Cross-Impact of Order Flow for Optimal Market-making

Aslanyan Artak

aaaslanyan_2@edu.hse.ru

Оглавление

Abstract

In this study, we investigated the effects of order book imbalance (OBI) and trade flow imbalance (TFI) on multiple stocks using synthetic data generated by a simulator. Our analysis involved exploring various levels and windows of parameters to understand cross-impact relationships. We systematically examined how different levels of OBI and TFI influence price movements and market dynamics across a range of stocks. By utilizing advanced modeling techniques and simulations, we aimed to uncover insights into how order flow imbalances at different levels and time intervals impact the overall market behavior and individual stock performance. Our findings shed light on the intricate interplay between OBI, TFI, and cross-impact effects, providing valuable insights for market participants and researchers in understanding and modeling financial market dynamics.

1. Introduction

Accurately understanding and predicting the impact of trading activities on asset price movements is paramount for both industry practitioners and academic researchers. This encompasses areas such as analyzing trading costs and optimizing trade execution strategies. Over the years, extensive research has been conducted on price impact, which refers to the influence of trades on asset prices. In more intricate market settings involving multiple assets, researchers have delved into the concept of cross-impact, aiming to discern how trading one asset influences the prices of other assets.

Many studies have scrutinized the impact of order flow on returns, often investigating cross-correlation patterns to uncover insights. For instance, some studies have revealed that similarities in order flow can elucidate why certain stocks exhibit similar returns. However, it has been noted that the positive relationship between a stock's returns and order flow imbalances in other stocks may not consistently demonstrate cross-impact. Additionally, incorporating cross-impact terms into models might only marginally enhance performance, suggesting they could potentially be disregarded.

Our study builds upon this foundation by exploring the efficacy of cross-order flow analysis within a given sample and its predictive capabilities concerning future movements using diverse metrics. Particularly, there is limited research on how order flows influence price changes across multiple assets, especially when considering deeper levels within the limit order book (LOB). In this article, we delve into various metrics such as Order Book Imbalance (OBI) and Trade Flow Imbalance (TFI), which have been explored to a lesser extent in existing literature.

With the emergence of high-frequency trading (HFT), accurately gauging cross-impact on future intraday returns has gained increasing significance. Machine learning models, particularly deep neural networks, have exhibited promise in modeling stock returns and improving forecasting accuracy, paving the way for enhanced understanding and predictive capabilities in dynamic financial markets.

2. Description of strategy

2.1 Data

We are going to use simulator of trading environment to generate synthetic data. There was defined several types of trading agents and their behaviors in an agent-based market simulation environment:

- **Trader:** Represents a trading agent that operates within a single market or exchange. It places orders to buy and sell assets based on various strategies.
- **Multi Trader:** Extends trader’s functionality to operate in multiple markets or exchanges simultaneously. It manages trading activities across different markets.
- **Random:** A trading agent that places orders randomly, mimicking noise or random trading behavior in the market.
- **Fundamentalist:** This trading agent makes trading decisions based on fundamental analysis, considering factors like dividends and risk-free rates to determine the value of assets.
- **Chartist:** Uses sentiment analysis of price movements on a single exchange to make trading decisions. It may buy or sell based on perceived trends in asset prices.
- **Multi Chartist:** Extends the behavior of Chartist to operate across multiple exchanges. It analyzes price movements across different markets to make trading decisions.

- **Universalist:** Combines the strategies of Fundamentalist and Chartist, switching between them randomly. It adapts its trading behavior based on market conditions and the performance of different strategies.
- **Market Maker:** Acts as a liquidity provider on a single exchange by placing limit orders on both sides of the order book. It aims to narrow the bid-ask spread and improve market liquidity.
- **Multi Market Maker:** Extends the behavior of MarketMaker1D to operate across multiple exchanges. It provides liquidity in multiple markets by placing orders on different exchanges.

We will use metrics defined in the article [1] for further analysis

Order book Imbalance.

The literature extensively supports the idea that the orderbook's structure has a strong impact on the distribution of future returns [2]. This is simple to grasp: if there is a big excess of liquidity on the bid side versus the ask side, it requires less trading volume to increase the price than to decrease it. Hence, if the probability of receiving a sell order is not substantially greater than that of a buy order, the probability of a price increase in the future is greater than that of a decrease. The same logic applies when the orderbook structure is reversed as well.

Article [1] provides advanced Order Book Imbalance metric which shows better results compared to the classical notion that is given by

$$\mathrm{OBI}_{\mathrm{n,t}} = \frac{b_{n,t} - a_{n,t}}{b_{n,t} + a_{n,t}} \in [-1, 1], \tag{1}$$

However, we are going to use classical notion to investigate its correlation with price changing.

Where *a* and *b* represent the sum of the top ask and bid liquidity for the top *n*-prices at the time *t*, respectively. For example, we are going to use top 10 prices. In some articles was observed the deeper prices forecast more long-term dependencies between metric and price changing.

Trade Flow Imbalance.

Article [1] also defines metric which is called trade imbalance and given by

$$TFI_t^{i,\delta} = B^i_{[t-\delta,t]} - S^i_{[t-\delta,t]}, \tag{2}$$

Where $B^i_{[t-\delta,t]}$ represents volume of buy orders on market i in the time interval $[t - \delta, t]$. Similarly, the quantity $S^i_{[t-\delta,t]}$ represents the total volume of sell orders on market i over the same time interval.

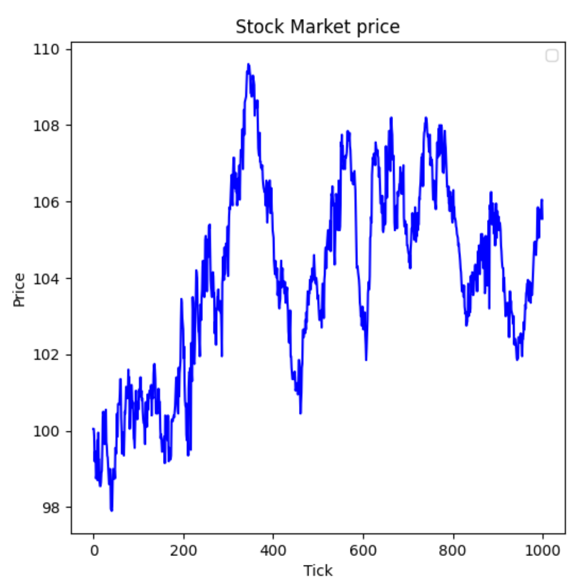
This feature was demonstrated in [2] to have significant explanatory power over contemporaneous returns.

We are going to generate a training dataset with 5000 iterations and a test dataset with 2500 iterations, respectively. Configuration of market is represented below:

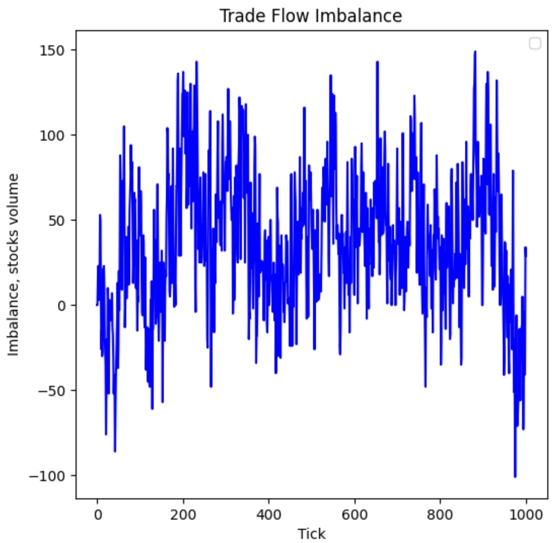
Stocks quantity	Risk free rate	Start price	Dividends
10	5e-4	100	5e-2

Figure 1: Here are represented the features generated by simulator on 1000 ticks for one stock with the same configuration.

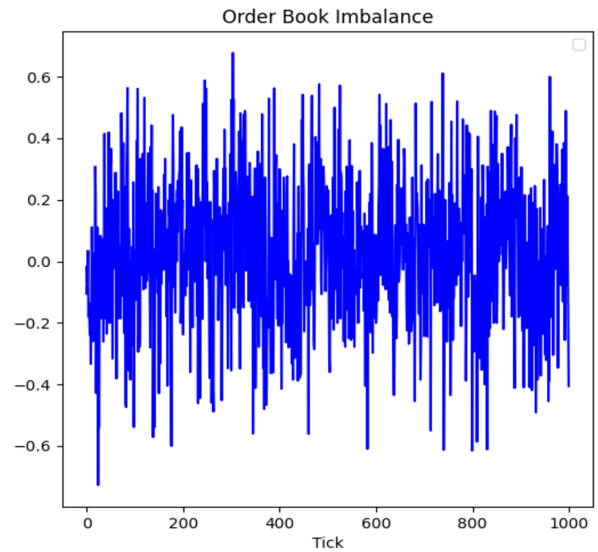
a) Stock market price changing



(b) TFI of 1 level



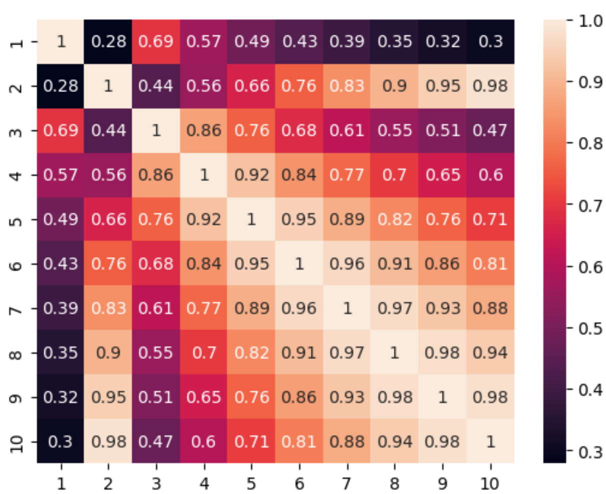
(c) OBI with 1 tick window



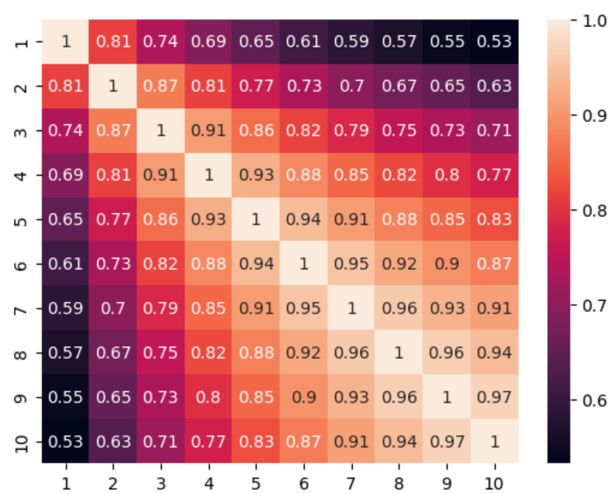
2.2 Summary Statistics

Figure 1 represents correlation matrix of multi-window OBIs and multi-level TFIs, respectively. The OBI correlation matrix doesn't show strong relationship between long distance between further windows (almost 30%), but at the same time TFI's

first level exhibits over 50% correlation with any of the remaining nine levels.



a) OBI



b) TFI

Figure 1: Correlation matrix of multi-levels and multi-windows

As we can observe from Table 1 metrics show that the first two principal components (PCA) explain more than 85 percent of the total variance, so we will use only them in further analysis to reduce overfitting.

Table 1: Average percentage and the standard deviation of variance attributed to each principal component considering various levels and windows for TFIs and OBIs, respectively.

a) OBI

PCA	1	2	3	4	5	6	7	8	9	10
Explained Variance Ratio	0,68	0.2	0,07	0.024	9e-3	4e-3	2e-3	1e-3	6e-3	3e-3
Explained Variance	0.46	0.14	0.04	1e-2	6e-3	2e-3	1e-3	7e-4	4e-4	2e-4

b) TFI

PCA	1	2	3	4	5	6	7	8	9	10
Explained Variance Ratio	0,88	0.05	0,02	0.01	7e-3	6e-3	5e-3	4e-3	3e-3	3e-3
Explained Variance	156	10.4	3.5	1.8	1.2	1	0.86	0.7	0.64	0.63

3. Contemporaneous cross-impact

In this section, we study the results of different models including basic ML approaches by comparing cross impact using various metrics of remaining stocks for predicting price of the first one.

3.1 Models

In the section below, we will pay attention to the impact of metrics separately and together using cross-impact of remaining stock's features.

Price and Cross impact of integrated OBIs.

Table 2: R^2 score using features of first n stocks (0 means we use only initial stock's features) for OBI.

	0	1	2	3	4	5	6	7	8	9
Linear Regression	-0.003	-0.004	-0.003	-0.003	-0.003	-0.001	-0.003	-0.003	-0.004	-0.004
Random Forest	-0.008	-0.003	-0.010	-0.011	-0.010	-0.008	-0.003	-0.000	-0.009	-0.005
Gradient Boosting	-0.023	-0.023	-0.037	-0.036	-0.024	-0.029	-0.032	-0.039	-0.031	-0.017

In Table 2, the R^2 scores for different machine learning models (Linear Regression, Random Forest, and Gradient Boosting) are presented using features from the first n stocks, where n ranges from 0 to 9 (0 means only the initial stock's features are used).

The negative R^2 scores indicate that the models are not performing well in capturing the dependency between the metrics and the changes in price. This lack of correlation suggests that the features being used are not effectively explaining the variation in the target variable (price changes).

The problem with classical order book imbalance is highlighted when there's significant uncertainty about future returns, especially during times of high volatility. When effective spreads are large (the difference between bid and ask prices), the classical order book imbalance may not accurately reflect market dynamics.

To illustrate this problem, consider an example of an order book during non-volatile times and another shortly after a volatility spike. In non-volatile times, the classical order book imbalance might show a value like -0.98, indicating a strong bias towards a price move down due to the overwhelming ask liquidity compared to bid liquidity.

However, during high volatility (as seen in the second order book with a classical imbalance of 0.69 indicating a potential price move up), the imbalance value may not align with market intuition. This discrepancy occurs because the classical imbalance calculation doesn't consider the deeper levels of the order book where more significant liquidity may be present, thus affecting the probability of future price movements.

Price and Cross impact of integrated TFIs.

Table 3: R^2 score using features of first n stocks for TFI.

	0	1	2	3	4	5	6	7	8	9
Linear Regression	0.022	0.120	0.131	0.211	0.212	0.228	0.228	0.246	0.246	0.246
Random Forest	0.010	0.181	0.309	0.384	0.413	0.492	0.512	0.569	0.576	0.599
Gradient Boosting	-0.003	0.165	0.267	0.323	0.362	0.452	0.562	0.517	0.593	0.549

The scores of Linear Regression show a gradual increase, indicating that including more features (more stocks) of independent variables(stocks) improves the predictive power of the linear model. Furthermore, TFI shows impressive results of R^2 score compared to the results of article [1]. Our model outperformed model defined in the article mentioned before.

Nevertheless, looking at the R^2 scores for Gradient Boosting across different numbers of features (n stocks), we observe fluctuations in the scores:

- The scores start from -0.003 and gradually increase, reaching a peak of 0.593 with 9 features (9 stocks).
- However, there are fluctuations in the scores, such as the drop from 0.562 to 0.517 when moving from 7 to 9 features.

These fluctuations mean that Gradient Boosting and Random Forest started overfitting.

Price and Cross impact of integrated OBIs and TFIs.

Table 4: R^2 score using features of first n stocks for both metrics.

	0	1	2	3	4	5	6	7	8	9
Linear Regression	0.022	0.120	0.131	0.211	0.212	0.228	0.228	0.246	0.246	0.247
Random Forest	0.011	0.175	0.307	0.383	0.419	0.500	0.526	0.571	0.581	0.609
Gradient Boosting	0.002	0.165	0.271	0.314	0.349	0.470	0.525	0.517	0.536	0.557

The results are almost same we got from Table 3, observed spread in last two models is clearly explained by architectures of Random Forest and Gradient Boosting. Therefore, this model did not outperform previous model, because of OBI problem described after Table 2.

3.2 Discussion about contemporaneous cross-impact

4. Forecasting future returns

References

[1] Jakob Albers, Mihai Cucuringu, Sam Howison, Alexander Y. Shestopalof. Fragmentation, Price Formation, and Cross-Impact in Bitcoin Markets

[2] Rama Cont, Arseniy Kukanov, and Sasha Stoikov. The price impact of order book events. Journal of financial econometrics, 12(1):47–88, 2014.

[3] Rama Cont. Mihai Cucuringu and Chao Zhang. Cross-Impact of Order Flow Imbalance in Equity Mar kets