Pipeline Architecture

The platform implements a multi-strategy cryptocurrency trading system with the following stages: data processing, signal generation, portfolio construction, backtesting, and performance evaluation.

Data Pipeline

Price Matrix Construction

Given raw OHLCV data for N symbols over T time periods, construct the price matrix:

$$P \in \mathbb{R}^{T \times N}, \quad P_{t,i} = \text{close price of asset } i \text{ at time } t$$

Data preprocessing includes removing duplicates, handling missing values via forward-filling, and resampling to hourly frequency.

Signal Generation Strategies

The platform implements three complementary signal strategies:

Momentum Strategy

Captures trending behavior using past returns:

$$\text{Signal}_{t,i}^{mom} = \frac{P_{t-1,i} - P_{t-1-L,i}}{P_{t-1-L,i}}$$

where L is the lookback period (typically 168 hours = 1 week). The shift by 1 period prevents look-ahead bias.

Mean Reversion Strategy

Exploits price extremes using z-score of returns:

$$r_{t,i} = \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}}$$

$$\mu_{t,i} = \frac{1}{L} \sum_{k=0}^{L-1} r_{t-k,i}$$

$$\sigma_{t,i} = \sqrt{\frac{1}{L}\sum_{k=0}^{L-1}(r_{t-k,i} - \mu_{t,i})^2}$$

$$\text{Signal}_{t,i}^{mr} = -\frac{r_{t-1,i} - \mu_{t-1,i}}{\sigma_{t-1,i}}$$

The negative sign implements mean reversion: oversold assets (negative z-score) produce positive signals.

EWMA Crossover Strategy

Identifies trend changes via exponential moving average divergences:

$$\text{EWMA}_{t,i}^{fast} = \alpha_{fast} P_{t,i} + (1 - \alpha_{fast})\text{EWMA}_{t-1,i}^{fast}$$

$$\text{EWMA}_{t,i}^{slow} = \alpha_{slow} P_{t,i} + (1 - \alpha_{slow})\text{EWMA}_{t-1,i}^{slow}$$

where $\alpha = 2/(span + 1)$. The signal is volatility-normalized:

$$\text{Signal}_{t,i}^{ewma} = \frac{\text{EWMA}_{t-1,i}^{fast} - \text{EWMA}_{t-1,i}^{slow}}{\sigma_{t-1,i}^{rolling}}$$

where $\sigma_{t,i}^{rolling}$ is the rolling standard deviation over a specified window.

Cross-Sectional Z-Score Transformation

For comparability across strategies, raw signals are cross-sectionally standardized:

$$\bar{s}_t = \frac{1}{N}\sum_{i=1}^{N}\text{Signal}_{t,i}$$

$$\sigma_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{Signal}_{t,i} - \bar{s}_t)^2}$$

$$Z_{t,i} = \frac{\text{Signal}_{t,i} - \bar{s}_t}{\sigma_t}$$

Quantile Filtering

Filter signals by cross-sectional quantiles to select top/bottom performers:

$$F_{t,i} = \begin{cases} Z_{t,i} & \text{if } Z_{t,i} \geq Q_t^{top} \text{ (long)} \\ Z_{t,i} & \text{if } Z_{t,i} \leq Q_t^{bottom} \text{ (short)} \\ 0 & \text{otherwise} \end{cases}$$

where Q_t^{top} and Q_t^{bottom} are the top and bottom quantile thresholds (e.g., 0.8 and 0.2).

Re-Z-Score on Active Names

After filtering, re-standardize on active (non-zero) positions only:

Let $A_t = \{i : F_{t,i} \neq 0\}$ be the active set at time t:

$$\bar{f}_t = \frac{1}{|A_t|} \sum_{i \in A_t} F_{t,i}$$

$$\text{Signal}_{t,i}^{final} = \begin{cases} \frac{F_{t,i} - \bar{f}_t}{\sigma_{A_t}} & \text{if } i \in A_t \\ 0 & \text{otherwise} \end{cases}$$

Portfolio Construction

Combine multiple strategies using weighted averaging:

$$Z_{t,i}^{combined} = \sum_{s=1}^{S} w_s \cdot Z_{t,i}^{(s)}$$

where w_s are strategy weights satisfying $\sum_s w_s = 1$, and $Z_{t,i}^{(s)}$ is the unfiltered z-scored signal from strategy s.

The combined signal undergoes filtering and re-z-scoring (same process as above) to produce final portfolio weights $W_{t,i}$.

Target Engineering

Compute forward returns for evaluation:

$$\text{Target}_{t,i} = \frac{P_{t+h,i} - P_{t,i}}{P_{t,i}}$$

where h is the forward period (typically 1 hour).

Backtesting

Timing Convention: Weights determined at close of day t are held through day $t+1$, realizing returns from close t to close $t+1$.

Asset returns:

$$R_{t,i} = \frac{P_{t,i} - P_{t-1,i}}{P_{t-1,i}}$$

Portfolio return at time t:

$$R_t^{portfolio} = \sum_{i=1}^{N} W_{t-1,i} \cdot R_{t,i}$$

Turnover (portfolio rebalancing):

$$\text{Turnover}_t = \frac{1}{2} \sum_{i=1}^{N} |W_{t,i} - W_{t-1,i}|$$

Transaction costs:

$$\text{Cost}_t = \text{Turnover}_t \times (\text{TC}_{bps} + \text{Slippage}_{bps})$$

Net return:

$$R_t^{net} = R_t^{portfolio} - \text{Cost}_t$$

Cumulative return:

$$\text{CR}_t = \prod_{k=1}^{t} (1 + R_k^{net})$$

Performance Metrics

Annualized Return (geometric mean):

$$R_{annual} = \left(\prod_{t=1}^{T} (1 + R_t^{net}) \right)^{\frac{8760}{T}} - 1$$

where 8760 = 365 × 24 hours per year.

Annualized Volatility:

$$\sigma_{annual} = \sqrt{8760} \cdot \sqrt{\frac{1}{T}\sum_{t=1}^{T}(R_t^{net} - \bar{R})^2}$$

Sharpe Ratio:

$$\text{Sharpe} = \frac{R_{annual} - R_f}{\sigma_{annual}}$$

Maximum Drawdown:

$$\text{MDD} = \min_t \left(\frac{\text{CR}_t - \max_{k \le t} \text{CR}_k}{\max_{k \le t} \text{CR}_k} \right)$$

Calmar Ratio:

$$\text{Calmar} = \frac{R_{annual}}{|\text{MDD}|}$$

Win Rate:

$$\text{WR} = \frac{\sum_{t=1}^{T} 1(R_t^{net} > 0)}{T}$$

Profit Factor:

$$\text{PF} = \frac{\sum_{t:R_t^{net}>0} R_t^{net}}{|\sum_{t:R_t^{net}<0} R_t^{net}|}$$