Market Microstructure Metrics

Ayush Singh

April 6, 2025

1 Introduction

In modern electronic markets, practitioners and researchers often rely on order book (LOB) data to measure key liquidity, price discovery, and volatility dynamics. High-frequency data reveal subtle aspects of the market that aid in forecasting near-future volatility and understanding risk. This document details common market microstructure metrics, their mathematical definitions, and how they can be applied to volatility prediction. Each section focuses on one feature, providing standard formulas and commentary on its relation to intraday volatility.

2 Definitions of Key Metrics

Below we list the relevant metrics or columns often extracted from high-frequency limit order book data. All references to "bid" and "ask" pertain to the first level of the order book unless otherwise specified (e.g., deeper levels for certain metrics).

2.1 Midpoint

The *midpoint* (also called the mid-price) is an average of the best (highest) bid price and the best (lowest) ask price. Denoting time by t,

$$\text{Midpoint}(t) \;=\; \frac{\text{BestBid}(t) + \text{BestAsk}(t)}{2}.$$

Relation to Volatility: The midpoint is commonly used for computing returns, which are then used to estimate realised or integrated volatility. Significant changes or jumps in the midpoint often reflect shifts in supply-demand dynamics, contributing to sudden changes in volatility.

2.2 Bid–Ask Spread

The *Bid–Ask Spread* is the difference between the lowest available ask price and the highest available bid price:

$$\text{Spread}(t) \;=\; \text{BestAsk}(t) \;-\; \text{BestBid}(t).$$

Relation to Volatility: A widening spread may indicate increased uncertainty or lower liquidity, often coinciding with heightened volatility. Conversely, a tight spread suggests a more liquid and stable market with typically lower short-term volatility.

2.3 LOB Spread Curve

The *LOB Spread Curve* generalizes the concept of the spread to multiple price levels in the order book. For level i, define:

$$\text{LOB Spread}_i(t) \;=\; \text{ask\_price}_i(t) - \text{bid\_price}_i(t).$$

Collecting these values across levels $i = 1, \ldots, L$ forms a spread curve describing how the spread behaves at progressively deeper price levels.

Relation to Volatility: If deeper levels also exhibit large spreads, it suggests market makers and traders demand more compensation for risk; this often correlates with higher perceived future volatility or event risk.

2.4 Realised Volatility

Realised Volatility is commonly estimated using the sum of squared log-returns over a fixed time horizon $(0, T)$. Let $M(\tau)$ be the midpoint (or some other reference price) at time τ. The log-return from τ_{k-1} to τ_k is:

$$r_k = \ln\Big(M(\tau_k)\Big) - \ln\Big(M(\tau_{k-1})\Big).$$

Then the discrete-time *realised variance* is:

$$\widehat{\sigma}^2 \;=\; \sum_{k=1}^{T} (r_k)^2,$$

and the *realised volatility* is the square root of realised variance:

$$\widehat{\sigma} \;=\; \sqrt{\sum_{k=1}^{T} (r_k)^2}.$$

Relation to Volatility: Realised volatility is a retrospective measure of how turbulent the price path has been in the past interval. It is also used as a proxy or forecast for near-future volatility under various modeling assumptions.

2.5 Integrated Volatility

In continuous-time finance models, the *Integrated Volatility* over the interval $[0, T]$ is:

$$\int_0^T \sigma_s^2 \, ds,$$

where σ_s is the instantaneous volatility. In practice, this quantity is approximated by the realised volatility in the limit of very high-frequency observations:

$$\sum_k \Big[\ln\Big(M(\tau_{k+1})\Big) - \ln\Big(M(\tau_k)\Big)\Big]^2 \;\approx\; \int_0^T \sigma_s^2 \, ds.$$

Relation to Volatility: In empirical applications, *integrated* and *realised* volatility are often used interchangeably (with slight nuances). Large integrated volatility indicates wide fluctuations in price over the period, signifying a high-volatility environment.

2.6 Rolling Realised Volatility

Rolling Realised Volatility is similar to the realised volatility metric but computed over a moving (or rolling) window of length W. For each time t:

$$\widehat{\sigma}_{\text{rolling}}(t) \;=\; \sqrt{\sum_{\tau \in (t-W,t]} \Big[\ln\Big(M(\tau)\Big) - \ln\Big(M(\tau^-)\Big)\Big]^2}.$$

Relation to Volatility: Rolling estimates provide an evolving view of short-horizon volatility. Spikes in rolling realised volatility signal abrupt changes in market conditions, often triggered by new information or liquidity shocks.

2.7 Normalised Bid–Ask Spread

To make spreads comparable across different price levels, a *Normalised Spread* is often used:

$$\text{Normalised Spread}(t) \;=\; \frac{\text{BestAsk}(t) - \text{BestBid}(t)}{\text{Midpoint}(t)}.$$

Relation to Volatility: A relatively large normalised spread typically indicates that liquidity providers anticipate higher near-future volatility (or risk). When normalised spreads contract, it often reflects calmer conditions and hence lower anticipated volatility.

2.8 Adversarial (Adverse Selection) Cost

Adverse Selection Cost (or *Adversarial Cost*) reflects the hidden cost of a trade if future prices move against you. For a buy at time t (fill price Fill(t)):

$$\text{AdverseCost}(t) \;=\; \mathbb{E}\Big[M(t + \Delta t)\Big] \;-\; \text{Fill}(t).$$

If the midpoint later drops below your fill price, you essentially "overpaid." For a sell, a symmetric definition applies.

Relation to Volatility: Periods of higher volatility tend to elevate adverse selection costs, as prices can move quickly. Traders adjust quotes to protect themselves, leading to larger bid–ask spreads and higher transaction costs.

2.9 Order Book Imbalance (OBI)

Order Book Imbalance measures the difference in liquidity on the bid and ask side, typically at the best price level. A simple version is:

$$\mathrm{OBI}(t) \;=\; \frac{\text{bid\_size1}(t) - \text{ask\_size1}(t)}{\text{bid\_size1}(t) + \text{ask\_size1}(t)}.$$

Relation to Volatility: When the market is heavily skewed to one side (e.g., a large imbalance on the bid side), a small shift in sentiment or arrivals of large orders can cause rapid price moves. Greater imbalance often coincides with more pronounced mid-price volatility, especially if the imbalance flips or quickly erodes.

2.10 Cumulative Order Book Imbalance

This extends OBI to deeper levels. Let $\text{bid\_size}_i(t)$ and $\text{ask\_size}_i(t)$ be the sizes at level i. One approach is:

$$\text{Cumulative OBI}(t) \;=\; \sum_{i=1}^{L} \Big[\text{bid\_size}_i(t) - \text{ask\_size}_i(t)\Big].$$

Alternatively, a ratio form is also used:

$$\text{Cumulative OBI}(t) \;=\; \frac{\sum_{i=1}^{L} \text{bid\_size}_i(t)}{\sum_{i=1}^{L} \text{ask\_size}_i(t)} \quad \text{or variations thereof.}$$

Relation to Volatility: A large imbalance across multiple levels signals a one-sided order book. Such a market is prone to higher short-term volatility if large trades or new limit orders arrive and quickly push the price through multiple levels.

2.11 LOB Slope

LOB Slope attempts to quantify how quickly prices move outward as you consume volume in the order book. One approach is to gather the pairs $(Q_{b,i}, P_{b,i})$ on the bid side, where $Q_{b,i}$ is the cumulative bid size up to level i and $P_{b,i}$ is the corresponding bid price. Then perform a simple linear regression:

$$P_{b,i} \;=\; \alpha_b + \beta_b\, Q_{b,i} \;+\; \epsilon_i,$$

where β_b is the slope for the bid side. Similarly, on the ask side we get a slope β_a.
A simplified version uses:

$$\text{LOB Slope}_b(t) \;=\; \frac{P_{b,L}(t) - P_{b,1}(t)}{Q_{b,L}(t)},$$

$$\text{LOB Slope}_a(t) \;=\; \frac{P_{a,L}(t) - P_{a,1}(t)}{Q_{a,L}(t)}.$$

Relation to Volatility: Steeper slopes (where small changes in volume consumption move price significantly) generally coincide with thinner markets and potentially higher volatility. Flatter slopes imply the market can absorb volume with less price impact, often indicating lower near-term volatility.

2.12 LOB Entropy

LOB Entropy measures how *dispersed* the liquidity is across different levels. Using a Shannon entropy framework, let:

$$p_i(t) \;=\; \frac{\text{Volume at level } i}{\sum_{j=1}^{L} \text{Volume at level } j}.$$

Then

$$\text{LOB Entropy}(t) \;=\; -\sum_{i=1}^{L} p_i(t)\, \ln\Big(p_i(t)\Big).$$

Relation to Volatility: Higher entropy means more even distribution of volume across levels, potentially lowering the chance of abrupt price jumps (suggesting lower volatility). Lower entropy means volume is concentrated at fewer levels, which can be more susceptible to large volatility swings if those levels are rapidly consumed.

2.13 Microprice

The *microprice* is a volume-weighted version of the midpoint that accounts for the relative balance of liquidity on each side. For the best level:

$$\text{Microprice}(t) \;=\; \frac{\text{BestAsk}(t)\,\text{bid\_size1}(t) + \text{BestBid}(t)\,\text{ask\_size1}(t)}{\text{bid\_size1}(t) + \text{ask\_size1}(t)}.$$

Relation to Volatility: A sudden shift in microprice (e.g., if the bid side or ask side becomes dominant) can signal an impending move in the midpoint, often foreshadowing short-term volatility surges.

3 Using These Metrics to Predict Volatility

3.1 Intuitive Explanations

- **Spread-based Metrics** (§2.2, §2.3, §2.7): High spreads often imply more uncertainty or lower liquidity, which correlates with greater volatility.
- **Order Book Imbalance** (§2.9, §2.10): Extreme imbalance can lead to rapid price moves if order flow arrives against the weaker side, thus preceding volatility spikes.
- **LOB Slope** (§2.11): Steep slopes imply that even moderate trades can shift prices, a hallmark of markets that can become more volatile.
- **LOB Entropy** (§2.12): Low entropy (volume concentrated at few price levels) can result in quick breaks and higher volatility.
- **Microprice** (§2.13): The microprice can lead the midpoint. If it strongly diverges from the midpoint, it suggests near-future price movement (and potential volatility).

3.2 Example Modeling Approaches

Common volatility prediction models using these features include:

- **Machine Learning Regression**: Use a rolling window of features (spreads, imbalances, microprice shifts, etc.) to predict the next period's realised volatility.
- **Structural Models**: Incorporate microstructure variables (e.g., order imbalance, slope, microprice) in a continuous-time stochastic volatility model.
- **Event-Driven Models**: Track how changes in LOB features (like a sudden shift in imbalance or a jump in spread) trigger volatility events in the next time steps.

Empirically, increases in spread, increases in imbalance, a steep LOB slope, or a strong tilt in microprice relative to midpoint often precede spikes in short-horizon volatility.

4 Conclusion

This document presented fundamental LOB-based metrics (*Midpoint*, *Bid–Ask Spread*, *LOB Spread Curve*, *Volatility Estimates*, *Imbalance*, *Microprice*, etc.) and explained how each links to volatility. These metrics are crucial for short-term trading, risk management, and optimal execution, as they provide insight into market conditions, immediate liquidity, and future volatility potential. Market participants often combine these signals in quantitative models to adapt trading strategies in real time, especially in high-frequency or intraday contexts.