Non-linear Signals

Quantitative Strategies – Traders@SMU

April 29, 2026

Team

Anders Bisgaard

Ameen Zia

Connor Reed

Arjan Oberoi

Demetrios Lahiri

Rebecca Preston

Zabe Nash

Junaid Khan

Non-Linear Signals

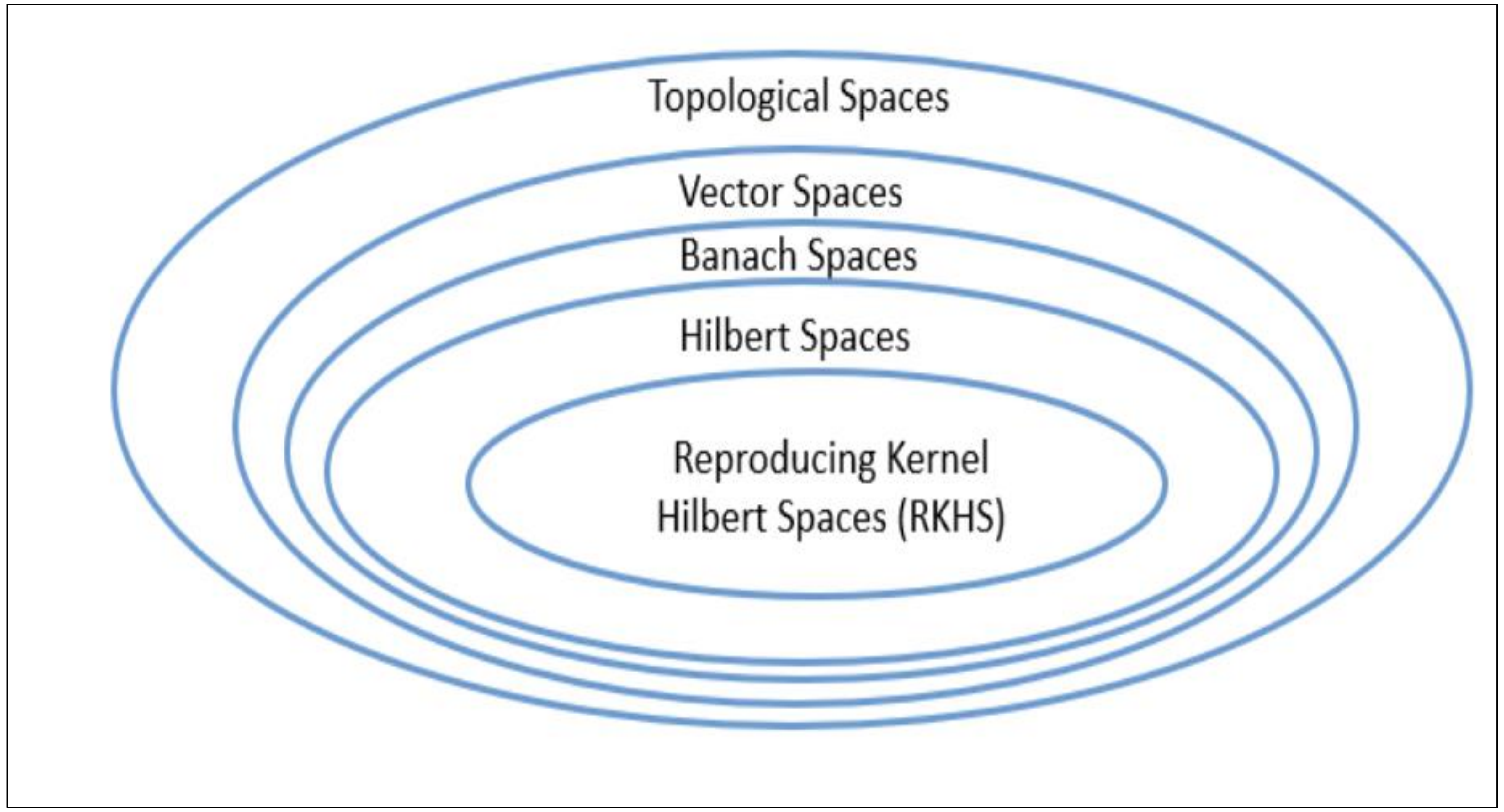
Agenda

- Introductions
- Spaces, Hilbert Spaces, and RKHS
- Layers Overview
- Back Testing
- Next Steps

Spaces, Hilbert Spaces, and RKHS

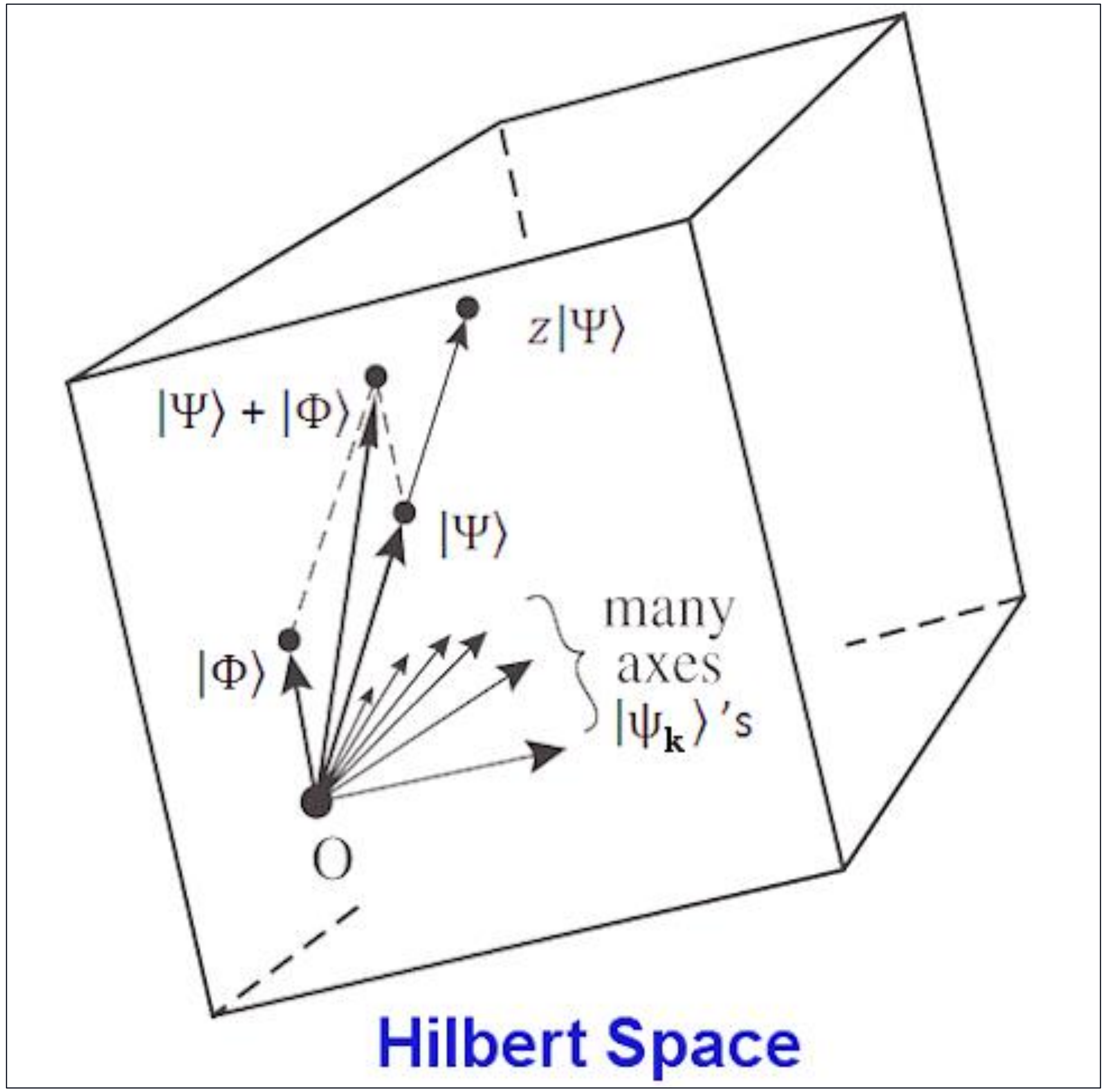
Mathematical Spaces: The Playground of Data

- A "space" is a set of objects + rules for how they relate
- Defines what exists (points, vectors, functions) and what you can do with them
- Euclidean space (R^n): Finite dimensions (X, Y, Z). Our intuitive 3D world
- Hilbert space (H): Infinite dimensions. Points can be functions or curves



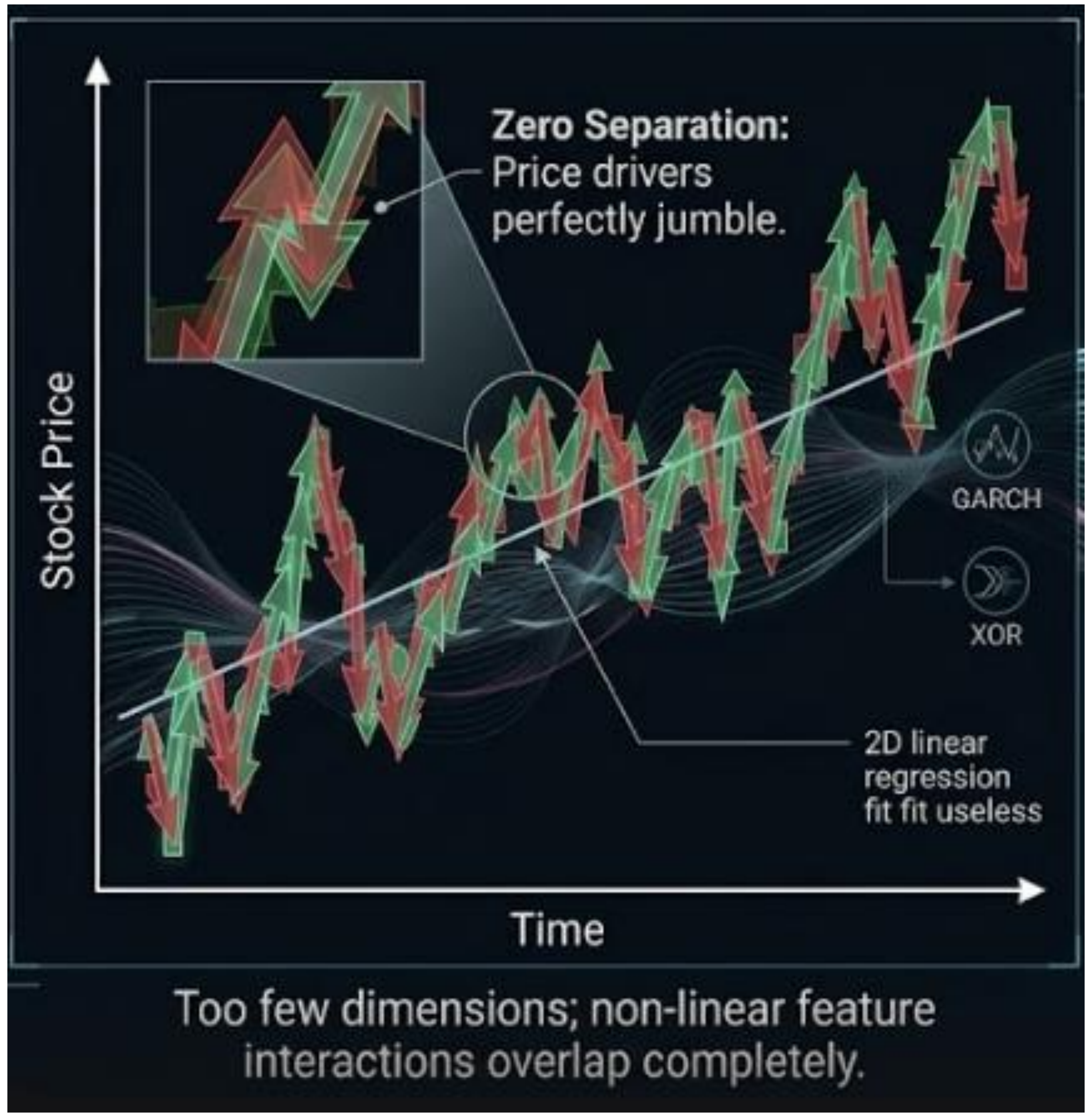
Hilbert Spaces: The Infinite Playground

- Hilbert space (H): Infinite dimensions. Points can be functions or curves
- A single "point" can be a function, signal, or time series
- Examples in our model:
- Limit order book snapshots
- VPIN sequences
- Tick-level order flow
- Powerful structure for complex data

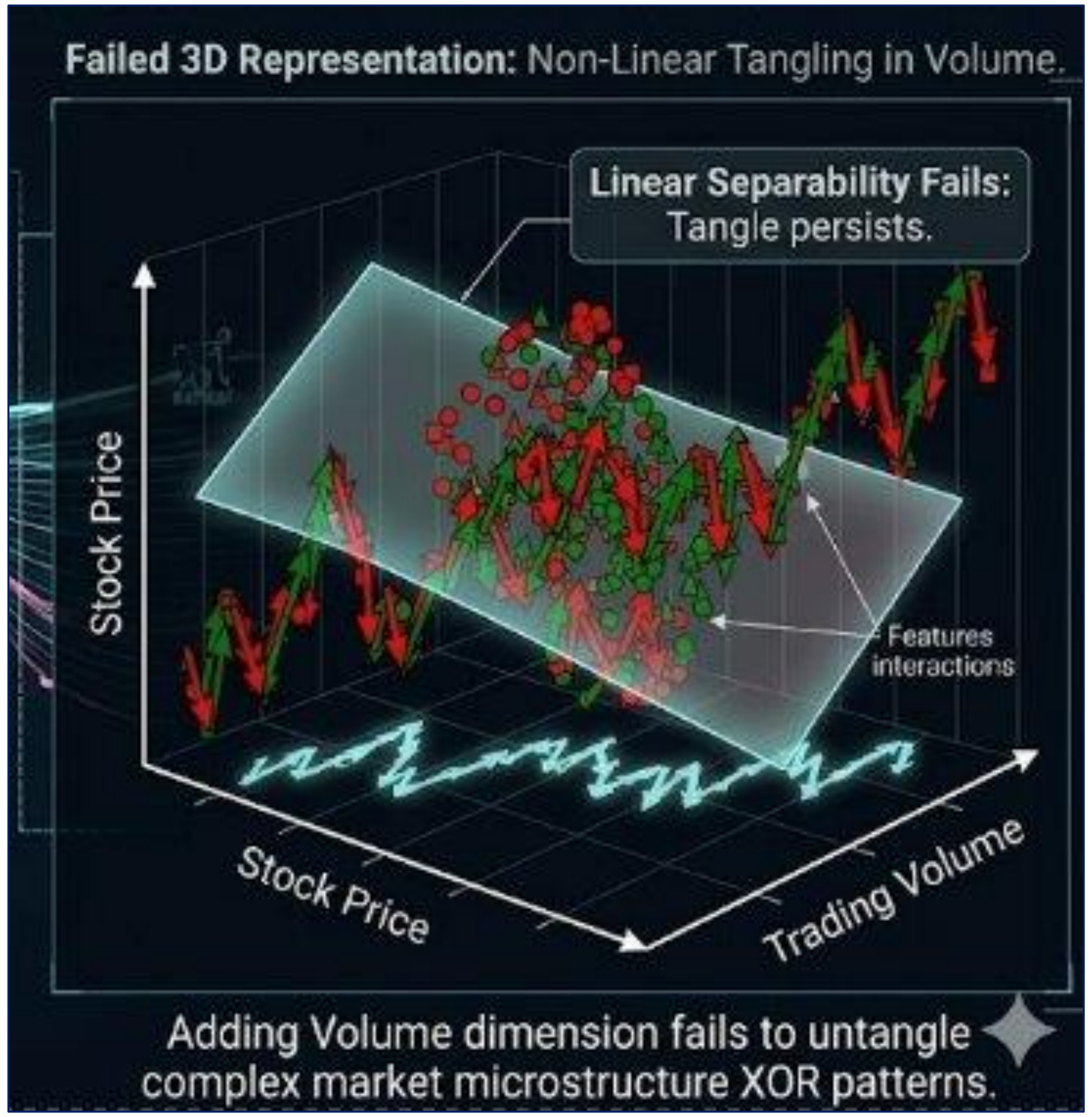


Setbacks of a Low Dimensional Model

Two Dimensional Setbacks

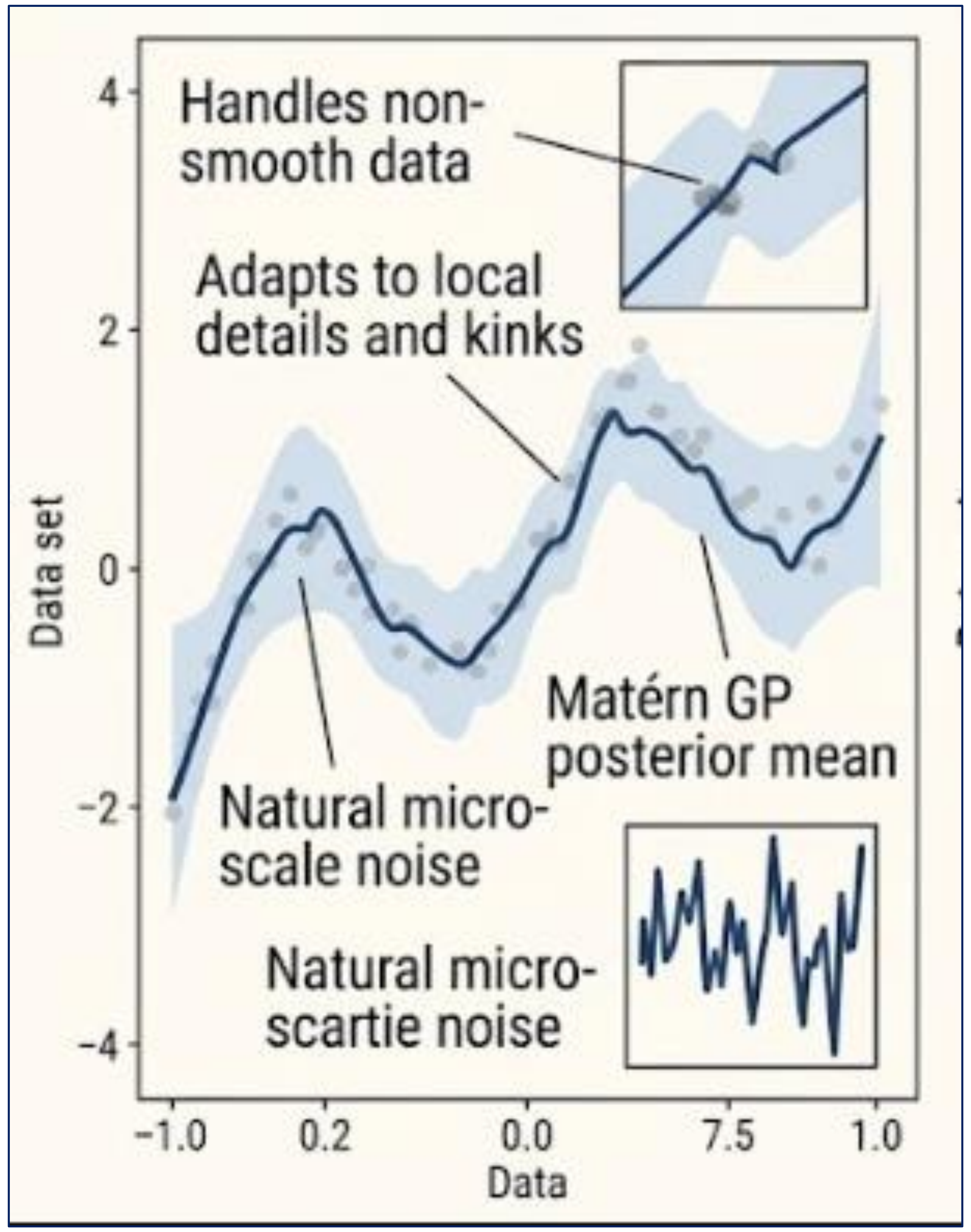


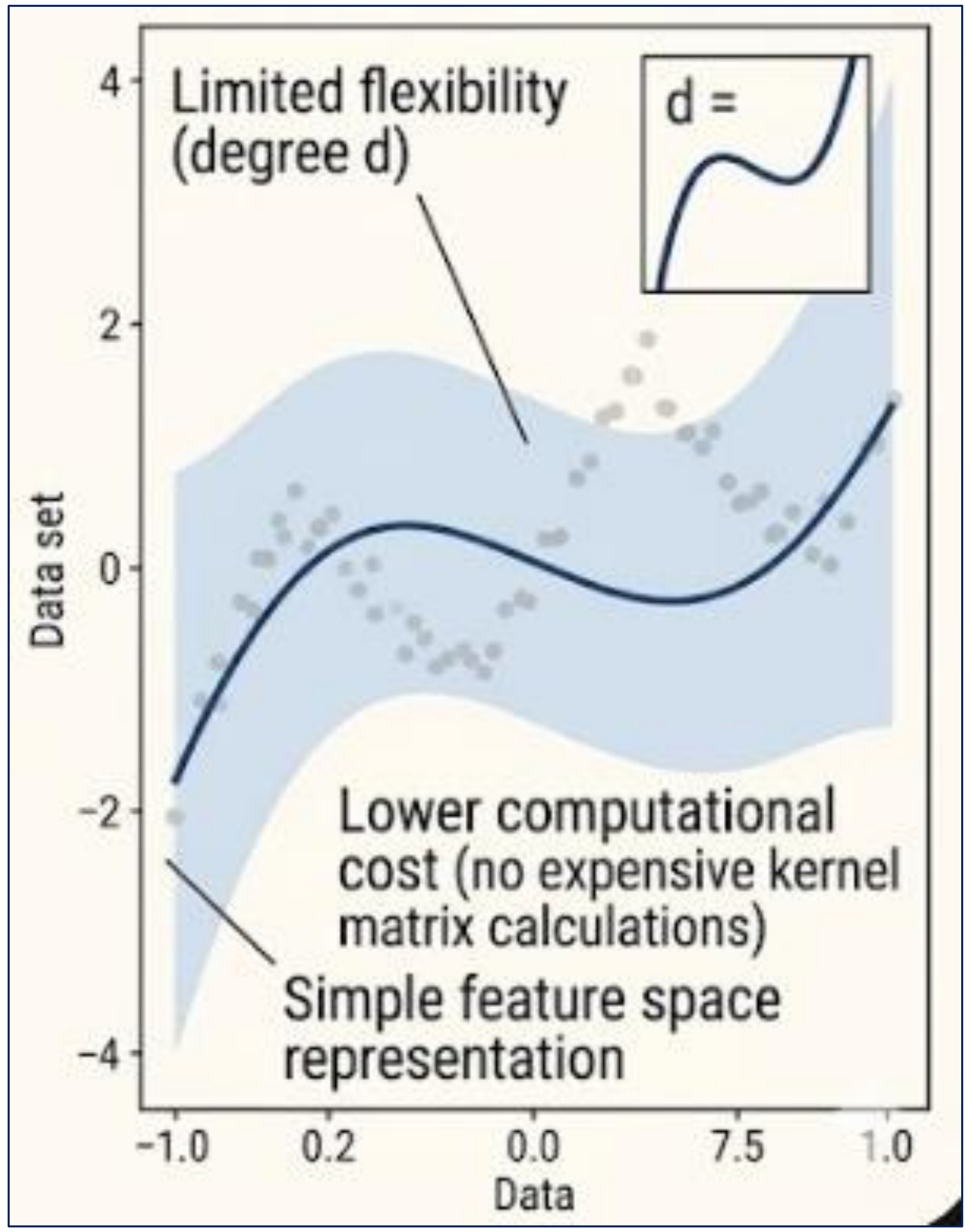
Three Dimensional Setbacks

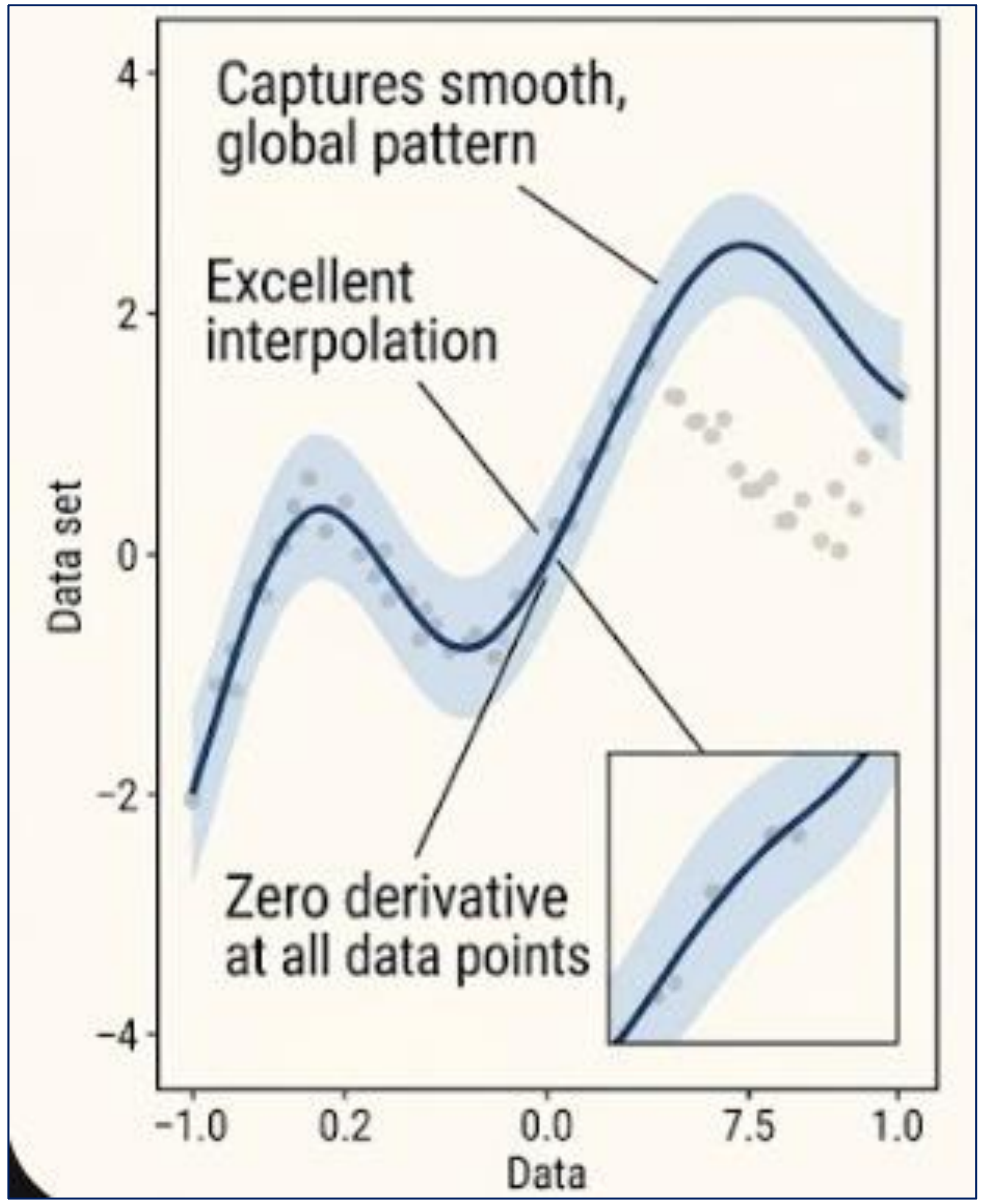


Kernels: The Similarity Engine

Matérn Kernel	Polynomial Kernel	RBF/Gaussian Kernel

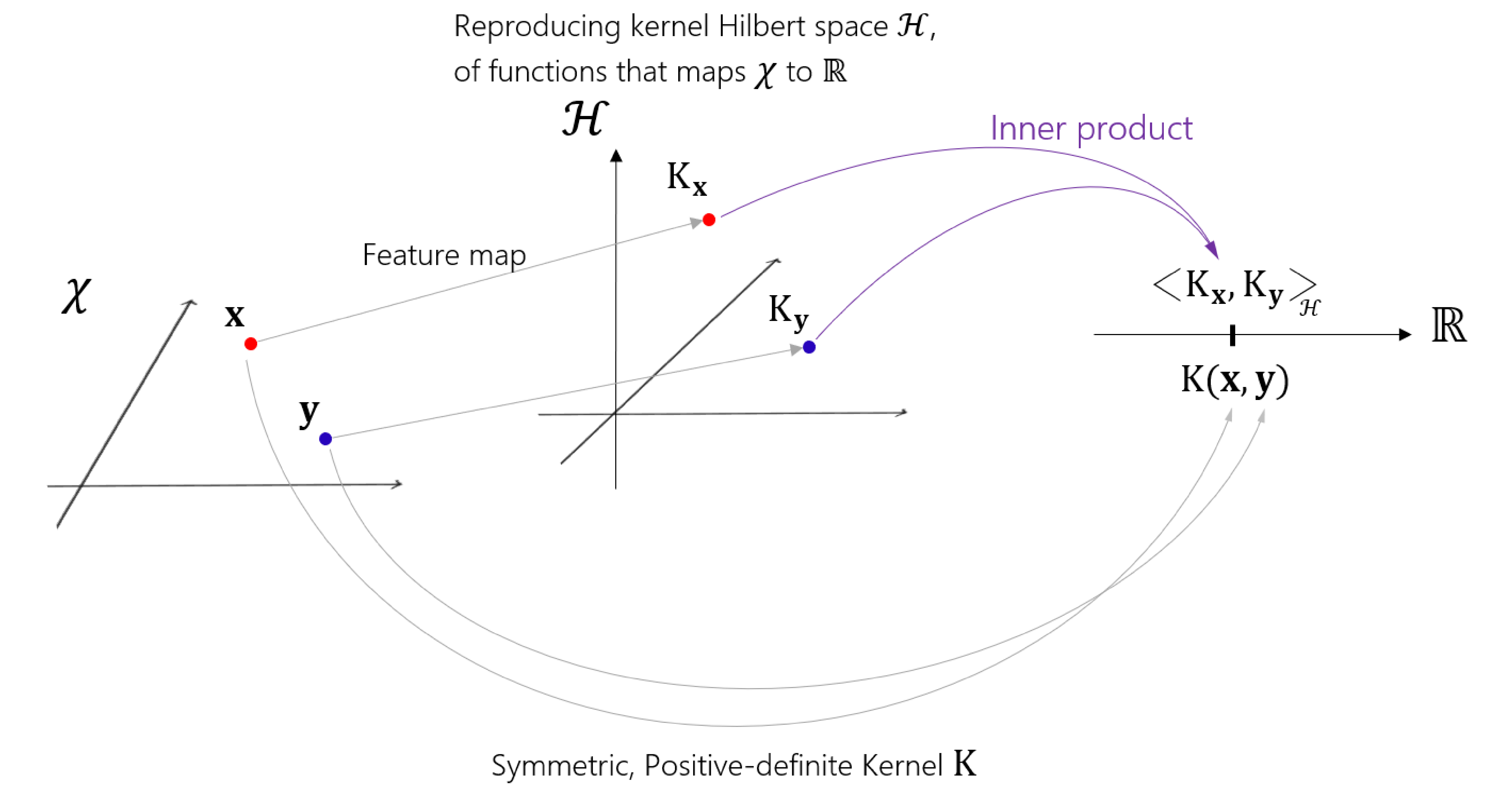






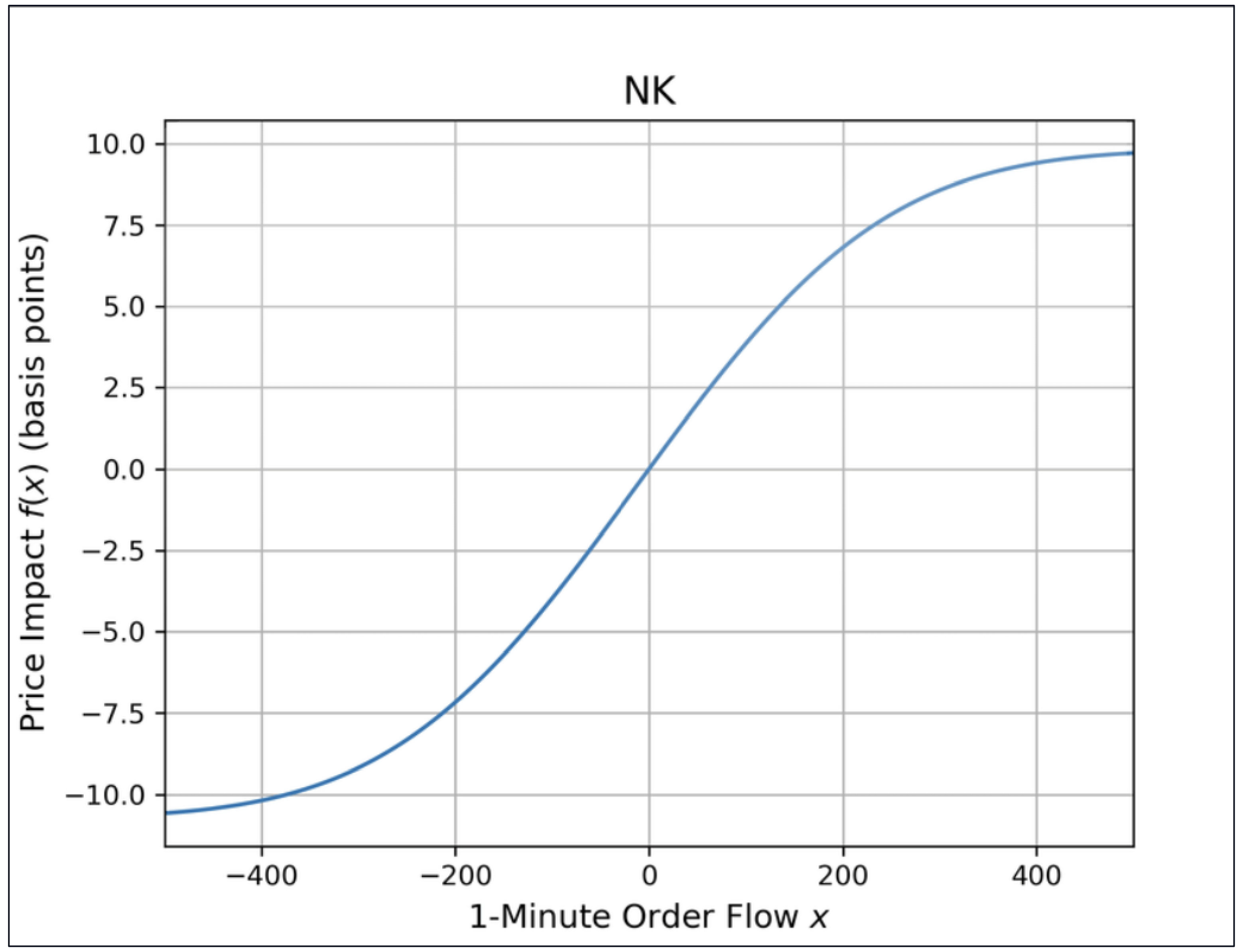
Reproducing Kernel Hilbert Space (RKHS) - Making the Infinite Computable

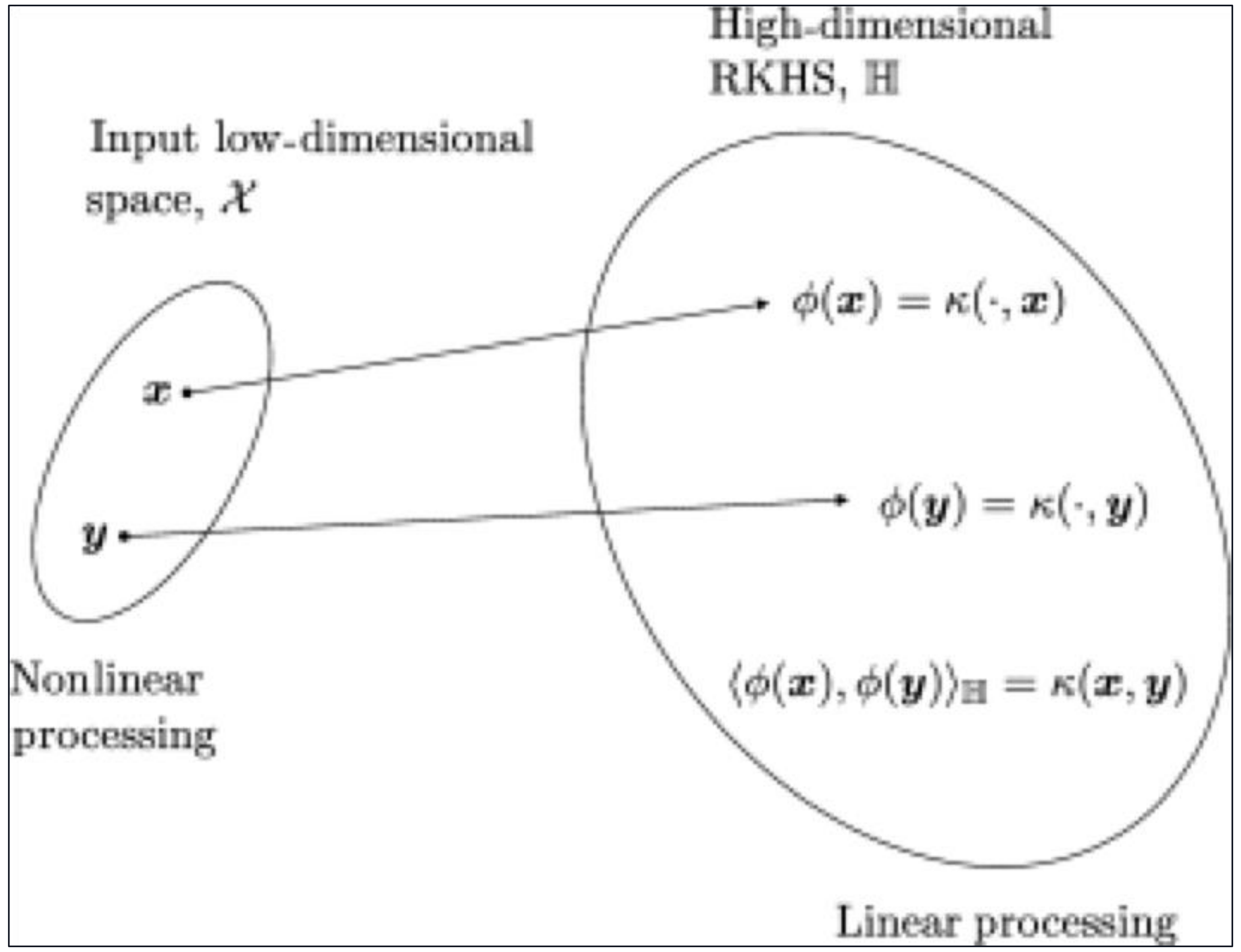
- A special type of Hilbert space with a reproducing kernel
- RKHS maps points that are not linearly separable in low dimensions to a high-dimensional space that could reveal hidden relationships
- Kernel Trick: Don’t have to use high degree polynomials
- Avoids high computation costs and risk of overfitting



Nonlinear signals in RKHS

- Nonlinear signals in finance are everywhere:
- Example: Price impact vs. order flow (curved, not straight)
- Find nonlinear signals ➡ lift them into Hilbert space ➡ make them linear.
- Compute the "infinite" via kernel trick





Layer Overview

Mapping the Market's Clocks – Fast Kernels

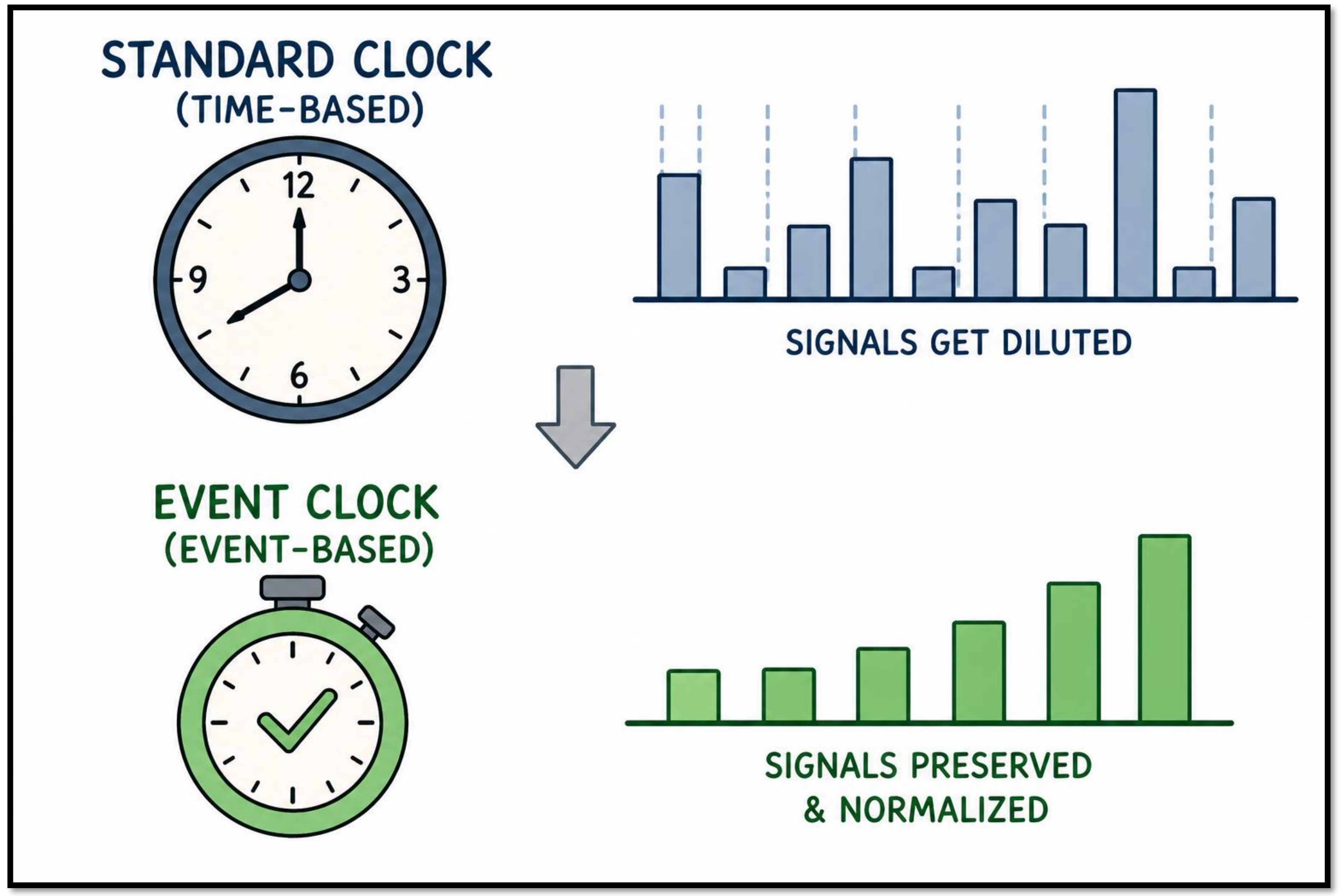
Moving from Abstract space to Market Reality

Event Time Over Calendar Time – Fast Kernal

- Financial markets don't run on a standard clock but on **events**
- Standard clocks dilute high-frequency signals

Example:

- **Dollar Bars:** Sampling data only once a fixed dollar volume transacts
- **Result:** Normalizes the arrival date of information and stabilizes calculations



Mapping the Market's Clocks – Fast/Slow Interaction

Multi-scale Data Integration

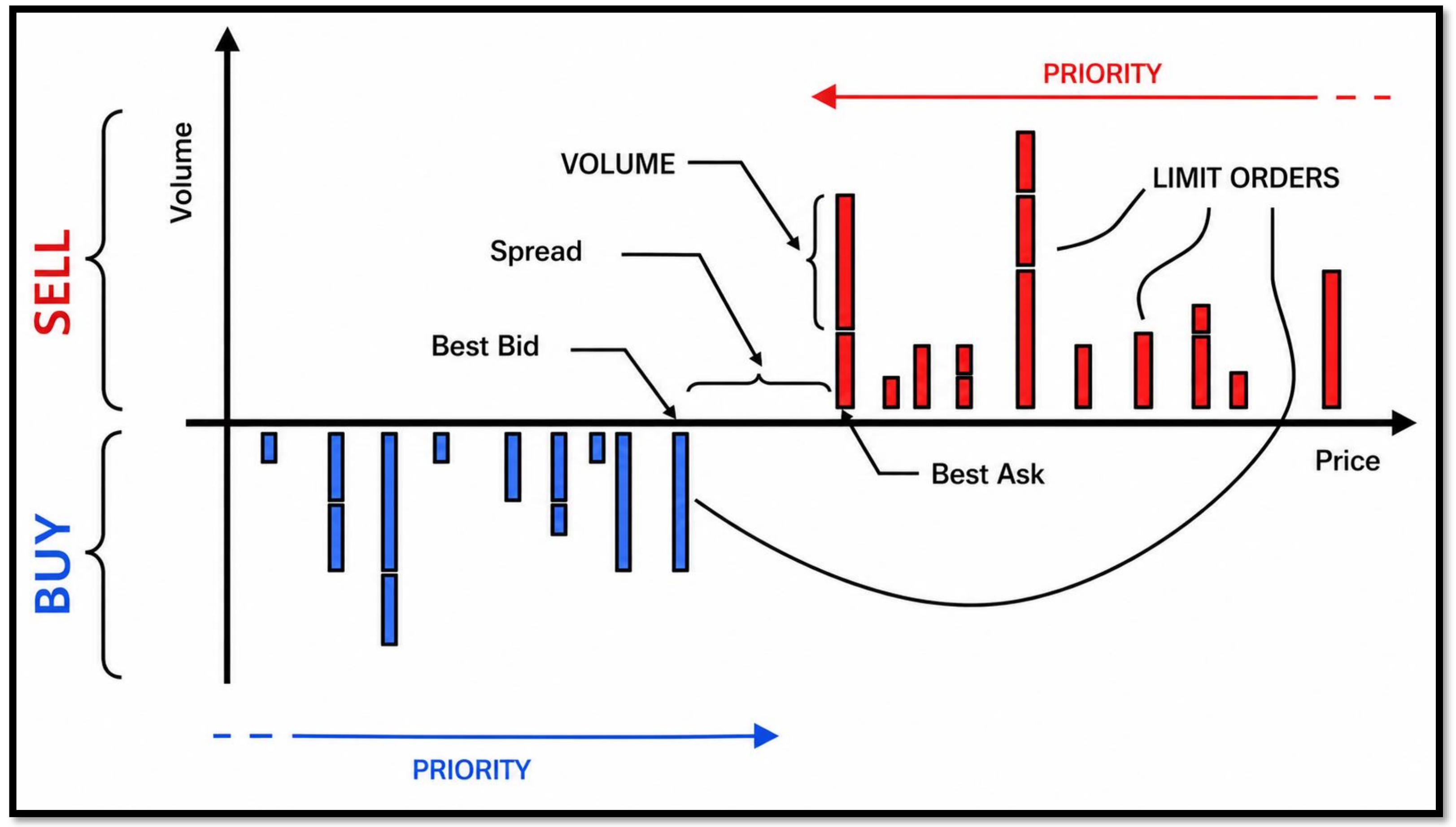
- **Mechanics:**
  - **Fast Kernels:** Microstructure and order flow (Event-time)
  - **Slow Kernels:** Macro and volatility regimes (Calendar-time)
  - **Fast/Slow Interaction:** Tensor Products force RKHS to capture non-linear relationships

$$K_{\text{total}} = K_{\text{fast}} + \beta \cdot (K_{\text{fast}} \otimes K_{\text{slow}})$$

Kernel 1 – Limit Order Book (LOB)

- **Static snapshot of resting liquidity**
- **Measures liquidity depth, market resilience, and execution cost**
- **Samples in event time (dollar bars)**
- **Frequency:** High-frequency (milliseconds to minutes)
- **Key Intuition:**

  Thick Book **-->** stable prices
  Thin Book **-->** fragile market & high price impact

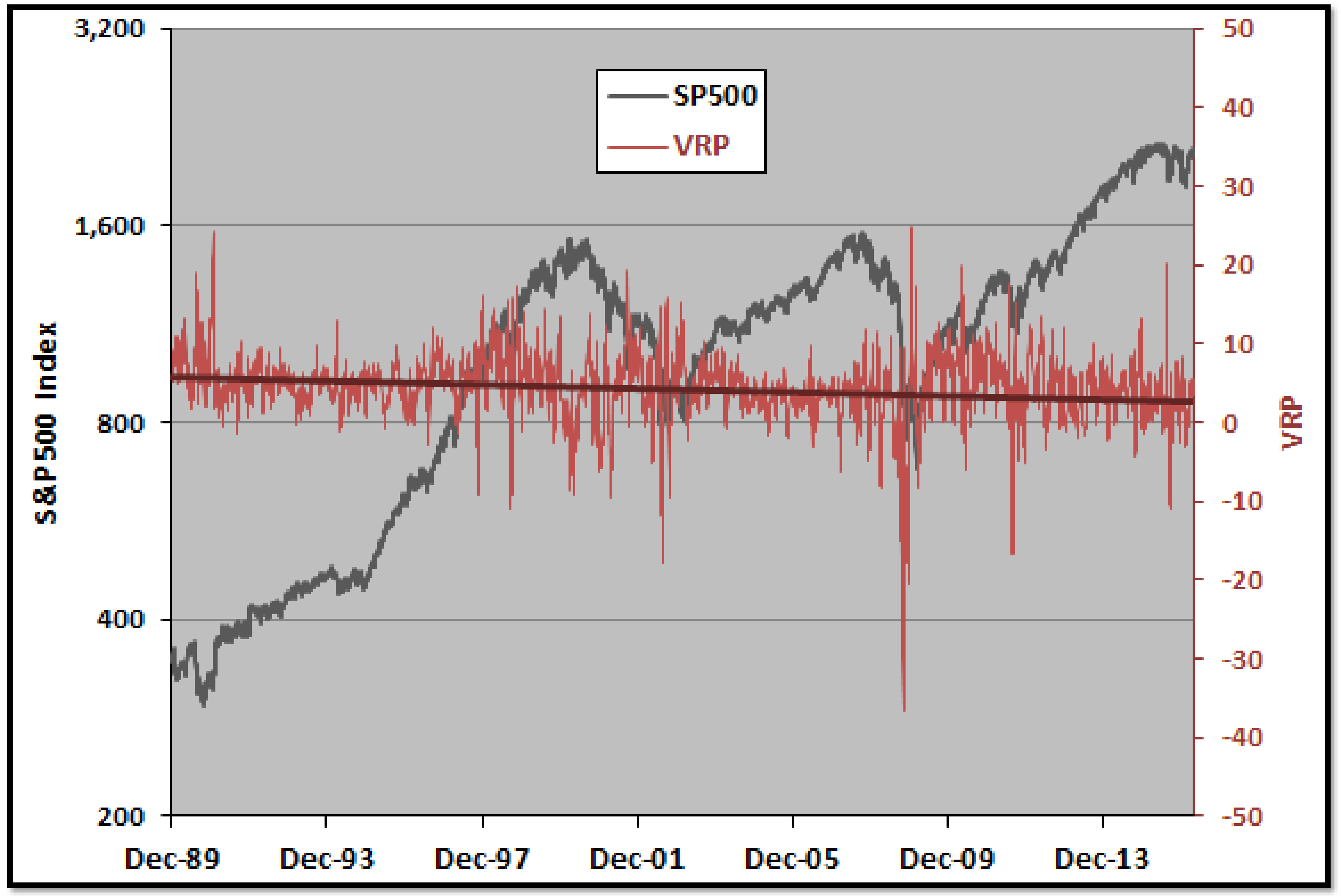


LOB directly captures supply–demand pressure.

Kernel 2 – Variance Risk Premium (VRP)

- **Bridges the gap between option market and the underlying spot market**
- **Samples in calendar time (daily)**
- **Frequency:** Medium-frequency (days to weeks)
- **Key Intuition:**

  Low VRP --> indicates market complacency
  High VRP --> predicts higher future returns



VRP acts as an early-warning radar for systemic risk appetite.

Kernel 3 – Macro Motion

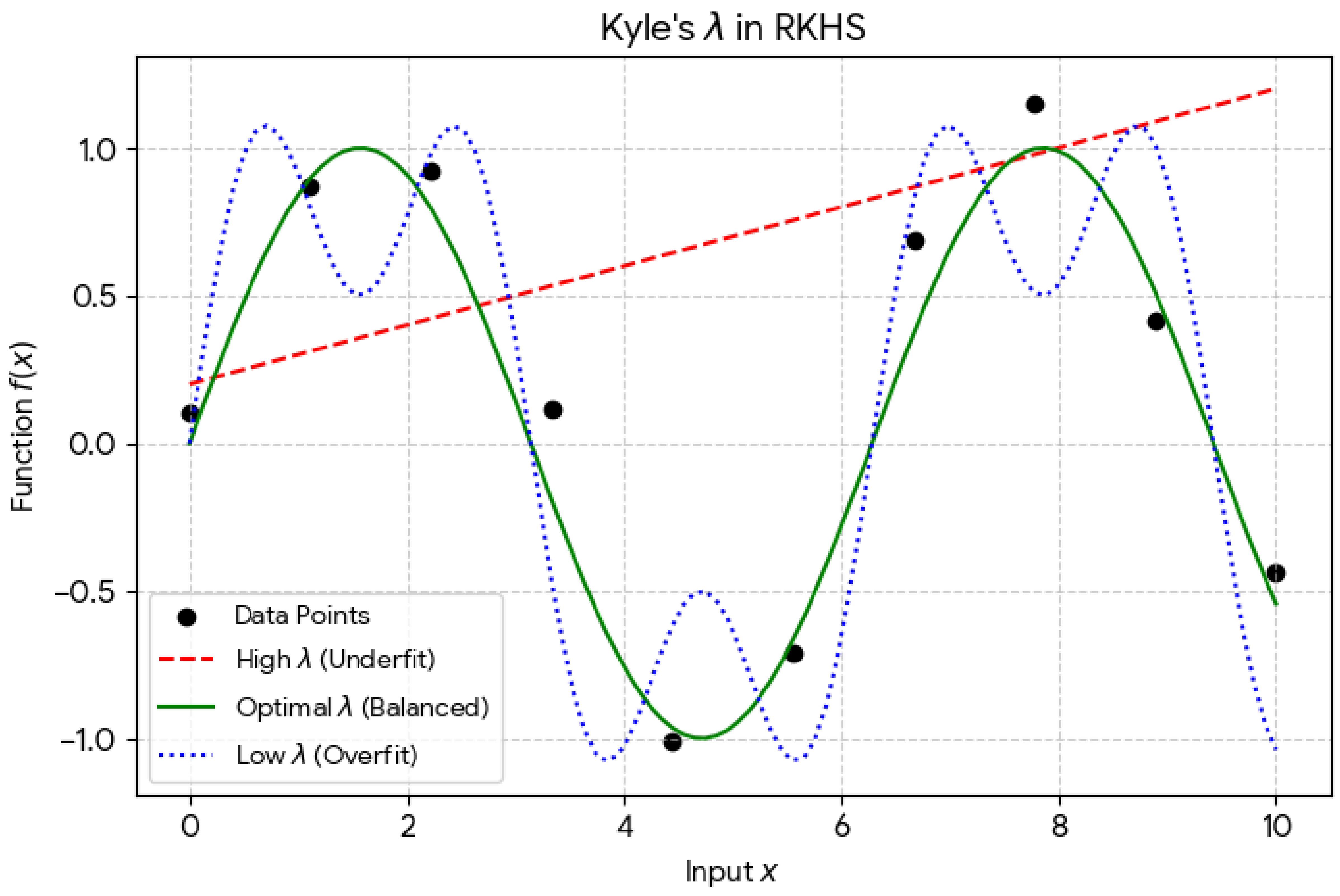
- **Factors in growth, inflation, policy, and credit conditions**
- **Samples in calendar time**
- **Frequency:** Low-frequency (months to years)
- **Key Intuition:**

High Smoothness **-->** implies a stable long-term baseline
Regime Filter **-->** modulates high-frequency signals to prevent mismatch errors

Macro Motion provides the baseline economic reality in our model.

Kyle’s Lambda

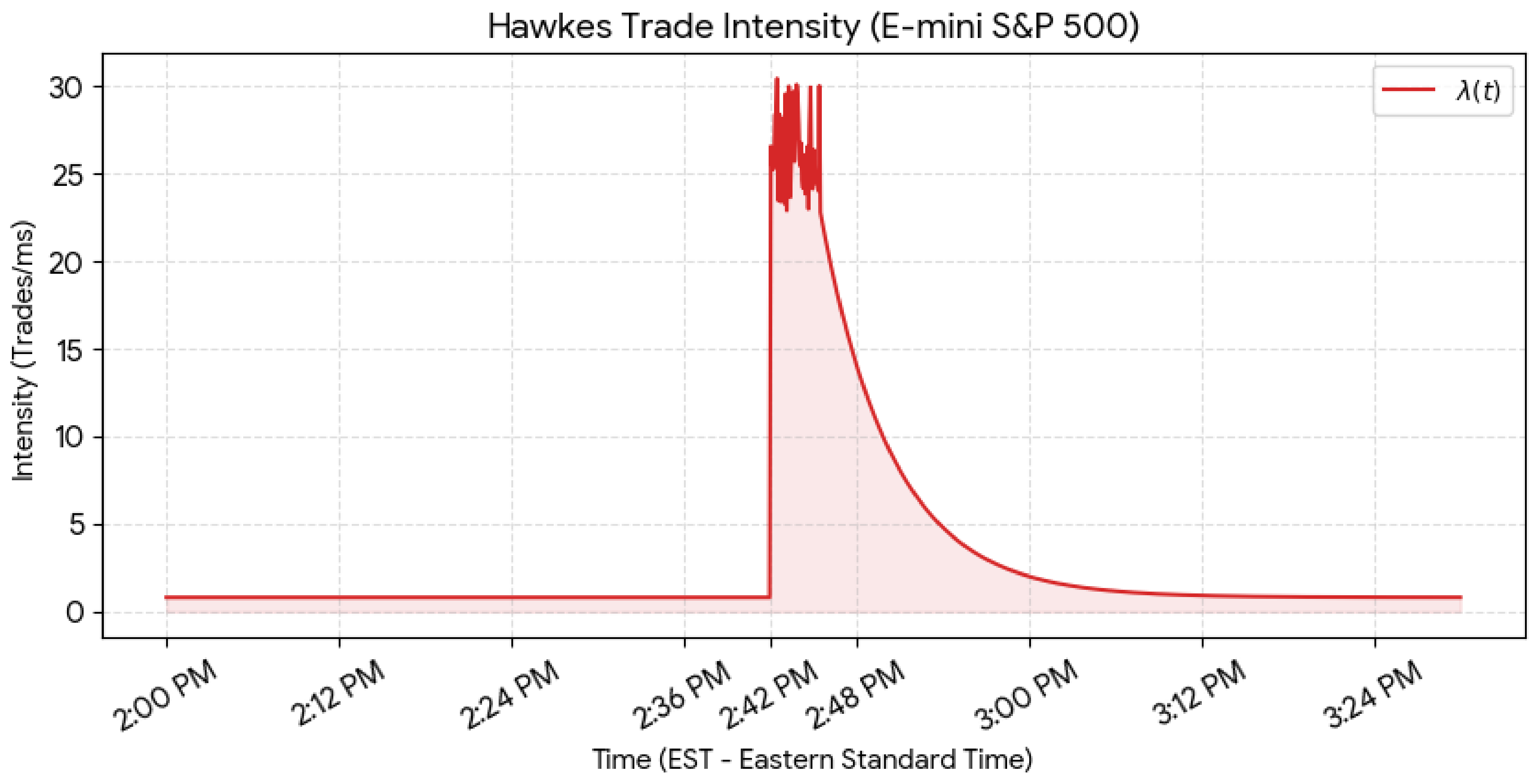
- In an RKHS model, Kyle’s Lambda operates as a way to measure the market illiquidity which we then can utilize to regularize the RKHS model.
- Simply put, Kyle’s Lambda helps show how trades move the market, avoiding the movements that can be distilled to “noise”.

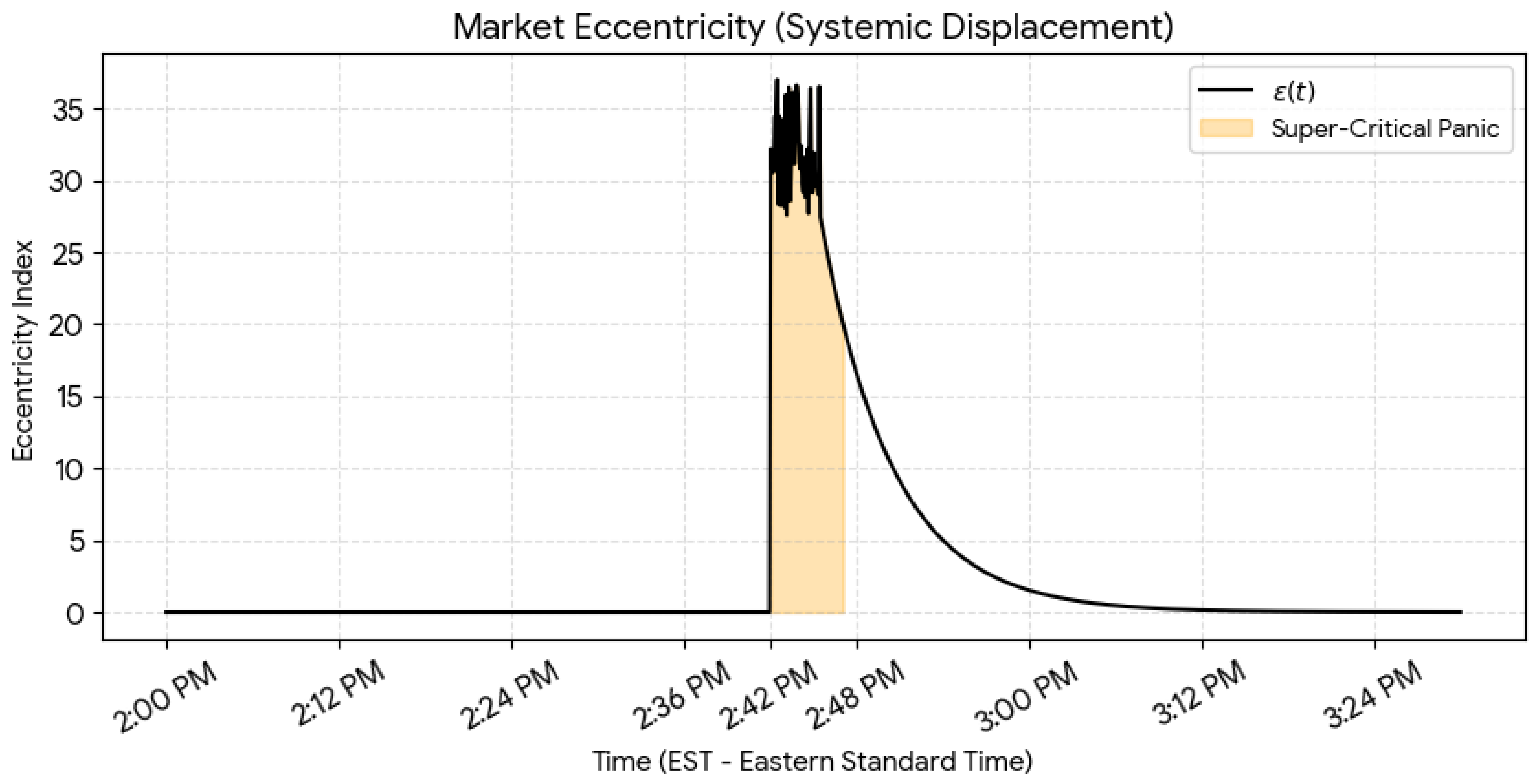


Hawkes Process Intensity

- The Hawkes Process model looks at how a trade can lead to a burst of follow-on trades, enacting a “panic” or “boom” after a trade, and an increase in the eccentricity of the market.

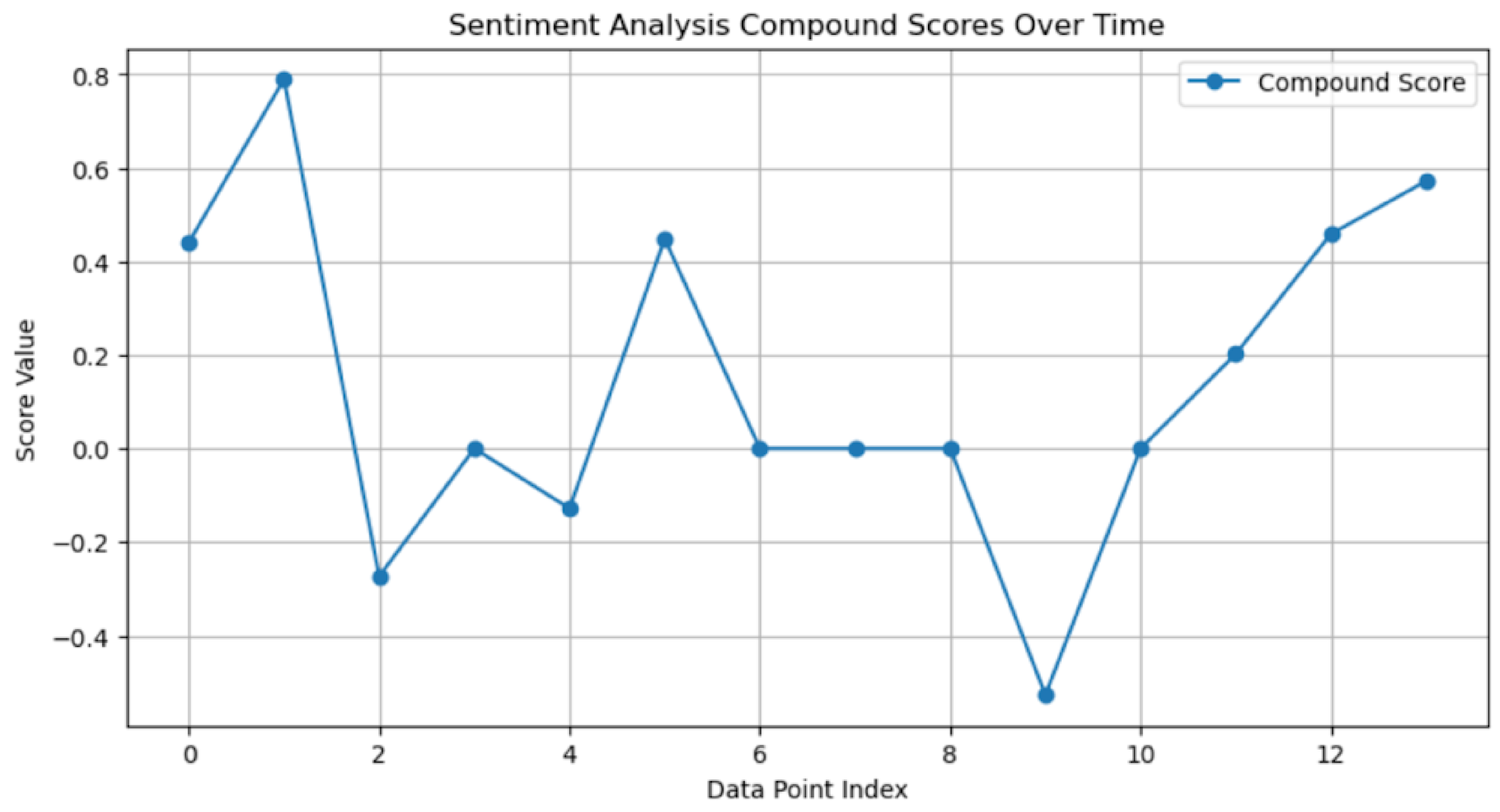
$$\lambda_i(t) = \mu_i(t) + \sum_{j=1}^{D} \sum_{t_k < t} g_{ij}(t - t_k)$$

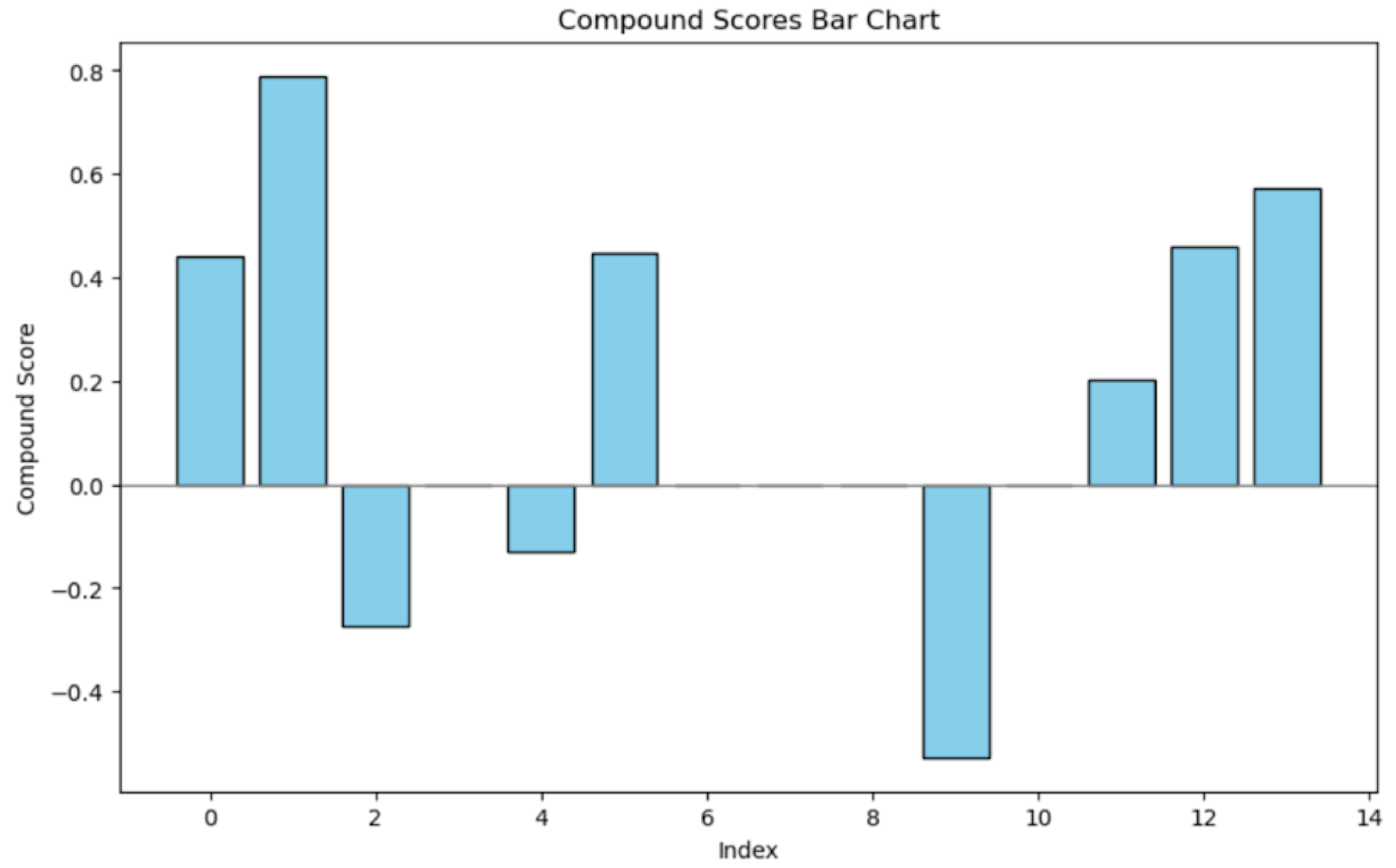




Sentiment Analysis

- Sentiment Analysis consists of utilizing text to make a quantifiable assessment of a market or stock.
- Our model consists of taking news headlines, utilizing FinBERT, to give us an understanding of the reactions the market had to news in this sector.





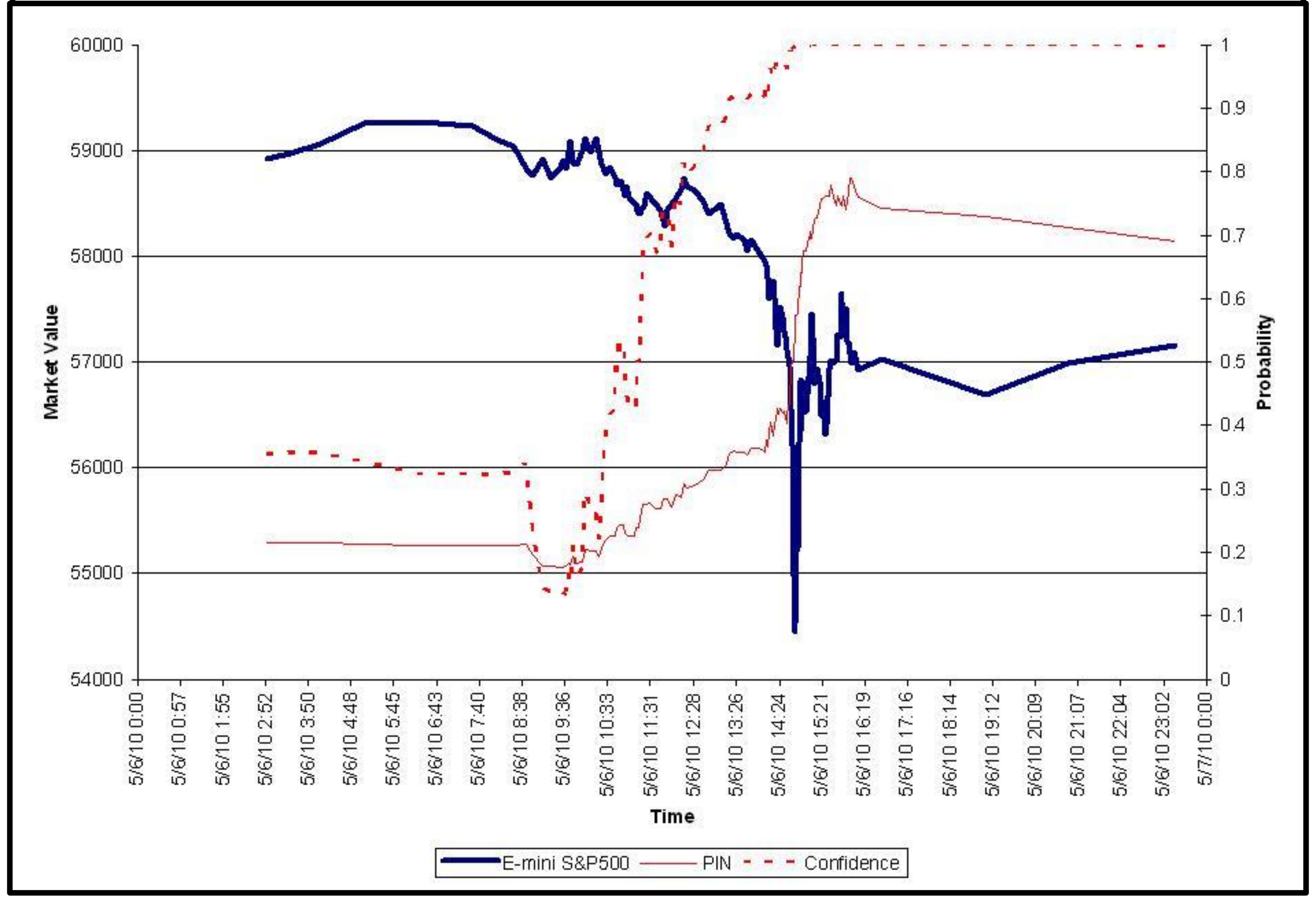
VPIN – Volume Synchronized Probability of Informed Trading

- **Helps measure order flow toxicity – describes the likelihood others are trading on unknown information**
- **Measured on dollar bars**

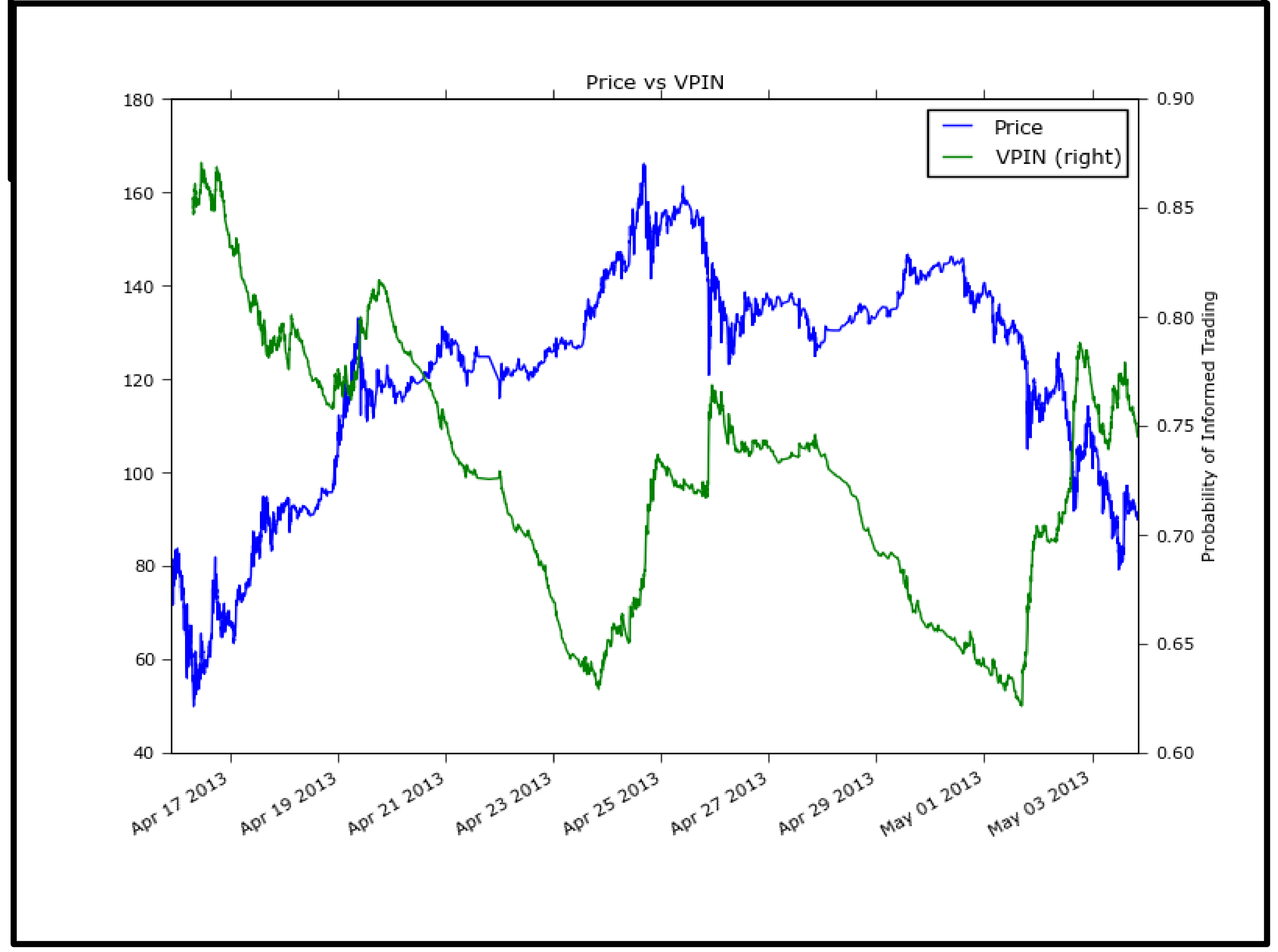
VPIN Equation

$$VPIN = \frac{\sum_{i=1}^{n} |V_{i,s} - V_{i,b}|}{\sum_{i=1}^{n} V_i}$$

2010 Crash Flash



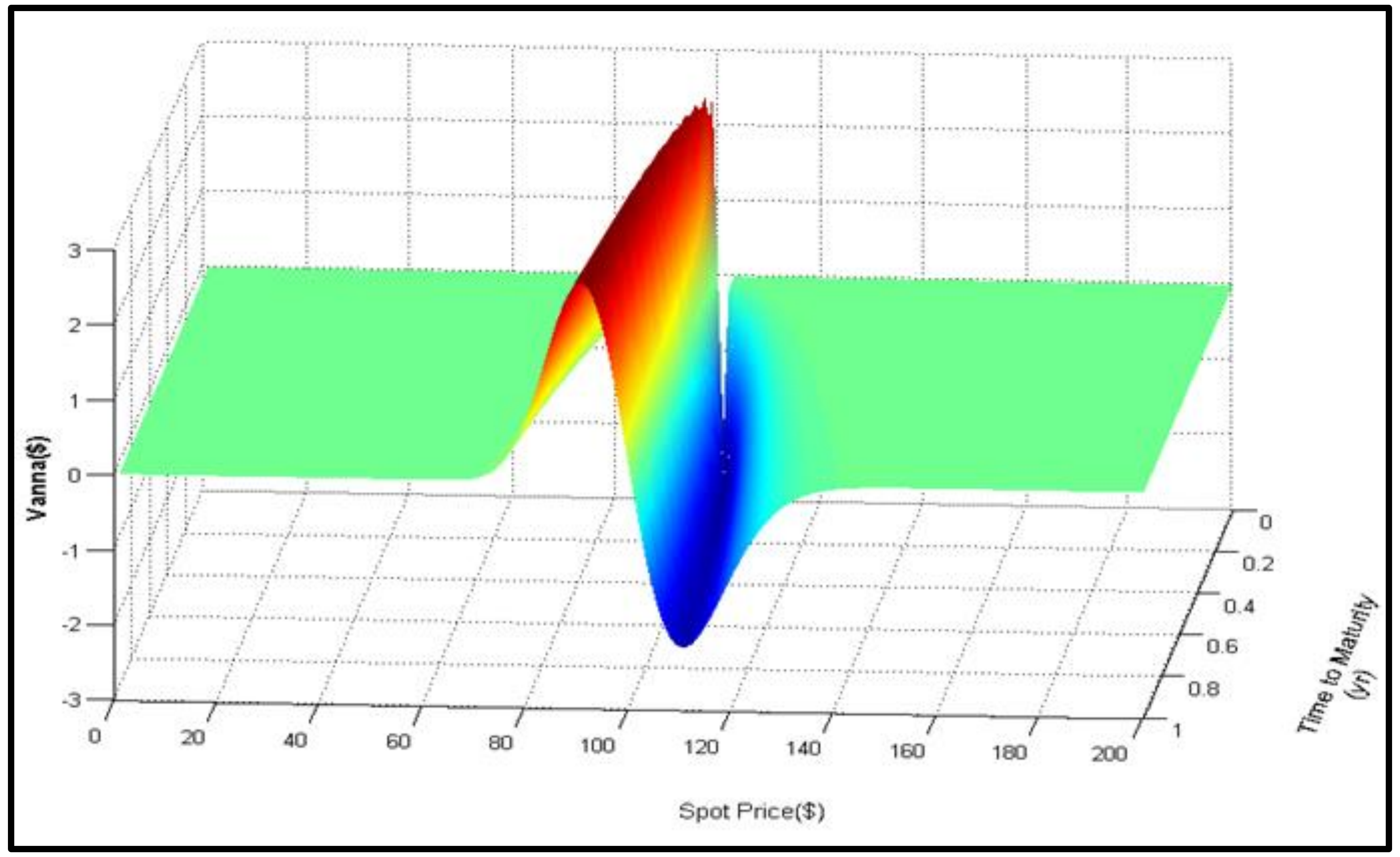
2013 Bitcoin Crash

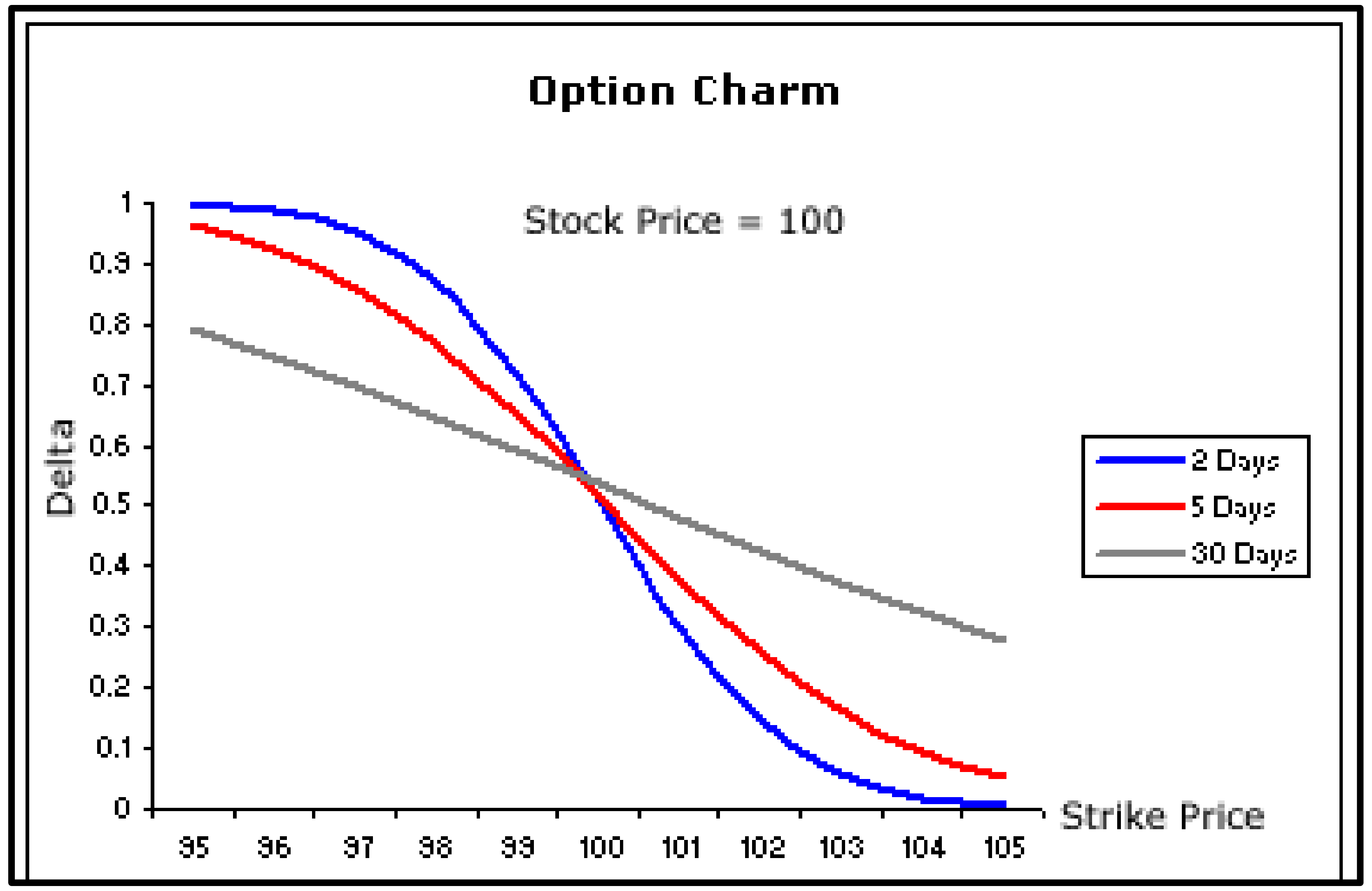


VannaCharm – Delta Derivatives

Vanna:
- Derivative of Delta with respect to Vega (Implied Vol)
- Combine to measure option prices response to various market conditions
- Measured in Dollar Bars

Charm:
- Derivative of Delta with respect to time
- Combine to measure option prices response to various market conditions
- Measured in Dollar Bars

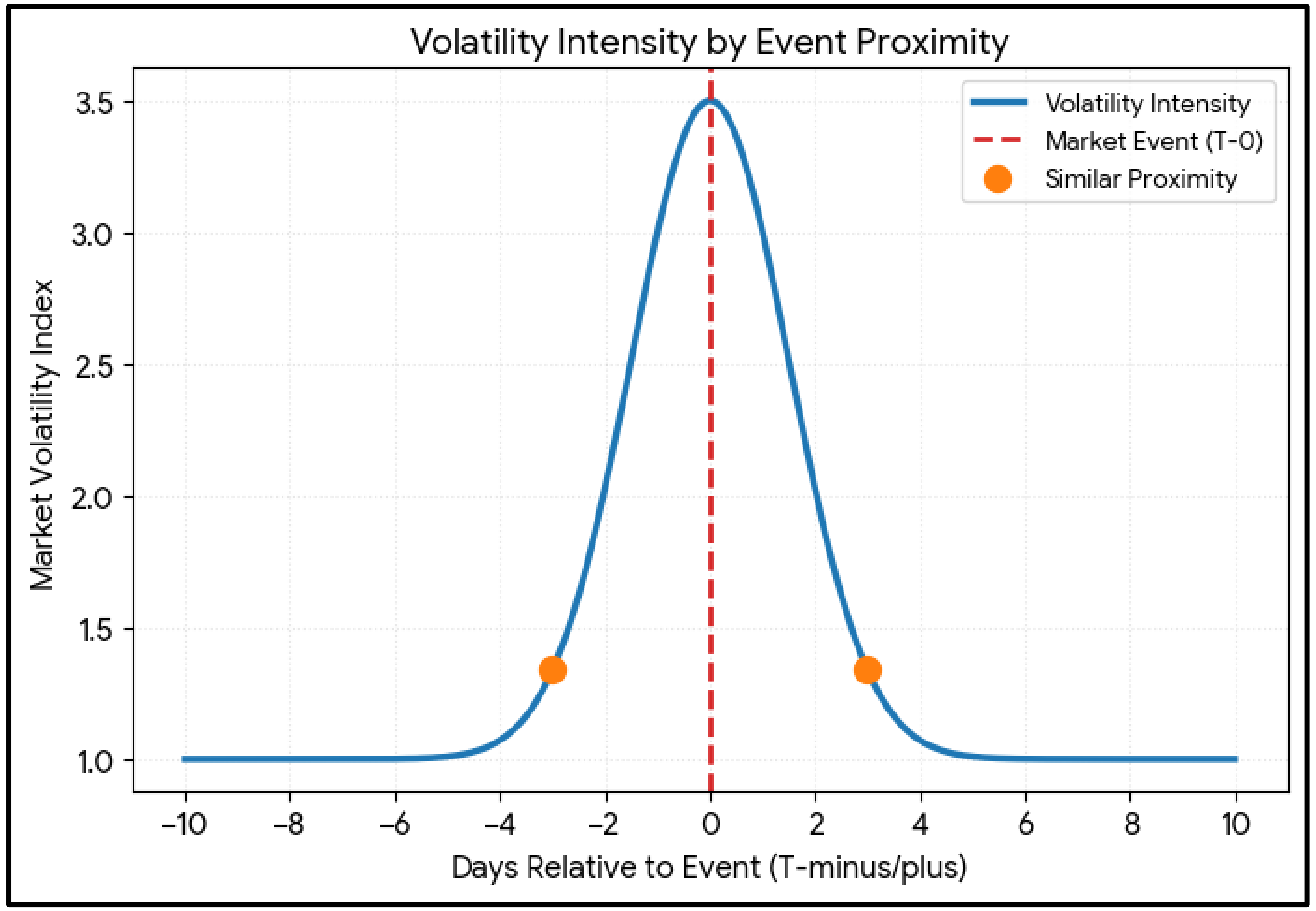


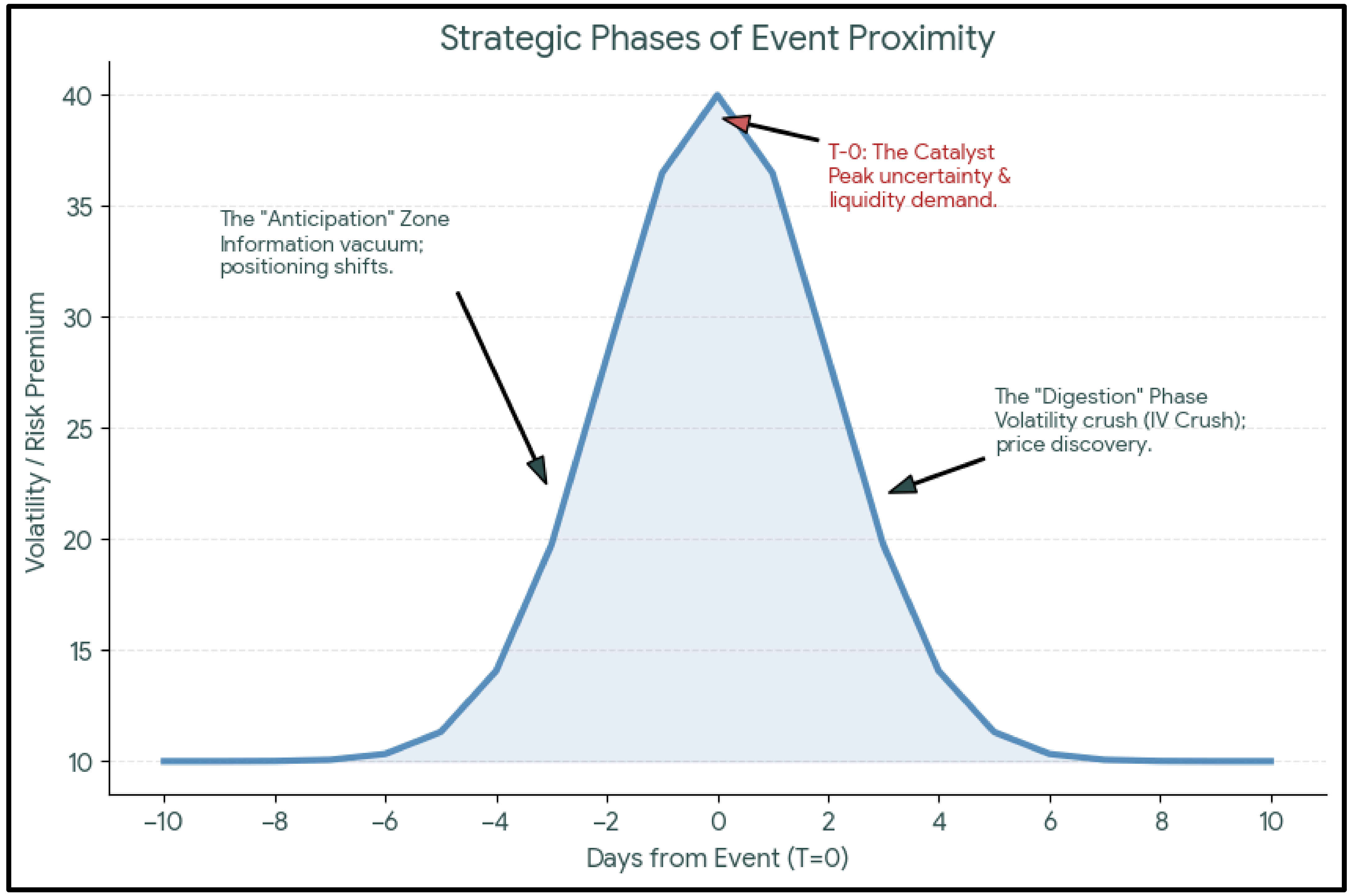


Event Proximity

- Utilizes exponential proximity decay
- Calendar utilizes Sin/Cosine Encoding

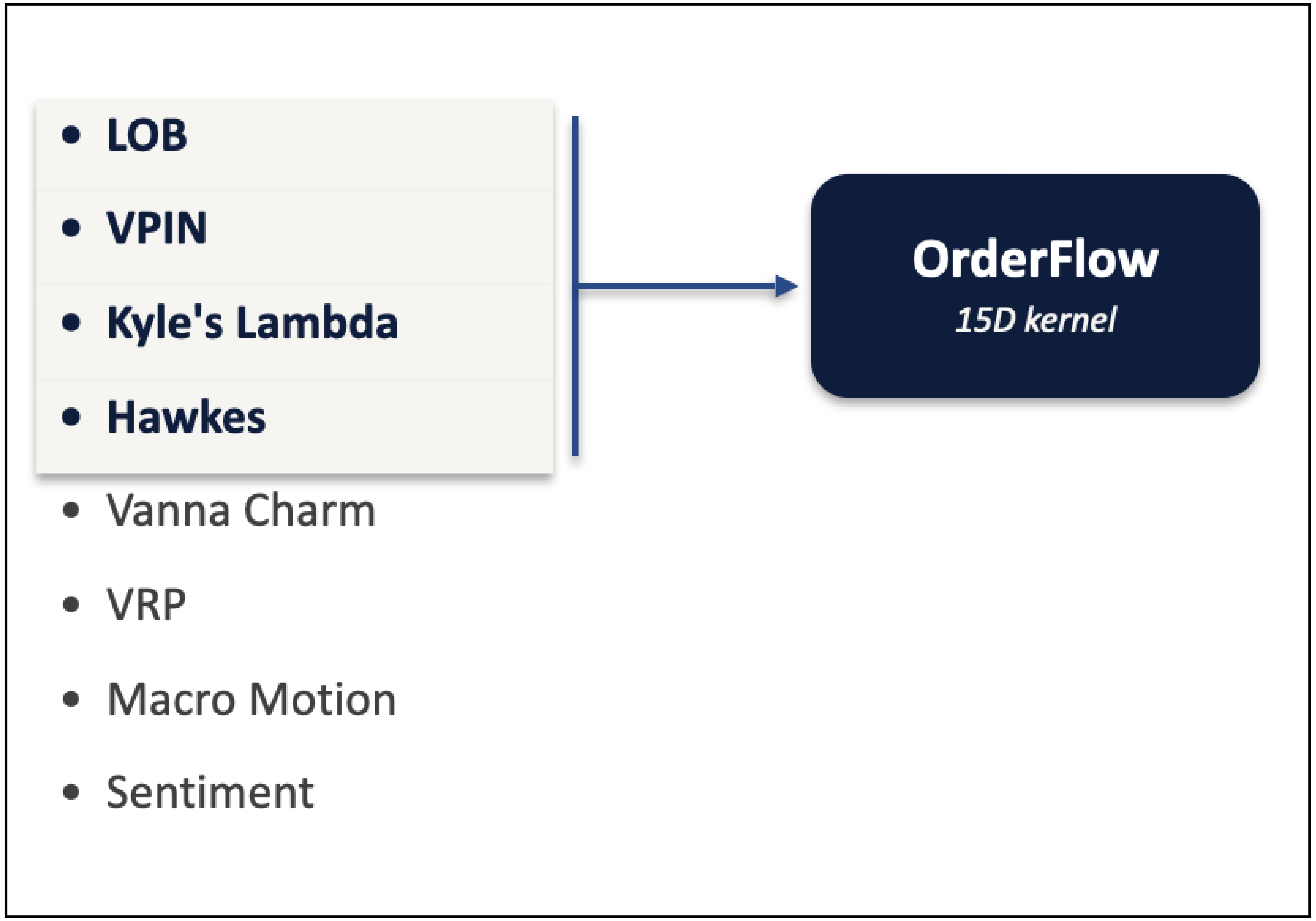
- Measures proximity to notable event dates – Earnings, FOMC, FPI, PPI, NFP, Options Expiry, Fed Speech, GDP





Back Testing

Kernel Redundancy



THE PROBLEM

Four order-flow kernels had 97% variance overlap (PCA)

THE FIX

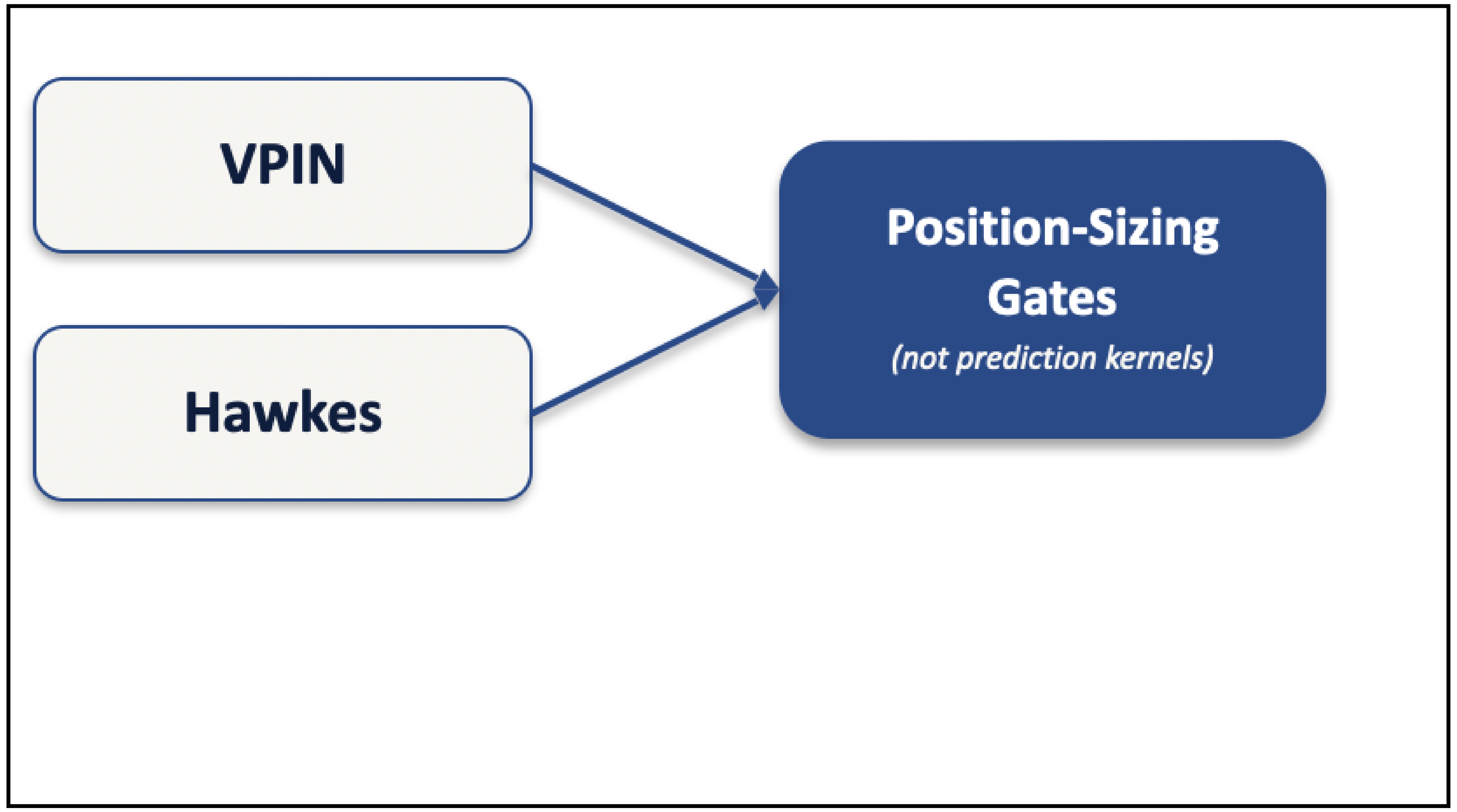
Collapsed **61D → 15D** via PCA
Four redundant kernels → **one OrderFlow kernel**

KEY FINDING

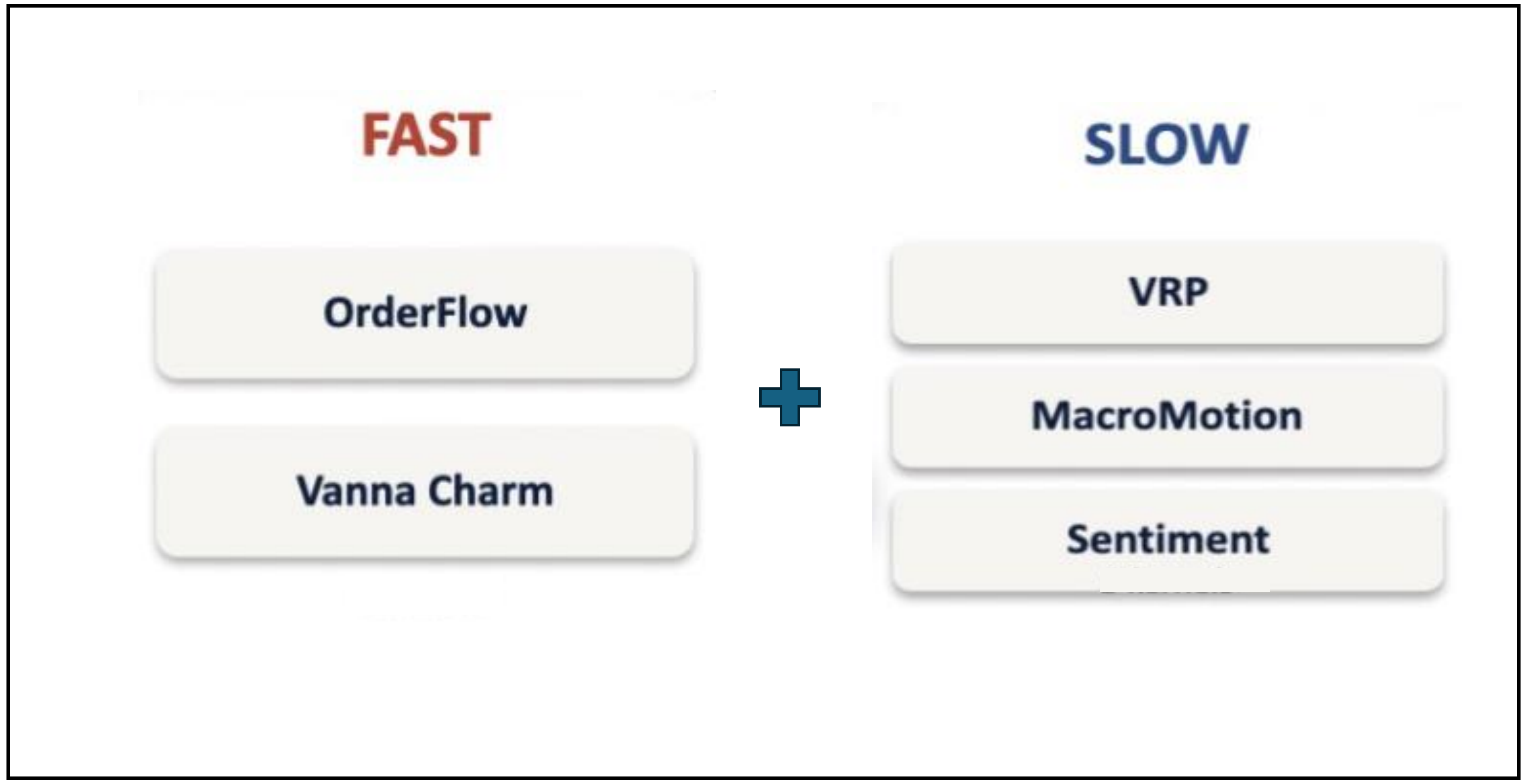
Five clean kernels beat eight noisy ones — fewer hyperparameters, cleaner signal separation.

Final Architecture

Repurposed



Final Architecture



–0.016 to +0.023

OOS correlations between kernels

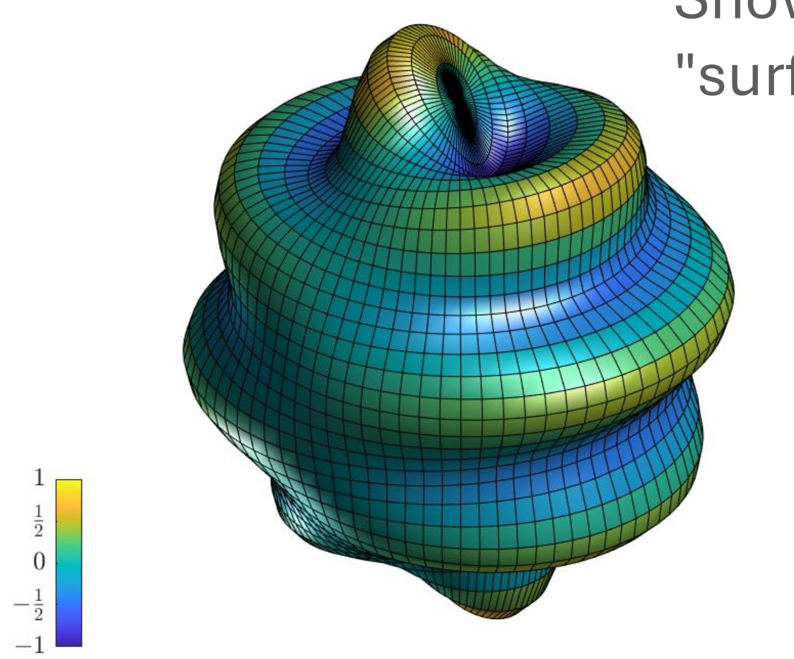
Near-zero correlation across kernels → the architecture is clean and the kernels are ***genuinely independent.***

Implementation: 3-stage pipeline

A Kernel Ridge Regression

Stage 1: Per-kernel prediction

- Each kernel → Matern Random Fourier Features (500D)
- Walk-forward trained; each fold → OOS predictions
- Key: individual kernel correlations near-zero (−0.016 to +0.023)



Shows us the "surface"

B ElasticNet Combiner

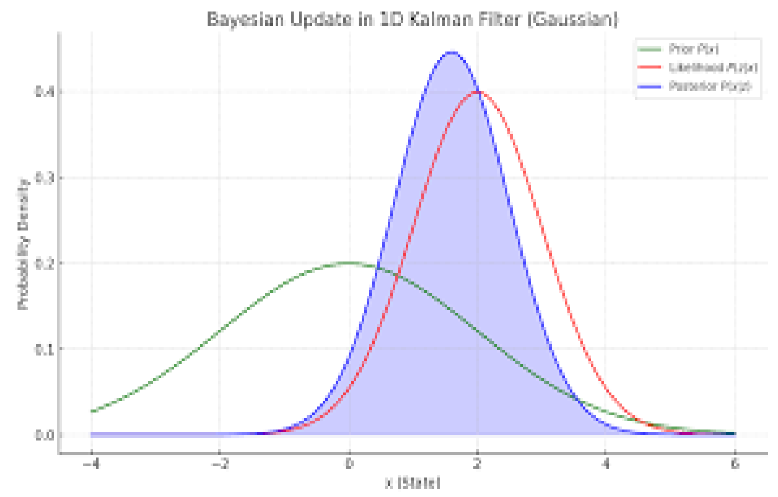
Stage 2: Combining kernel outputs

- Takes 6 kernel predictions as inputs
- Walk-forward ElasticNet (8 expanding folds, L1+L2, positive weights)
- Finding: all weights zeroed — profitability lives in the time-varying intercept
- Signal = intercept drift, not kernel weights

C Kalman Filter

Stage 3: Signal extraction

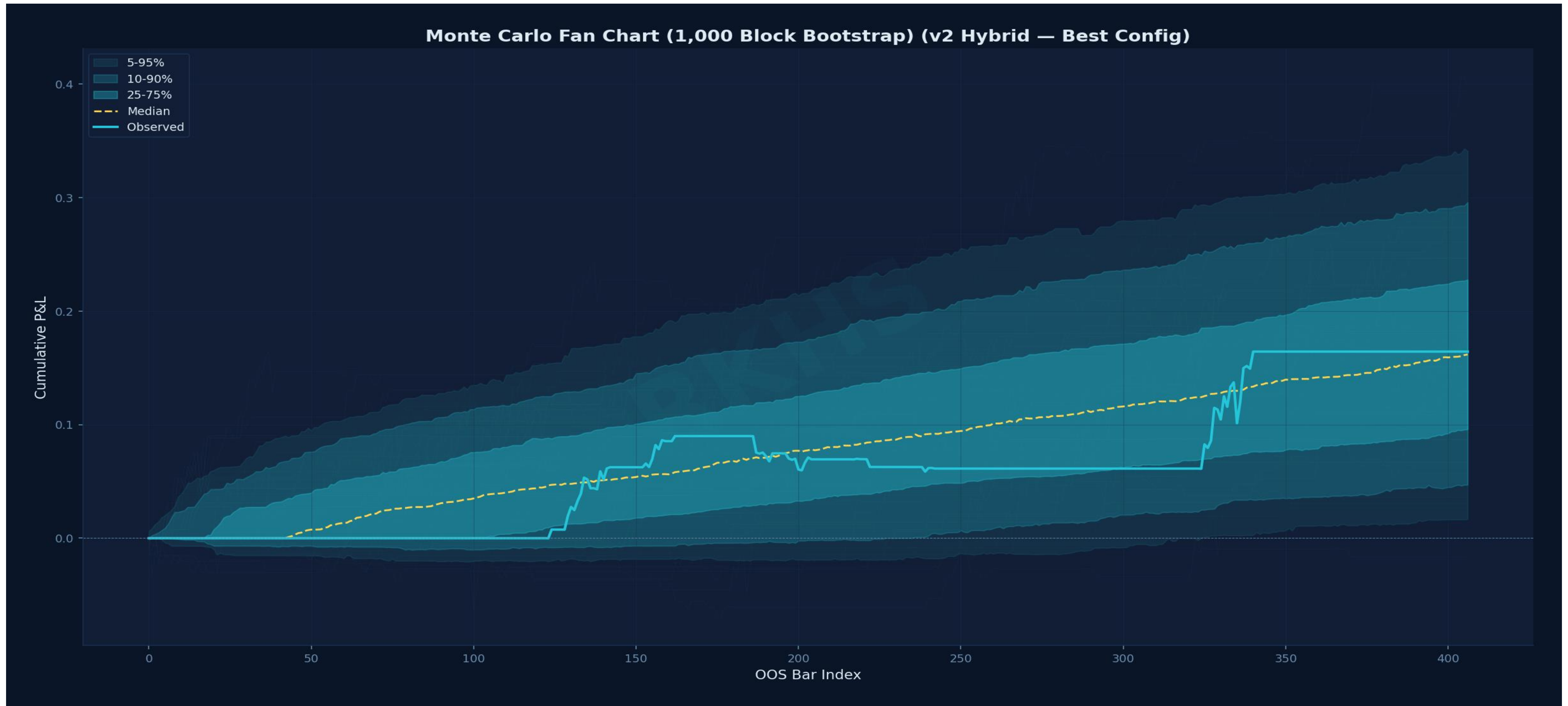
- State-space model: filters noisy ElasticNet output → clean signal
- Aggressive Q (10× auto-estimate) was the unlock — tracks regime shifts faster
- Outputs s-score (rolling z-score of filtered state)
- Regime-conditional thresholds: ±0.50 trending, ±1.50 choppy markets

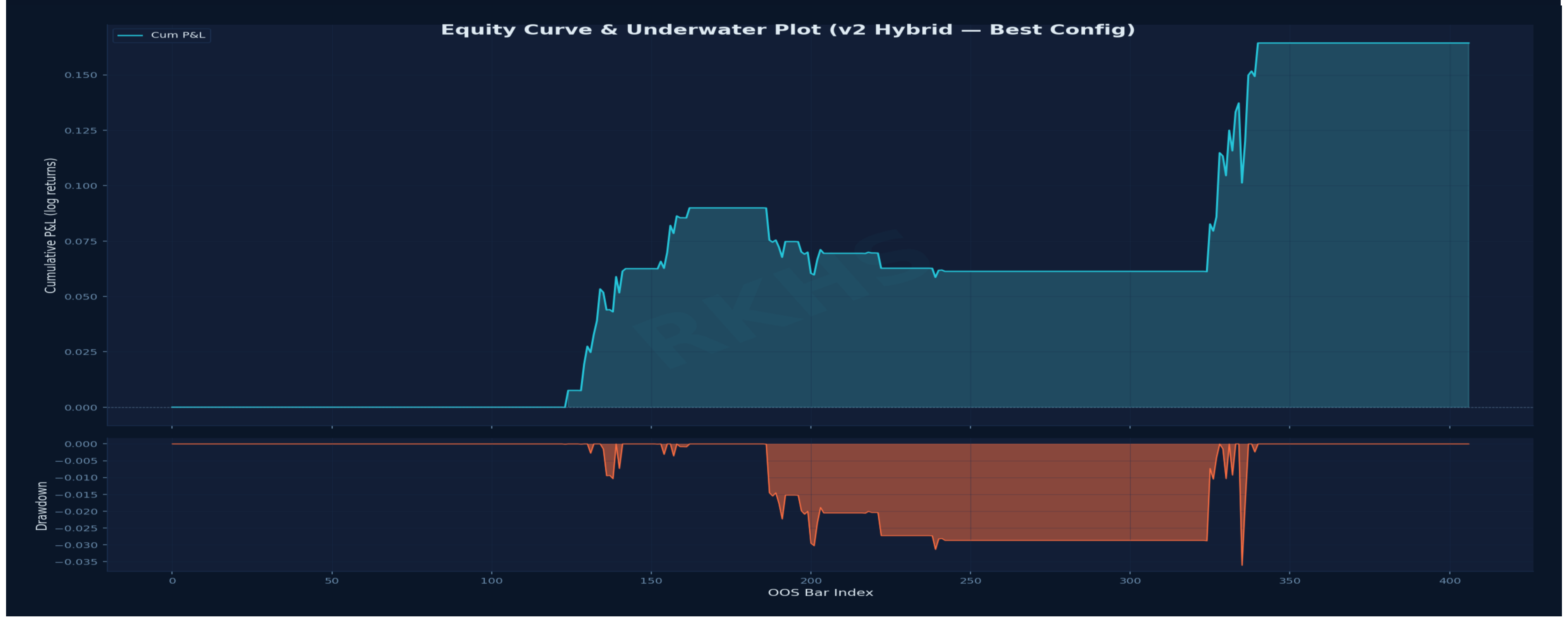


Finds the true state by filtering signals

Back Testing

First Results





RKHS Strategy Scorecard

CL Futures | Jul 2023 - Dec 2024 | Two-Stage + Kalman

Sharpe Ratio	1.5522
Sortino Ratio	2.7648
Max Drawdown	0.0814
Hit Rate (active)	54.7%
Win/Loss Ratio	1.46x
Wins / Losses	47 / 39
Total P&L (net)	0.328178
Skewness	1.20
Kurtosis	10.82
Trading Days	407
Days Active	86 (21.1%)
Kernels Active	6 / 10 + 2 gates
Stage	Two-Stage + Kalman

Too Good To Be True

BUG #1: Fast Feature Time Misalignment

After arr[:n_fast] trimming to 4,980 rows, LOB row 4,000 referred to July 2023 while Kyle row 4,000 was late 2024. The v2 OrderFlow PCA was concatenating features from completely different time periods.

Kernel	Total Bars	First 4,980 Cover
LOB	372,893	~6 days (Jul 2–10, 2023)
Kyle	5,000	All 463 days
Hawkes	157,824	~17 days (Jul 2–20, 2023)
VPIN	5,546	~95% of range

BUG #2: Degenerate OOS Target

- MultiResolutionAligner only covered bar indices 0–2,330
- Every OOS bar (2,580–4,979) clipped to day 233
- Forward-return target was identically zero over entire OOS window
- ElasticNet collapsed to intercept-only
- Kalman filtered fold-boundary intercept shifts as "regime transitions"
- These correlated with CL's trend 83% of the time — by coincidence
- Fix inverted the hit rate to 16.7%

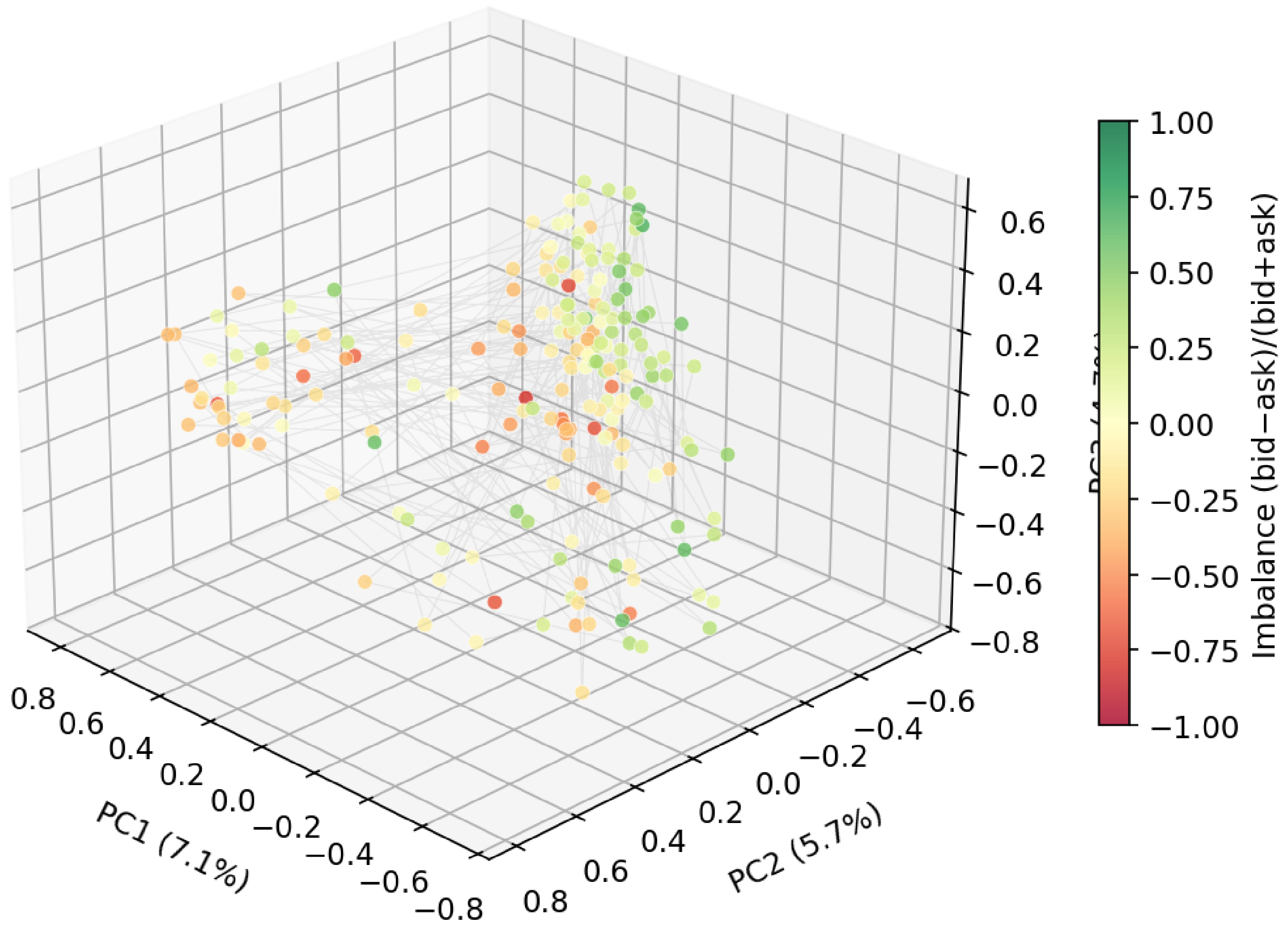
VERDICT: Post-fix, the kernels hit rate dropped to 16.7% and we are left with zero directional edge on CL.

Back Testing

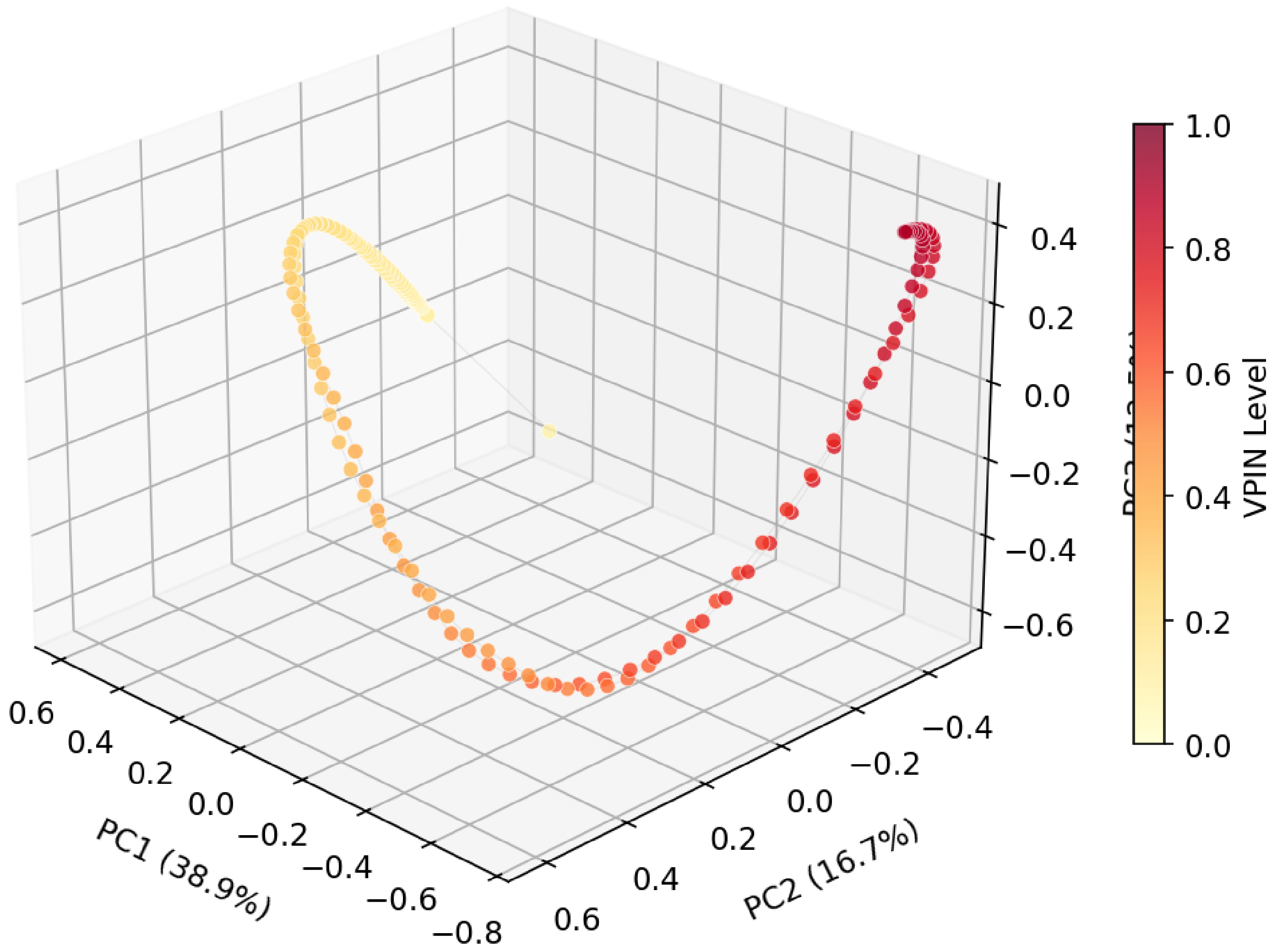
What We Do Have: Real Signals

RKHS Feature Space — PCA Projection (200 Synthetic Time Points)
Each kernel organises the same market states in its own Hilbert space

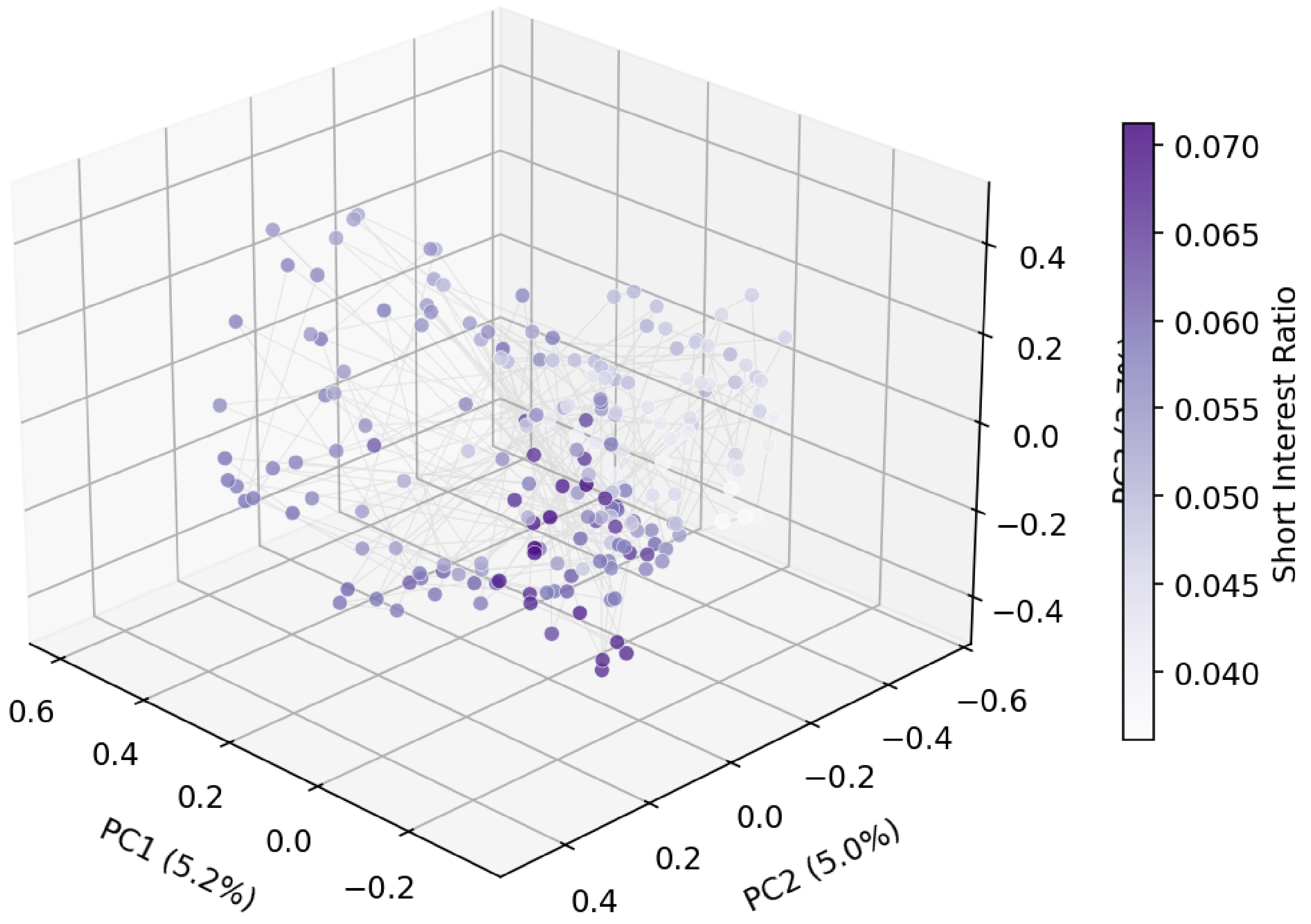
Limit Order Book Signal



VPIN Signal



Short Interest Signal



What Quant Firms Do

Model → **Expanded Universe** → **Leverage**

Model

Trend Following
Time-series momentum (TSMOM)

Mean Reversion
Pairs trading, stat arb

Machine Learning
Random forests, neural nets

Factor Models
Value, carry, quality

Volatility
VRP, options strategies

Expanded Universe

Energy
Crude Oil · Natural Gas · Heating Oil

Agriculture
Corn · Wheat · Soybeans · Sugar

Metals
Gold · Silver · Copper · Platinum

Equities
S&P · Nasdaq · DAX · Nikkei futures

Fixed Income
10yr Treasury · Bunds · Gilts

FX
EUR · JPY · GBP · AUD · CAD

50–100+ instruments across uncorrelated asset classes

Leverage

Typical CTA / quant fund targets:

Lev	Return	Vol	Sharpe	DD
1×	5–8%	10%	0.5–0.8	15%
2×	10–16%	20%	0.5–0.8	30%
3×	15–24%	30%	0.5–0.8	45%

Note: Sharpe is leverage-invariant — returns and drawdowns scale, edge does not.

Real-world examples
AQR · Man AHL · Winton
Millennium · Two Sigma · D.E. Shaw

Next Steps

Cross Assets

Integrating fixed income, currency, and commodity features would give the model a fuller picture of what is driving price movement.

The model could detect Risk-On and Risk-Off shifts in market mood before they hit individual stock prices.

This would let us identify whether a stock's move is driven by a broader global trend or something specific to that company.

Next Steps

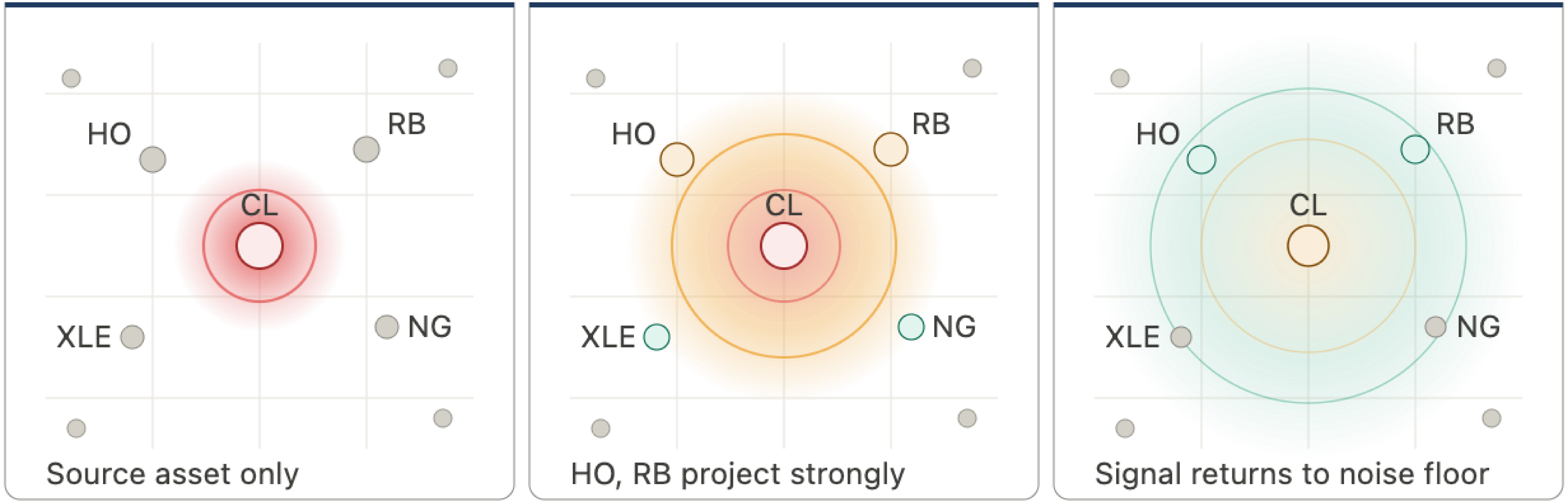
More Data

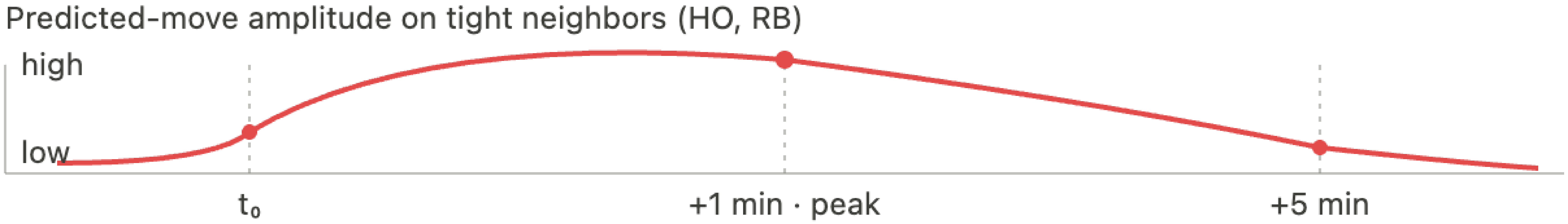
Pulling in real-time economic indicators and SEC filings would give the model a broader set of signals on top of what it already uses.

Market-By-Order data would show every individual order placed in the market rather than aggregated summaries.

These additional layers would allow the model to detect structural changes in a stock before they show up in the price.

Cross Asset Goal





Traders@SMU