Statistical Outlier Detection Notes

Outlier Detection Using Mean and Standard Deviation (Z-Score Based Outlier Detection)

To detect outliers in a dataset Δ, we use the mean and standard deviation:

- $\mu(\Delta)$: Mean of the data
- $\sigma(\Delta)$: Standard deviation of the data

Normal Range

The normal range is defined as:

$$\mu(\Delta) \pm 2\sigma(\Delta)$$

This means most data points (about 95% if normally distributed) are expected to lie within this range.

Outlier Condition

A value is considered an outlier if:

$$\Delta < \mu(\Delta) - 2\sigma(\Delta) \quad \text{or} \quad \Delta > \mu(\Delta) + 2\sigma(\Delta)$$

- Δ - Orderbook Delta Depth of 5% from Coinbase
- $\mu(\Delta)$ - Mean of Δ
- $\sigma(\Delta)$ - Standard deviation of the dataset

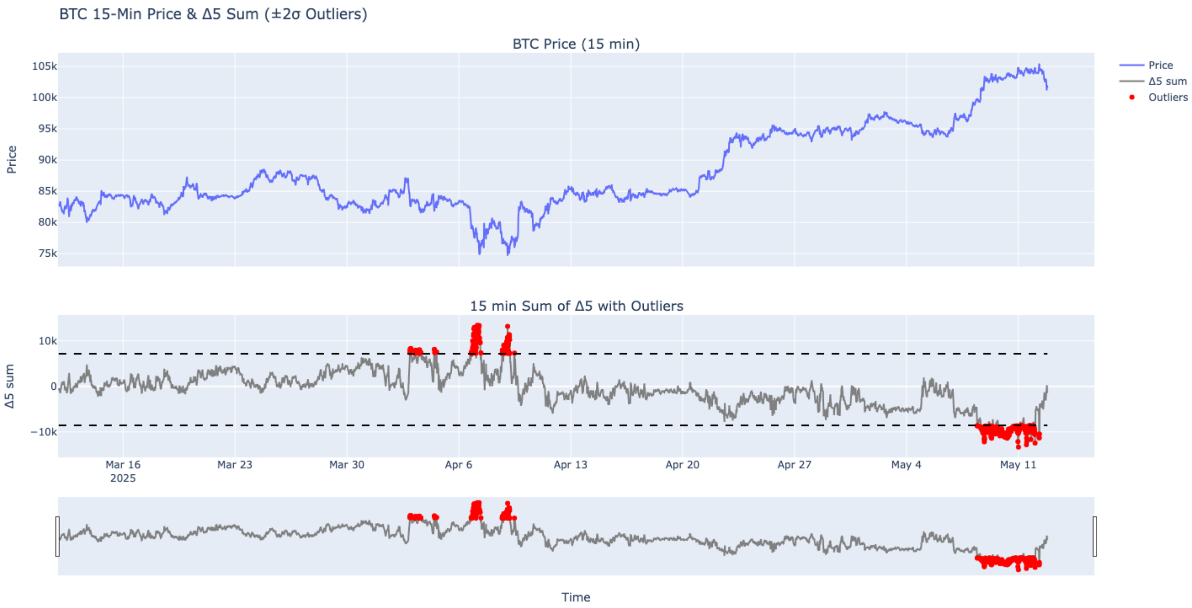

Figure 1: results over a two month data set

Idea behind

- This method assumes data is roughly normally distributed.
- Using 2σ captures approximately 95% of data points under a normal distribution.
- You can adjust the multiplier (e.g., 3σ) for stricter or looser thresholds.

Future Plans

- Test on more data
- use rolling windows (e.g. 1 day or 1 week) for local context.
- Compare sensitivity with +- 1.5σ or +-2.5σ

Order Book Delta RSI (14-period)

The Relative Strength Index (RSI) applied to the order book delta Δ is defined as:

$$\mathrm{RSI}_{\Delta}(t) = 100 - \frac{100}{1 + RS(t)}$$

where

$$RS(t) = \frac{\text{Average Gain over 14 periods}}{\text{Average Loss over 14 periods}}$$

Dictionary of Terms

- $\Delta(t)$ — Order book delta at time t (e.g., 5% depth imbalance)
- Gain — Positive change in Δ: $\Delta(t) - \Delta(t-1) > 0$