

# Limit Order Book Feature Engineering Reference

---

Complete guide to market microstructure features for HFT machine learning models.

---

## Basic Features

### Spread

**What It Is:** The difference between the best ask price and the best bid price, representing the immediate cost of a round-trip trade.

**Equation:**

$$\text{spread} = \text{ask\_px\_1} - \text{bid\_px\_1}$$

**Motivation:** Measures the cost of immediate execution and market impact. The spread compensates market makers for providing liquidity and bearing inventory risk.

**Market Interpretation:** Wider spreads indicate lower liquidity, higher volatility, or increased information asymmetry. Narrow spreads suggest competitive, liquid markets with many active market makers. During calm periods, spreads tighten; during volatile or uncertain periods, spreads widen as market makers protect themselves.

---

### Mid Price

**What It Is:** The midpoint between the best bid and best ask prices, serving as a reference "fair value" for the asset.

**Equation:**

$$\text{mid\_price} = (\text{ask\_px\_1} + \text{bid\_px\_1}) / 2$$

**Motivation:** Provides a central tendency price that averages out the bid-ask spread. Used as the baseline for computing returns, price movements, and other derived features.

**Market Interpretation:** Represents the theoretical fair value in a frictionless market. Mid-price movements indicate genuine price discovery rather than bid-ask bounce noise. Most ML models predict mid-price direction rather than actual trade prices.

---

### Relative Spread

**What It Is:** The bid-ask spread expressed as a percentage of the mid price, enabling cross-asset and cross-time comparisons.

**Equation:**

$$\text{relative\_spread} = \text{spread} / \text{mid\_price}$$

**Motivation:** Normalizes spread by price level. A \$1 spread on a \$100 asset is very different from a \$1 spread on a \$50,000 asset. Relative spread makes comparisons meaningful.

**Market Interpretation:** Typical values for liquid crypto: 0.01%-0.05% (1-5 basis points). Values >0.1% signal illiquidity or stress. Rising relative spread indicates deteriorating market quality. Useful for comparing liquidity across different price regimes.

---

## Weighted Mid Price

**What It Is:** A volume-weighted midpoint that adjusts for liquidity imbalance at the top of the book, providing a more accurate fair value estimate than simple mid price.

**Equation:**

$$\text{weighted\_mid} = (\text{ask\_px}_1 \times \text{bid\_qty}_1 + \text{bid\_px}_1 \times \text{ask\_qty}_1) / (\text{bid\_qty}_1 + \text{ask\_qty}_1)$$

**Motivation:** When one side of the book has significantly more volume, the simple mid-price doesn't reflect the true balance of supply and demand. Weighted mid shifts toward the side with deeper liquidity.

**Market Interpretation:** If  $\text{bid\_qty} \gg \text{ask\_qty}$ ,  $\text{weighted\_mid}$  shifts below simple mid (downward pressure expected). If  $\text{ask\_qty} \gg \text{bid\_qty}$ ,  $\text{weighted\_mid}$  shifts above simple mid (upward pressure expected). The difference between  $\text{weighted\_mid}$  and  $\text{mid\_price}$  is itself a predictive signal.

---



## Volatility & Price Movement

### Rolling Volatility

**What It Is:** The standard deviation of returns over a rolling time window, capturing recent price variability and risk.

**Equation:**

$$\text{volatility\_Ns} = \text{std}(\text{return\_1s}, \text{window}=\text{N})$$

where  $\text{return\_1s}$  are 1-second log returns.

**Motivation:** Markets transition between calm and volatile regimes. Volatility clustering means high volatility persists. Capturing recent volatility helps identify regime changes and adjust trading strategies.

**Market Interpretation:** Low volatility (< 10 bps/sec) indicates stable, predictable markets suitable for aggressive market making. High volatility (> 50 bps/sec) signals information arrival, uncertainty, or potential regime shift. Market makers widen spreads and reduce depth during high volatility.

---

## Returns (Multiple Lags)

**What It Is:** Logarithmic price changes over various time horizons, capturing momentum and mean-reversion patterns at different scales.

### Equation:

$$\text{return\_Ns} = \ln(\text{mid\_price\_t} / \text{mid\_price\_}\{t-N\})$$

**Motivation:** Returns are more stationary than prices and capture relative changes. Multiple lags reveal momentum (trending) vs mean-reversion (bouncing) behavior. Essential features for price direction prediction.

**Market Interpretation:** Short lags (1-5s) capture microstructure noise and mean-reversion. Medium lags (10-30s) capture momentum. Long lags (60s+) capture trend. Positive autocorrelation suggests momentum; negative suggests mean-reversion. High-frequency returns exhibit mean-reversion due to bid-ask bounce.

---

## Order Book Shape & Slope

### Book Slope / Volume Decay Rate

**What It Is:** A measure of how quickly liquidity deteriorates (volume decays) as you move away from the best bid/ask. Quantifies the steepness of the order book depth profile.

### Equation:

#### Linear slope (price-volume):

$$\text{slope\_bid} = (\text{bid\_px\_1} - \text{bid\_px\_N}) / \text{sum}(\text{bid\_qty\_1}, \dots, \text{bid\_qty\_N})$$

#### Exponential decay rate:

$$\begin{aligned} \text{qty\_i} &\approx \text{qty\_1} \times \exp(-\alpha \times i) \\ \text{slope} &= -\alpha \quad (\text{fitted decay parameter}) \end{aligned}$$

#### Log-volume gradient:

$$\text{gradient} = \Delta \log(\text{qty}) / \Delta \text{level} = \text{mean}(\log(\text{qty\_}\{i\}) / \text{qty\_}\{i+1\}))$$

**Motivation:** Book slope reveals market depth quality and price impact. Steep slopes indicate thin liquidity where large orders have high execution costs. Flat slopes indicate deep, resilient liquidity. Market makers adjust book slope based on their confidence and market conditions.

**Market Interpretation: Steep slope (high  $\alpha > 0.5$ ):** Liquidity concentrated at top, high price impact, fragile market. Seen during volatile periods or defensive market making. Signals caution and reduced depth provision.

**Flat slope (low  $\alpha < 0.2$ ):** Liquidity spread evenly, low price impact, resilient market. Seen during calm periods with competitive market making. Signals confidence and aggressive depth provision.

**Asymmetric slopes (bid\_slope  $\neq$  ask\_slope):** Steeper ask slope suggests market makers expect upward move (pulling ask liquidity). Steeper bid slope suggests expected downward move (pulling bid liquidity). Directional predictor.

---

## Order Book Thickness

**What It Is:** The average quantity available per level in the order book, indicating overall book density and liquidity provision intensity.

**Equation:**

```
thickness_bid = mean(bid_qty_1, ..., bid_qty_N)
thickness_ask = mean(ask_qty_1, ..., ask_qty_N)
```

**Motivation:** Thickness reveals whether market makers are providing dense, continuous liquidity or sparse, patchy liquidity. Complements slope by measuring average depth rather than decay rate.

**Market Interpretation:** Thick books (high average qty) absorb orders without large price moves. Thin books are fragile to shocks. Increasing thickness suggests growing market maker confidence. Decreasing thickness signals caution or inventory constraints. Asymmetric thickness (bid\_thick  $\neq$  ask\_thick) provides directional signals.

---

## Liquidity Measures

### VWAP (Volume-Weighted Average Price)

**What It Is:** The average price weighted by volume across multiple levels, representing realistic execution cost for larger orders that walk the book.

**Equation:**

```
vwap_bid = sum(bid_px_i × bid_qty_i) / sum(bid_qty_i) for i=1..N
vwap_ask = sum(ask_px_i × ask_qty_i) / sum(ask_qty_i) for i=1..N
```

**Motivation:** Simple mid-price or best bid/ask don't reflect execution cost for large orders. VWAP accounts for the need to trade across multiple levels, providing realistic cost estimates.

**Market Interpretation:** VWAP deviates from mid-price when the book is imbalanced or thin. Large VWAP\_ask - mid\_price spread indicates high cost to buy. Large mid\_price - VWAP\_bid spread indicates high cost to sell. VWAP spread > simple spread indicates depth is shallow beyond top level.

---

## Effective Spread

**What It Is:** The difference between volume-weighted average prices on bid and ask sides, representing actual trading cost for realistic-sized orders rather than just top-of-book spread.

**Equation:**

$$\text{effective\_spread} = \text{vwap\_ask} - \text{vwap\_bid}$$

**Motivation:** Simple spread only considers the best bid/ask, but real execution requires trading across multiple levels. Effective spread accounts for depth effects and provides accurate cost measurement.

**Market Interpretation:** Effective spread > simple spread indicates thin book beyond top level (high depth-dependent costs). Effective spread  $\approx$  simple spread indicates uniform, deep book. Rising effective spread signals deteriorating depth. Important for comparing liquidity quality across time periods or assets.

---

## Microprice

**What It Is:** A volume-weighted mid-price that uses quantities at the best bid and ask to predict the next mid-price movement. More responsive to imbalances than simple mid-price.

**Equation:**

$$\text{microprice} = (\text{ask\_px\_1} \times \text{bid\_qty\_1} + \text{bid\_px\_1} \times \text{ask\_qty\_1}) / (\text{bid\_qty\_1} + \text{ask\_qty\_1})$$

**Motivation:** When bid\_qty >> ask\_qty, the next price movement is more likely upward (buying pressure), so microprice shifts above mid. When ask\_qty >> bid\_qty, selling pressure dominates, and microprice shifts below mid. This creates a more accurate "fair value" estimate.

**Market Interpretation:** Microprice - mid\_price reveals directional pressure. Positive difference = buying pressure (microprice above mid). Negative difference = selling pressure (microprice below mid). Research shows microprice has lower variance than mid and better short-term forecasting properties.

---

## Order Book Pressure & Flow

### Order Flow Imbalance (OFI)

**What It Is:** The net change in bid and ask quantities between consecutive snapshots, capturing buying and selling pressure from order arrivals, cancellations, and executions.

**Equation:**

$$\text{ofi}_t = \Delta \text{bid\_qty} - \Delta \text{ask\_qty}$$

where  $\Delta$  denotes change from previous snapshot, summed across levels.

**Motivation:** OFI captures order flow—the actual trading activity and liquidity changes. Positive OFI means more bids arrived or asks were consumed (bullish). Negative OFI means more asks arrived or bids were consumed (bearish). This flow information is highly predictive of short-term price movements.

**Market Interpretation:** OFI > 0: Net buying pressure, expect upward price movement. OFI < 0: Net selling pressure, expect downward movement. Large |OFI| indicates aggressive trading or information arrival. OFI clustering (sustained positive or negative) suggests directional momentum. Mean-reverting OFI suggests noise trading.

---

## Depth Imbalance

**What It Is:** The normalized difference between total bid and ask quantities, measuring current supply/demand imbalance across multiple price levels.

**Equation:**

$$\text{depth\_imbalance} = (\text{total\_bid\_qty} - \text{total\_ask\_qty}) / (\text{total\_bid\_qty} + \text{total\_ask\_qty})$$

where totals are summed across N levels.

**Motivation:** Captures the overall state of the order book. Unlike OFI (which measures changes), depth imbalance measures the current snapshot of supply vs demand. Persistent imbalances often precede price movements.

**Market Interpretation:** Values near +1: Strong buy-side liquidity, potential resistance to downward moves, or trapped long positions. Values near -1: Strong sell-side liquidity, potential resistance to upward moves, or trapped short positions. Values near 0: Balanced book, fair price discovery. Extreme values often revert as market makers rebalance.

---

## Liquidity Imbalance at Multiple Depths

**What It Is:** Depth imbalance computed separately at each depth level (cumulative from level 1 to level N), revealing how imbalance evolves as you go deeper into the book.

**Equation:**

```
liquidity_imbalance_depth_N = (sum_bid_1_to_N - sum_ask_1_to_N) /
(sum_bid_1_to_N + sum_ask_1_to_N)
```

**Motivation:** Different depths reveal different information. Imbalance at level 1 shows immediate pressure. Imbalance at level 5 shows near-term positioning. Imbalance at level 10 shows institutional or informed positioning. Creating a depth profile reveals **where in the book** the pressure lies.

**Market Interpretation: Top-heavy imbalance (high at level 1-3, neutral deeper):** High-frequency traders or aggressive market makers creating short-term pressure. Often mean-reverts quickly.

**Deep imbalance (neutral at top, high at levels 5-10):** Institutional orders or informed traders positioning for larger moves. More persistent signal.

**Uniform imbalance (same across all depths):** Strong directional consensus. Most reliable for prediction.

**Divergent imbalance (opposite signs at different depths):** Conflicting signals, often noisy or transitional state.

## Order Book Pressure (Cumulative Depth-Weighted Imbalance)

**What It Is:** A sophisticated variant of depth imbalance that weights levels by their proximity to the mid-price, emphasizing actionable liquidity near the top of the book while including deeper levels with reduced influence.

**Equation: Inverse distance weighting:**

$$OBP = \sum_{i=1}^N (1/i) \times (bid\_qty\_i - ask\_qty\_i)$$

**Exponential decay weighting:**

$$OBP = \sum_{i=1}^N \exp(-\lambda \times i) \times (bid\_qty\_i - ask\_qty\_i)$$

where  $\lambda$  controls decay rate (typical values: 0.3-1.0)

**Normalized version (for [-1, 1] range):**

$$OBP\_norm = \sum w_i(bid\_qty\_i - ask\_qty\_i) / \sum w_i(bid\_qty\_i + ask\_qty\_i)$$

**Motivation:** Unlike simple depth imbalance (which treats all levels equally), order book pressure recognizes that levels closer to mid-price are more actionable and informationally relevant. A 100 BTC order at level 1 has far more immediate price impact than 100 BTC at level 10. OBP captures this economic reality by weighting imbalances by depth.

**Market Interpretation: Positive OBP:** Weighted buying pressure, particularly at actionable levels near mid-price. Suggests upward price movement more likely. Larger magnitude indicates stronger directional pressure.

**Negative OBP:** Weighted selling pressure concentrated where it matters most. Suggests downward price movement.

**High magnitude OBP:** Strong directional signal in the executable book. More reliable than simple depth imbalance for prediction.

**Low magnitude OBP:** Balanced book with no clear pressure. Neutral or ranging market.

**Key insight:** OBP correctly emphasizes liquidity at the top of the book. A book with 50 BTC at level 1 bid and 10 BTC at level 1 ask will show strong positive OBP even if deeper levels are balanced, because the top-level imbalance is what drives immediate price action.

---

## Queue Imbalance (Multiple Levels)

**What It Is:** Depth imbalance computed at specific aggregation levels (top 1, top 5, cumulative), revealing how supply/demand imbalance manifests at different depths of the order book.

**Equation:**

$$\text{queue\_imbalance\_topN} = (\text{queue\_bid\_topN} - \text{queue\_ask\_topN}) / (\text{queue\_bid\_topN} + \text{queue\_ask\_topN})$$

where `queue_bid_topN` is sum of bid quantities from level 1 to N.

**Motivation:** Similar to liquidity imbalance at multiple depths, but focuses on standard aggregation levels (1, 5, all) that correspond to different trading contexts. Top 1 = immediate executable liquidity, Top 5 = near-term resilience, Cumulative = total visible depth.

**Market Interpretation: `queue_imbalance_top1`:** Immediate pressure at best bid/ask. Most volatile, responds to every order arrival/cancellation. High-frequency signal.

**`queue_imbalance_top5`:** Near-term pressure within ~\$1-2 of mid. More stable than top1. Captures medium-frequency positioning.

**`queue_imbalance_cumulative`:** Overall book state across all visible levels. Most stable, captures long-term positioning and institutional orders.

**Divergence patterns:**

- **Top1 positive, cumulative negative:** Short-term buying pressure against deeper selling—expect mean reversion
  - **All levels positive:** Uniform buying pressure—expect continuation
  - **Top1 neutral, cumulative positive:** Deep hidden demand—potential breakout setup
-



## Cumulative Volume Imbalance

**What It Is:** The absolute (not normalized) difference in total volume between bid and ask sides across all levels, measuring raw supply/demand asymmetry in quantity terms.

**Equation:**

$$cvi = \sum_{i=1}^N (bid\_qty\_i - ask\_qty\_i)$$

**Motivation:** Unlike normalized depth imbalance (which scales to  $[-1,1]$ ), cumulative volume imbalance preserves the magnitude of the imbalance. A book with 100 BTC bid vs 50 BTC ask is different from 10 BTC bid vs 5 BTC ask, even though both have the same normalized imbalance ratio.

**Market Interpretation: Large positive CVI:** Significant excess bid-side liquidity. Could indicate: (1) trapped long positions unable to exit, (2) institutional accumulation, (3) defensive market making after downward move.

**Large negative CVI:** Excess ask-side liquidity. Could indicate: (1) trapped shorts, (2) institutional distribution, (3) defensive market making after upward move.

**Near-zero CVI:** Balanced book regardless of total size. Healthy two-sided market.

**Scaling with book size:** CVI should be considered relative to total book depth. 50 BTC imbalance is huge for a thin book but small for a deep book.

---

## Order Flow Toxicity

**What It Is:** A measure of adverse selection risk, quantifying how much informed traders (who move prices with small volume) are active versus uninformed traders (large volume, small price impact).

**Equation:**

$$toxicity \approx |\Delta price| / volume\_executed$$

or using VPIN (Volume-Synchronized Probability of Informed Trading):

$$VPIN = |buy\_volume - sell\_volume| / total\_volume$$

over recent time window.

**Motivation:** When informed traders are active, small orders have large price impacts (high toxicity). When uninformed traders dominate, large orders have small price impacts (low toxicity). Market makers widen spreads and reduce depth when toxicity is high to protect against adverse selection.

**Market Interpretation: High toxicity (>threshold):** Informed trading detected. Price movements are driven by small volumes. High information asymmetry. Market makers should be defensive (wide spreads, shallow depth).

**Low toxicity:** Uninformed order flow. Price movements require large volumes. Low information asymmetry. Market makers can be aggressive.

**Rising toxicity:** Information event approaching or occurring. Signals regime transition to volatile/informed state.

## Queue Depth Features

Queue Depth (Top 1, Top 5, Cumulative)

**What It Is:** The total quantity available at specific depth levels of the order book, measuring immediate, near-term, and total liquidity respectively.

**Equation:**

```
queue_bid_top1 = bid_qty_1
queue_bid_top5 = sum(bid_qty_1, ..., bid_qty_5)
queue_bid_cumulative = sum(bid_qty_1, ..., bid_qty_N)

queue_total_top1 = queue_bid_top1 + queue_ask_top1
queue_total_top5 = queue_bid_top5 + queue_ask_top5
queue_total_cumulative = queue_bid_cumulative + queue_ask_cumulative
```

**Motivation:** Measures available liquidity at different scales. Top 1 = immediate execution capacity (can I fill my order right now?). Top 5 = near-term resilience (will the book absorb moderate flow?). Cumulative = total visible depth (what's the maximum visible liquidity?).

**Market Interpretation: High queue\_total\_top1:** Dense liquidity at best prices. Low immediate execution cost. Competitive market making.

**Low queue\_total\_top1:** Thin top of book. High execution cost. Cautious or defensive market making. Potential for large bid-ask bounce.

**queue\_total\_top5 >> queue\_total\_top1:** Liquidity concentrated just beyond best prices. "Hidden depth" or stepped pricing strategy.

**queue\_total\_cumulative >> queue\_total\_top5:** Deep book with liquidity spread across many levels. Resilient to large orders.

**Declining queue depths over time:** Liquidity withdrawal, regime transition to volatile state, or pre-announcement positioning.

## Depth Concentration

**What It Is:** The fraction of total liquidity concentrated at the very top of the book, measuring how front-loaded the order book is.

**Equation:**

$$\text{depth\_concentration} = \text{queue\_total\_top1} / \text{queue\_total\_cumulative}$$

**Motivation:** Reveals whether liquidity is concentrated at best prices (aggressive, competitive market making) or spread across many levels (defensive, cautious provision). Different market maker strategies and market conditions produce different concentration patterns.

**Market Interpretation: High concentration (>0.4):** Most liquidity at top of book. Could indicate: (1) aggressive market making competing for rebates, (2) fleeting liquidity that cancels quickly, (3) high-frequency traders dominating, (4) narrow spread environment.

**Low concentration (<0.15):** Liquidity spread evenly or concentrated deeper. Could indicate: (1) defensive market making expecting volatility, (2) institutional orders placed away from market, (3) wide spread environment, (4) low confidence in current price.

**Rising concentration:** Market makers becoming more aggressive or confident. Potentially transitioning to calmer regime.

**Falling concentration:** Liquidity withdrawal from top, defensive positioning. Warning sign of potential volatility or stress.

---

## Price Impact Features

### Price Impact (Multiple Volumes)

**What It Is:** The estimated cost of executing an order of given size, measured as the deviation from mid-price to the volume-weighted average execution price when walking the book.

**Equation:**

$$\begin{aligned} \text{execution\_price\_V} &= \text{VWAP to execute volume V} \\ \text{price\_impact\_V} &= (\text{execution\_price\_V} - \text{mid\_price}) / \text{mid\_price} \end{aligned}$$

For buy order: walk up the ask side until cumulative volume  $\geq V$  For sell order: walk down the bid side until cumulative volume  $\geq V$

**Motivation:** Captures market depth quality and resilience. Shows how much price degrades as you execute larger orders. Directly related to execution cost and optimal trading strategies. Different volumes reveal depth at different scales.

**Market Interpretation: Low price\_impact:** Deep, liquid book. Large orders can be executed with minimal cost. Resilient market.

**High price\_impact:** Thin book. Even moderate orders significantly move the execution price. Fragile, illiquid market.

**Increasing impact with volume:** Normal concave relationship. Impact grows but at decreasing rate (square-root law).

**Super-linear impact:** Impact grows faster than square-root. Indicates very thin book beyond a certain depth threshold.

**Asymmetric impact (buy\_impact  $\neq$  sell\_impact):** Directional signal. Higher buy impact = thin ask side = resistance to upward moves. Higher sell impact = thin bid side = resistance to downward moves.

---

## Multi-Level Spreads

### Spread at Multiple Levels

**What It Is:** The price difference between bid and ask at each depth level (not just level 1), showing how spread widens as you go deeper into the book.

#### Equation:

$$\text{spread\_level\_N} = \text{ask\_px\_N} - \text{bid\_px\_N}$$

**Motivation:** While the best bid-ask spread (level 1) is most commonly quoted, deeper spreads reveal book quality and liquidity provision strategies. Widening spreads at deeper levels indicate thinning liquidity. Consistent spreads across levels indicate robust market making.

**Market Interpretation: Constant spread across levels:** Market makers maintaining tight, uniform spreads throughout the book. High confidence, aggressive liquidity provision. Seen in calm, competitive markets.

**Widening spreads deeper:** Normal pattern. Market makers less willing to commit capital far from current price. Indicates uncertainty about fair value at those levels.

**Dramatically widening spreads:** Thin, fragile book. Liquidity dries up quickly beyond top few levels. High execution risk for larger orders.

**Spread contraction at deeper levels:** Unusual pattern, possibly institutional resting orders or temporary depth at specific price targets.

---

### Average Spread (Top N Levels)

**What It Is:** The mean spread across multiple levels, summarizing overall book tightness beyond just the best bid-ask.

#### Equation:

$$\text{avg\_spread\_topN} = \text{mean}(\text{spread\_level\_i for } i=1..N)$$

---

**Motivation:** Single number summarizing spread quality across the book. More robust than level-1 spread alone, which can be spoofed or fleeting. Average spread captures true depth of competitive pricing.

**Market Interpretation: Low average spread:** Competitive market making across multiple levels. Good liquidity depth. Favorable execution environment.

**High average spread:** Wide spreads throughout the book. Poor liquidity. High execution costs at all sizes.

**avg\_spread >> spread\_level\_1:** Spread blows out beyond top level. Top-of-book liquidity is misleading—don't trust it for large orders.

**avg\_spread ≈ spread\_level\_1:** Uniform tight spreads. Robust, deep market. Safe for larger execution.

---

## Time-Based Features

### Time Since Event

**What It Is:** Elapsed time since the last occurrence of a specific market event (spread change, significant price move, depth change), capturing temporal patterns in information arrival and market activity.

#### Equation:

$$\text{time\_since\_event} = t_{\text{current}} - t_{\text{last\_event}}$$

where event can be: spread widening/tightening beyond threshold, mid-price move > X bps, depth change > Y%, volatility spike, etc.

**Motivation:** Information doesn't arrive uniformly. Events cluster temporally (Hawkes process). Long gaps suggest quiet periods; short gaps indicate active trading or information flow. Time since last event helps models understand current market state.

**Market Interpretation: Short time since last event (<5 seconds):** Active market, information flow, potentially elevated volatility. Events are clustering. Hawkes self-exciting behavior.

**Long time since last event (>60 seconds):** Quiet market, low information arrival, typically lower volatility. Mean-reversion regime.

**Accelerating events (decreasing gaps):** Information cascade or volatility feedback loop beginning. Potentially leading to regime transition.

**Decelerating events (increasing gaps):** Market calming down after information event. Transitioning back to quiet regime.

---

### Intraday Patterns (Hour, Minute)

**What It Is:** Time-of-day indicators capturing intraday seasonality in volatility, liquidity, and trading activity patterns.

**Equation:**

```
hour = timestamp.hour
minute = timestamp.minute
seconds_since_midnight = (hour × 3600) + (minute × 60) + second
```

Can also encode cyclically:

```
hour_sin = sin(2π × hour / 24)
hour_cos = cos(2π × hour / 24)
```

**Motivation:** Markets exhibit strong intraday seasonality. Volatility and spreads follow U-shaped patterns (high at open/close, low mid-day). Liquidity varies by time of day. European vs US trading hours affect crypto differently. Capturing time-of-day helps models adapt to predictable patterns.

**Market Interpretation: Market open (9:30-10:00 AM EST):** High volatility, wide spreads, information from overnight news. For crypto (24/7): US market open still creates activity spike.

**Mid-day (11:00 AM - 2:00 PM):** Lower volatility, tighter spreads, more predictable. Lowest activity period.

**Market close (3:30-4:00 PM EST):** High volatility, portfolio rebalancing, index replication trades. For crypto: US close creates activity change.

**Overnight/Asian hours (crypto):** Different liquidity providers, often thinner books, different volatility patterns.

**Weekend (crypto):** Significantly different patterns—lower institutional participation, thinner liquidity, higher retail percentage.

## Higher-Order Statistics

### Book Skewness

**What It Is:** The third moment of the quantity distribution across order book levels, measuring asymmetry in how volume is distributed.

**Equation:**

```
skewness_bid = (1/N) × sum((qty_i - mean_qty)3) / std_qty3
```

for levels  $i = 1$  to  $N$ .

**Motivation:** Skewness reveals the shape of liquidity distribution. Positive skew = few large orders dominate. Negative skew = many small orders. Captures order placement strategies and presence of large resting orders (institutional, icebergs).

**Market Interpretation: Positive skewness (right-skewed):** Few levels have very large quantities, most have small. Suggests: (1) large institutional orders at specific levels, (2) iceberg orders revealing themselves, (3) concentrated liquidity provision.

**Negative skewness (left-skewed):** Many small orders, few outliers. Suggests: (1) order splitting behavior, (2) distributed market making, (3) retail-dominated flow.

**Near-zero skewness (symmetric):** Uniform distribution or normal-like. Balanced, competitive market making.

**Asymmetric skewness (bid\_skew  $\neq$  ask\_skew):** Different strategies on different sides. High bid skew + low ask skew might indicate institutional buyer with large resting bid vs distributed ask-side liquidity.

---

## Book Kurtosis

**What It Is:** The fourth moment of the quantity distribution, measuring tail heaviness (presence of extreme values/outliers) in order quantities.

### Equation:

$$\text{kurtosis\_bid} = (1/N) \times \text{sum}((\text{qty\_i} - \text{mean\_qty})^4) / \text{std\_qty}^4 - 3$$

(excess kurtosis, where 3 is subtracted so normal distribution has kurtosis=0)

**Motivation:** Kurtosis captures tail behavior—how often extreme order sizes appear. High kurtosis = fat tails = frequent large orders. Low kurtosis = thin tails = uniform order sizes. Reveals presence of outlier orders that might signal informed trading or institutional positioning.

**Market Interpretation: High kurtosis (>3):** Fat tails, frequent large orders. Suggests: (1) presence of institutional orders, (2) heterogeneous market participants, (3) occasional large informed trades, (4) iceberg orders periodically revealing size.

**Low/negative kurtosis (<0):** Thin tails, uniform distribution. Suggests: (1) homogeneous market making, (2) order size limits or regulations, (3) algorithmic order splitting creating uniform sizes.

**Increasing kurtosis:** Large orders entering the book. Potential information event or institutional positioning. May precede price movement.

**Kurtosis spikes:** Occasional very large orders. Monitor for execution or cancellation—can signal informed activity.

---

## Market Resilience Features

### Book Resilience

**What It Is:** The rate at which order book depth replenishes after being depleted by trades or cancellations, measuring how quickly market makers refill liquidity.

**Equation:**

$$\text{resilience} = \Delta \text{depth} / \Delta \text{time}$$

specifically: rate of liquidity recovery after a depletion event.

Or: time constant  $\tau$  from exponential recovery model:

$$\begin{aligned} \text{depth}(t) &= \text{depth\_final} - (\text{depth\_final} - \text{depth\_min}) \times \exp(-t/\tau) \\ \text{resilience} &= 1/\tau \end{aligned}$$

**Motivation:** Resilient markets quickly recover from shocks. Fragile markets take long to refill depth. Resilience measures market maker competition, confidence, and willingness to provide liquidity. Critical for understanding how markets absorb large orders.

**Market Interpretation: High resilience (fast recovery):** Competitive market making, multiple liquidity providers, high market maker confidence. Market can absorb repeated large orders without lasting damage.

**Low resilience (slow recovery):** Few market makers, low confidence, risk aversion. Large orders create lasting depth holes. Vulnerable to predatory trading.

**Decreasing resilience over time:** Market makers becoming more cautious, potentially sensing information asymmetry or inventory risk. Warning sign of regime transition.

**Asymmetric resilience (bid  $\neq$  ask):** One side recovers faster. Faster bid recovery = market makers eager to buy (bullish signal). Faster ask recovery = eager to sell (bearish signal).

## Depth Depletion Rate

**What It Is:** The rate at which total liquidity is being consumed or withdrawn from the book, indicating aggressive order flow or liquidity withdrawal.

**Equation:**

$$\text{depletion\_rate} = -\Delta(\text{queue\_total}) / \Delta \text{time}$$

measured when queue\_total is decreasing (negative during depletion).

Can also measure separately:

$$\begin{aligned} \text{bid\_depletion\_rate} &= -\Delta(\text{queue\_bid\_cumulative}) / \Delta \text{time} \\ \text{ask\_depletion\_rate} &= -\Delta(\text{queue\_ask\_cumulative}) / \Delta \text{time} \end{aligned}$$



**Motivation:** Rapid depletion signals aggressive trading, informed order flow, or market maker withdrawal. Helps identify liquidity crises before they become severe. Predicts potential price movements when one side depletes faster than the other.

**Market Interpretation: High depletion rate:** Aggressive order flow consuming liquidity faster than market makers replenish. Often precedes price movement in depletion direction. Can signal information event.

**Low/zero depletion:** Stable book, balanced flow, market makers keeping up with demand. Calm market state.

**Accelerating depletion:** Positive feedback loop—market makers withdrawing as aggression increases. Flash crash precursor pattern.

**Asymmetric depletion (bid >> ask or vice versa):** Strong directional flow. Depleting bid side = aggressive selling pressure, expect downward price move. Depleting ask side = aggressive buying, expect upward move.

**Depletion without price movement:** Potentially trapped orders or market makers testing. Can precede reversal if depletion stops and refills.

## Volume Concentration (Herfindahl Index)

**What It Is:** A measure of how concentrated liquidity is across order book levels, borrowed from economics (market concentration index), applied to order book depth distribution.

**Equation:**

```
HHI_bid = sum((bid_qty_i / total_bid_qty)2 for i=1..N)
HHI_ask = sum((ask_qty_i / total_ask_qty)2 for i=1..N)
```

Range: **1/N to 1**

- **1/N:** Perfect uniform distribution (all levels have equal quantity)
- **1:** Complete concentration (all quantity at one level)

**Motivation:** Reveals whether liquidity is concentrated at a few levels (fragile, vulnerable) or spread evenly (resilient, robust). Different market maker strategies and market conditions produce different concentration patterns. Concentration affects execution risk and market stability.

**Market Interpretation: High HHI (>0.3):** Volume concentrated at few levels. Could indicate: (1) aggressive market making competing for rebates at best prices, (2) fleeting HFT liquidity that may vanish, (3) large resting orders at specific prices, (4) defensive concentration after uncertainty.

**Low HHI (~1/N, e.g., 0.1 for 10 levels):** Volume evenly distributed. Could indicate: (1) defensive market making expecting volatility, (2) institutional orders spread across levels, (3) algorithmic liquidity provision, (4) high confidence in current price range.

**Rising concentration:** Liquidity pulling toward top of book. May signal: (1) increased confidence (competitive aggressive quotes), (2) increased caution (pulling deep liquidity), (3) HFT dominance

increasing.

**Falling concentration:** Liquidity spreading out. May signal: (1) defensive positioning for volatility, (2) institutional depth building, (3) reduced confidence in current price.

**Asymmetric concentration ( $\text{bid\_HHI} \neq \text{ask\_HHI}$ ):** Different strategies on each side. High bid concentration + low ask concentration might indicate aggressive market making on bids (expect support) vs distributed asks (expect resistance).

---