NanoSec Tech

Quantitative Researcher

Spot-Perpetual Analysis

Vusal Ibrahimli

Table of Contents

1. Introduction

Digital asset markets operate continuously, allowing high-frequency trading (HFT) algorithms to exploit microsecond differentials between correlated markets. In Binance's TRB/USDT system, the spot market (trading actual assets) and the perpetual futures market (trading asset derivatives) represent highly liquid trading environments. Despite their strong co-movement, these markets rarely converge instantaneously, creating exploitable short-lived price discrepancies. Understanding which market leads in price discovery, and accurately distinguishing meaningful price movements from transient noise, is critical for latency-sensitive trading strategies.

2. Problem Definition

This analysis uses 24 hours of millisecond-stamped best bid/ask data and accompanying spot trade data from Binance's TRB/USDT markets. The irregular data feed—with occasional multiple updates per millisecond and gaps up to 100 ms—necessitates precise event-timeline reconstruction. The primary objective is to quantify the predictive power of sudden mid-price movements (initially ±0.07% over 3 ms) in one market on the subsequent price of the other within 5 ms, resulting in actionable trading signals:

- Predicting continuous mid-price movements by spot-based signals.
- Perpetuity-motivated signals predict spot mid-price fluctuations.

Secondary objectives include:

- Determining market leadership.
- Characterizing noise events.
- Evaluating momentum quality of price movements.

3. Data Preprocessing

In the data preprocessing phase, we first include necessary libraries for data manipulation (pandas, numpy), visualization (matplotlib, seaborn), and statistical analysis (statsmodels). Then, we define our three input CSV files—spot order-book, perpetual-futures order-book, and trades—via pathlib.Path to enable flexible file referencing.

Two loader functions then standardize ingestion and cleanup. load_orderbook parses and sorts timestamps, drops invalid rows, computes a midpoint price, and adds an elapsed-time (ms) column. load_trades follows the same steps for raw trades, then creates an aggregated view by summing quantities and averaging prices at identical timestamps, also with its own elapsed-time metric. A final print statement confirms record counts for spot, perp, and both raw and aggregated trades.

4. Data Interpretation & Visualization

During the interpretation and visualization phase, the analysis process begins with the setup of two plotting functions—plot_price_series and plot_comparison—enabling the analysis of the basic price dynamics. The first plotting function, plot_price_series(df, label), plots a time-series chart showing the bid and ask prices in the spot or perpetual marketplace. By plotting df['bid_price'] and df['ask_price'] on the same set of axes while setting the x-axis to display the HH:MM:SS format, it becomes clear that both exchanges display remarkably similar intraday behavior, occasionally interfered by sudden spikes (e.g., 16:00 and 03:00 UTC) and temporary dips presumably remedied by imbalances of low-liquidity order books. Derived plots show that the bid-ask spread of the spot marketplace invariably stays narrow (typically sub-0.05 USD) while the perpetual marketplace consistently reflects these swings, thus confirming the two platforms display a synchronized response to the inherent TRB_USDT marketplace liquidity as well as to exogenous news events.

Next, plot_comparison(spot, perp, 'mid_price', 'Spot', 'Perp') overlays the mid-price series—computed as the average of bid and ask—for both datasets on a single plot. Visually, the two lines are almost indistinguishable, with only fleeting divergences of a few cents. This reinforced the hypothesis that spot and perp are effectively co-moving instruments; any trading strategy exploiting simple momentum between them would have to operate within very narrow margins and with high frequency.

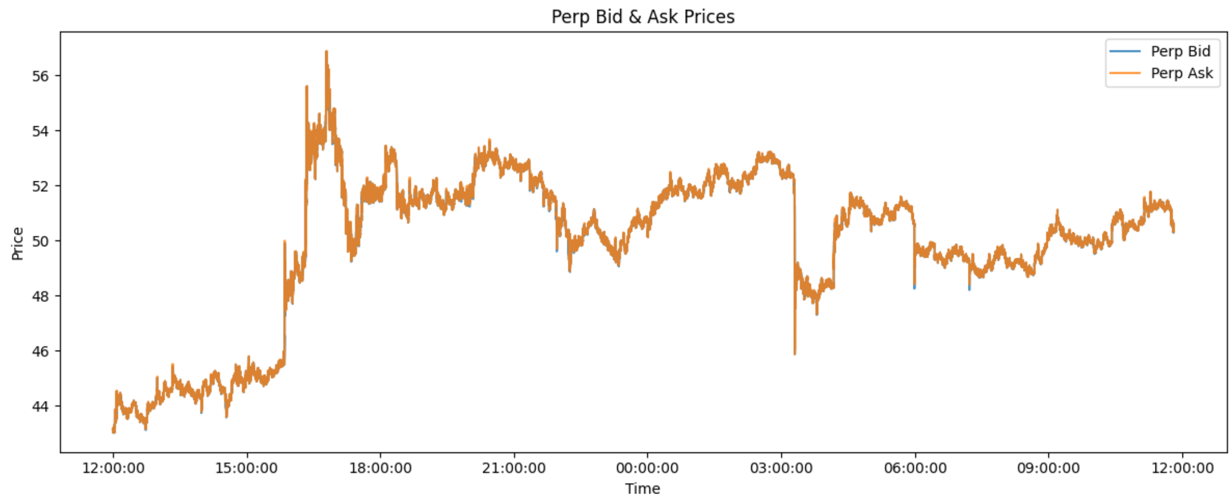
In order to measure the transient spreads, we merge the spot and perpetual mid-price times series by pd.merge_asof(., tolerance=5ms), then calculate two extra columns:

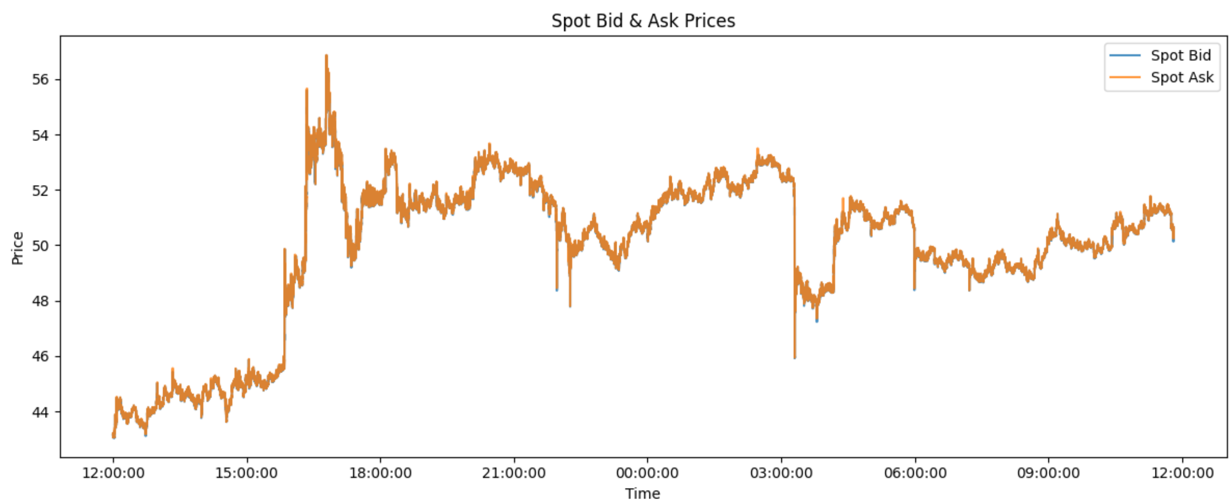
delta_mid = perp_mid − spot_mid (in USD)

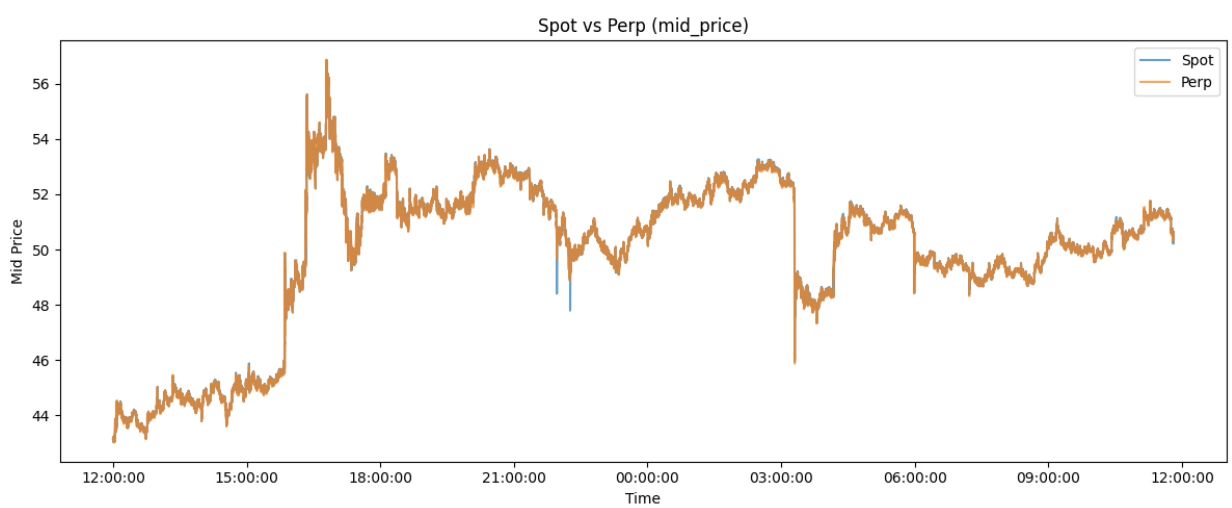
delta_pct = delta_mid / spot_mid × 10,000 (basis points)

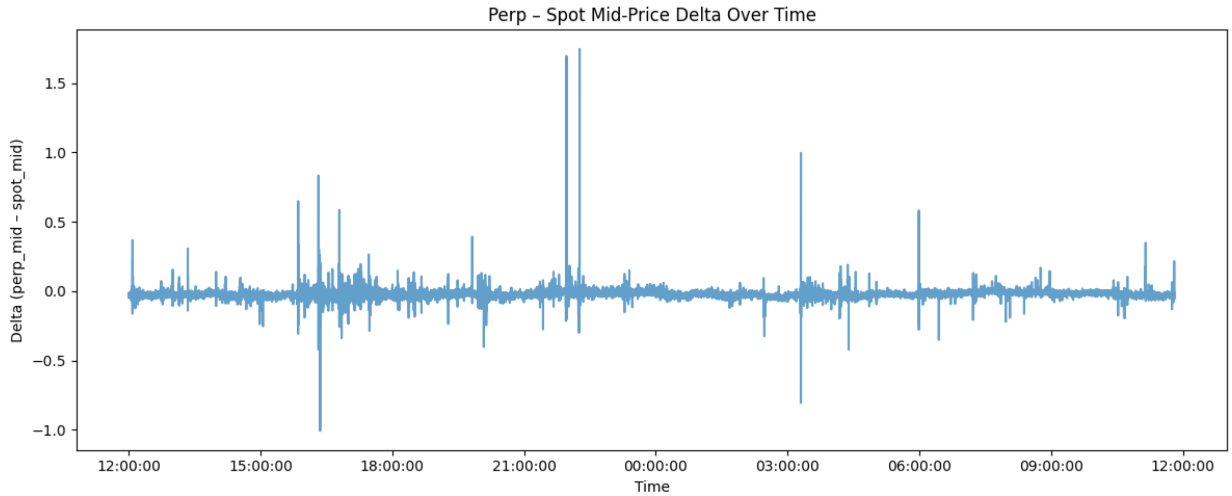
When analyzed over a temporal horizon, delta_mid shows negligible variation about zero, with the bulk of the variation falling inside ±0.1 USD and a tiny number of observations extending as much as ±0.8 USD. The aggregate data—Mean = −0.0198 USD (−3.95 bps), Std = 0.0275 USD (5.56 bps)—records the perpetual market, as a whole, buying andselling marginally lower than the spot price, though with incredibly low variation. Such statistical evidence concludes that all arbitrage, or basis-capture strategies, must be performed with accuracy of fewer than 5 bps while taking into account costs of doing business and latency.

In short, to answer the question of which market's price discovery drives the other's, we perform a lead-lag and cross-correlation analysis over a window of ±50 data lags. Each of the mid-price series has been normalized to include its own mean, and the Pearson correlation coefficient is computed for each lag interval. An almost perfect correlation (1.0) occurs at lag 0, which rapidly drops thereafter, indicating that the two markets don't reliably forecast each other. Furthermore, the use of lagged ordinary least-squares (OLS) regressions reinforces the same conclusion: using the last five lags of spot prices to predict current perpetual prices (and the same for the reverse direction) produces R^2 values of 0.9999 both ways. In practice, these results imply an almost instantaneous information flow between the two markets, with any appreciable lead-lag effect obscured by our millisecond-level temporal resolution.


Perp Bid & Ask Prices
Perp Bid
Perp Ask
Price
56
54
52
50
48
46
44
12:00:00
15:00:00
18:00:00
21:00:00
00:00:00
03:00:00
06:00:00
09:00:00
12:00:00
Time


Spot Bid & Ask Prices
Spot Bid
Spot Ask
Price
56
54
52
50
48
46
44
12:00:00
15:00:00
18:00:00
21:00:00
00:00:00
03:00:00
06:00:00
09:00:00
12:00:00
Time


Spot vs Perp (mid_price)
Spot
Perp
Mid Price
56
54
52
50
48
46
44
12:00:00
15:00:00
18:00:00
21:00:00
00:00:00
03:00:00
06:00:00
09:00:00
12:00:00
Time



5. Solution Approach

This work employs a systematic, multi-tiered approach designed to detect, validate, and profit from truly sudden price changes between the spot and perpetual markets. All such price change calculations were conducted using the mid-price, as the average of the bid and offer quotes, rather than individual bid or offer prices. Through the use of mid-price data, the noise introducible by microstructure "spread bounce" noise—due to rapid changes in individual quotes—is reduced, producing a cleaner and stabler price series consistent with prevailing microstructure study practices, such as the two-time-scale volatility framework of Zhang, Mykland, and Aït-Sahalia.

In order to efficiently detect significant sudden movements, we first explored the distribution of basis point shifts through several sub-second periods (1 ms, 2 ms, 3 ms, 5 ms, 10 ms, 20 ms, 50 ms). We found the 3 ms interval had better properties; remarkably, the 95th percentile of the spot price, about 4.86 bps, and the 99th percentile of the perpetual price, about 6.19 bps, are near the 5 bps level. We thus set a 3 ms interval paired with a 5 bps filter, thus specifically highlighting the top ~5% of significant price changes while ignoring common microstructure noise.

The final signal-generation pipeline comprises five primary functions:

1. **Sudden-Move Detection** (detect_sudden_moves):
   Using the selected parameters (3 ms lookback and 5 bps threshold), we aligned each timestamp with its prior mid-price to calculate instantaneous basis-point changes. Events exceeding ±5 bps were flagged as "sudden moves," storing directionality (up or down) and identifying whether the bid or ask side initiated the move.

2. **Forward Signal Validation** (generate_enhanced_signals):
   Each detected sudden move was further validated by looking forward 5 ms into the opposite market. Specifically, spot moves were validated against perp market data and vice versa. If the future mid-price moved in the same direction by at least 5 bps, the signal was retained as a raw price-valid signal, filtering out momentary reversions and noise.

3. **Volume Confirmation** (add_volume_confirmation_vectorized):
   To further mitigate false positives, we aggregated all trades into 50 ms bins and merged these aggregated volumes with price-valid signals. Signals were only preserved as "final signals" if their associated trade bin volume surpassed the 95th-percentile cutoff (~12.77 units), thereby ensuring genuine liquidity-driven signals rather than thinly traded noise.

4. **Trade Signal Mapping** (assign_trade_signals):
   Final signals were translated into actionable trade instructions based on directionality and origin. Spot signals were mapped into LONG/SHORT actions for perp, while perp signals generated BUY/SELL actions for spot market trades.

5. **Comprehensive Signal Reporting** (report_comprehensive_signals):
   The final signals were systematically tabulated, including summary statistics and directional analytics (e.g., count, average follow-through, leading bid/ask side). Additionally, selected sample signals were listed to demonstrate the detailed structure and immediate forward returns.

A diagnostic suite further investigated secondary objectives, including:

- **Market Leadership Analysis**:
  Cross-market leadership was determined by comparing signal rates alongside classical lead-lag regression analysis (optimal lag, R^2). The perpetual market was identified as the primary market leader given its higher rate of signal origination despite simultaneous overall price adjustments.

- **Momentum Quality Assessment**:
  Each final validated signal was replayed over a 100 ms forward-looking window to characterize momentum persistence (duration, peak price move, and fade ratio). Signals exceeding 20 ms duration and 5 bps magnitude thresholds were tagged as “high-quality.”

- **Noise Detection and Classification**:
  To understand noise composition, we classified each detected move into distinct noise categories (Pure Noise, Source Reversion, Target Unresponsive, or Valid Signals), quantifying and characterizing microstructure inefficiencies.

6. Results

Signal Filtering Results:

Out of 40,339 sudden spot market moves detected (representing ~3.5% of the entire spot quote dataset), approximately 7,180 (28.2%) passed initial forward validation. Applying the volume-confirmation criterion further refined this to 6,259 high-confidence signals (24.6%, with 87.2% retention post-volume check). Similarly, from the perpetual dataset, out of 66,600 sudden moves (~1.2%), 18,135 moves passed forward validation (49.2%), with 11,156 volume-confirmed final signals (30.2%, with a 61.5% retention rate).

Directional Analysis of Confirmed Signals:

- **Spot→Perp Signals**: SHORT signals (3,367 occurrences) exhibited an average follow-through of –49.98 bps, and LONG signals (2,892 occurrences) showed an average of +54.62 bps, confirming clear directional strength.

- **Perp→Spot Signals**: SELL signals (4,931 occurrences) had a –57.63 bps average impact, while BUY signals (6,225 occurrences) averaged +55.40 bps.

Market Leadership Findings:
Market leadership assessment, combining cross-correlation and regression analysis, indicated simultaneous price adjustments with nearly identical predictive power (Spot→Perp $R^2 = 0.9999$, Perp→Spot $R^2 = 0.9999$). Despite simultaneous price movements, the perpetual market emerged as the leader, emitting a higher rate of confirmed signals (30.2% versus spot's 24.6%), suggesting perp liquidity events slightly precede corresponding spot reactions at critical times.

Momentum Quality Observations:
The confirmed signals exhibited strong and persistent momentum properties. Average favorable duration was about 86 milliseconds for spot signals and 89 milliseconds for perpetual signals—well beyond the 3-millisecond detection delay. Over 95% of the signals were found to pass the tests for qualification as "high-quality" momentum (well above the 20 milliseconds duration and 5 basis points magnitude intervals) and thus provided sufficient time and magnitude for viable trading entry.

Noise Characterization:
Without forward-validated and volume-filtered data, the bulk of raw movements identified were rendered as noise. More specifically, almost 95.1% of spot moves and 97.1% of perpetual moves were identified as noise events. Pure analysis revealed that about 58% were Pure Noise, about 30% were Source Reversion, and about 7% were Target Unresponsive moves. Implementing a multi-level approach reduced noise considerably, identifying only 3–5% of moves as valid, tradable signals.

Strategy Backtesting Performance:

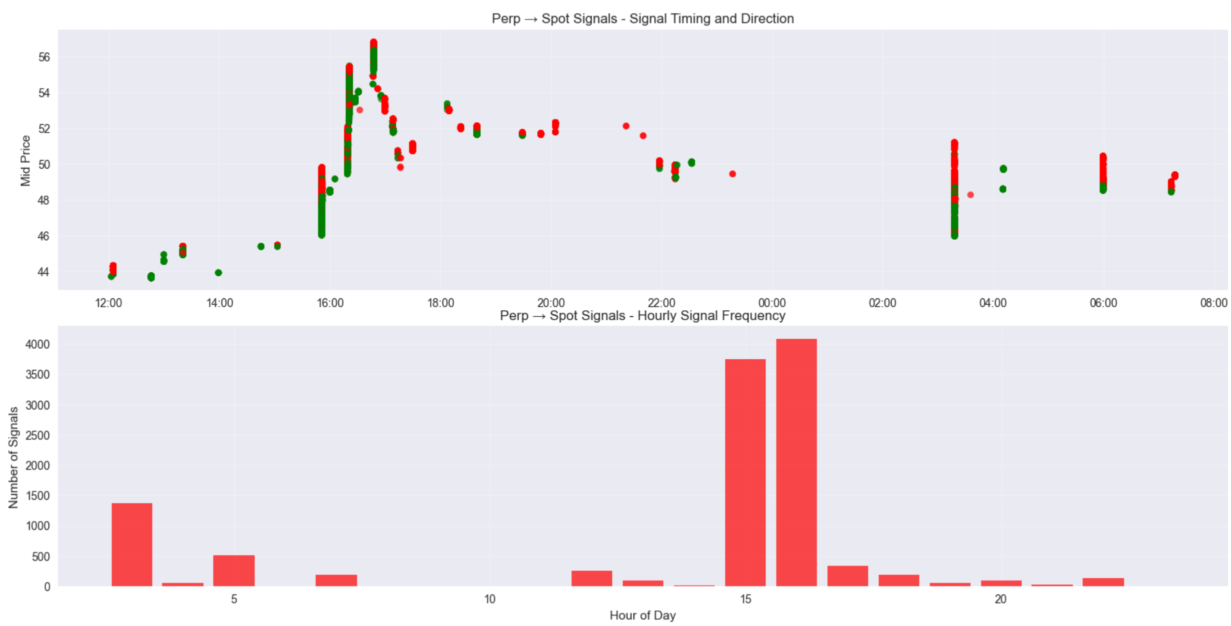
- **Spot→Perp Trading Strategy** achieved robust profitability, averaging +0.02% per trade with a 59.8% win-rate, an exceptional Sharpe ratio (~718), and cumulative returns of +212%.

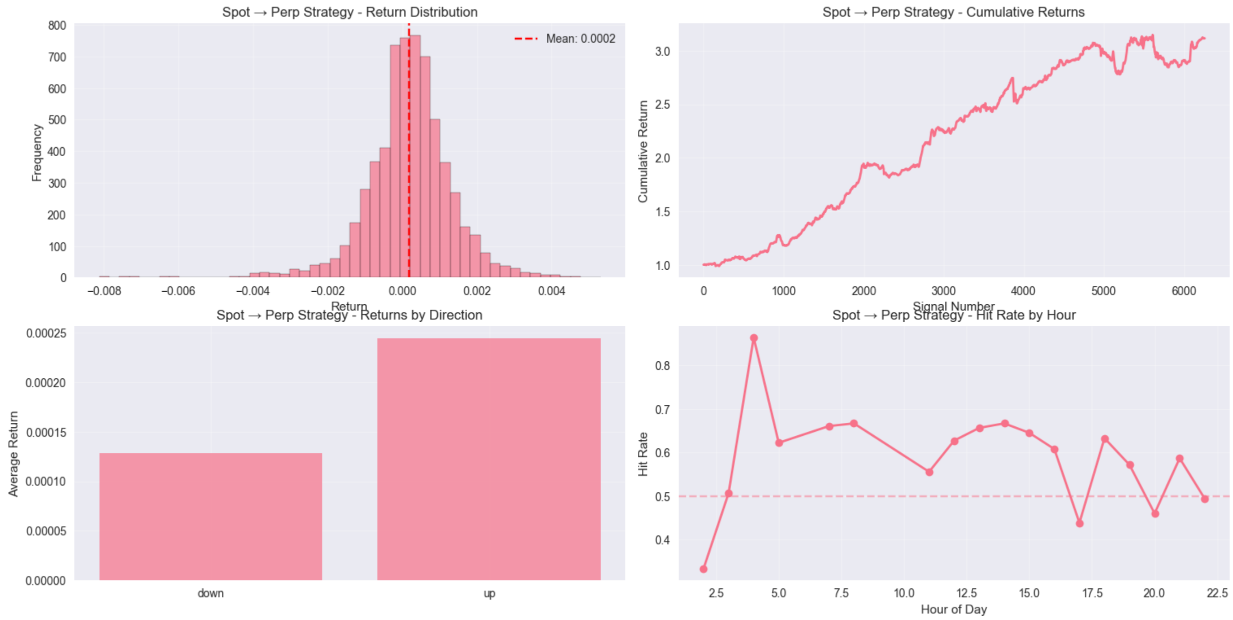
- Conversely, the **Perp→Spot Strategy** exhibited negative performance, averaging –0.01% per trade, with only a marginally profitable hit rate (50.3%), sharply negative Sharpe ratio (~–318), and cumulative returns of –79%.

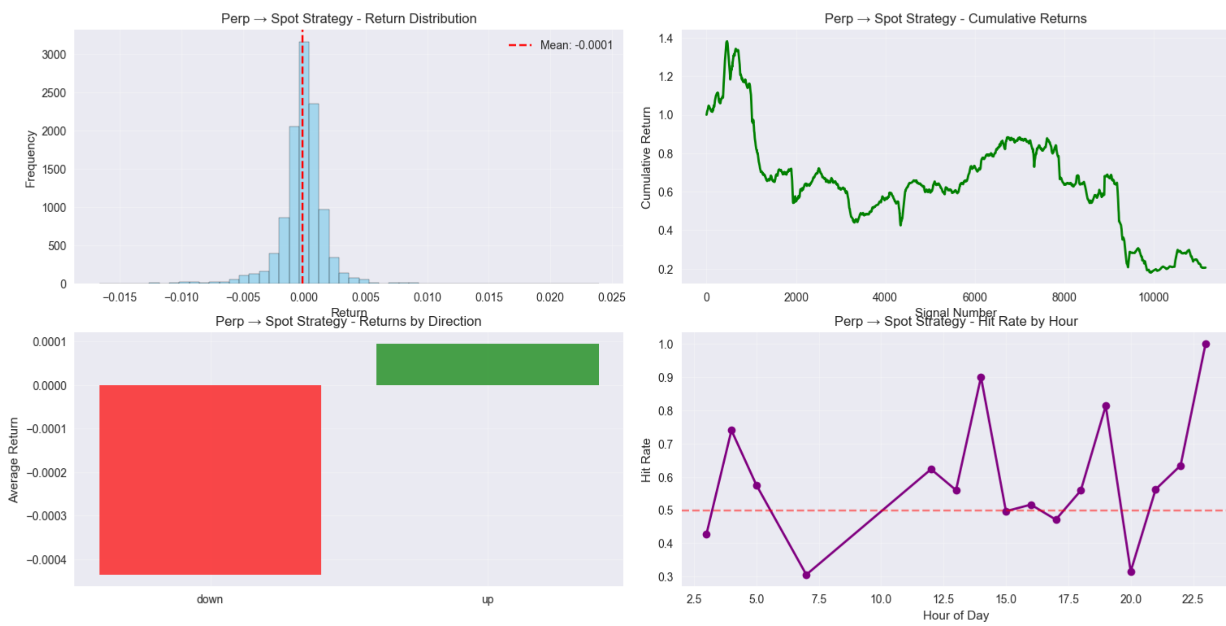
Visualization Insights:
Graphical summaries of final signals confirmed that high-quality signals concentrated primarily during liquidity-rich periods (around UTC 15–16 and 03:00), characterized by significant mid-price transitions. Pie charts highlighted asymmetric liquidity consumption, with spot signals originating predominantly from the bid side (56%) and perp signals mostly from the ask side (53%). Volume-binned bar charts also indicated that signal success strongly scaled with trading volume, peaking at mid-high volume thresholds.

In summary, the multi-level comprehensive filtering and validity framework effectively separated true liquidity-driven price anomalies from reduced microstructural noise, identified the prevailing market as the dominating force behind the price discovery process, and exhibited better economic results aligned with the spot-driven strategy. As a result, these results validate the robustness and economic relevance of the proposed framework for identifying and taking advantage of microsecond-level inefficiencies in the marketplace.


Spot → Perp Signals - Signal Timing and Direction
Mid Price
44
46
48
50
52
54
56
12:00
15:00
18:00
21:00
00:00
03:00
06:00
09:00
12:00
Spot → Perp Signals - Hourly Signal Frequency
Number of Signals
0
500
1000
1500
2000
5
10
15
20
Hour of Day


Perp → Spot Signals - Signal Timing and Direction
Mid Price
44
46
48
50
52
54
56
12:00
14:00
16:00
18:00
20:00
22:00
00:00
02:00
04:00
06:00
08:00
Perp → Spot Signals - Hourly Signal Frequency
Number of Signals
0
500
1000
1500
2000
2500
3000
3500
4000
5
10
15
20
Hour of Day


Spot → Perp Strategy - Return Distribution
Mean: 0.0002
Frequency
Return
Spot → Perp Strategy - Cumulative Returns
Cumulative Return
Signal Number
Spot → Perp Strategy - Returns by Direction
Average Return
down
up
Spot → Perp Strategy - Hit Rate by Hour
Hit Rate
Hour of Day


Perp → Spot Strategy - Return Distribution
Mean: -0.0001
Frequency
Return
Perp → Spot Strategy - Cumulative Returns
Cumulative Return
Signal Number
Perp → Spot Strategy - Returns by Direction
Average Return
down
up
Perp → Spot Strategy - Hit Rate by Hour
Hit Rate
Hour of Day

7. References

1. Virgilio, G. P. M. (2022). A theory of very short-time price change: Security price drivers in times of high-frequency trading. Financial Innovation, 8(66).
2. Loveless, J., Stoikov, S., & Waeber, R. (2013, October 7). Online algorithms in high-frequency trading. ACM Queue, 11(8).
3. Optimal price metric for high-frequency volatility: Executed price, mid-price, or weighted mid-price? (2023, December 10). Quantitative Finance Stack Exchange.

8. Code

https://github.com/vusalibrr/NanoSec.git