Conceptual Questions on Cross-Impact and Order-Flow Imbalance

Babak Badnava
Department of Electrical Engineering and Computer Science
University of Kansas, Lawrence, KS, USA
babak.badnava@ku.edu

***Abstract*—This note answers three conceptual questions posed by Blockhouse Capital: (1) the motivation for measuring order-flow imbalance (OFI) across multiple depth levels, (2) the motivation for using least absolute shrinkage and selection operator (Lasso) rather than ordinary least squares (OLS) in cross-impact estimation, and (3) why OFI outperforms trade volume in forecasting short-horizon returns. A succinct explanation of each point is provided on the following pages. Implementation details and reproducible code are documented and can be accessed via this Github repository.**

I. Motivation for Multi-Level OFI

Order flow does not concentrate solely at the inside quote; traders continuously add and cancel liquidity several ticks deep in the book. Those deeper queues contain information about *latent* supply and demand that level-1 alone cannot reveal. If we can measure the pressure (i.e., buying power vs. selling power) at each depth, we obtain an early signal of which side is likely to dominate the next price move.

Following Cont et al., the order-flow imbalance at depth level m for stock i in a bar of length h ending at t is:

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

where $\mathrm{OF}_{i,n}^{m,b}$ and $\mathrm{OF}_{i,n}^{m,a}$ denote the signed bid- and ask-side order-flow at message n. The imbalance quantifies net buying pressure at that particular depth.

Best-level OFI (level 1) reflects the *current* state of the market, but empirical results show that deeper-level OFIs anticipate future price changes and exhibit low correlation with level 1. Aggregating the first ten levels therefore supplies a richer, forward-looking feature (*multi-level OFI*) that lifts contemporaneous R^2 from $\sim 71\%$ to $\sim 87\%$ and improves short-horizon return forecasts.

II. Why Lasso for Cross-Impact

Estimating cross-impact means regressing each stock's return on the 10-level OFI vector of every other stock. With $N = 100$ names this yields $p = 10N \approx 1{,}000$ predictors, whereas a single trading day provides only $n \approx 390$ one-minute observations. Because $p > n$, the OLS normal equations are rank-deficient—no unique solution exists.

Even in multi-day panels where $p < n$, the predictors are strongly correlated: Fig. 3 of the paper shows that roughly 10% of cross-asset OFI pairs have correlations exceeding 0.30, creating severe *multicollinearity*. Under multicollinearity OLS remains unbiased but its variance explodes, rendering many coefficients statistically insignificant.

Lasso regression remedies both issues by adding an L^1 penalty to the least-squares objective:

$$\hat{\beta} \;=\; \arg\min_{\beta} \left\{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\}.$$

The L^1 term (i) drives many coefficients exactly to zero, automatically selecting a *sparse* subset of influential peers; (ii) shrinks the remaining coefficients, trading a small bias for a large variance reduction; and (iii) delivers a unique, well-posed solution even when $p \gg n$. Cross-validation chooses λ to maximise out-of-sample R^2, yielding a robust yet interpretable estimate of cross-impact.

III. OFI vs. Trade Volume

Trade volume is *directionless*: equal buying and selling generates the same turnover yet exerts no net price pressure. OFI, defined as bid-side flow minus ask-side flow, is inherently *signed*; it also counts limit-order additions and cancellations *before* they turn into trades. Because each signed update is weighted by its *size*, OFI captures *both* the direction *and* the magnitude of trading pressure, acting as a volume-weighted measure of aggressiveness. Hence OFI tracks the true supply–demand imbalance that actually moves the mid-price, whereas raw volume only records how much changed hands.

References

[1] R. Cont, M. Cucuringu, and C. Zhang, "Cross-impact of order flow imbalance in equity markets," 2023. [Online]. Available: https://arxiv.org/abs/2112.13213