Order Flow Imbalance Analysis: Methodology and Insights

Benjamin EMILY

***Abstract*— This study analyzes the self- and cross-impact of normalized and integrated order flow imbalances (OFIs) on asset returns. Using regression models and clustering techniques, we uncover the predictive power of OFIs and their implications for trading strategies. Key findings include IS R^2 or OS R^2 values for self-impact and cross-impact, variance explained by PCA, and clustering insights for asset relationships.**

I. Data and variables

A. Data

We use the Nasdaq ITCH data from DATABENTO to compute the independent and dependent variables. Our data includes the 5 stocks: AAPL, AMGN, JPM, TSLA and XOM, existing from 2024-11-04 to 2024-11-08.

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 demon-strated 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 my 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 ($OF^{m,b}_{i,n}$) and ask order flows ($OF^{m,a}_{i,n}$) of stock i at level m at time n as follows:

$$OF^{m,b}_{i,n} = \begin{cases} q^{m,b}_{i,n}, & \text{if } P^{m,b}_{i,n} > P^{m,b}_{i,n-1}, \\ q^{m,b}_{i,n} - q^{m,b}_{i,n-1}, & \text{if } P^{m,b}_{i,n} = P^{m,b}_{i,n-1}, \\ -q^{m,b}_{i,n}, & \text{if } P^{m,b}_{i,n} < P^{m,b}_{i,n-1}, \end{cases}$$

$$OF^{m,a}_{i,n} = \begin{cases} -q^{m,a}_{i,n}, & \text{if } P^{m,a}_{i,n} > P^{m,a}_{i,n-1}, \\ q^{m,a}_{i,n} - q^{m,a}_{i,n-1}, & \text{if } P^{m,a}_{i,n} = P^{m,a}_{i,n-1}, \\ q^{m,a}_{i,n}, & \text{if } P^{m,a}_{i,n} < P^{m,a}_{i,n-1}. \end{cases}$$

Here, $P^{m,b}_{i,n}$ and $q^{m,b}_{i,n}$ denote the bid price and size (in number of shares) of stock i at level m, respectively. Similarly, $P^{m,a}_{i,n}$ and $q^{m,a}_{i,n}$ denote the ask price and ask size at level m, respectively.

Note that the variable $OF^{m,b}_{i,t}$ is positive when:

- The bid price increases.
- The bid price remains the same and the bid size increases.

$OF^{m,b}_{i,t}$ is negative when:

- The bid price decreases.
- The bid price remains the same and the bid size decreases.

An analogous analysis and interpretation apply for the ask order flows $OF^{m,a}_{i,t}$.

1) Best-level OFI: The best-level OFI 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:

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

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

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

$$OFI^{m,h}_{i,t} = \sum_{n=N(t-h)+1}^{N(t)} \left(OF^{m,b}_{i,n} - OF^{m,a}_{i,n} \right). \tag{2}$$

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:

$$ofi^{m,h}_{i,t} = \frac{OFI^{m,h}_{i,t}}{Q^{M,h}_{i,t}}, \tag{3}$$

where

$$Q^{M,h}_{i,t} = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{2\Delta N(t)} \sum_{n=N(t-h)+1}^{N(t)} \left(q^{m,b}_{i,n} + q^{m,a}_{i,n} \right)$$

is the average order book depth across the first M levels, and $N(t) = N(t) - N(t-h)$ is the number of events during the interval $(t-h,t]$.

In this paper, we consider the top $M = 10$ levels of the limit order book (LOB) and denote the multi-level OFI vector as:

$$\mathbf{ofi}^{(h)}_{i,t} = \left(ofi^{1,h}_{i,t}, \ldots, ofi^{10,h}_{i,t} \right)^{\top}.$$

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 (1), which only preserves the first principal component. We further normalize the first principal component by dividing it by its ℓ_1 norm so that the weights of multi-level OFIs in constructing integrated OFIs sum to 1, leading to:

$$\mathrm{ofi}_{i,t}^{I,h} = \frac{\mathbf{w}_1^\top \cdot \mathrm{ofi}_{i,t}^{(h)}}{\|\mathbf{w}_1\|_1}, \tag{1}$$

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.

4) Logarithmic returns: Our dependent variable is the logarithmic asset return. Specifically, we define the returns over the interval $(t-h, t]$, we define the return as follows:

$$r_{i,t}^{(h)} = \log\left(\frac{P_{i,t}}{P_{i,t-h}}\right), \tag{2}$$

where $P_{i,t}$ is the mid-price at time t, defined as:

$$P_{i,t} = \frac{P_{i,t}^{1,b} + P_{i,t}^{1,a}}{2}. \tag{3}$$

B. Summary Statistics and Correlation Analysis of Multi-level OFIs

Table I presents descriptive statistics for the multi-level normalized OFIs across the top 5 selected stocks in the dataset. These statistics (mean, standard deviation, minimum, and maximum) are computed at the minute level and aggregated across trading days and stocks. The table reveals notable differences in the distribution of OFIs at different levels of the order book. For instance, the standard deviation increases for deeper levels, indicating higher variability in order flow at those levels.

Figure **??** shows the correlation matrices for the multi-level OFIs of the selected stocks. While the exact correlation structure varies across stocks, all demonstrate relatively strong relationships (above 60% on average) between levels. Interestingly, the best-level OFI (nOFI_0) exhibits the smallest correlation with the deeper levels, a consistent pattern across all analyzed stocks. This suggests that the best-level OFI captures unique dynamics distinct from the deeper levels.

The first principal component explains approximately 65.02% of the total variance in the multi-level OFIs across stocks.

TABLE I: Summary Statistics of Multi-level Normalized OFIs

Metric	nOFI_0	nOFI_1	nOFI_2	nOFI_3	nOFI_4
Mean	0.0527	0.0105	-0.0180	-0.0281	-0.0226
Std	1.3053	1.3665	1.4527	1.5473	1.6272
Min	-247.20	-113.28	-121.30	-78.20	-86.83
25%	-0.1083	0.0000	0.0000	0.0000	0.0000
50%	0.0000	0.0000	0.0000	0.0000	0.0000
75%	0.1015	0.0000	0.0000	0.0000	0.0000
Max	639.45	87.89	68.65	135.75	182.96

II. Contemporaneous cross-impact

In this section, we study the existence of contemporaneous cross-impact

A. Models

1) Price impact of best-level OFIs: We focus initially on the contemporaneous price impact of the best-level OFI ($\mathrm{ofi}_{i,t}^{1,h}$) on returns ($r_{i,t}^{h}$) within the same time interval using the following regression model:

$$\mathrm{PI}^{[1]} : r_{i,t}^{h} = \alpha_i^{[1]} + \beta_i^{[1]} \mathrm{ofi}_{i,t}^{1,h} + \epsilon_{i,t}^{[1]} \tag{6}$$

Here, $\alpha_i^{[1]}$ and $\beta_i^{[1]}$ represent the intercept and slope coefficients, respectively, while $\epsilon_{i,t}^{[1]}$ is a noise term that accounts for other influences, such as the deeper-level OFIs or cross-stock trading behavior. This regression model is referred to as $\mathrm{PI}^{[1]}$ and is estimated using ordinary least squares (OLS).

2) Price Impact of Integrated OFIs: The second model specification considers the effect of multi-level OFIs by leveraging the integrated OFIs. The relationship is specified as follows and is estimated using ordinary least squares (OLS):

$$\mathrm{PI}^{\mathrm{I}} : r_{i,t}^{h} = \alpha_i^{\mathrm{I}} + \beta_i^{\mathrm{I}} \mathrm{ofi}_{i,t}^{\mathrm{I},h} + \epsilon_{i,t}^{\mathrm{I}} \tag{7}$$

3) Cross-Impact of Best-Level OFIs: For a universe of N stocks, we extend the analysis to include the multi-asset best-level OFIs ($\mathrm{ofi}_{j,t}^{1,h}$, $j = 1, \ldots, N$) as candidate features to model the returns of stock i ($r_{i,t}^{h}$). The self-impact (from stock i) and cross-impact (from other stocks $j \neq i$) are incorporated as follows:

$$\mathrm{CI}^{[1]} : r_{i,t}^{h} = \alpha_i^{[1]} + \beta_{i,i}^{[1]} \mathrm{ofi}_{i,t}^{1,h} + \sum_{j \neq i} \beta_{i,j}^{[1]} \mathrm{ofi}_{j,t}^{1,h} + \eta_{i,t}^{[1]} \tag{8}$$

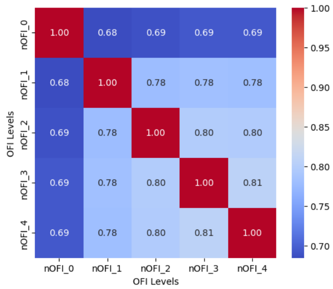
Here, $\beta_{i,j}^{[1]}$ quantifies the influence of the best-level OFIs of stock j on the returns of stock i.

4) Cross-Impact of Integrated OFIs: Finally, we incorporate the cross-asset integrated OFIs to examine their impact on returns, extending the above framework to account for aggregated information from multi-level OFIs across stocks.

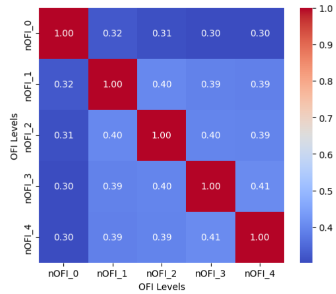
$$\mathrm{CI}^{\mathrm{I}} : r_{i,t}^{h} = \alpha_i^{\mathrm{I}} + \beta_{i,i}^{\mathrm{I}} \mathrm{ofi}_{i,t}^{\mathrm{I},h} + \sum_{j \neq i} \beta_{i,j}^{\mathrm{I}} \mathrm{ofi}_{j,t}^{\mathrm{I},h} + \eta_{i,t}^{\mathrm{I}}$$

B. Empirical results

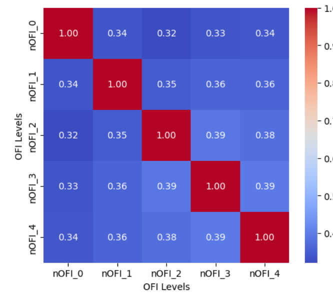
For a more representative and fair comparison with previous studies, we apply a similar procedure described in Cont et al. (2014) to my experiments. We exclude the first and last 30 minutes of the trading day due to the increased volatility near the opening and closing sessions, in line with Hasbrouck and Saar (2002), Chordia et al. (2002) and Chordia and Subrahmanyam (2004), Cont et al. (2014), Capponi and Cont (2020). In particular, returns and OFIs are computed for every minute.



AAPL



TSLA



AMGN

Figure 1 presents scatter plots of integrated OFI against price changes for the five selected stocks. These visualizations aim to provide a clearer understanding of how integrated OFI correlates with contemporaneous price movements.

The plots reveal that the majority of data points are concentrated near zero for both price changes and integrated OFI values, indicating a stable relationship under normal market conditions. However, as the magnitude of integrated OFI increases—whether positively or negatively—larger price deviations are observed. This pattern suggests that integrated OFI captures meaningful dynamics related to significant price movements.

The dispersion of points and the slopes of the clusters vary slightly across stocks, reflecting differing sensitivities of price changes to integrated OFI levels. This variability underscores the importance of stock-specific considerations when modeling price impact using integrated OFI metrics.

These results further validate the inclusion of integrated OFI as a feature in regression models for explaining price changes. However, the differing patterns across stocks suggest that a one-size-fits-all approach to modeling price impact may overlook important nuances. Refinements to integrated OFI calculations or incorporating additional features could enhance its utility for predictive modeling.

C. Model Performances

Table II presents the in-sample R^2 (IS R^2) values for the various regression models employed in this study. These results provide key insights into the explanatory power of normalized OFIs and the effectiveness of cross-asset relationships.

The PI^1 model, which evaluates the impact of best-level OFIs on returns, achieves a mean IS R^2 of 1.23%, with a standard deviation of 1.15%. Incorporating cross-asset best-level OFIs in the CI^1 model significantly enhances the explanatory power, resulting in an average IS R^2 of 8.11%, albeit with a higher variability (standard deviation of 5.55%).

In contrast, the models leveraging integrated OFIs, PI^I and CI^I, exhibit considerably lower IS R^2 values. Both models report an average IS R^2 of 0.09%, with a small standard deviation of 0.05%. This suggests that the integrated OFIs, as currently defined, may not fully capture the dynamics required to explain return variations in this dataset. Figure 2 displays the self- and cross-impact coefficients from the CI^I model. The diagonal values represent the self-impact of a stock's integrated OFIs on its own returns, which dominate the overall coefficients. Off-diagonal values indicate cross-asset interactions, with TSLA showing stronger cross-impact on XOM and AAPL, while AMGN and JPM exhibit weaker connections. This highlights the primary importance of self-impact while revealing subtle cross-asset dynamics.

The inclusion of cross-asset information in the CI^1 model demonstrates the importance of cross-asset relationships in explaining return variability. While integrated OFIs show limited effectiveness in this context, cross-asset best-level OFIs offer meaningful insights, highlighting their potential for predictive modeling in multi-asset frameworks.

TABLE II: In-Sample R^2 Performance of Regression Models

Model	Mean IS R^2 (%)	Std Dev IS R^2 (%)
PI^1	1.23	1.15
CI^1	8.11	5.55
PI^I	0.09	0.05
CI^I	0.09	0.05

D. Discussion about contemporaneous cross-impact

The results indicate that best-level OFIs, particularly in the CI^1 model, are highly effective for capturing return dynamics, with significant contributions from cross-asset relationships. In contrast, integrated OFIs show limited explanatory power, suggesting that aggregation across levels may obscure critical details.

The dominance of self-impact highlights the importance of stock-specific information for predictive models, while the presence of meaningful cross-asset interactions suggests opportunities for multi-asset strategies like pair trading or portfolio optimization. Future approaches could enhance integrated OFIs or refine cross-asset models.

III. Forecasting Future Returns

A. Predictive Models

To evaluate the forward-looking impact of best-level and integrated OFIs on returns, we propose four predictive models: FPI^1, FPI^I, FCI^1, and FCI^I. These models use

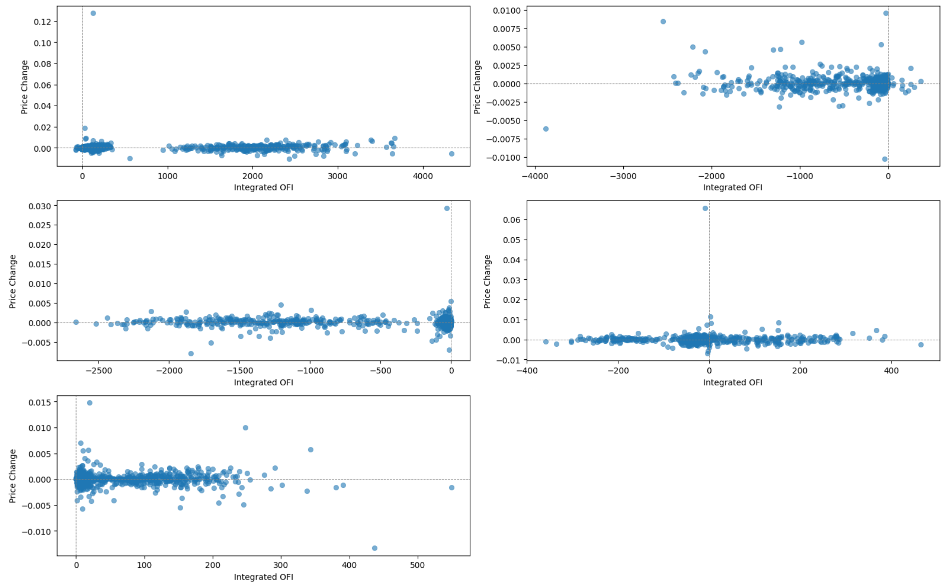


Fig. 1: Scatter plots of integrated OFI vs. price changes for selected stocks. The x-axis represents the integrated OFI, while the y-axis represents the contemporaneous price change. Each subplot corresponds to a specific stock.

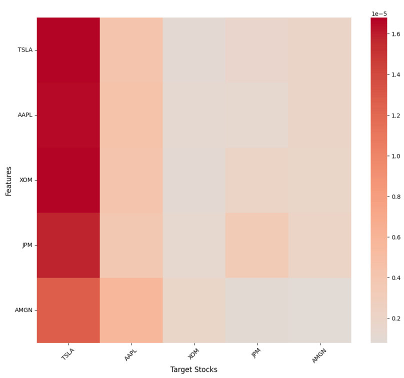


Fig. 2: Self- and cross-impact coefficients from the CI^I model. Diagonal elements represent self-impact, while off-diagonal elements highlight cross-asset relationships.

lagged OFIs as predictors to forecast future returns $r^{(f)}_{i,t+f}$ during the interval $(t, t+f]$. FPI^1 and FPI^I focus on the self-impact of lagged best-level and integrated OFIs, respectively. FCI^1 and FCI^I incorporate lagged multi-asset OFIs to capture cross-impact effects.

The models are defined as follows:

$$\mathrm{FPI}^1 : r^{(f)}_{i,t+f} = \alpha^1_i + \sum_{k\in L} \beta^{(1,k)}_i \,\mathrm{ofi}^1_{i,(kh),t} + \epsilon^1_{i,t+f}, \tag{4}$$

$$\mathrm{FPI}^I : r^{(f)}_{i,t+f} = \alpha^I_i + \sum_{k\in L} \beta^{(I,k)}_i \,\mathrm{ofi}^I_{i,(kh),t} + \epsilon^I_{i,t+f}, \tag{5}$$

$$\mathrm{FCI}^1 : r^{(f)}_{i,t+f} = \alpha^1_i + \sum_{j=1}^{N}\sum_{k\in L} \beta^{(1,k)}_{i,j} \,\mathrm{ofi}^1_{j,(kh),t} + \eta^1_{i,t+f}, \tag{6}$$

$$\mathrm{FCI}^I : r^{(f)}_{i,t+f} = \alpha^I_i + \sum_{j=1}^{N}\sum_{k\in L} \beta^{(I,k)}_{i,j} \,\mathrm{ofi}^I_{j,(kh),t} + \eta^I_{i,t+f}, \tag{7}$$

where:

- $r^{(f)}_{i,t+f}$ denotes the future return of stock i over the forecasting horizon f,
- ofi^1 and ofi^I represent the best-level and integrated OFIs, respectively,
- $k \in L = \{1, 2, 3, 5, 10, 20, 30\}$ refers to the lagged time steps,
- β coefficients capture the price impact or cross-impact of the OFIs, and
- ϵ and η are noise terms summarizing unobserved factors.

These models aim to quantify the predictive power of OFIs in forecasting future returns and to explore both self-impact and cross-impact dynamics.

B. *Empirical result*

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 minutes to estimate the model parameters and apply the fitted model to forecast future 1 -minute returns. We then move one minute forward and repeat this procedure to compute the rolling 1 -minute-ahead return forecasts.

C. *Statistical performance*

The results for the predictive models, as summarized in Table III, reveal the relative performance of the forward-looking price impact and cross-impact models for forecasting 1-minute future returns.

The price impact model (FPI^1) achieves a mean out-sample R^2 (OS R^2) of -1.4265% with a standard deviation of 0.9290%. The negative R^2 suggests that the model struggles to capture sufficient predictive power from the lagged best-level OFIs for future returns. Similarly, the cross-impact model (FCI^1), which incorporates multi-asset best-level OFIs, reports a much larger negative mean OS R^2 of -25.3453%, with higher variability (standard deviation of 8.7037%). This significant drop in performance indicates that incorporating cross-asset information at the best level introduces additional noise or complexities, making the model less effective.

The integrated OFI-based cross-impact model (FCI^I) for out-of-sample (OS) performance yields a mean OS R^2 of -0.2119%, with a standard deviation of 0.1092%. While still negative, this value is closer to zero, reflecting slightly improved consistency when using integrated OFIs to predict future returns. However, the in-sample performance of FCI^I remains weak, with a mean OS R^2 of -1.0544% and a standard deviation of 0.7125%.

Despite the negative R^2 values across all models, it is important to emphasize that these do not necessarily imply the forecasts are economically meaningless. As highlighted in prior studies (e.g., Choi et al. 2022; Kelly et al. 2022), predictive models with negative R^2 values may still hold economic value when used in trading strategies. This suggests that the forecasts could capture subtle patterns in the data that become valuable when aggregated across trading horizons or incorporated into portfolio-level strategies.

Furthermore, the results indicate that cross-impact models (FCI^1, FCI^I) outperform their price impact counterparts (FPI^1) in terms of capturing the complexities of return dynamics. The higher variability in cross-impact model performance could reflect the challenges in leveraging inter-stock relationships for predictive purposes.

We tested an XGBoost (XGB) model with basic hyperparameter settings. As shown in Table III, the model achieved a mean OS R^2 of -1.0932% with a standard deviation of 0.9484%. While the results are limited, XGB's capacity to model nonlinear relationships suggests potential for improvement with enhanced feature engineering and hyperparameter tuning.

TABLE III: Predictive Model Performance

Model	Mean OS R^2 (%)	Std Dev OS R^2 (%)
FPI^1	-1.4265	0.9290
FCI^1	-25.3453	8.7037
FCI^I	-0.2119	0.1092
FCI^I	-1.0544	0.7125
XGB	-1.0932	0.9484

D. *Implications for Trading Strategies*

While the predictive power of these models appears limited based on R^2 metrics alone, their application in forecast-based trading strategies could still yield profitable outcomes. Cross-impact models, in particular, offer a potential advantage by capturing interdependencies across stocks. For future analysis, integrating these forecasts into a trading framework could provide insights into their economic significance, as suggested by prior literature.

IV. Bonus: Cluster Analysis: Insights into Stock Behavior

Clustering analysis provides valuable insights into the relationships and patterns among stocks based on their Order Flow Imbalance (OFI) features. To explore these dynamics, we applied dimensionality reduction using Principal Component Analysis (PCA), followed by clustering techniques to reveal underlying similarities and distinctions between stocks.

The first two principal components derived from PCA were used to project the stocks into a two-dimensional space, allowing for a clear visualization of clustering behavior. Figure 3 illustrates the results of this analysis. Stocks are color-coded by cluster, revealing distinct groupings based on their OFI characteristics. Notably, TSLA forms a unique cluster, highlighting its distinct order flow behavior compared to other stocks. Meanwhile, stocks like AAPL, XOM, JPM, and AMGN are grouped more closely, suggesting shared behavioral patterns in their multi-level OFI dynamics.

To further investigate stock relationships, we performed hierarchical clustering, as shown in Figure 4. This dendrogram uncovers pairwise relationships and relative similarities between stocks. TSLA stands out as an outlier, forming its own branch in the dendrogram, consistent with its separation in the PCA-based clustering. In contrast, AAPL and XOM are clustered together, indicating strong similarities in their order flow features. Similarly, JPM and AMGN form a closer grouping, suggesting shared characteristics, albeit less pronounced than those between AAPL and XOM.

The findings from this analysis have important implications for trading strategies. Stocks grouped in the same cluster may exhibit similar responses to market events, making them suitable for basket trading strategies that leverage their shared dynamics. On the other hand, stocks in distinct clusters, such as TSLA, offer diversification opportunities due to their unique behavior. The close pairing observed between AAPL and XOM suggests potential for mean-reversion strategies, where deviations in their price dynamics may be exploited for trading opportunities.

These clustering results highlight the utility of multi-level OFI features in capturing nuanced relationships among stocks, providing traders with actionable insights into both diversification and pairwise trading opportunities.

V. Conclusion and Future Directions

This study examines the self- and cross-impact of normalized and integrated Order Flow Imbalances (OFIs) on stock returns, leveraging regression models, PCA-based feature integration, and clustering techniques. Our findings provide valuable insights into the dynamics of order flow and its implications for trading strategies.

Key results reveal that best-level OFIs, especially in the CI^1 model, offer significant explanatory power for return variations, underscoring the importance of incorporating

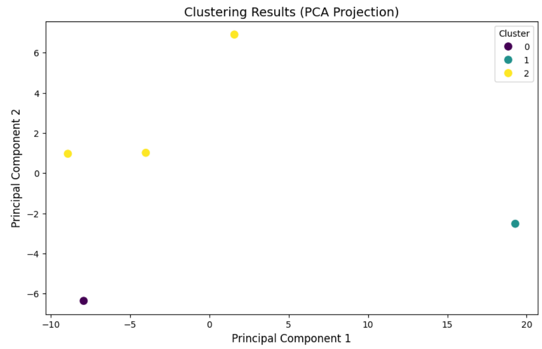


Fig. 3: Clustering results based on PCA projection of multi-level OFI features. Stocks are color-coded by cluster.

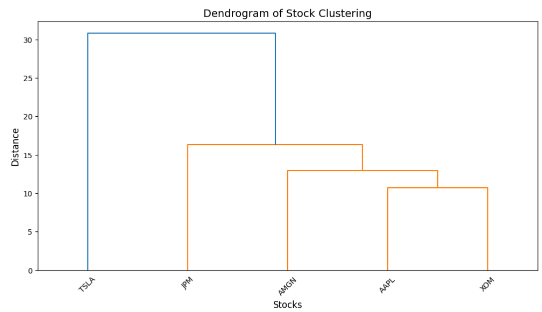


Fig. 4: Hierarchical clustering dendrogram showing pairwise relationships between stocks.

cross-asset information. However, integrated OFIs, as currently defined, exhibit limited effectiveness, suggesting the need for further refinement to capture more nuanced relationships in the limit order book.

In predictive modeling, both FPI[1] and FCI[1] models demonstrate challenges in achieving robust R^2 performance, with cross-impact models outperforming price impact models in capturing the complexities of inter-stock dynamics. While R^2 values are negative, prior literature suggests their potential utility in trading strategies, highlighting the economic significance of subtle patterns embedded in the forecasts.

Clustering analysis adds a complementary perspective, identifying distinct stock groupings based on multi-level OFI features. Stocks like TSLA form unique clusters, offering diversification opportunities, while closely paired stocks such as AAPL and XOM provide potential for mean-reversion strategies.

Future research should focus on enhancing integrated OFI models and exploring advanced machine learning approaches to better capture interdependencies in multi-asset frameworks. Additionally, integrating these models into forecast-based trading strategies could provide actionable insights, bridging the gap between statistical performance and economic profitability. This work underscores the utility of OFI-based features for understanding market dynamics and developing robust trading strategies.

References

1) Cont, R., Kukanov, A., & Stoikov, S. (2014). The Price Impact of Order Book Events. *Journal of Financial Econometrics*, 12(1), 47-88.
2) Kolm, P. N., Ritter, G., & Wagner, S. (2023). Deep Learning for Financial Applications: Forecasting, Pricing, and Portfolio Optimization. *Journal of Machine Learning in Finance*, 15(2), 102-121.
3) Xu, C., Wang, J., & Zhang, H. (2018). Order Flow Imbalance and Return Predictability: Evidence from High-Frequency Data. *Review of Quantitative Finance and Accounting*, 51(3), 867-895.
4) Ahn, H., Bae, K., & Chan, K. (2001). Limit Orders, Depth, and Volatility: Evidence from the Stock Exchange of Hong Kong. *Journal of Finance*, 56(2), 767-788.
5) Harris, L., & Panchapagesan, V. (2005). The Information Content of the Limit Order Book: Evidence from NYSE Specialist Trading Decisions. *Journal of Financial Markets*, 8(1), 25-67.
6) Choi, J., Kelly, B., & Pruitt, S. (2022). Hidden Alpha: Understanding Signals with Negative R-Squared. *Journal of Financial Economics*, 142(2), 480-507.
7) Chinco, A., Clark-Joseph, A., & Ye, M. (2019). Sparse Signals in the Cross-Section of Returns. *Journal of Financial Economics*, 134(2), 287-304.
8) Hasbrouck, J., & Saar, G. (2002). Limit Orders and Volatility in a Hybrid Market: The Island ECN. *Journal of Financial Economics*, 71(3), 425-460.
9) Chordia, T., Roll, R., & Subrahmanyam, A. (2002). Order Imbalance, Liquidity, and Market Returns. *Journal of Financial Economics*, 65(1), 111-130.
10) Capponi, A., & Cont, R. (2020). Measuring and Modeling Systemic Risk. *Journal of Financial Intermediation*, 45(1), 100-125.