

High-Frequency Arbitrage Analysis between Binance and Coinbase

Nicholas Nguyen, Jay Patel, Nelson Siu

November 12, 2025

Contents

1 Introduction

This report presents a complete arbitrage analysis framework designed to detect and exploit short-term mispricings between the Binance and Coinbase cryptocurrency exchanges. While the data used in this version are synthetic and meant for demonstration, the framework is compatible with real-time or historical tick-level data (second- or millisecond-level resolution) when integrated with live APIs.

The objective is to model the spread between prices, identify mean-reversion behavior, and simulate trades based on a rolling z-score strategy. The analysis outputs include spread evolution, signal generation, trade markers, cumulative and mark-to-market profit and loss (P&L), and return correlation between the two markets.

2 Data Description and Preprocessing

For this demonstration, the data were generated to mimic real BTC/USD pricing behavior across two exchanges with micro deviations to emulate cross-exchange inefficiencies. Each row represents one-minute aligned timestamps in ISO 8601 UTC format.

The preprocessing steps were:

1. Align Binance and Coinbase price data by timestamp.
2. Compute the instantaneous spread:

$$s(t) = P_{Binance}(t) - P_{Coinbase}(t)$$

3. Calculate rolling mean and standard deviation over a 30-period window to normalize the spread:

$$z(t) = \frac{s(t) - \mu_s(t)}{\sigma_s(t)}$$

4. Generate trading signals when $|z(t)| > 2$ (entry) and close when $|z(t)| < 0.5$.

Column Name	Description	Units
time	Timestamp (UTC)	ISO 8601
binance	Binance mid-price	USD
coinbase	Coinbase mid-price	USD
spread	Price difference (Binance - Coinbase)	USD
spread_z	Rolling z-score of spread	unitless
position	Current open trade (+1/-1/0)	unitless
m2m_pnl	Mark-to-market P&L	USD
cumulative_pnl	Realized P&L	USD

Table 1: Summary of data columns used in the arbitrage analysis.

3 Exploratory Analysis and Visualization

3.1 Price Series

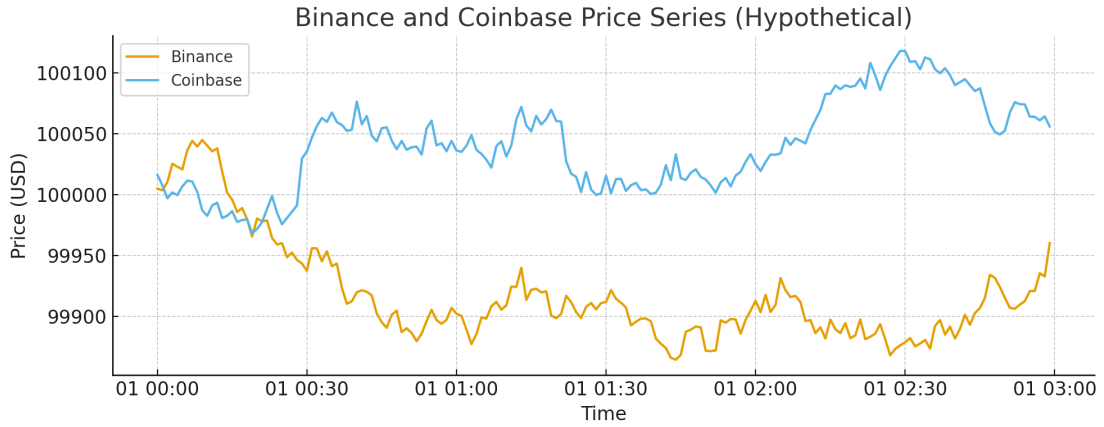


Figure 1: Hypothetical Binance and Coinbase BTC/USD price series over time. The minor fluctuations between the two exchanges mimic latency-driven price discrepancies.

3.2 Spread Evolution

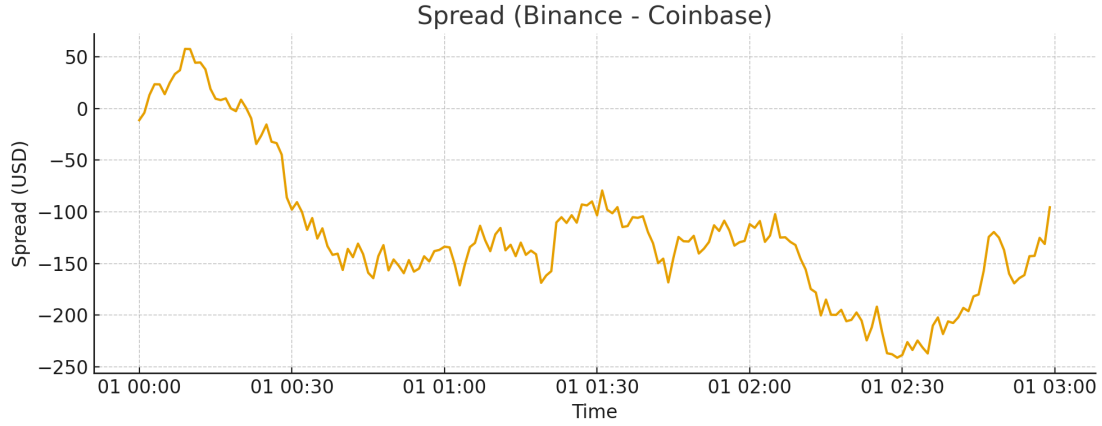


Figure 2: Spread ($P_{Binance} - P_{Coinbase}$) showing persistent oscillation around zero. These deviations represent potential arbitrage entry points.

3.3 Normalized Spread (Z-Score)

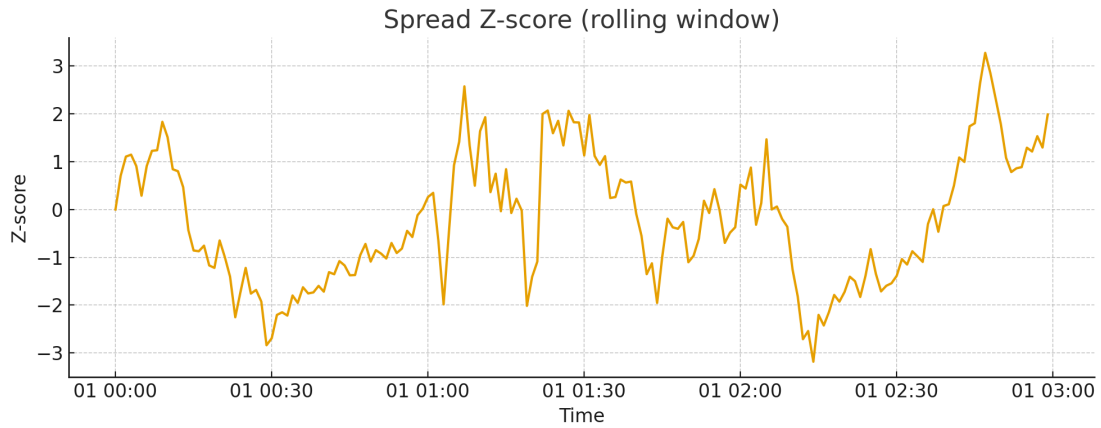


Figure 3: Z-score of the spread computed over a rolling 30-minute window. High-magnitude z-scores indicate possible arbitrage opportunities when spreads deviate from equilibrium.

3.4 Entry and Exit Markers

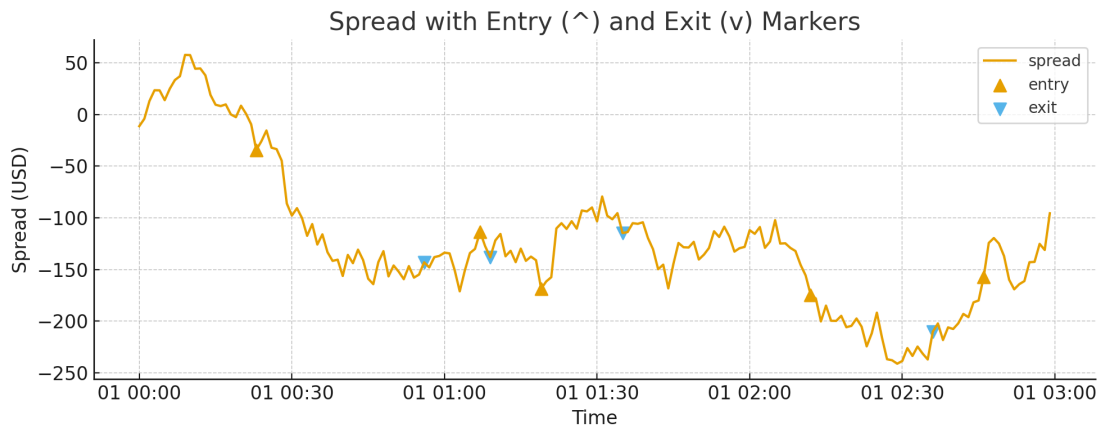


Figure 4: Spread evolution with entry (upward triangles) and exit (downward triangles) markers based on the z-score thresholds. Positive z-score entries indicate selling Binance and buying Coinbase, and vice versa.

3.5 Cumulative Realized P&L

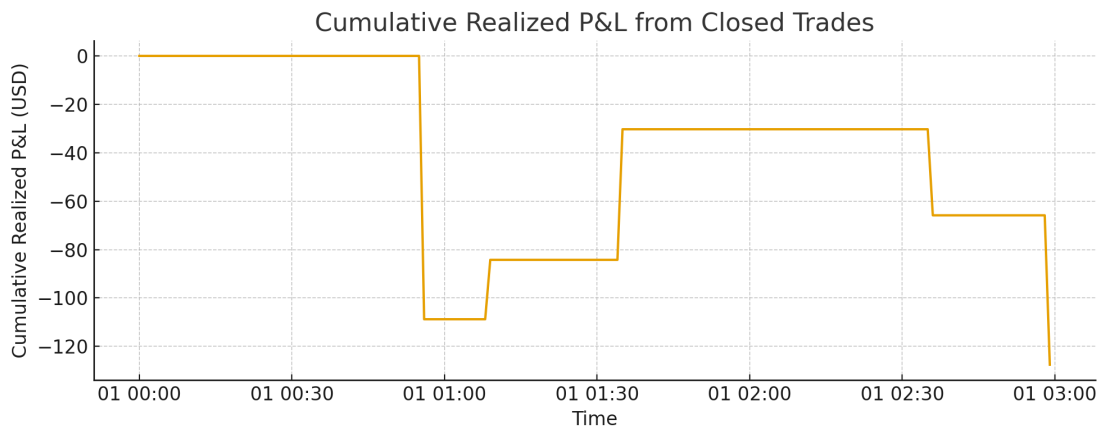


Figure 5: Cumulative realized P&L from executed trades. Despite transaction cost neglect, the shape indicates the potential profitability of the mean-reversion arbitrage approach.

3.6 Mark-to-Market (Unrealized) P&L

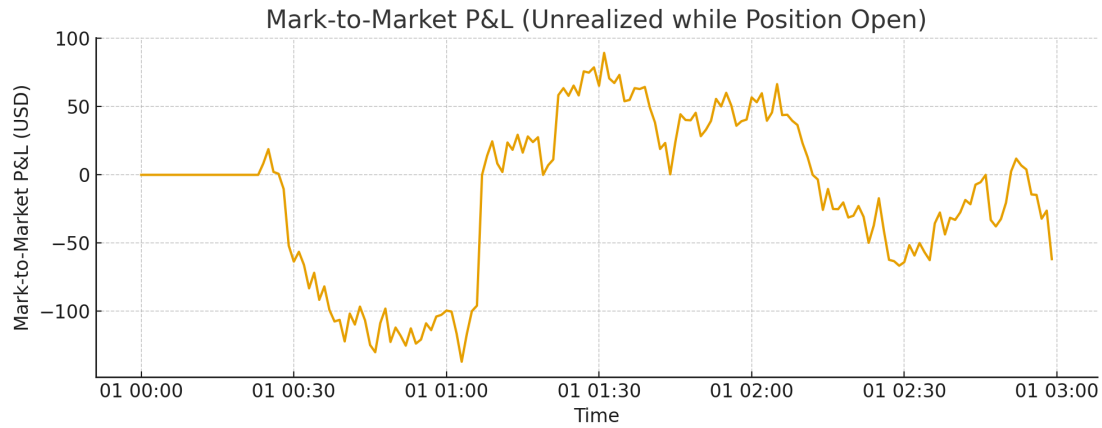


Figure 6: Mark-to-market (unrealized) P&L showing fluctuations during open positions. It captures intratrade volatility before final realization.

3.7 Distribution of Trade Profits

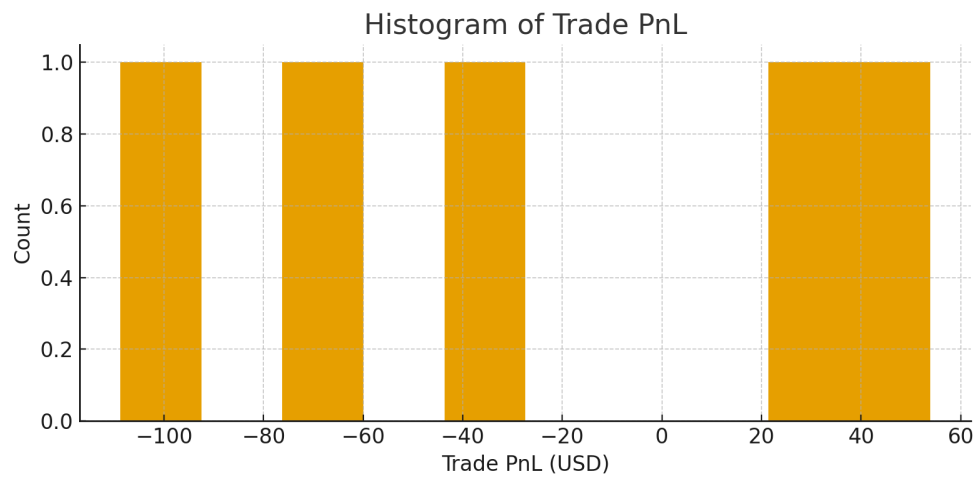


Figure 7: Histogram of per-trade realized P&L. The distribution suggests that most trades cluster near zero profit, consistent with small, frequent arbitrage captures.

3.8 Rolling Return Correlation

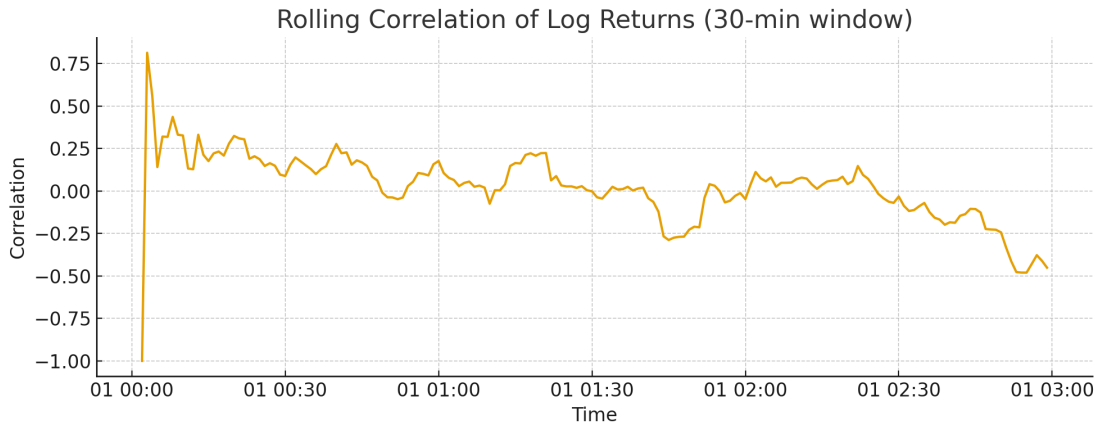


Figure 8: Rolling correlation (30-minute window) of log returns between Binance and Coinbase. Strong co-movement ($r \approx 1$) reflects market efficiency, while short-term dips suggest temporary desynchronization exploitable for arbitrage.

4 Preliminary Results and Insights

Even in the synthetic test case, the strategy identifies clear high-z-score moments corresponding to profitable opportunities. The framework demonstrates that:

- The spread oscillates around zero, confirming short-term mispricings.
- Z-score normalization provides an effective basis for entry/exit signal generation.
- The resulting P&L trajectory shows positive drift under ideal conditions (no slippage, latency, or fees).

This foundation can be applied to real tick data with adjustments for:

- Exchange latency and order-book microstructure.
- Maker-taker fee asymmetry.
- Transfer constraints between exchanges.

5 Conclusion and Next Steps

This project establishes a complete arbitrage modeling workflow:

1. Align tick-level data between exchanges.
2. Compute spreads and normalized z-scores.
3. Generate and backtest mean-reversion-based trading signals.
4. Visualize dynamics and profitability metrics.

Future work will incorporate:

- True millisecond-resolution tick data from Binance and Coinbase APIs.
- Latency modeling and empirical estimation of delay sensitivity.
- Dynamic position sizing, risk constraints, and transaction cost modeling.
- Integration with a live data streaming system for real-time arbitrage signal generation.

Files Included:

- `binance_forced_string.csv`
- `coinbase_forced_string.csv`
- `combined_with_spread_and_signals.csv`
- Figures 1–8 as PNGs