

Functional Extreme Partial Least Squares

Unraveling the Intuition and Empirical Validation

Janis Aiad Simon Elis

Master MVA - Statistical Learning with Extreme Values
ENS Paris Saclay

Outline

1 Introduction

2 Theory

3 Practice

PCA vs PLS: The Fundamental Distinction

PCA (Unsupervised)

- Operates solely on X
- Maximizes *variance*
- Ignores relationship with Y
- May prioritize irrelevant directions

PLS (Supervised)

- Uses both X and Y
- Maximizes *covariance* with Y
- Finds predictive directions
- Directly relevant to prediction

FEPLS = PLS for extremes: Focus on covariance with Y *in the tail*

The FEPLS Model

Problem: Given (Y, X) , find direction w that best explains extreme Y values.

FEPLS Optimization

$$w(y) := \arg \max_{\|w\|=1} \text{Cov}(\langle w, X \rangle, Y \mid Y \geq y)$$

Inverse Model Assumption:

$$X = g(Y)\beta + \varepsilon, \quad \beta \in H, \quad \|\beta\| = 1$$

What's fixed and what's chosen?

- $g(\kappa)$: Real-world relationship (data-generating process)
- $\varphi(\tau)$: User-chosen test function (tuning parameter)

Parameter Tuning: What's Estimated, What's Chosen?

Estimated from data:

- γ : Tail index of Y (Hill estimator)
- ρ : Second-order parameter (2RV)

Chosen/Tuned:

- τ : Test function index (tau tuning)
- k : Number of extremes (k tuning)

Moment Condition

$$0 < 2(\kappa + \tau)\gamma < 1$$

Ensures moments exist. Tuning τ :

- Large τ : Focus on deepest extremes
- Negative τ : Better stability

Signal Dominance: $q\kappa\gamma > 1$ (ensures identifiability)

Convergence: Choice of k and Bias-Variance Tradeoff

Optimal Choice of k

Balancing bias and variance:

$$\text{error}(n, k) \approx C_1 (n/k)^\rho + C_2 k^{-1/2}$$

- $(n/k)^\rho$: Bias (2RV approximation error)
- $k^{-1/2}$: Variance (statistical noise)

Optimal: $k_n \sim c n^{-2\rho/(1-2\rho)}$ for $\rho < 0$

Bias-Variance Dilemma:

- Small k : Low bias, high variance
- Large k : Lower variance, higher bias

2RV Interpretation: ρ quantifies how fast the tail deviates from pure Pareto.
More negative $\rho =$ straighter tail = easier estimation.

Convergence Rate: Scaling Law

Theorem (FEPLS Consistency)

Under regularity conditions, the FEPLS estimator satisfies:

$$\|\hat{\beta}_\varphi(Y_{n-k+1,n}) - \beta\| = O_P(\delta_{n,k})$$

where

$$\delta_{n,k} \sim n^{(1/q - \gamma\kappa)/(1-2\rho)}$$

Key Insights

- Convergence rate depends on γ , κ , q , and ρ
- For $\rho < 0$: Explicit rate $O(n^{(1/q - \gamma\kappa)/(1-2\rho)})$
- Optimal k : $k_n \sim c n^{-2\rho/(1-2\rho)}$ balances bias and variance

Empirical Results: Overview

Two Settings:

① Low-frequency (daily): Hungarian stock market (Stooq data)

- 4IG paired with multiple stocks (3500 pairs)
- Good Hill estimation: $\gamma \in [0.2, 0.4]$ (very consistent)
- Limited samples: $k < 10$

② High-frequency (5-min): AAPL intraday data

- 5-minute windows
- Large sample: k up to 800
- Microstructure effects challenge Hill estimation

Low-Frequency Results: Hungarian Stock Market

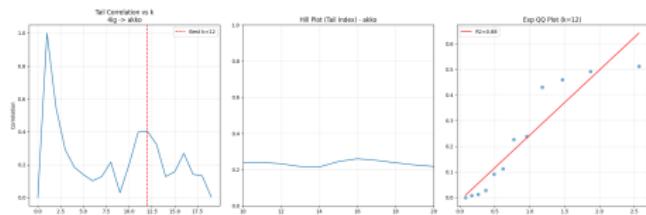
Key Findings:

- Hill estimator: $\gamma \in [0.2, 0.4]$ (very consistent)
- Limited extremes: $k < 10$
- τ dependency crucial for noisy, small-sample regime

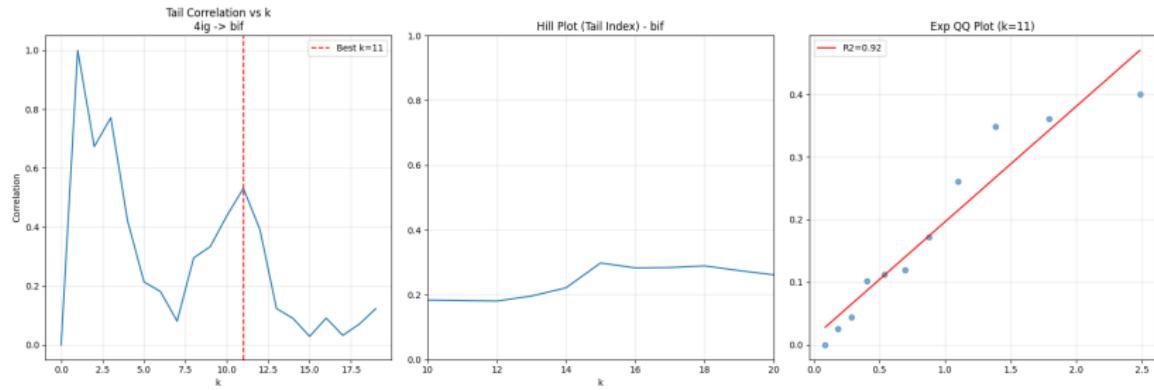
Stock Groups:

- **Sensitive:** End-of-day volatility drives next day's max
- **Insensitive:** No clear relationship

Interpretation: Cross-asset dependencies and volatility spillover mechanisms



Diagnostic Plots: Hill, Q-Q, and Tail Analysis



Multiple stock pairs show:

- Consistent Hill estimates ($\gamma \in [0.2, 0.4]$)
- Good Q-Q plot fit for low-frequency data
- Clear tail behavior suitable for FEPLS

High-Frequency Results: AAPL 5-Minute Data

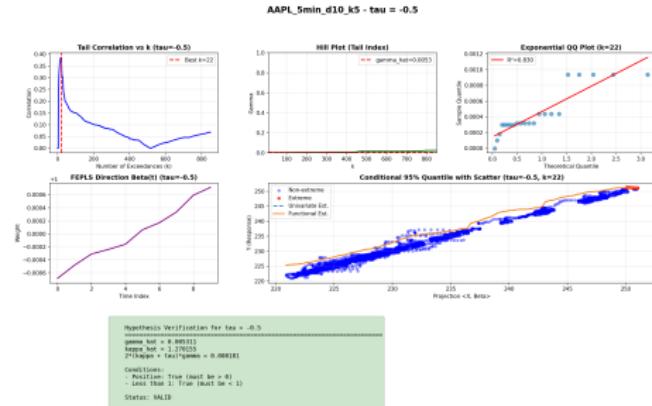
Setup:

- X : 5-minute return window
- Y : Max of next 5-minute window
- Large sample: k from 0 to 800

Challenges:

- Hill estimator very low
(microstructure/0.01\$ effects)
- Q-Q plot shows poor fit

Key Finding: Clear increasing relationship \Rightarrow **Mid-frequency is trend-driven!**



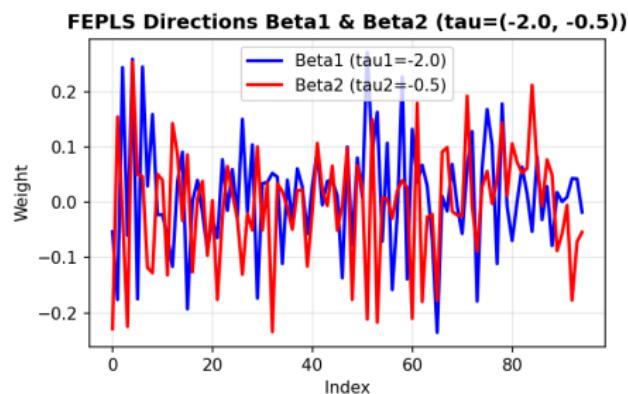
Beta Stability and Conclusions

Beta Comparison:

- Stability across τ values
- Sensitivity analysis for test function
- Identifies robust directions

Main Conclusions:

- ① FEPLS works well for heavy-tailed data
- ② τ tuning crucial in small-sample regime
- ③ Clear patterns: trend-driven (HF) vs volatility-driven (LF)



Practical Impact: Identifies predictive features for extreme events in financial markets