High-Frequency Microstructure Strategy for Perpetual Futures Markets

Aslan Akhmetov

October 2025

Abstract

This paper presents the architecture and methodology of an algorithmic trading system developed to exploit microstructural inefficiencies in the BTC/USDT Perpetual Futures market. The strategy is based on the hypothesis of a linear relationship between Order Flow Imbalance (OFI) and short-term price changes, adjusted for Market Depth (Liquidity Depth). The architecture employs a two-level ensemble of gradient boosting models (LightGBM) within a Meta-Labeling framework to filter out false signals. Particular attention is paid to the problem of overfitting: the *Purged K-Fold Cross-Validation* methodology is applied on information-driven Dollar Bars. The system is optimized for a High-Frequency Low-Variance (HFLV) regime to minimize equity curve volatility and comply with strict Daily Loss Limits.

Contents

1 Introduction and Economic Rationale

Modern high-frequency cryptocurrency derivatives markets are characterized by a high degree of efficiency at the macro level but retain significant inefficiencies at the microstructural level. This work formalizes an approach to extracting alpha from information asymmetry contained within the order flow.

1.1 Price Impact Hypothesis

The core of the strategy relies on the price impact model proposed by Cont et al. (2014). According to this model, short-term price change ΔP is a function of aggressive order pressure (OFI), inversely proportional to the order book depth (Market Depth):

$$\Delta P_t \propto \frac{\mathrm{OFI}_t}{\mathrm{Depth}_t} \tag{1}$$

Where OFI_t (Order Flow Imbalance) represents the net volume of aggressive buy and sell orders initiated by market orders.

1.2 Flow Toxicity and Volatility Regimes

To identify high-risk regimes (Liquidity Cascades), the VPIN (Volume-Synchronized Probability of Informed Trading) metric, developed by Easley and López de Prado (2012), is used. High VPIN values signal the presence of toxic flow, which precedes spread widening and volatility spikes.

2 Data Processing and Feature Engineering

The data architecture is built considering the asymmetry of historical data availability: L1 (trades) are available from 2020, while high-quality L2 (order book) data is available only from 2024.

2.1 Aggregation: Dollar Bars

To eliminate heteroscedasticity and synchronize with market activity, volume-based sampling is used.

- **Method:** Bar formation upon reaching a cumulative turnover D_{thresh}.
- **Parameter:** $D_{thresh} = \$5,000,000$.
- **Result:** The average bar formation frequency is 3-5 minutes, balancing noise reduction with reaction speed.

2.2 Feature Vector Construction (X)

The feature vector is formed by synchronizing L1 and L2 streams:

1. **L1 Features (2020-2025):** Based on trade data, OFI (using the `isBuyerMaker` flag to determine the aggressor side) and VPIN are calculated on a rolling window.
2. **L2 Features (2024-2025):** Based on order book snapshots, Liquidity Imbalance (LOB Imbalance) and depth pressure are calculated. Time-Weighted Average is applied for synchronization with Dollar Bars.
3. **Composite Signal:** The key predictor `ofi_x_depth` is calculated, representing the ratio of aggression to resistance.

3 Modeling Architecture

A two-level Machine Learning Pipeline is utilized, implemented using the LightGBM library.

3.1 Data Labeling: Triple Barrier Method

To generate the target variable Y, the Triple Barrier Method is applied, adapted for the HFLV (High-Frequency Low-Variance) profile to satisfy Prop-Firm constraints:

- **Volatility:** ATR is calculated on a rolling window of 100 Dollar Bars (approx. 5 hours).
- **Barriers:** Symmetric $TP = 1.0 \times ATR$ and $SL = 1.0 \times ATR$.
- **Timeout:** The vertical barrier is set to 10 bars ($\approx 30 - 50$ min).

The goal of this configuration (R:R 1:1) is to maximize Win Rate and minimize Exposure Time, preventing unrealized P&L from breaching Daily Loss Limits.

3.2 Level 1: Primary Model

A multi-class LightGBM classifier trained on microstructural features (X).

- **Objective:** Prediction of volatility breakout direction (Label: +1, -1, 0).
- **Characteristic:** High sensitivity (Recall), tasked with identifying all potential anomalies.

3.3 Level 2: Meta-Model

A binary classifier implementing the **Meta-Labeling** concept.

- **Input:** Probabilities from the Primary Model + Contextual Features (VPIN, Spread, Realized Volatility).
- **Objective:** Estimation of the probability of success ($P_{success}$) for a given signal in the current market regime.
- **Result:** Filtration of low-confidence signals, increasing Precision to $> 55 - 60\%$.

4 Validation and Execution Protocol

To ensure the robustness of results and compliance with prop-firm rules, a strict protocol is applied.

4.1 Purged K-Fold Cross-Validation

Standard cross-validation is not applicable to time series due to autocorrelation. The *Purged K-Fold CV* method (López de Prado, 2018) is implemented:

- **Purging:** Removal of observations from the training set whose labels overlap in time with the test set.
- **Embargo:** Introduction of a time buffer after the test period to eliminate data leakage.

4.2 Execution and Risk Management

- **Risk per Trade:** Fixed at 0.2% of initial capital.
- **Chaos Filter:** Trading is halted if current VPIN $> 80^{th}$ percentile.
- **Order Type:** Market orders for entry; immediate placement of OCO (One-Cancels-Other) Limit and Stop-Market orders for exit.

5 Conclusion

The presented **HFLV System** implements a scientifically grounded approach to trading microstructural inefficiencies. The use of Dollar Bars and VPIN filtering ensures statistical data stability. The two-level architecture with Meta-Labeling allows for dynamic adaptation of position sizing and entry thresholds based on model confidence.

Strict adherence to validation protocols (Purged CV) and the structural design (short holding period, narrow stops) confirm the strategy's viability for passing prop-firm challenges by minimizing the volatility of the equity curve.

References

[1] Cont, R., Kukanov, A., Stoikov, S. (2014). *The Price Impact of Order Book Events*. Journal of Financial Econometrics.

[2] Easley, D., Lopez de Prado, M., O'Hara, M. (2012). *Flow Toxicity and Liquidity in a High-Frequency World*. The Review of Financial Studies.

[3] López de Prado, M. (2018). *Advances in Financial Machine Learning*. Wiley.