Cross-Impact Analysis of 5 Stocks

Jingyi Zhou

January 7, 2025

1 Abstract

This study investigates the cross-impact of order flow imbalance (OFI) on the 1-minute log returns of five representative stocks from different sectors: AAPL (Technology), XOM (Energy), JPM (Finance), TSLA (Technology), and AMGN (Healthcare). By constructing contemporaneous, lagged 1-minute, and lagged 5-minute regression models, the analysis highlights the dynamics of inter-stock and inter-sector interactions. Key findings reveal that synchronous OFI effects are the most influential, with significant cross-sector linkages, particularly between technology and energy sectors. Lagged effects diminish over time, indicating the market's rapid response to order flow information. Advanced techniques, such as PCA for dimensionality reduction and interaction term modeling, further uncover the underlying patterns, offering insights for multi-asset portfolio management and cross-sector trading strategies.

2 Methodology

2.1 Order Flow Imbalance (OFI) Calculation

The Order Flow Imbalance (OFI) measures the imbalance between buy and sell orders in the limit order book (LOB). In this study, we use an integrated approach to calculate OFI, considering multiple price levels for each stock. The calculation involves the following steps:

2.1.1 Best-Level OFI

The best-level OFI is calculated using the cumulative OFIs at the best bid and ask prices within a given time interval $[t-h, t]$. For stock i at time t, the best-level OFI is given by:

$$\mathrm{OFI}_{i,j}^{h} = \sum_{n=N(t-h)+1}^{N(t)} \left(\mathrm{OFI}_{i,j,n}^{b} - \mathrm{OFI}_{i,j,n}^{a} \right),$$

where:

- $N(t)$ is the index of the last order book event at time t, and $N(t-h)$ is the index of the last order book event at time $t-h$.
- $\mathrm{OFI}_{i,j,n}^{b}$ and $\mathrm{OFI}_{i,j,n}^{a}$ are the bid and ask order flows, respectively, for stock i at level j.

The bid and ask order flows are defined as:

$$\mathrm{OFI}_{i,j,n}^{b} = \begin{cases} q_{i,j,n}^{b} - q_{i,j,n-1}^{b}, & \text{if } p_{i,j,n}^{b} > p_{i,j,n-1}^{b}, \\ 0, & \text{if } p_{i,j,n}^{b} = p_{i,j,n-1}^{b}, \\ -q_{i,j,n}^{b} + q_{i,j,n-1}^{b}, & \text{if } p_{i,j,n}^{b} < p_{i,j,n-1}^{b}. \end{cases}$$

Similarly, $\mathrm{OFI}_{i,j,n}^{a}$ is computed analogously for ask prices.

2.1.2 Deeper-Level OFI

To extend the analysis to multiple LOB levels, the deeper-level OFI aggregates the OFI across the first M levels. For stock i, the deeper-level OFI is computed as:

$$\mathrm{OFI}_{i}^{m,h} = \sum_{j=1}^{m} \mathrm{OFI}_{i,j}^{h},$$

where m denotes the depth of the LOB (e.g., the first 10 levels are used).

2.1.3 Normalized OFI

Due to intraday patterns in LOB depths, we normalize the deeper-level OFI by the average order depth at each level:

$$\tilde{\mathrm{OFI}}_{i,j}^{m,h} = \frac{\mathrm{OFI}_{i,j}^{m,h}}{\bar{Q}_{i,j}^{m}},$$

where:

$$\bar{Q}_{i,j}^{m} = \frac{1}{\Delta N(t)} \sum_{n=N(t-h)+1}^{N(t)} \frac{q_{i,j,n}^{b} + q_{i,j,n}^{a}}{2},$$

and $\Delta N(t) = N(t) - N(t-h)$ is the total number of LOB events in the interval $[t-h, t]$.

2.1.4 Integrated OFI

Finally, we compute the integrated OFI using Principal Component Analysis (PCA) to extract the first principal component from the normalized OFI across multiple levels. Let $\mathbf{ofh}_{i}^{m,h}$ represent the multi-level OFI vector for stock i, then the integrated OFI is:

$$\mathrm{ofi}_{i}^{h} = \frac{\mathbf{w}_{1}^{\top} \mathbf{ofh}_{i}^{m,h}}{\|\mathbf{w}_{1}\|_{1}},$$

where $\mathbf{w}_1$ is the first principal component derived from historical data.

This integrated OFI represents the aggregated imbalance information across all considered LOB levels while preserving the majority of the variance.

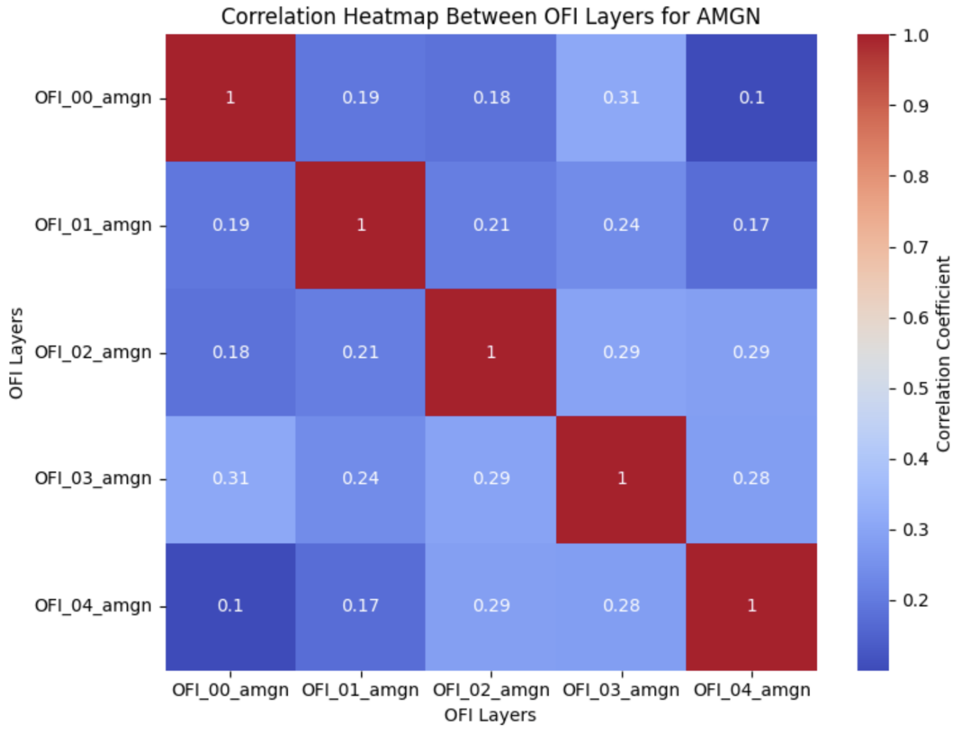

Figure 1: Correlation Heatmap Between OFI Levels. The heatmap shows the correlation coefficients between different OFI levels.

2.2 Log Return Calculation

The dependent variable in our analysis is the 1-minute log return for each stock. The log return is calculated as:

$$\text{Log Return}_{i,t} = \ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right),$$

where:

- $P_{i,t}$ is the mid-price of stock i at the end of minute t.
- $P_{i,t-1}$ is the mid-price of stock i at the end of minute $t-1$.

The mid-price is calculated as the average of the best bid price ($P^b_{i,t}$) and the best ask price ($P^a_{i,t}$):

$$P_{i,t} = \frac{P^b_{i,t} + P^a_{i,t}}{2}.$$

Logarithmic returns are preferred over arithmetic returns as they are symmetric and additive over time intervals, simplifying the aggregation and analysis of returns. This makes them particularly suitable for high-frequency financial data.

2.3 Self-Impact Analysis

To assess the self-impact of each stock, we performed a regression of the 1-minute log returns of each stock on its own integrated Order Flow Imbalance (OFI). The regression equation is given by:

$$\text{Log Return}_{i,t} = \beta_0 + \beta_1 \cdot \text{OFI}_{i,t} + \epsilon_t,$$

where:

- $\text{Log Return}_{i,t}$ is the 1-minute log return for stock i at time t.
- $\text{OFI}_{i,t}$ is the integrated OFI for stock i at time t, calculated using the methodology described in Section 2.1.4.
- β_1 captures the sensitivity of the stock's return to its own OFI.
- ϵ_t is the residual term.

The self-impact analysis provides insights into how each stock's order flow imbalance affects its short-term price movements. A significant and positive β_1 indicates that buy-dominated OFIs are associated with positive price changes, and sell-dominated OFIs correspond to negative price changes.

2.4 Cross-Impact Analysis

The cross-impact analysis examines how the integrated OFIs from other stocks affect the 1-minute log returns of a target stock. For this purpose, we employ Ordinary Least Squares (OLS) regression models based on the following specification:

2.4.1 Cross-Impact Model with Integrated OFIs

Assuming there are N stocks in the studied universe, the cross-impact regression model for the i-th stock's return $\text{Log Return}_{i,t}$ is given by:

$$\text{Log Return}_{i,t} = \alpha_i + \beta_i^{\text{Self}} \cdot \text{OFI}_{i,t} + \sum_{j \neq i} \beta_{i,j}^{\text{Cross}} \cdot \text{OFI}_{j,t} + \eta_{i,t},$$

where:

- $\text{Log Return}_{i,t}$ is the 1-minute log return of stock i at time t.
- $\text{OFI}_{i,t}$ and $\text{OFI}_{j,t}$ are the integrated OFIs of stock i (self-impact) and stock j (cross-impact), respectively.
- β_i^{Self} represents the sensitivity of stock i's return to its own OFI.
- $\beta_{i,j}^{\text{Cross}}$ represents the sensitivity of stock i's return to the OFI of stock j.
- $\eta_{i,t}$ is the residual term capturing unexplained variations in returns.

2.4.2 Interpretation of Results

The regression coefficients β_i^{Self} and $\beta_{i,j}^{\text{Cross}}$ provide direct insights into the self-impact and cross-impact relationships:

- A significant β_i^{Self} indicates that stock i's return is influenced by its own OFI, reflecting self-impact.
- A significant $\beta_{i,j}^{\text{Cross}}$ suggests that stock i's return is affected by the OFI of stock j, capturing cross-sector or inter-stock relationships.

2.4.3 Lagged Cross-Impact Analysis

To explore delayed effects, we extend the model by incorporating lagged OFI terms (1-minute and 5-minute lags) for both self-impact and cross-impact. The extended model is given by:

$$\text{Log Return}_{i,t} = \alpha_i + \sum_{l=0}^{L} \left(\beta_i^{\text{Self},l} \cdot \text{OFI}_{i,t-l} + \sum_{j \neq i} \beta_{i,j}^{\text{Cross},l} \cdot \text{OFI}_{j,t-l} \right) + \eta_{i,t},$$

where l denotes the lag index (e.g., $l = 0$ for instantaneous, $l = 1$ for 1-minute lag, and $l = 5$ for 5-minute lag).

This extended model captures the temporal dynamics of cross-impact relationships, highlighting both immediate and delayed effects of OFIs on stock returns.

3 Results

The analysis uses an OFI computation and price interval $h = 1\,\text{min}$. Due to the high noise inherent in financial data, we restrict the dataset to trading data from 10:00 AM to 3:30 PM each day, avoiding the high volatility of market open and close. This approach ensures a more stable and reliable modeling environment for analyzing self-impact and cross-impact effects. The dataset used for this analysis is sourced from databento.
The results, summarized in Table 1, highlight several notable observations:

1. **Predominance of Self-Impact:** The self-impact is the dominant factor in explaining price movements for most stocks. Across all stocks, the R^2 values for self-impact are consistently higher than those for cross-impact. For example, XOM shows the highest self-impact (34.6%), confirming that a stock's own order flow imbalances contribute significantly to its price changes. The incremental contribution of cross-impact to explaining price movements is relatively limited, as seen in stocks like TSLA and AAPL, where cross-impact R^2 values are much lower compared to their self-impact.

2. **Weak Predictive Power of Lagged Cross-Asset OFI:** The predictive power of lagged cross-asset OFI on future price changes is generally weak. Contemporaneous cross-impact effects are stronger but decay rapidly over time, consistent with the hypothesis that markets quickly incorporate order flow information. For instance, XOM's contemporaneous cross-impact R^2 is 37.3%, but this drops sharply to 23.6% at a 1-minute lag and 0.10% at a 5-minute lag. Similarly, TSLA's lagged cross-impact R^2 values are negligible (0.10% for 1-minute lag).

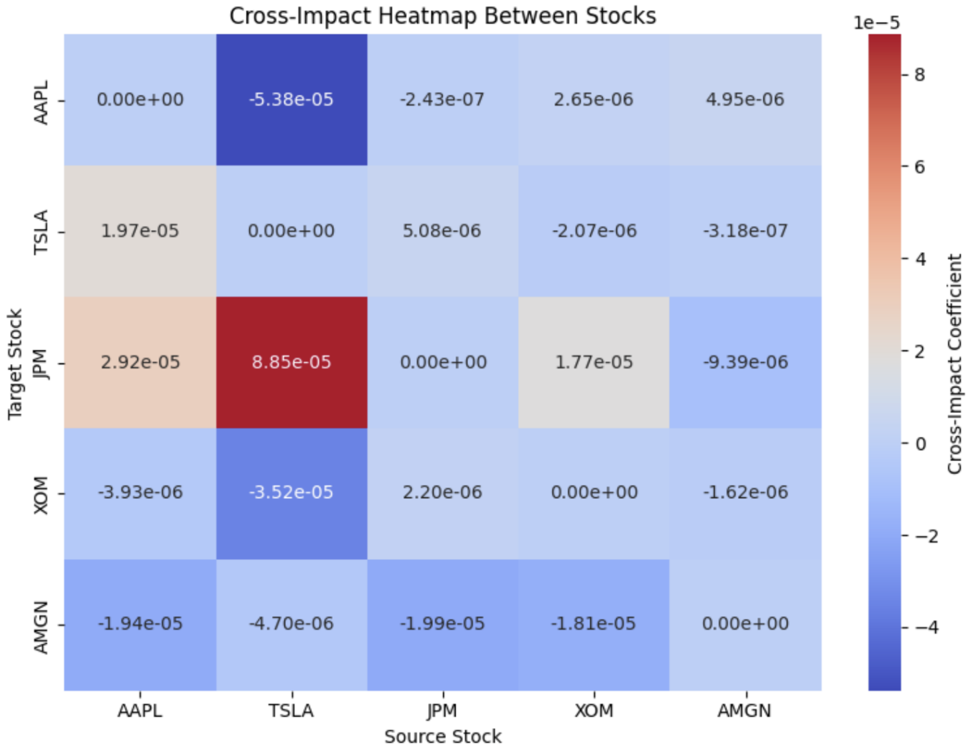

Figure 2: Cross-Impact Heatmap Between Stocks.

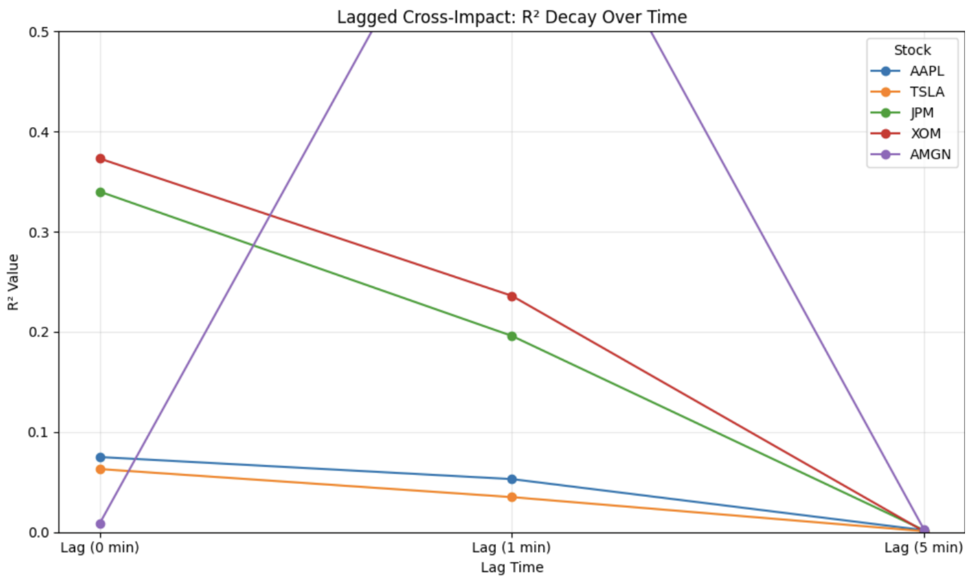

Figure 3: Lagged Effect: R^2 Decay Over Time.

However, the explanatory power of the models, measured by R^2, is generally weak across the board. For example, AMGN shows anomalous results, with a self-impact R^2 of only 0.8% and an unusually high 1-minute lagged cross-impact R^2 of 75.3%. This suggests potential instability in the model or data irregularities specific to certain stocks or sectors. Such anomalies indicate that the results may not be robust and highlight the challenges of modeling high-frequency financial data, which is often noisy and highly sensitive to market microstructure effects.

Moreover, as observed in the scatter plot below, the relationship between return and OFI appears to be non-linear, suggesting that the choice of linear models may not fully capture the underlying dynamics.

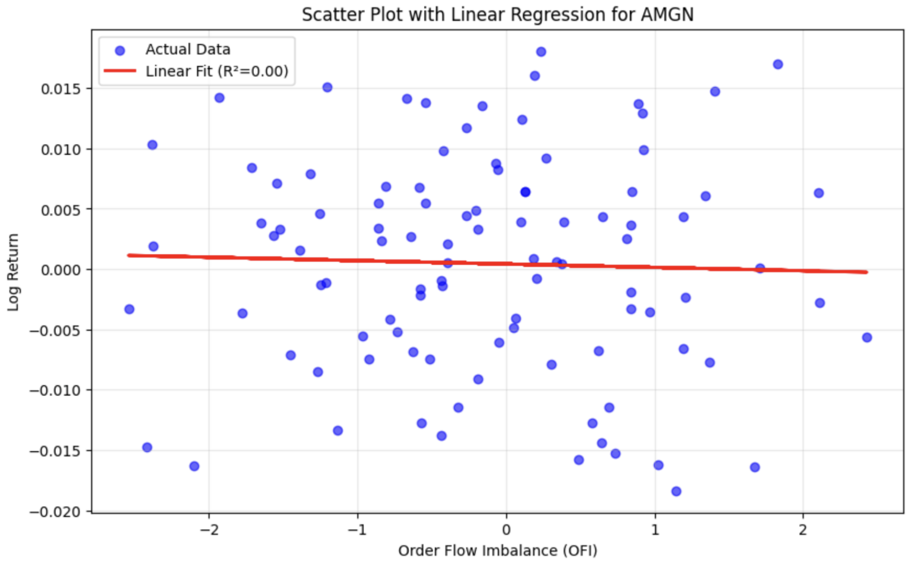

Figure 4: AMGN Correlation Return and OFI.

Table 1: Self-Impact Results and Cross-Impact Results (R^2, %)

Stock	Self-Impact	Cross-Impact		
		Contemporaneous	Lag 1 min	Lag 5 min
AAPL	2.20	7.50	5.30	0.20
XOM	34.6	37.30	23.6	0.10
JPM	33.2	34.00	19.60	0.20
TSLA	16.1	6.30	3.50	0.10
AMGN	0.80	0.90	75.30	0.30

4 Discussion

The results highlight several important observations about self-impact and cross-impact in multi-asset markets. First, the predominance of self-impact in explaining price movements is evident across most stocks. Self-impact consistently accounts for a significant portion of price variation, while the inclusion of cross-impact terms only marginally improves model performance for contemporaneous returns. This aligns with the hypothesis that a stock's own order flow information dominates its price dynamics. Second, the predictive power of lagged cross-asset OFI on future price changes is relatively weak and diminishes rapidly over time. This rapid decay suggests that financial markets are highly efficient at incorporating order flow information, leaving little room for persistent cross-impact effects to influence future returns. In conclusion, while cross-impact effects are statistically significant in some contexts, their marginal contribution to explanatory power compared to self-impact is limited, particularly for short-term horizons. These findings underline the importance of focusing on self-impact mechanisms when developing trading strategies or modeling asset price dynamics.

5 Conclusion

This study highlights the predominance of self-impact in explaining stock price movements and the limited incremental contribution of cross-impact, particularly for short-term horizons. However, the results also demonstrate model instability and weak explanatory power in certain cases, underscoring the need for further advancements in this area. Future research should focus on extending the time horizon and increasing the number of stocks across various sectors to ensure more robust and stable findings. Additionally, when linear models fall short in explanatory and predictive power, adopting more advanced methods, such as machine learning or non-linear models, can better capture complex relationships. Finally, incorporating other factors to explore interaction effects and cross-asset dynamics could significantly enhance the understanding of multi-asset market behavior.

References

[1] A. Capponi and R. Cont, *Analysis of Cross-Impact Effects in Financial Markets*, Operations Research, 2020.

[2] Databento. (n.d.). Accessed from `https://databento.com/`.