Regime Detection in Financial Markets using Signature Methods

Update 01

Zhe GUAN

The University of Southamtpon
zg2u24@soton.ac.uk

15th June 2025

Overview

Repository: Link

- **Path Construction**
  - Bitcoin daily market from 2014-09-17 to 2024-11-07
  - Path: ’close price’ and ‘volume’ as a 2D array representing price-volume trajectory
  - windows = 7 days

- **Transformation**
  - Time-integrated Transformation added as additional dimension
  - Invisibility Reset Transformation added with two terminal points to the path
  - Lead-Lag Transformation added to create a forward and delayed version
  - Cumulative Sum transformation added

- **Signature** extracted with *iisignature* library truncated at level 4

  The final feature vector has ~1,554 dimensions per window segment.

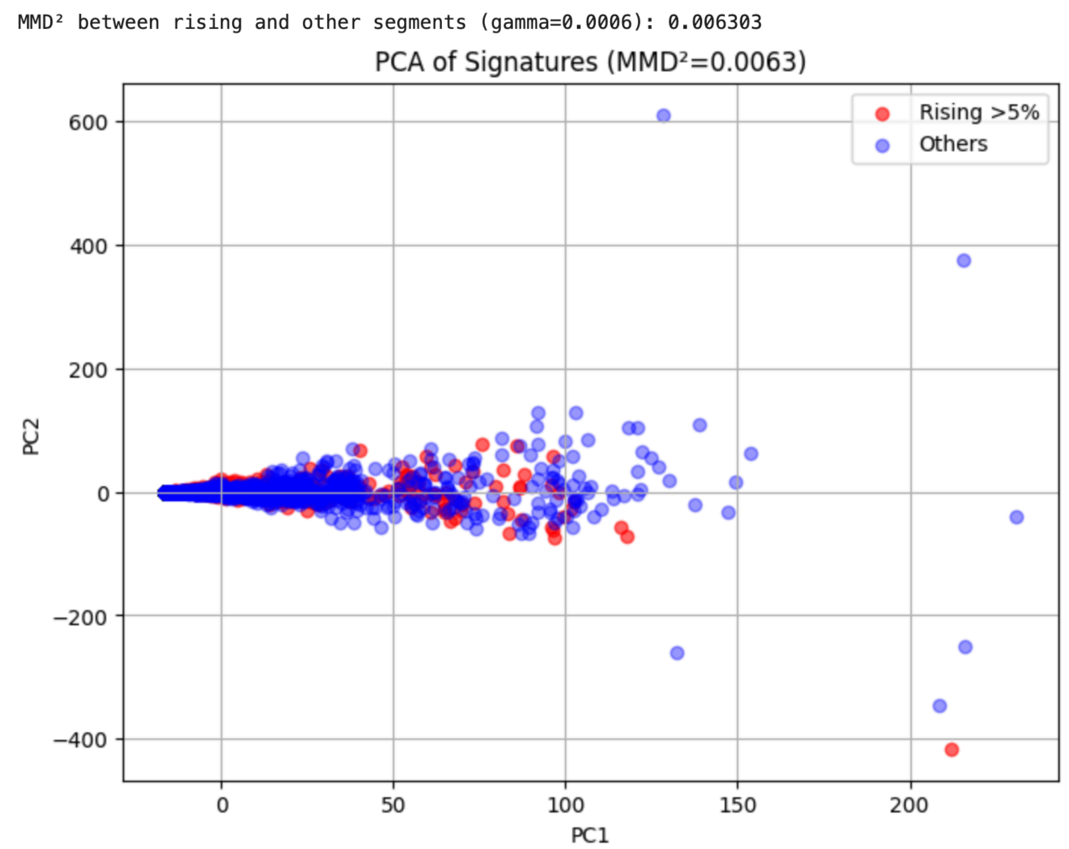
MMD and distribution differences

Goal:

Use MMD² to measure how different the **signature distributions** are between **rising (≥5%)** and **non-rising** path segments.

Steps:

1. Extract signature features from transformed price-volume segments.
2. Identify reference set: all windows with ≥5% price increase.
3. Compute **MMD² using RBF kernel**
4. **PCA** projects high-dimensional signature vectors to 2D



MMD:

$$\mathrm{MMD}^2(P,Q) = \mathbb{E}_{x,x'}\left[k\left(x,x'\right)\right] + \mathbb{E}_{y,y'}\left[k\left(y,y'\right)\right] - 2\mathbb{E}_{x,y}[k(x,y)]$$

- We use the full MMD² formulation to compare the signature distributions of rising vs. non-rising segments, and use PCA to visualise. It shows with current settings, it is difficult to distinguish between rising and non-rising patterns and maybe more input features should be used.

Kernel Scoring Rule for Distribution Similarity

Goal:

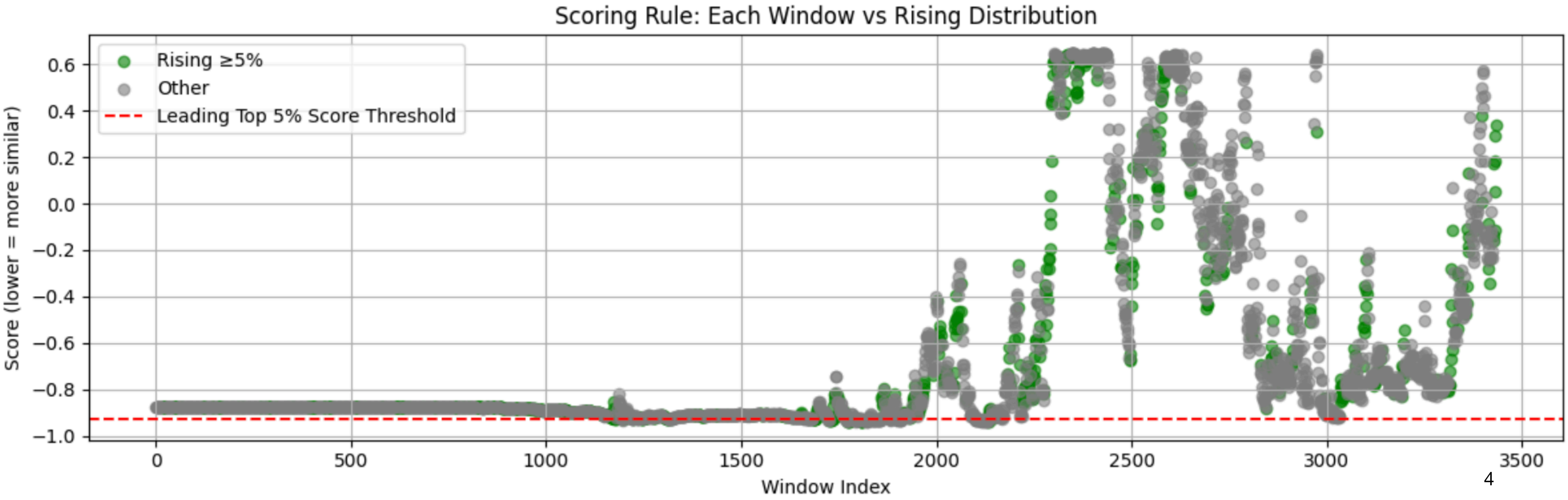
Use a kernel scoring rule to quantify how similar **each window's signature** is to **the distribution of rising segments (≥5% price increase)**.

Steps:

1. Extract signature features from price-volume segments.
2. Identify reference set: all windows with ≥5% price increase.
3. Fit RBF kernel in scaled signature space.
4. Compute a score for each window that estimates

Kernel Scoring Rule

$$s_k(\mathbb{P}, y) = \mathbb{E}_{x,x'\sim\mathbb{P}}\left[k\left(x, x'\right)\right] - 2\mathbb{E}_{x\sim\mathbb{P}}[k(x, y)]$$



Finding Segments Similar to a Given Reference

Goal:

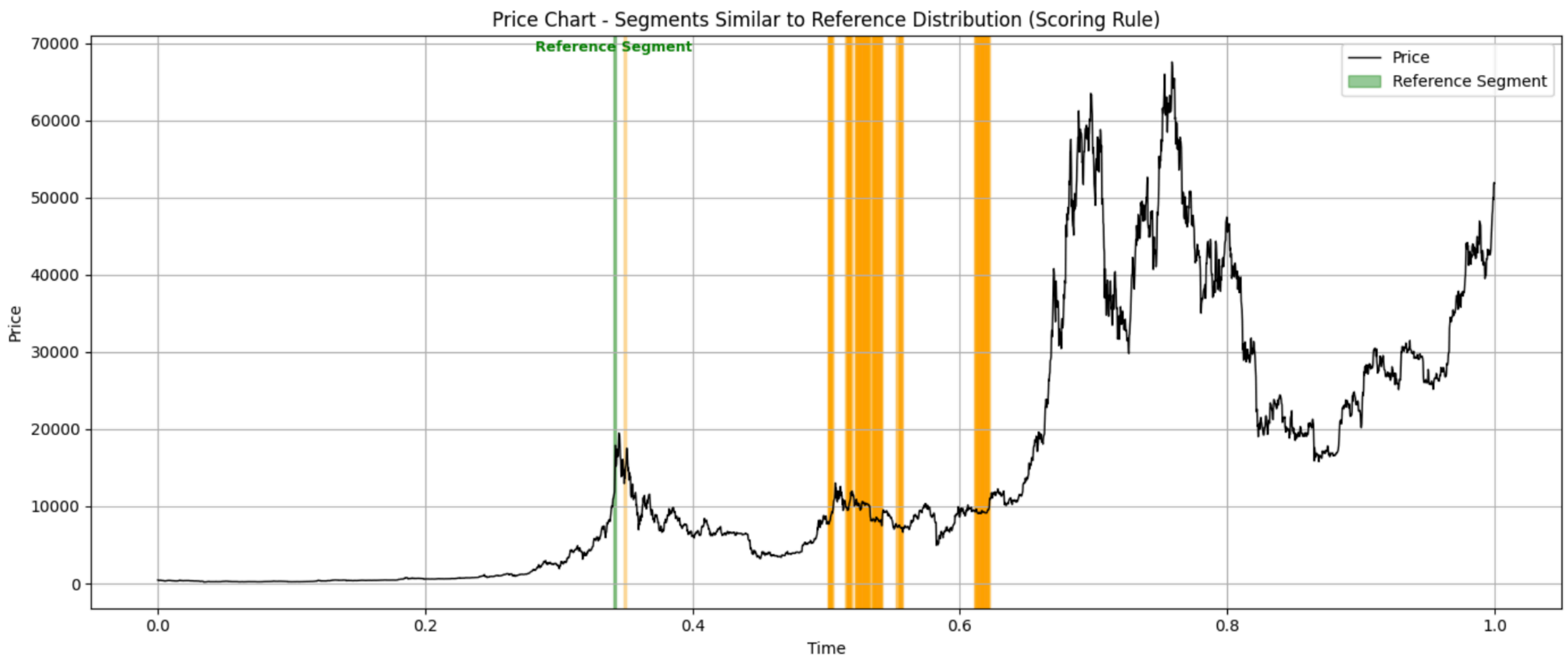
Given a specific 7-day window (e.g., one with the maximum 5%+ rise), find other windows whose **path signature features are most similar to it**.

Steps:

All steps similar with previous slides, but only pick the lowest 3% as **most similar segments**.

Kernel Scoring Rule

$$s_k(\mathbb{P}, y) = \mathbb{E}_{x,x' \sim \mathbb{P}}\left[k\left(x, x'\right)\right] - 2\mathbb{E}_{x \sim \mathbb{P}}[k(x, y)]$$



Summary and Next Step

- Reviewed the Bloomberg Quant Seminar (May 2025) and the repository tutorial *"signatures_introduction"* to understand signature-based modeling techniques.
- Based on that, I prepared an initial framework for path signature extraction and distribution-based analysis.
- completed the three initial studies, but the results should be improved

Next Steps:

- Study *An Empirical Analysis of Path Signatures* (arXiv:2306.15835) to refine modeling strategies and gain more insights to improve results
- Experiment with: Richer input features and Longer or multi-scale time windows, etc
- Extend current methods to match a given regime with similar historical regimes