Order Flow Imbalance Analysis

Conceptual Questions

1. Motivation for Multi-Level OFI Measurement

The motivation for measuring Order Flow Imbalance (OFI) at multiple depth levels comes from recognizing that the best bid/ask prices alone don't capture the full market picture. Think of the limit order book as an iceberg - the top level is just the visible tip, while significant liquidity and trading intentions lie beneath the surface.

When we only consider the top level, we're missing crucial information about large institutional orders that are deliberately placed deeper in the book to avoid immediate price impact. These hidden orders represent substantial buying or selling pressure that hasn't yet reached the visible market.

By incorporating multiple levels, we get a more complete view of true market sentiment. Research shows this multi-level approach explains price movements significantly better than top-level OFI alone - capturing nearly 90% of price impact versus about 70% for best-level OFI. Essentially, deeper levels contain valuable predictive signals that would otherwise be overlooked.

2. Why LASSO Instead of OLS?

The choice of LASSO regression over Ordinary Least Squares (OLS) comes down to practical realities of financial markets. When modeling cross-asset impacts, we face two main challenges:

First, markets contain hundreds of assets, creating a "too many variables" problem. If we try to model how every asset affects every other asset, we quickly end up with more variables than data points - a mathematical impossibility for OLS.

Second, asset order flows tend to move together, especially within sectors. This correlation creates multicollinearity that makes OLS coefficients unstable and difficult to interpret.

LASSO solves both problems by automatically selecting only the most important relationships. It imposes a penalty that drives unimportant coefficients to zero, creating a sparse model that focuses only on meaningful cross-asset impacts. This approach aligns with market reality - most assets don't meaningfully impact each other, but a few key relationships matter significantly.

3. OFI vs. Trade Volume for Short-Term Forecasting

OFI outperforms trade volume as a short-term predictor because it captures the *direction* and *quality* of market pressure, not just quantity. Consider two scenarios with identical trading volume:

- Scenario A: Aggressive buyers lifting offers (taking liquidity)
- Scenario B: Patient sellers posting limit orders (providing liquidity)

Volume alone can't distinguish these fundamentally different market conditions. OFI does - it quantifies the net *imbalance* between buy and sell pressure.

This directional insight matters because:

- Persistent buying pressure (positive OFI) typically precedes price increases

- Sustained selling pressure (negative OFI) often foreshadows price drops
- The *way* orders are executed (aggressive vs. passive) signals trader urgency

Additionally, OFI incorporates valuable microstructure details - price improvements, size changes at fixed prices, and depth information - that raw volume ignores. This makes OFI a more nuanced and responsive indicator for short-term price movements.

Implementation Approach

The Python implementation closely follows the research methodology:

Best-Level OFI

Computed by tracking changes at the best bid/ask prices:

- Price increase at bid: + Size added
- Price decrease at bid: - Size removed
- Size change at same price: Net size difference

Net OFI = (Bid changes) - (Ask changes)

Multi-Level OFI

Extends the same logic to multiple price levels:

- Computes OFI at each depth level (up to 10 levels)
- Applies depth-based weighting (deeper levels have less impact)
- Sums contributions across all levels

Integrated OFI

Uses Principal Component Analysis (PCA) to intelligently combine levels:

1. Computes OFI at each level separately
2. Finds dominant pattern across levels using PCA
3. Combines levels using PCA weights (normalized)

This creates a single powerful signal capturing the most important information across all depths.

Cross-Asset OFI

Models relationships between assets:

1. Computes best-level OFI for all assets
2. Uses LASSO regression to identify significant cross-asset impacts
3. Models asset returns as: Own OFI + Weighted impact of other assets' OFI

GitHub Repository

The complete implementation code is available at:
`https://github.com/ojausmane/ofi-feature-construction`