocks, at the 1% confidence level.

Most of the empirical literature in return prediction focuses its evaluations on out-of-sample R^2 . However, we remark that negative R^2 values do not imply that the forecasts are economically meaningless (see more discussions in Choi *et al.* 2022, Kelly *et al.* 2022).[‡] To emphasize this point,

to work out the cross-impact between stocks on a trading time scale, also see Wang *et al.* (2016a) and Wang *et al.* (2016b). The choice of a 1-minute bin size allows us to abstract away from microstructure effects which are not the focus of the present mesoscopic study, as is the case in Benzaquen *et al.* (2017) and Chinco *et al.* (2019).

[‡] A simple example can be framed as follows. Consider a model with one predictor and suppose that the estimated predictive coefficient is a significantly large multiple of the actual value. In this case, the R^2 will become negative. However, the predictions will be perfectly correlated with the true expected return, resulting in a positive expected return for our strategy. Proposition 4 in Kelly *et al.* (2022) further proposed that we can worry less about the positivity of out-of-sample R^2 from a prediction model and focus more on the out-of-sample performance of specific trading strategies based on predicted returns.

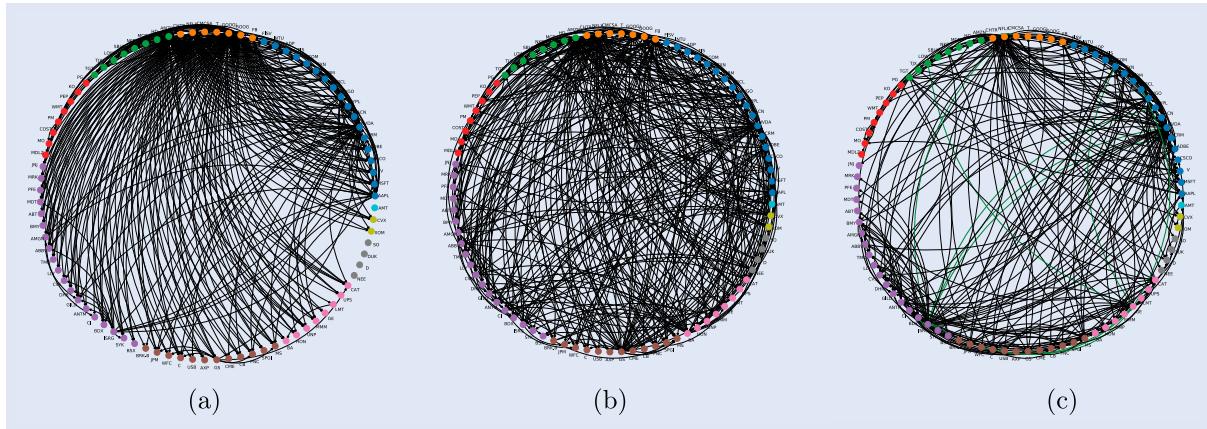


Figure 8. Network structure of the coefficient matrix constructed from forward-looking cross-impact models. (a) Based on best-level OFIs. (b) Based on integrated OFIs and (c) Based on returns.

Note: The coefficients are averaged over 2017–2019. To render the networks more interpretable and for ease of visualization, we only plot the top 5% largest in magnitude coefficients. Nodes are colored by the GICS structure and sorted by market capitalization. Green links represent positive values while black links represent negative values. The width of edges is proportional to the absolute values of their respective coefficients.

Table 9. Group degree centrality for each GICS sector.

	Group In-degree Centrality			Group Out-degree Centrality		
	Best-level OFIs	Integrated OFIs	Returns	Best-level OFIs	Integrated OFIs	Returns
Information Technology	0.12	0.36	0.26	0.46	0.62	0.59
Communication Services	0.06	0.24	0.20	0.85	0.74	0.60
Consumer Discretionary	0.09	0.20	0.15	0.86	0.51	0.17
Consumer Staples	0.03	0.15	0.09	0.00	0.11	0.01
Health Care	0.10	0.37	0.19	0.12	0.22	0.59
Financials	0.12	0.21	0.17	0.03	0.41	0.08
Industrials	0.10	0.19	0.20	0.00	0.39	0.27
Utilities	0.00	0.07	0.04	0.00	0.06	0.00
Energy	0.06	0.07	0.04	0.00	0.14	0.00
Real Estate	0.00	0.05	0.02	0.00	0.00	0.01

Note: According to the threshold networks as shown in figure 8, we compute the fraction of stocks outside of a specific sector connected to stocks in this specific sector. The color of each sector in this table corresponds to the color in figure 8.

we will incorporate these return forecasts into a forecast-based trading strategy, and showcase their profitability in the following subsection.

Considering the different magnitudes of the OFIs and returns, we first normalize the coefficient matrix of each model by dividing by the average of the absolute coefficients. Figure A2 (deferred to Appendix 5) shows the average coefficient matrices of $\text{FCI}^{[1]}$, FCI^I , and CAR . For example, as revealed in figure A2(a) ($\text{FCI}^{[1]}$), for a specific stock, the main influence comes from its own OFI, i.e. the absolute values of diagonal elements are significantly larger than the off-diagonal ones. We observe that cross-impact is often negative, consistent with Pasquariello and Vega (2015). Except for the self-impact, most stocks are also influenced by stocks in Communication Services, Consumer Discretionary and Information Technology.

To better illustrate the interactions between different stocks, we construct a network for each normalized coefficient matrix and only preserve the cross-asset edges (i.e. off-diagonal elements) larger than the 95th percentile of coefficients. Figure 8 illustrates some of the main characteristics of the coefficient

networks for $\text{FCI}^{[1]}$, FCI^I , and CAR . For example, we again observe that there are more edges from Communication Services, Consumer Discretionary and Information Technology, indicating they may contain more predictive power for others.

To gain a better understanding of the structural properties of the resulting network, we aggregate node centrality measures (see Everett and Borgatti 1999) at the sector level, and also perform a spectral analysis of the adjacency matrix. From table 9, we observe that the out-degree centrality of Communication Services, Consumer Discretionary and Information Technology is significantly larger than that of others, consistent with previous findings. Figure 8(c) also shows that the network based on returns contains more inner-sector connections than the other two counterparts, thus implying a sectorial structure. Table 10 presents the top five most significant stocks in terms of out-degree centrality in each network, which exhibit more impact on the prices of other stocks.

Figure 9 shows a barplot with the average value for the top 20 largest singular values of the network adjacency matrix, for best-level OFIs, integrated OFIs, and returns, where the average is performed over all constructed networks. For ease

Table 10. Top 5 stocks according to node out-degree centrality in threshold networks.

Best-level OFIs	Integrated OFIs	Returns
AMZN	NFLX	NVDA
GOOG	AMZN	NFLX
GOOGL	NVDA	ISRG
NVDA	GS	AVGO
NFLX	FB	GE

Note: The out-degree centrality for a node is the fraction of nodes its outgoing edges are connected to.

of visualization and comparison, we first normalize the adjacency matrix before computing the top singular values, which exhibit a fast decay. In addition to the significantly large top singular value revealing that the networks have a strong rank-1 structure, the next 6-8 singular values are likely to correspond to the more prominent industry sectors.

4.2.2. Economic gains. On the basis of return forecasts, we employ a portfolio construction method, proposed by Chinco *et al.* (2019), to evaluate the economic gains of the aforementioned predictive models.

Forecast-implied portfolio. For a specific forecasting model F , the motivations of portfolio construction can be summarized as follows.

- It only executes an order when the one-minute-ahead return forecast exceeds the bid-ask spread.
- It buys/sells more shares of the i th stock when the absolute value of one-minute-ahead return forecast for i th stock is higher.
- It buys/sells more shares of the i th stock when the one-minute-ahead return forecasts for the i th stock tend to be less volatile throughout the trading day.

This strategy allocates a fraction $w_{i,t}$ of its capital to the i th stock

$$w_{i,t} \stackrel{\text{def}}{=} \frac{1\{|f_{i,t}^F| > sprd_{i,t}\} \cdot f_{i,t}^F / \sigma_{i,t}^F}{\sum_{n=1}^N 1\{|f_{n,t}^F| > sprd_{n,t}\} \cdot |f_{n,t}^F| / \sigma_{n,t}^F}, \quad (18)$$

where $f_{i,t}^F$ represents the one-minute-ahead return forecast according to model F for minute $(t+1)$, $sprd_{i,t}$ represents the

Table 11. Economic performance of the forecast-implied trading strategy.

	Best-level OFIs		Integrated OFIs		Returns	
	FPI ^[1]	FCI ^[1]	FPI ^I	FCI ^I	AR	CAR
PnL	0.21 (0.12)	0.43 (0.17)	0.23 (0.13)	0.39 (0.19)	0.23 (0.13)	0.40 (0.18)

Note: The table reports the mean values and standard deviations (in parentheses) of annualized PnLs of forecast-implied trading strategy of various models for forecasting one-minute-ahead returns. The predictive models include **FPI**^[1] (equation (12)), **FCI**^[1] (equation (13)), **FPI**^I (equation (14)), **FCI**^I (equation (15)), **AR** (equation (16)) and **CAR** (equation (17)). These statistics are averaged over 2017-2019.

relative bid-ask spread at time t , $\sigma_{i,t}^F$ represents the standard deviation of the model's one-minute-ahead return forecasts for the i th stock during the previous 30 minutes of trading, i.e. the standard deviation of in-sample fits. The denominator is the total investment so that the strategy is self-financed. If there are no stocks with forecasts that exceed the spread in a given minute, then we set $w_{i,t} = 0, \forall i$.

Finally, we compute the *profit and loss* (PnL) of the resulting portfolios on each trading day by summing the strategy's minutely returns as in Chinco *et al.* (2019).

Table 11 compares the performance (annualized PnL) of the forecast-implied strategies, based on forecast returns from various predictive models. It is worth noting that in the following analysis, the strategy ignores trading costs, as this is not the focus of our paper. Table 11 shows that portfolios based on forecasts of the forward-looking cross-impact model outperform those based on forecasts of the forward-looking price impact model.

4.3. Longer forecasting horizons

One-minute-ahead return forecasts are not the only time horizon of interest to practitioners and academics. Additionally, we evaluate the performance of the above models and examine the forecasting ability of cross-impact terms over longer prediction horizons.

Figure 10 illustrates the model predictability from the perspective of raw annualized PnL across multiple

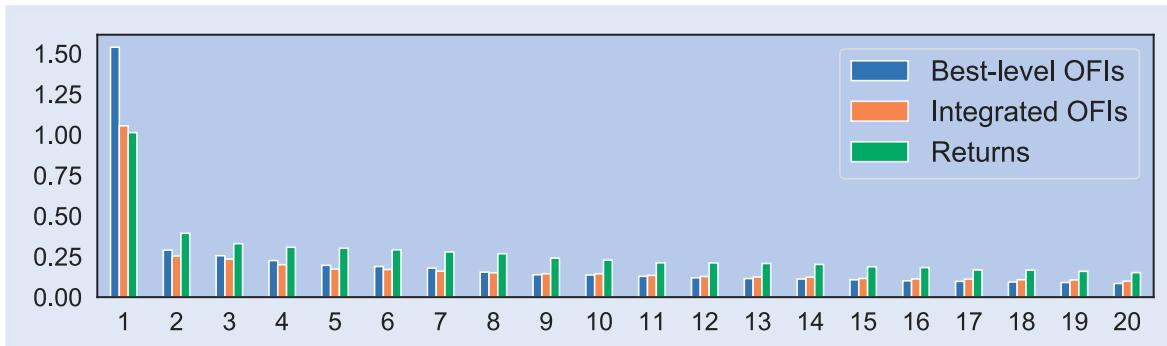


Figure 9. Barplot of normalized singular values for the average coefficient matrix in forward-looking cross-impact models.

Note: We perform Singular Value Decomposition (SVD) on the coefficient matrix and obtain the singular values. The x -axis represents the singular value rank, and the y -axis shows the normalized singular values. The coefficients are averaged over 2017–2019.

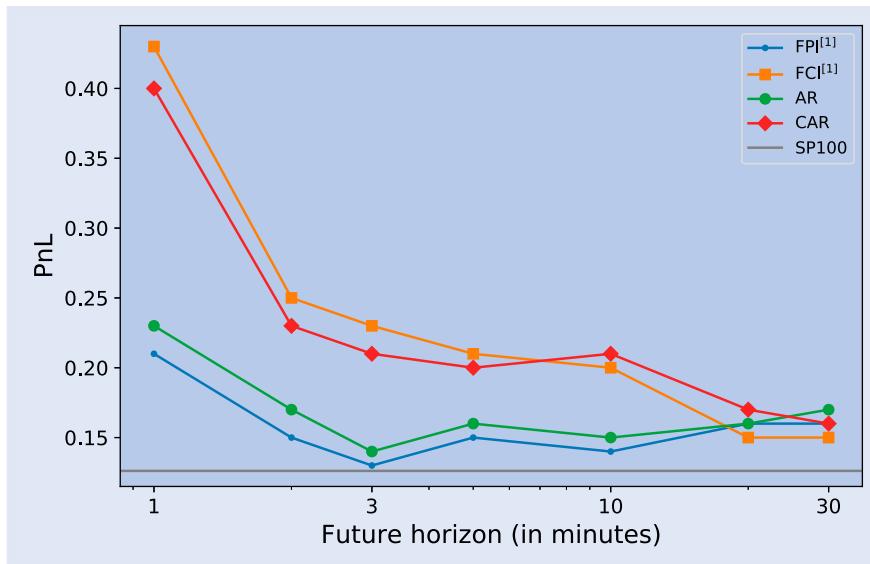


Figure 10. Annualized PnL as a function of the forecasting horizon.

Note: The x-axis represents the prediction horizon (in minutes), while the y-axis represents the annualized PnL. The grey horizontal line is the performance of the S&P100 ETF index.

horizons.[†] Due to the similar performance of **FPI^[1]** and **FPI^[l]** (respectively, **FCI^[1]** and **FCI^[l]**) over longer horizons, we only show the curves of **FPI^[1]**, **FCI^[1]**, **AR**, **CAR**, and a benchmark (S&P100 ETF). It appears that superior forecasting ability arises from cross-asset terms at short horizons. However, the PnL of cross-asset models declines more quickly over longer horizons. A further study with more focus on the reasons for the predictability of cross-asset OFIs over multiple horizons is therefore suggested. Finally, the models in which each stock only relies on its own returns/OFIs marginally outperform their counterparts which use the entire cross-sectional predictors.

4.4. Discussion about predictive cross-impact

Tables 8 and 11 reveal that, in contrast to the price impact model, multi-asset OFIs can provide considerably more additional explanatory power for *future returns* compared to *contemporaneous returns*. A possible explanation for this asymmetric phenomenon is that there exists a time lag between when the OFIs of a given stock are formed (a so-called *flow formation period*) and the actual time when traders notice this change of flow and incorporate it into their trading model (see Buccheri *et al.* 2021).[‡] For example, assume a trader submitted an unexpectedly large amount

of buy limit orders of Apple (AAPL) at 10:00 am, at either the first level or potentially deeper in the book. Other traders may notice this anomaly and adjust their portfolios (including Apple) at a later time, for example, 10:01 am. In this case, the OFIs of Apple may indicate future price changes of other stocks.

Consistent with our explanation, Hou (2007) argued that the gradual diffusion of industry information is a leading cause of the lead-lag effect in stock returns. Cohen and Frazzini (2008) found that certain stock prices do not promptly incorporate news pertaining to economically related firms, due to the presence of investors subject to attention constraints. Further research should be undertaken to investigate the origins of the success of multi-asset OFIs in predicting future returns.

It is also interesting to note that forward-looking models using integrated OFIs cannot significantly outperform models using the best-level OFIs. This phenomenon might stem from the fact that the integrated OFIs do not explicitly take into account the level information (distance of a given level to the best bid/ask) of multi-level OFIs, and are agnostic to different sizes resting at different levels on the bid and ask sides of the book. Previous studies (such as Hasbrouck and Saar 2002, Cao *et al.* 2009, Cenesizoglu *et al.* 2022) demonstrated that traders might strategically choose to place their orders in different levels of the book depending on various factors, therefore limit orders at different price levels may contain different information content with respect to predicting future returns. A further study with more focus on the impact of multi-level OFIs over different time horizons is suggested.

5. Conclusion

We have systematically examined the impact of OFIs from multiple perspectives. The main contributions can be summarized as follows.

[†] To plot this figure, we only accumulate the PnLs between [10:31, 15:30], which is the shared trading period for the studied forecasting horizons.

[‡] A closely-related phenomenon is the Epps effect documented by Epps (1979), which showed that the empirical correlation estimates tend to decrease when the sampling is done at high frequencies. Previous research (including Renò 2003, Tóth and Kertész 2009, Zhang 2011) demonstrated that the Epps effect might be explained by the non-synchronicity, the possible lead-lag relationship between stock returns, etc. Tóth and Kertész (2009) also described that the Epps effect might be caused by the reaction time of traders to news and events, which is usually spread out over a time interval of a few minutes.

First, we verify the effects of multi-level and cross-asset OFIs on contemporaneous price dynamics. We introduce a new procedure to examine the cross-impact on contemporaneous returns. Under the sparsity assumption of cross-impact coefficients, we use LASSO to describe such a structure and compare the performances with the price impact model which only utilizes a stock's own OFIs. We implement models with the best-level OFIs and integrated OFIs, respectively. The results first demonstrate that our integrated OFIs provide higher explanatory power for price movements than the widely-used best-level OFIs. More interestingly, in comparison with the price impact model using best-level OFIs, the cross-impact model exhibits additional explanatory power. However, the cross-impact model with integrated OFIs cannot provide extra explanatory power to the price impact model with integrated OFIs, indicating the effectiveness of our integrated OFIs.

In addition, we apply the price impact and cross-impact models to the challenging task of predicting future returns. The results reveal that involving cross-asset OFIs can increase out-of-sample R^2 . We subsequently demonstrate that this increase in out-of-sample R^2 leads to additional economic profits, when incorporated in common trading strategies, thus providing evidence of cross-impact over short future horizons. We also find that predictability of cross-impact terms vanishes quickly over longer horizons.

Future research directions. There are a number of interesting avenues to explore in future research. One such direction pertains to the assessment of whether cross-asset multi-level OFIs can improve the forecast of future returns (in the present work, we only considered the best-level OFI and integrated OFI due to limited computing power). Another interesting direction pertains to performing a similar analysis as in the present paper, but for the last 15–30 minutes of the trading day, where a significant fraction of the total daily trading volume occurs. For example, for the first few months of 2020 in the US equity market, about 23% of trading volume in the 3000 largest stocks by market value has taken place after 3:30 pm, compared with about 4% from 12: pm to 1 pm (Banerji 2020). It would be an interesting study to explore the interplay between the OFI dynamics and this surge of trading activity at the end of U.S. market hours.

Acknowledgments

We thank Robert Engle, Álvaro Cartea, Slavi Marinov and seminar participants at the 11th Bachelier World Congress 2022; the 14th Annual SoFiE Conference; University of Oxford for helpful comments. We also thank the Oxford Suzhou Centre for Advanced Research for providing the computational facilities. Earlier versions of this paper circulated under the titles ‘Price Impact of Order Flow Imbalance: Multi-level, Cross-asset and Forecasting’ and ‘Cross-Impact of Order Flow Imbalance: Contemporaneous and Predictive’. First draft: December 2021.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Chao Zhang has been supported by the Clarendon Fund Scholarship and G-Research PhD prize in Maths and Data Science.

ORCID

Rama Cont  <http://orcid.org/0000-0003-1164-6053>
 Mihai Cucuringu  <http://orcid.org/0000-0002-8464-2152>
 Chao Zhang  <http://orcid.org/0000-0002-0013-336X>

References

- Ahn, H.-J., Bae, K.-H. and Chan, K., Limit orders, depth, and volatility: Evidence from the stock exchange of Hong Kong. *J. Finance*, 2001, **56**(2), 767–788.
- Ait-Sahalia, Y., Fan, J., Xue, L. and Zhou, Y., How and when are high-frequency stock returns predictable? Available at SSRN 4095405, 2022.
- Avellaneda, M. and Lee, J.-H., Statistical arbitrage in the US equities market. *Quant. Finance*, 2010, **10**(7), 761–782.
- Banerji, G., The 30 minutes that can make or break the trading day. 2020. https://www.wsj.com/articles/the-30-minutes-that-can-make-or-break-the-trading-day-11583886131?relink=desktopwebshare_permalink.
- Benzaquen, M., Mastromatteo, I., Eisler, Z. and Bouchaud, J.-P., Dissecting cross-impact on stock markets: An empirical analysis. *J. Stat. Mech. Theory Exp.*, 2017, **2017**(2), 023406.
- Buccheri, G., Corsi, F. and Peluso, S., High-frequency lead-lag effects and cross-asset linkages: A multi-asset lagged adjustment model. *J. Bus. Econ. Stat.*, 2021, **39**(3), 605–621.
- Cao, C., Hansch, O. and Wang, X., The information content of an open limit-order book. *J. Futures Markets*, 2009, **29**(1), 16–41.
- Capponi, F. and Cont, R., Multi-asset market impact and order flow commonality. SSRN. 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3706390.
- Cartea, Á., Donnelly, R. and Jaimungal, S., Enhancing trading strategies with order book signals. *Appl. Math. Finance*, 2018, **25**(1), 1–35.
- Cenesizoglu, T., Dionne, G. and Zhou, X., Asymmetric effects of the limit order book on price dynamics. *J. Empir. Finance*, 2022, **65**, 77–98.
- Chakrabarty, B., Hendershott, T., Nawn, S. and Pascual, R., Order exposure in high frequency markets. 2022. Available at SSRN 3074049.
- Chinco, A., Clark-Joseph, A.D. and Ye, M., Sparse signals in the cross-section of returns. *J. Finance*, 2019, **74**(1), 449–492.
- Choi, D., Jiang, W. and Zhang, C., Alpha go everywhere: Machine learning and international stock returns. 2022. Available at SSRN 3489679.
- Chordia, T. and Subrahmanyam, A., Order imbalance and individual stock returns: Theory and evidence. *J. Financ. Econ.*, 2004, **72**(3), 485–518.
- Chordia, T., Roll, R. and Subrahmanyam, A., Order imbalance, liquidity, and market returns. *J. Financ. Econ.*, 2002, **65**(1), 111–130.
- Cohen, L. and Frazzini, A., Economic links and predictable returns. *J. Finance*, 2008, **63**(4), 1977–2011.
- Cont, R., Kukanov, A. and Stoikov, S., The price impact of order book events. *J. Financ. Econom.*, 2014, **12**(1), 47–88.
- Corsi, F., A simple approximate long-memory model of realized volatility. *J. Financ. Econom.*, 2009, **7**(2), 174–196.
- Curato, G. and Lillo, F., How tick size affects the high frequency scaling of stock return distributions. In *Financial Econometrics and Empirical Market Microstructure*, pp. 55–76, 2015 (Springer).

- Curme, C., Tumminello, M., Mantegna, R.N., Stanley, H.E. and Kenett, D.Y., Emergence of statistically validated financial intra-day lead-lag relationships. *Quant. Finance*, 2015, **15**(8), 1375–1386.
- Epps, T.W., Comovements in stock prices in the very short run. *J. Am. Stat. Assoc.*, 1979, **74**(366a), 291–298.
- Everett, M.G. and Borgatti, S.P., The centrality of groups and classes. *J. Math. Sociol.*, 1999, **23**(3), 181–201.
- Giacomini, R. and White, H., Tests of conditional predictive ability. *Econometrica*, 2006, **74**(6), 1545–1578.
- Gu, S., Kelly, B. and Xiu, D., Empirical asset pricing via machine learning. *Rev. Financ. Stud.*, 2020, **33**(5), 2223–2273.
- Harris, L.E. and Panchapagesan, V., The information content of the limit order book: Evidence from NYSE specialist trading decisions. *J. Financ. Mark.*, 2005, **8**(1), 25–67.
- Hasbrouck, J. and Saar, G., Limit orders and volatility in a hybrid market: The Island ECN. Stern School of Business Dept. of Finance Working Paper FIN-01-025, 2002.
- Hasbrouck, J. and Seppi, D.J., Common factors in prices, order flows, and liquidity. *J. Financ. Econ.*, 2001, **59**(3), 383–411.
- Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2009 (Springer Science & Business Media: New York).
- Hautsch, N. and Huang, R., The market impact of a limit order. *J. Econ. Dyn. Control*, 2012, **36**(4), 501–522.
- Hou, K., Industry information diffusion and the lead-lag effect in stock returns. *Rev. Financ. Stud.*, 2007, **20**(4), 1113–1138.
- Huck, N., Large data sets and machine learning: Applications to statistical arbitrage. *Eur. J. Oper. Res.*, 2019, **278**(1), 330–342.
- Kelly, B.T., Malamud, S. and Zhou, K., The virtue of complexity in return prediction. *J. Finance*, 2022, forthcoming.
- Kenett, D.Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R.N. and Ben-Jacob, E., Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PLoS. ONE.*, 2010, **5**(12), e15032.
- Kolm, P.N., Turiel, J. and Westray, N., Deep order flow imbalance: Extracting alpha at multiple horizons from the limit order book. *Math. Finance*, 2023, forthcoming.
- Krauss, C., Do, X.A. and Huck, N., Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *Eur. J. Oper. Res.*, 2017, **259**(2), 689–702.
- Kyle, A.S., Continuous auctions and insider trading. *Econometrica*, 1985, **53**(6), 1315–1335.
- Laloux, L., Cizeau, P., Potters, M. and Bouchaud, J.-P., Random matrix theory and financial correlations. *Int. J. Theor. Appl. Finance*, 2000, **3**(03), 391–397.
- Lillo, F., Farmer, J.D. and Mantegna, R.N., Master curve for price-impact function. *Nature*, 2003, **421**(6919), 129–130.
- Menzly, L. and Ozbas, O., Market segmentation and cross-predictability of returns. *J. Finance*, 2010, **65**(4), 1555–1580.
- Pasquariello, P. and Vega, C., Strategic cross-trading in the US stock market. *Rev. Financ.*, 2015, **19**(1), 229–282.
- Renò, R., A closer look at the epps effect. *Int. J. Theor. Appl. Finance*, 2003, **6**(01), 87–102.
- Rosenbaum, M. and Tomas, M., A characterisation of cross-impact kernels. arXiv preprint arXiv:2107.08684. 2021.
- Schneider, M. and Lillo, F., Cross-impact and no-dynamic-arbitrage. *Quant. Finance*, 2019, **19**(1), 137–154.
- Sirignano, J.A., Deep learning for limit order books. *Quant. Finance*, 2019, **19**(4), 549–570.
- Tashiro, D., Matsushima, H., Izumi, K. and Sakaji, H., Encoding of high-frequency order information and prediction of short-term stock price by deep learning. *Quant. Finance*, 2019, **19**(9), 1499–1506.
- Tomas, M., Mastromatteo, I. and Benzaquen, M., How to build a cross-impact model from first principles: Theoretical requirements and empirical results. *Quant. Finance*, 2022, **22**(6), 1017–1036.
- Tóth, B. and Kertész, J., The epps effect revisited. *Quant. Finance*, 2009, **9**(7), 793–802.
- Wang, S., Schäfer, R. and Guhr, T., Average cross-responses in correlated financial markets. *Eur. Phys. J. B*, 2016a, **89**(9), 207.
- Wang, S., Schäfer, R. and Guhr, T., Cross-response in correlated financial markets: Individual stocks. *Eur. Phys. J. B*, 2016b, **89**(4), 105.
- Wang, S., Neusüß, S. and Guhr, T., Statistical properties of market collective responses. *Eur. Phys. J. B*, 2018, **91**, 1–11.
- Ward, M.D. and Ahlquist, J.S., *Maximum Likelihood for Social Science: Strategies for Analysis*, 2018 (Cambridge University Press: Cambridge).
- Xu, K., Gould, M.D. and Howison, S.D., Multi-level order-flow imbalance in a limit order book. *Mark. Microstruct. Liq.*, 2018, **4**(3–4), 1950011.
- Zhang, L., Estimating covariation: Epps effect, microstructure noise. *J. Econom.*, 2011, **160**(1), 33–47.

Appendices

Appendix 1. Aggregation of multi-level OFIs

Table 2 presents evidence of the effectiveness of PCA in selecting weights for combining multi-level OFIs. However, figure 2 shows that the weights derived from PCA are not extremely different. This prompts us to consider a simpler method, namely the simple average (SA), to achieve similar performance. Table A1 reveals that the explained variance of the SA increases as more levels are included. Nonetheless, the SA across 10 levels is inferior to the first principal component (PC) in terms of EVR, i.e. 85.07% of SA vs 89.06% of PC in table 2. Additionally, table A2 provides further evidence that PC consistently performs better than SA across different subsets grouped by stock-specific characteristics.

We then perform the price impact and cross-impact analysis with simple average multi-level OFIs, denoted as \mathbf{PI}^{SA} and \mathbf{CI}^{SA} , respectively. As shown in table A3, \mathbf{PI}^{SA} has a similar performance with \mathbf{CI}^{SA} , consistent with our main analysis. This again confirms that as long as multi-level orders are taken into account, adding cross-impact terms cannot significantly improve model performance. On the other hand, \mathbf{PI}^I is slightly better than \mathbf{PI}^{SA} . A future research direction might be to devise various weighting schemes that average the OFI information across the multiple levels, where the weights could be given, for example, by an inverse function of the distance of each price level to the mid-price or applying tensor-SVD/PCA on this data.

Appendix 2. Contemporaneous price impact of multi-level OFIs

To explicitly identify the impact of deeper-level OFIs, we also consider an extended version of $\mathbf{PI}^{[1]}$ by incorporating multi-level OFIs as features in the model

$$\mathbf{PI}^{[m]} : r_{i,t}^{(h)} = \alpha_i^{[m]} + \sum_{k=1}^m \beta_i^{[m],k} \text{ofi}_{i,t}^{k,(h)} + \epsilon_{i,t}^{[m]}. \quad (\text{A1})$$

Recall that $\text{ofi}_{i,t}^{k,(h)}$ is the OFI at level k . We refer to this model as $\mathbf{PI}^{[m]}$, and use OLS to estimate it.

The top panel of table A4 shows that the in-sample R^2 values increase as more multi-level OFIs are included as features, which is not surprising given that $\mathbf{PI}^{[m]}$ is a nested model of $\mathbf{PI}^{[m+1]}$. However the increments of the in-sample R^2 are descending, indicating that much deeper LOB data might be unable to provide additional information. This argument is confirmed by the models' performance on out-of-sample data, as shown at the bottom panel of table A4. Out-of-sample R^2 reaches a peak at $\mathbf{PI}^{[8]}$.

Impact comparison between multi-level OFIs. An interesting question is whether the OFIs at different price levels contribute

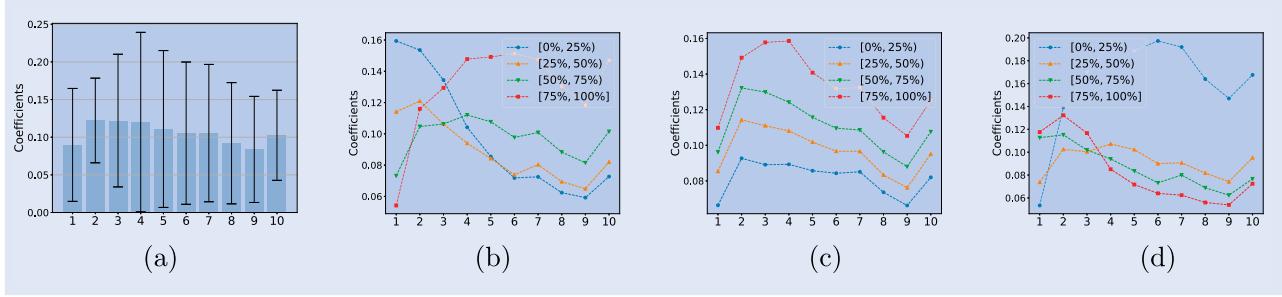


Figure A1. Coefficients of the model $\mathbf{PI}^{[10]}$. (a) Average. (b) Volume. (c) Volatility and (d) Spread.

Note: Plot (a) reports average coefficients and one standard deviation (error bars); Plots (b)-(d) show coefficients sorted by stock characteristics. Volume: trading volume on the previous trading day. Volatility: volatility of one-minute returns during the previous trading day. Spread: average bid-ask spread during the previous trading day. $[0\%, 25\%]$, respectively $[75\%, 100\%]$, denote the subset of stocks with the lowest, respectively highest, 25% values for a given stock characteristic. The x -axis represents different levels of OFIs and the y -axis represents the coefficients.

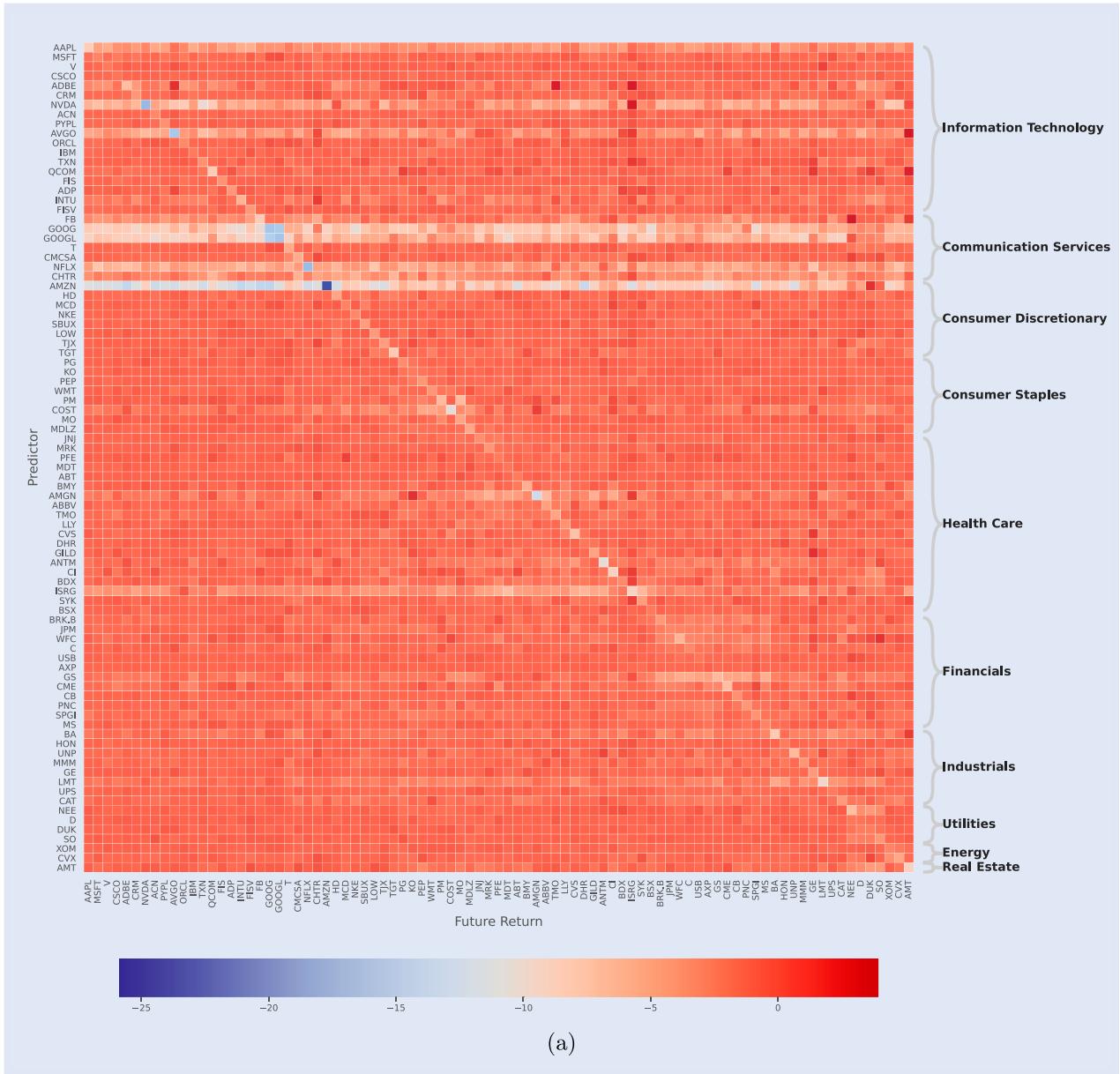


Figure A2. Average coefficient matrices constructed from forward-looking cross-impact models. (a) Coefficient matrix of $\mathbf{FCI}^{[1]}$.

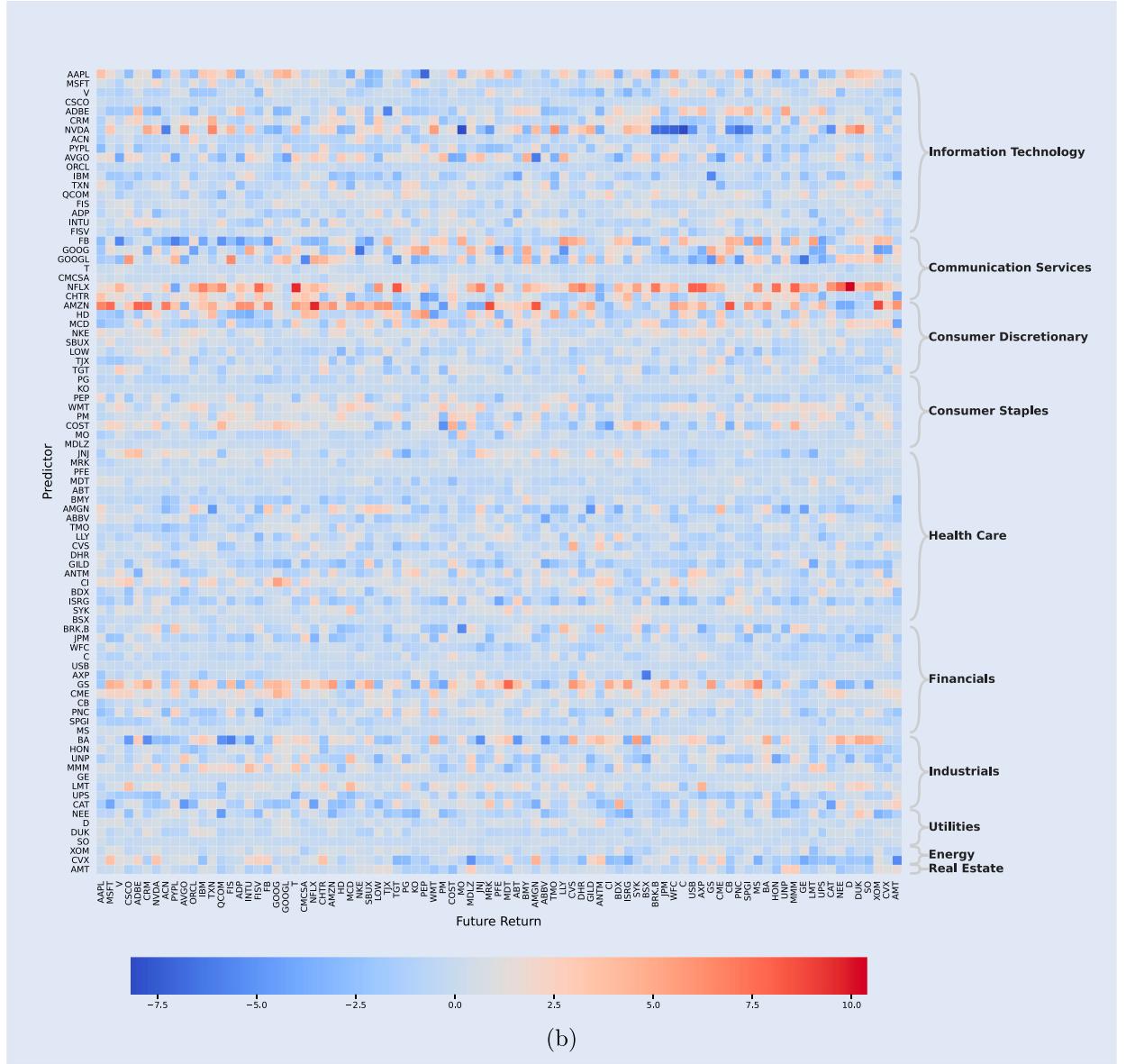


Figure A2. Continued. (b) Coefficient matrix of FCI^I

evenly in terms of price impact. Based on Figure 1(a), we conclude that multi-level OFIs have different contributions to price movements. Generally, OFIs at the second-best level manifest greater influence than OFIs at the best level in model $\text{PI}^{[10]}$, which is perhaps counter-intuitive, at first sight.

We further investigate how the coefficients vary across stocks with different characteristics, such as volume, volatility, and bid-ask spread. Figure A1(b-d) reveals that for stocks with *high-volume* and *small-spread*, order flow posted deeper in the LOB has more influence on price movements. The results regarding spread are in line with Xu *et al.* (2018), where it is observed that for large-spread stocks (AMZN, TSLA, and NFLX), the coefficients of ofi^m (OFIs at the m th level) tend to get smaller as the LOB level m increases, while for small-spread stocks (ORCL, CSCO, and MU), the coefficients of ofi^m may become larger as m increases.

Cont *et al.* (2014) concluded that the effect of ofi^m ($m \geq 2$) on price changes is only second-order or null. There are two likely causes for the differences between their findings and ours. First, the data used in Cont *et al.* (2014) includes 50 stocks (randomly picked from S&P500 constituents) for a single month in 2010, while we use the top 100 large-cap stocks for 36 months during 2017-2019. Second, Cont *et al.* (2014) considered the average of the coefficients across 50 stocks. In our work, we first group 100 stocks by

firm characteristics, and then study the average coefficients of each subset. Therefore, our results are based on a more granular analysis, across a significantly longer period of time.

Appendix 3. Comparison with Capponi and Cont (2020)

One closely related work is Capponi and Cont (2020) (CC hereafter), where the authors propose a two-step procedure to justify the significance of cross-impact terms and render a different conclusion about cross-impact.

In the first step, the authors use OLS to decompose each stock's OFIs ($\text{ofi}_{i,t}^{1,(h)}$) into the common factor of OFIs ($F_{\text{ofi},t}^{(h)}$), that is the first principal component of the multi-asset order flow imbalances, and obtain the idiosyncratic components ($\tau_{i,t}^{(h)}$) of the OFIs, for each individual stock.

$$\text{ofi}_{i,t}^{1,(h)} = \mu_i + \gamma_i F_{\text{ofi},t}^{(h)} + \tau_{i,t}^{(h)}. \quad (\text{A2})$$

In the second step, they regress returns ($r_{i,t}^{(h)}$) of stock i against (i) the common factor of OFIs ($F_{\text{ofi},t}^{(h)}$), (ii) the idiosyncratic components

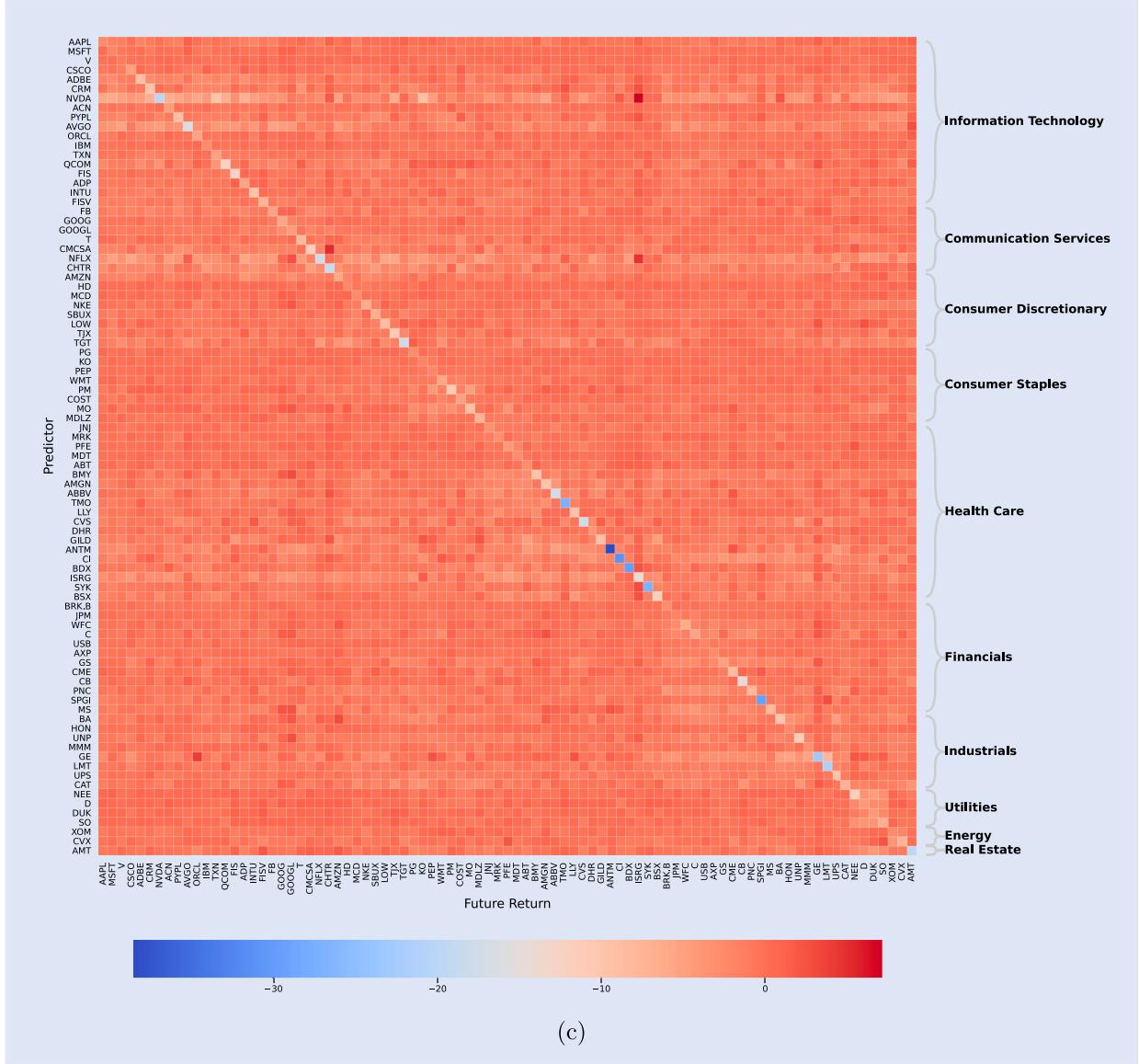


Figure A2. Continued. (c) Coefficient matrix of CAR.

of its own OFIs ($\tau_{i,t}^{(h)}$), and (iii) the idiosyncratic components of the OFIs of other stocks ($\tau_{j,t}^{(h)}, j \neq i$). Finally, we arrive at the cross-impact model proposed by Capponi and Cont (2020) in equation (A3), denoted as \mathbf{CI}^{CC} .

$$\mathbf{CI}^{CC} : r_{i,t}^{(h)} = \alpha_i^{CC} + \beta_{i0}^{CC} F_{\text{ofi},t}^{(h)} + \sum_{j=1}^N \beta_{ij}^{CC} \tau_{j,t}^{(h)} + \eta_{i,t}^{CC}. \quad (\text{A3})$$

\mathbf{CI}^{CC} is compared with a parsimonious model \mathbf{PI}^{CC} (equation (A4)), in which only the common order flow factor and a stock's own idiosyncratic OFI are utilized.

$$\mathbf{PI}^{CC} : r_{i,t}^{(h)} = \alpha_i^{CC} + \beta_{i0}^{CC} F_{\text{ofi},t}^{(h)} + \beta_{ii}^{CC} \tau_{i,t}^{(h)} + \epsilon_{i,t}^{CC}. \quad (\text{A4})$$

We estimate the \mathbf{PI}^{CC} and \mathbf{CI}^{CC} models on historical data, under the same setting as in Section 3.2. Given that there are more features than observations, we employ LASSO in the second step to testify the intraday cross-impact of the idiosyncratic OFIs.

Similarly, we present the both in-sample and out-of-sample R^2 values of \mathbf{PI}^{CC} and \mathbf{CI}^{CC} in table A5. We observe small improvements (1.37% in in-sample tests, 0.58% in out-of-sample tests) from \mathbf{PI}^{CC} to \mathbf{CI}^{CC} . From considering tables 3, 5, and A5, we also observe

that introducing the common factor leads to quite small changes in the model's explanatory power of price dynamics in the in-sample and out-of-sample tests. Moreover, the models employing integrated OFIs continually outperform others.

Capponi and Cont (2020) claim that the main determinants of impact are from idiosyncratic order flow imbalance as well as a market order flow factor common across stocks; we conclude that as long as the multi-level OFIs are included, additional cross-impact terms are not necessary. The results also reveal that the sparse price impact model with integrated (or multi-level) OFIs can explain the price dynamics better than the models proposed by Capponi and Cont (2020).

Appendix 4. High-frequency updates of contemporaneous models

In this experiment, we use a 30-minute window to estimate contemporaneous models. We then apply the estimated coefficients to fit data in the next one minute, and repeat this procedure every minute. The results summarized in table A6 reveal similar conclusions as in Section 3, illustrating the robustness of our findings.

Table A1. Average percentage and the standard deviation (in parentheses) of variance attributed to SA across multiple levels.

Average across	1	2	3	4	5	6	7	8	9	10
Explained Variance Ratio	13.82 (9.04)	22.81 (10.09)	32.53 (10.35)	42.73 (10.17)	52.58 (9.58)	61.45 (8.69)	68.82 (7.71)	75.10 (7.10)	80.49 (7.23)	85.07 (8.14)

Notes: The table reports the ratio (in percentage points) between the variance of SA across multiple levels and the total variance averaged across each stock and trading day.

Table A2. Explained variance ratio by different integration ways of multi-level OFIs, sorted by stock characteristics.

		Volume	Volatility	Spread
[0%, 25%)	SA	83.02	85.09	81.84
	PC	85.79	89.64	89.90
[25%, 50%)	SA	87.33	84.75	87.97
	PC	89.93	89.12	91.17
[50%, 75%)	SA	86.93	85.23	87.57
	PC	90.76	89.03	89.81
[75%, 100%]	SA	83.55	85.23	83.03
	PC	90.23	88.48	85.53

Notes: SA: taking the simple average of OFIs across 10 levels. PC: taking the first principal component of OFIs across 10 levels. Volume: trading volume on the previous trading day. Volatility: volatility of one-minute returns during the previous trading day. Spread: average bid-ask spread during the previous trading day. [0%, 25%), respectively [75%, 100%], denote the subset of stocks with the lowest, respectively highest, 25% values for a given stock characteristic.

Table A3. Out-of-sample performance of models based on different aggregations of multi-level OFIs.

	PI^{SA}	CI^{SA}	PI^I	CI^I
OS R^2	82.34 (18.02)	82.51 (14.27)	83.83 (16.90)	83.62 (14.53)

Notes: **PI^{SA}** (resp. **CI^{SA}**): price impact (resp. cross-impact) model using the simple average of OFIs across 10 levels.

Appendix 5. Additional results of Section 4

One interesting question to consider when examining the predictability of cross-impact is whether the lead-lag effect is linked to the frequency of LOB updates. As shown in table A7, the results indicate that assets with a higher frequency of book updates tend to lead ‘slower’ assets, which aligns with the findings reported by Kolm *et al.* (2023).

Table A5. Performance of CC’s models.

	PI^{CC}	CI^{CC}
IS R^2	72.58 (13.22)	73.95 (12.56)
OS R^2	64.78 (19.95)	65.36 (18.68)

Notes: The table reports the mean values and standard deviations (in parentheses) of both in-sample and out-of-sample R^2 (in percentage points) of **PI^{CC}** and **CI^{CC}** when modeling contemporaneous returns. These statistics are averaged across each stock and each regression window.

Table A6. Performance of various models under *one-minute update frequency*.

	Best-level OFIs		Integrated OFIs	
	PI^[1]	CI^[1]	PI^I	CI^I
IS R^2	70.80 (13.10)	73.55 (12.73)	86.10 (9.64)	86.84 (8.79)
OS R^2	59.67 (23.15)	61.46 (18.96)	78.88 (16.78)	78.91 (15.02)

Notes: The table reports the mean values and standard deviations (in parentheses) of both in-sample and out-of-sample R^2 (in percentage points) of various models when modeling contemporaneous returns in one-minute update frequency. The models include **PI^[1]** (equation (6)), **CI^[1]** (equation (8)), **PI^I** (equation (7)), and **CI^I** (equation (9)). These statistics are averaged across each stock and each regression window.

Table A7. Out-of-sample R^2 of various predictive models sorted by book updates.

	[0%, 25%)	[25%, 50%)	[50%, 75%)	[75%, 100%)
FPI ^[1]	-0.38	-0.37	-0.36	-0.34
FCI ^[1]	-0.12	-0.11	-0.10	-0.09
FPI ^I	-0.37	-0.35	-0.35	-0.33
FCI ^I	-0.12	-0.11	-0.10	-0.09

Notes: [0%, 25%), respectively [75%, 100%], denote the subset of stocks with the lowest, respectively highest, 25% values according to the frequencies of book updates.

Table A4. Performance of price impact models with multi-level OFIs.

	PI^[1]	PI^[2]	PI^[3]	PI^[4]	PI^[5]	PI^[6]	PI^[7]	PI^[8]	PI^[9]	PI^[10]
IS R^2	71.16 (13.80)	81.61 (11.80)	85.07 (10.76)	86.69 (10.30)	87.66 (10.05)	88.30 (9.86)	88.74 (9.71)	89.04 (9.57)	89.24 (9.45)	89.38 (9.34)
OS R^2	64.64 (21.82)	75.81 (19.83)	79.47 (18.87)	81.13 (18.61)	82.05 (18.58)	82.65 (18.65)	83.01 (18.78)	83.16 (18.93)	83.15 (19.49)	83.11 (20.93)

Notes: The table reports the mean values and standard deviations (in parentheses) of both in-sample and out-of-sample R^2 (in percentage points) of **PI^[m]** ($m = 1, \dots, 10$) when modeling contemporaneous returns. These statistics are averaged across each stock and each regression window.