

Conceptual Questions: Cross-Impact of Order Flow Imbalance

Monalisa Mondal
Pace University

May 4, 2025

1. What's the motivation behind measuring OFI at multiple depth levels of the order book?

The primary motivation for measuring Order Flow Imbalance (OFI) at multiple depth levels of the limit order book (LOB) is to capture a more comprehensive and informative signal of the buying and selling pressure that drives short-term price movements. The paper argues:

- **Limitations of Best-Level OFI:** Information relevant to price formation is not solely contained at the best bid and ask prices (P_t^b, P_t^a) and their corresponding sizes (S_t^b, S_t^a) . Relying only on the top level might ignore significant liquidity provision or consumption intentions deeper in the book.
- **Information Content of Deeper Levels:** Orders placed or canceled at levels $m > 0$ $(P_t^{b,m}, S_t^{b,m}, P_t^{a,m}, S_t^{a,m})$ also reflect market participants' strategies and expectations, contributing to price discovery. Large orders might be strategically placed deeper.
- **Improved Explanatory Power:** Integrating information from multiple levels is hypothesized to better explain price dynamics. The paper introduces an *Integrated OFI* (OFI_t^I) using Principal Component Analysis (PCA) on the vector of multi-level OFIs $\mathbf{OFI}_t = (OFI_t^{[1]}, \dots, OFI_t^{[M]})$:

$$OFI_t^I = \sum_{m=1}^M w_m OFI_t^{[m]} \quad (1)$$

where $w = (w_1, \dots, w_M)$ is the first principal component of the covariance matrix of \mathbf{OFI}_t . They empirically show this OFI^I significantly improves the explanation of contemporaneous returns compared to just $OFI^{[1]}$.

- **Understanding Cross-Impact:** Considering multi-level information helps disentangle true cross-asset impacts from effects that are simply better captured by a more complete view of the single asset's own order book dynamics.

Essentially, looking deeper into the LOB provides a richer signal about supply and demand than the top level alone.

2. Why do the authors use Lasso regression rather than OLS for estimating cross-impact?

The authors employ Lasso (Least Absolute Shrinkage and Selection Operator) regression for estimating cross-impact models (predicting return R_t from own OFI and cross-asset OFIs $OFI_{j,t}$) instead of standard Ordinary Least Squares (OLS) for several compelling reasons:

- **High Dimensionality** ($p \gg n$): The model includes OFIs from many other assets ($j = 1, \dots, N$, where $N \approx 100$) as potential predictors. When estimated over short time windows (e.g., 30 minutes with 1-minute data), the number of predictors N can easily exceed the number of observations, rendering OLS ill-posed or highly unstable.
- **Multicollinearity:** OFIs across different assets ($OFI_{j,t}$) are often significantly correlated due to market-wide factors or sector effects. High multicollinearity inflates the variance of OLS estimates, making them unreliable.

- **Sparsity Assumption:** The authors assume that, for any given stock, only a relatively small subset of other stocks exerts a significant cross-impact. Lasso is well-suited for this by performing automatic variable selection. It adds an L_1 penalty to the objective function, shrinking many coefficients towards zero and potentially setting some exactly to zero. The objective function for Lasso is:

$$\min_{\beta} \left\{ \frac{1}{2T} \sum_{t=1}^T (R_t - (\beta_0 + \beta_i OFI_{i,t} + \sum_{j \neq i} \beta_j OFI_{j,t}))^2 + \lambda \sum_{k=0}^N |\beta_k| \right\} \quad (2)$$

where λ is the regularization parameter.

- **Regularization and Overfitting Prevention:** The L_1 penalty inherent in Lasso helps to regularize the model, preventing overfitting that is common in high-dimensional OLS and potentially leading to better out-of-sample generalization.

Lasso provides a statistically robust framework to handle the challenges of dimensionality, multicollinearity, and model selection inherent in estimating sparse cross-impact structures.

3. Why is OFI considered a better predictor of short-term returns than trade volume?

While the paper doesn't directly compare OFI against trade volume empirically, it implicitly relies on the established superiority of OFI based on prior market microstructure literature (e.g., Cont et al., 2014) and its own findings. The rationale includes:

- **Capturing Net Directional Pressure:** OFI measures the net imbalance between liquidity supply and demand. The original OFI definition aggregates different event types:

$$OFI_t = (\text{New Bid Limits}) - (\text{Bid Limit Cancellations}) - (\text{Market Sell Orders}) \\ + (\text{New Ask Limits}) - (\text{Ask Limit Cancellations}) - (\text{Market Buy Orders})$$

(Simplified representation). This directly quantifies the direction of pressure. Trade volume, conversely, simply sums executed quantities ($V_t = \sum \text{trades}_t$) without distinguishing buyer vs. seller initiation, thus masking the net pressure driving prices. High volume can occur with balanced flow and little price impact.

- **Incorporating Latent Liquidity Dynamics:** OFI calculations incorporate changes in the state of the limit order book (LOB), including modifications to resting limit orders (prices $P_t^{b/a,m}$ and sizes $S_t^{b/a,m}$), even if they don't result in immediate trades. This provides information about shifts in supply and demand intentions that volume alone misses.
- **Stronger Empirical Link to Price Changes:** Seminal studies cited by the authors demonstrated that OFI exhibits a strong, approximately linear relationship with contemporaneous price changes ($\Delta P_t \approx \beta \cdot OFI_t + \epsilon_t$), making it a powerful explanatory variable. The predictive models in this paper further show that lagged OFI retains predictive power for future returns.

In essence, OFI provides a more granular and directional measure of the forces acting on the LOB, making it theoretically and empirically a more potent predictor of short-term price movements than aggregate trade volume.