Functional Extreme Partial Least Squares: Application to Financial Data Analysis

Analysis Report

December 4, 2025

Abstract

This report presents an empirical analysis of the Functional Extreme Partial Least Squares (FEPLS) method applied to financial time series data. We focus on pairs of stocks from the Hungarian market, specifically examining the relationship between 4IG and AKKO stocks. The analysis includes hypothesis verification for different values of the parameter τ, visualization of the FEPLS direction estimation, and assessment of the method's performance through train-test validation. The theoretical framework is based on the FEPLS methodology which extends Extreme-PLS to functional data in Hilbert spaces, providing dimension reduction for extreme value analysis.

1 Introduction

The Functional Extreme Partial Least Squares (FEPLS) method represents a powerful approach for dimension reduction in extreme value analysis when dealing with functional covariates. This method extends the classical Partial Least Squares (PLS) framework to the extreme regime, where the focus is on the conditional tail of a response variable Y given a high-dimensional or infinite-dimensional covariate X.

In this report, we apply the FEPLS methodology to analyze the relationship between pairs of financial assets, with particular emphasis on the 4IG and AKKO stock pairs. The analysis involves:

- Testing multiple values of the parameter τ that controls the weight function in the FEPLS estimator
- Verifying theoretical hypotheses required for consistency
- Visualizing the estimated FEPLS directions and their properties
- Assessing predictive performance through train-test splits

2 Notation and Mathematical Framework

2.1 Notation

We introduce the main notation used throughout this report:

2.1.1 Spaces and Norms

- H: A separable Hilbert space (possibly infinite-dimensional)
- $\langle \cdot, \cdot \rangle$: The inner product on H
- $\| \cdot \|$: The norm induced by the inner product on H

- $\mathbb{E}_{\mathcal{B}}$ or $\mathbb{E}^{B}$: The Bochner expectation for random variables in H
- $(\Omega, \mathcal{A}, \mathbb{P})$: A complete probability space

2.1.2 Random Variables and Distributions

- $Y : \Omega \to \mathbb{R}$: The real-valued response variable (heavy-tailed)
- $X : \Omega \to H$: The functional covariate in H
- F: The cumulative distribution function (cdf) of Y
- $\bar{F} = 1 - F$: The survival function
- $f = F'$: The density of Y (when it exists)
- F^{-}: The generalized inverse (quantile function) of F
- $U(t) := F^{-}(1 - 1/t)$: The tail quantile function
- $Y_{1,n} \leq Y_{2,n} \leq \cdots \leq Y_{n,n}$: The order statistics of the sample $(Y_i)_{1 \leq i \leq n}$
- $Y_{n-k+1,n}$: The $(n - k + 1)$-th order statistic (the k-th largest value)

2.1.3 Tail Moments and Regular Variation

- $m_W(y) := \mathbb{E}(W \mathbf{1}_{\{Y \geq y\}})$: The tail-moment of a random object W (when it exists)
- $\hat{m}_W(y) := \frac{1}{n} \sum_{i=1}^{n} W_i \mathbf{1}_{\{Y_i \geq y\}}$: The empirical tail-moment
- $\mathrm{RV}_{\tau}(+\infty)$: The class of regularly-varying functions at infinity with index τ
- $2\mathrm{RV}_{\gamma,\rho}(+\infty)$: The class of second-order regularly-varying functions
- $H_{\rho}(y) := y^{\gamma} \int_1^y u^{\rho-1} \mathrm{d}u$: The auxiliary function for second-order regular variation
- $A(t)$: An auxiliary function in the second-order condition, ultimately of constant sign with $A(t) \to 0$ as $t \to +\infty$

2.1.4 Model and Estimators

- φ: A test function such that $\varphi \in \mathrm{RV}_{\tau}(+\infty)$ for some $\tau \in \mathbb{R}$
- $w_{\varphi}(y)$: The theoretical FEPLS direction (Functional Extreme-PLS)
- $\hat{\beta}_{\varphi}(y)$: The empirical FEPLS estimator of the direction
- g: A deterministic link function such that $g \in \mathrm{RV}_{\kappa}(+\infty)$ for some $\kappa > 0$
- $\beta \in H$: The true index direction with $\|\beta\| = 1$
- $\varepsilon : \Omega \to H$: A noise random variable
- $k = k_n$: An intermediate sequence (integer such that $k \to +\infty$ and $k/n \to 0$)
- $y_{n,k}$: A deterministic sequence such that $y_{n,k} \sim U(n/k)$ as $n \to +\infty$
- $\delta_{n,k}$: The convergence rate, defined as $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$
- $q > 2$: The integrability order of the noise (independent of n)

2.1.5 Model Parameters

- $\gamma \in (0, 1)$: The tail index of Y (first-order parameter)
- $\rho \leq 0$: The second-order parameter controlling the rate of convergence
- $\kappa > 0$: The index of regular variation of the link function g
- $\tau \in \mathbb{R}$: The index of regular variation of the test function φ

3 Intuitive Understanding of the FEPLS Framework

3.1 The Core Intuition: Why FEPLS Works

The fundamental idea behind FEPLS is to find a **single direction** in the infinite-dimensional space of functional covariates that best explains extreme events. Think of it as finding the "recipe" for extreme market movements: if you could only look at one linear combination of all the intraday price movements, which combination would best predict tomorrow's crash?

Unlike classical dimension reduction methods (like PCA) that focus on average behavior, FEPLS specifically targets the tail. This is crucial because:

- **Extreme events are rare**: Most days are "normal," but the few extreme days contain the most important information for risk management.
- **Tail behavior is different**: The patterns that predict a 5% daily return are fundamentally different from those that predict a 0.1% return.
- **High-dimensional curse**: With thousands of intraday observations, we need to reduce dimension, but we want to preserve the information relevant to extremes.

3.2 The Mathematical Intuition: From Covariance to Direction

The optimization problem (1) asks: "Which direction w maximizes the covariance between the projection $\langle w, X \rangle$ and Y, *when we only look at extreme values of* Y?"

This is different from standard PLS because:

- Standard PLS: Maximizes covariance over all data points (average behavior).
- FEPLS: Maximizes covariance *conditionally* on $Y \geq y$ (tail behavior).

The key insight is that by conditioning on extremes, we focus the optimization on the regime that matters most for risk prediction.

3.3 Why the Inverse Model?

The model $X = g(Y)\beta + \varepsilon$ is called "inverse" because we write the covariate X in terms of the response Y, rather than the usual $Y = f(X) + \text{noise}$.

Why this makes sense:

- In extreme value theory, it's often easier to work with the inverse relationship.
- The function g captures how extreme values of Y "explode" the covariate X.
- When Y is very large, $g(Y)$ is also very large (since g is regularly varying with positive index), so X is dominated by the signal term $g(Y)\beta$ rather than the noise ε.

3.4 The Role of Regular Variation: Why Power Laws Matter

Regular variation is the mathematical language for describing "power law" behavior. A function f is regularly varying with index α if, for large t, it behaves like $f(t) \approx t^\alpha \times$slowly varying function.
Why this matters:

- Heavy-tailed distributions (like those in finance) have power-law tails: $P(Y > y) \approx y^{-1/\gamma}$.
- The link function g also has power-law behavior: $g(y) \approx y^\kappa$.
- This allows us to use the powerful machinery of extreme value theory, which is built on regular variation.
- The composition of regularly-varying functions is also regularly varying, which makes the mathematics tractable.

4 Theoretical Framework

4.1 The FEPLS Optimization Problem

Let $(H, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space and consider a random pair (Y, X) where $Y : \Omega \to \mathbb{R}$ and $X : \Omega \to H$. The FEPLS direction is defined as the solution to the optimization problem:

$$w(y) :=_{\|w\|=1} \operatorname{Cov}\big(\langle w, X\rangle, Y \mid Y \geq y\big), \quad y \in \mathbb{R}, \tag{1}$$

where y is a large threshold. This optimization seeks the direction in H that maximizes the covariance between the projection of X and the response Y, conditionally on Y being in the tail region.

4.2 Explicit Solution and Generalization

The solution to the optimization problem (1) can be explicitly characterized. More generally, we consider a family of estimators parameterized by a test function φ:

Theorem 4.1 (Explicit Solution of the FEPLS Optimization Problem)**.** *Suppose that Y is integrable and X is Bochner-integrable. Then, the unique solution of the optimization problem (1) is given by, for any $y \in \mathbb{R}$,*

$$w(y) = \frac{v(y)}{\|v(y)\|} \quad \textit{where} \quad v(y) = \bar{F}(y) m_{XY}(y) - m_X(y) m_Y(y), \tag{2}$$

where $\bar{F}(y) = 1 - F(y)$ is the survival function and $m_W(y) = \mathbb{E}(W \mathbf{1}_{\{Y \geq y\}})$ denotes the tail-moment of a random object W.

More generally, we consider the FEPLS direction:

$$w_\varphi(y) = \frac{v_\varphi(y)}{\|v_\varphi(y)\|} \quad \textit{with} \quad v_\varphi(y) := m_{X\varphi(Y)}(y), \tag{3}$$

where φ is a test function that is regularly-varying at infinity with index $\tau \in \mathbb{R}$, i.e., $\varphi \in RV_\tau(+\infty)$.

4.3 The Inverse Single-Index Model

To provide theoretical guarantees, we assume the following inverse single-index functional model:

$$X = g(Y)\beta + \varepsilon, \quad \beta \in H, \quad \|\beta\| = 1, \tag{4}$$

where $\varepsilon : \Omega \to H$ is a noise random variable and g is an unknown deterministic link function such that $g \in \mathrm{RV}_\kappa(+\infty)$ for some $\kappa > 0$.

The Distinction Between g and φ: Model vs. Method

A fundamental distinction exists between the link function g and the test function φ:
g is imposed (It's the nature of the data):

- g is part of the data-generating model ($X = g(Y)\beta + \varepsilon$)
- It describes the physical or economic reality of the relationship between Y and X
- You do not choose g; it is determined by the underlying process
- For example, if the market reacts to the square of past returns, then $g(y) = y^2$

φ is chosen (It's your tuning tool):

- φ is part of the estimator ($\hat{\beta}_\varphi$)
- It is a "tuning knob" that the statistician adjusts to optimize estimation
- It does not exist in nature; it is an "instrumental" or "test" function introduced to capture information

Why do we need both?
Role of φ (Mathematical): Ensuring moment existence (The "Compensator")
The crucial condition $0 < 2(\kappa + \tau)\gamma < 1$ involves:

- κ: The power index of g (fixed by nature)
- τ: The power index of φ (chosen by you)
- γ: The tail index (fixed by nature)

If the natural relationship g is too strong (e.g., κ is large, X explodes when Y is large), the empirical moments may explode (the integral diverges), making estimation impossible. φ serves to compensate g. If κ is too large, you can choose a negative τ (a decreasing function φ) to "calm" the integral and satisfy the condition.
Role of φ (Statistical): Weighting extremes (The "Focus")
The estimator computes a weighted covariance:

$$\sum X_i \cdot \varphi(Y_i) \cdot \mathbf{1}_{\{Y_i \geq y\}}$$

By choosing φ, you decide what weight to give to the most extreme values among the extremes:

- If $\varphi(y) = y^\tau$ with $\tau > 0$, you give enormous weight to the observations farthest in the tail. This may reduce bias (truly in the extreme) but increase variance (based on very few highly volatile points).
- If $\tau \approx 0$ or negative, you smooth the estimation more over all values exceeding the threshold.

Why not use only one or the other?

- If we only had g (and fixed $\varphi = 1$), we could not estimate the model when g grows too fast (moments would not exist).
- If we only had φ (assuming $g(y) = y$, linear model), we would make a systematic modeling error (bias) whenever the real relationship between X and Y is not linear.

4.4 Second-Order Regular Variation

We assume that the response variable Y is heavy-tailed to the second order. Specifically, the tail quantile function $U(t) := F^{-}(1 - 1/t)$ belongs to the class of second-order regularly-varying functions:

Definition 4.2 (Second-Order Regular Variation)**.** The function U belongs to $2\mathrm{RV}_{\gamma,\rho}(+\infty)$ if there exist $\gamma \in (0, 1)$, $\rho \leq 0$ and an auxiliary function A ultimately of constant sign with $A(t) \to 0$ as $t \to +\infty$ such that:

$$\lim_{t\to+\infty} \frac{1}{A(t)}\Big(\frac{U(ty)}{U(t)} - y^{\gamma}\Big) = y^{\gamma} H_{\rho}(y) := y^{\gamma} \int_1^y u^{\rho-1}\mathrm{d}u, \quad y > 0. \tag{5}$$

Interpretation of ρ and the Role of Second-Order Regular Variation

The parameter $\rho \leq 0$ is the **second-order parameter** that controls the rate of convergence in the second-order regular variation condition.

- When $\rho = 0$: The convergence is logarithmic, i.e., $H_{\rho}(y) = \log(y)$. This corresponds to a slow convergence rate.
- When $\rho < 0$: The convergence is polynomial, i.e., $H_{\rho}(y) = \frac{y^{\rho}-1}{\rho}$. More negative values of ρ indicate faster convergence.
- The condition $\sqrt{k}A(n/k) = O(1)$ in the consistency theorem ensures that the threshold selection is compatible with the second-order behavior, typically requiring $k \sim c \cdot n^{-2\rho/(1-2\rho)}$ for some $c > 0$ when $\rho < 0$.

In practice, ρ quantifies how well the distribution approximates a pure Pareto distribution in the tail. A more negative ρ means the tail behavior is closer to the first-order approximation.

Why is Second-Order Regular Variation (2RV) necessary?

The first-order regular variation (RV_{γ}) tells us that the distribution "resembles" a Pareto distribution asymptotically, but in finite samples, we are never truly "at infinity." The distribution is:

$$\text{True Distribution} = \text{Pareto} + \text{Error}$$

The 2RV quantifies this error through the auxiliary function $A(t)$. This becomes crucial when choosing k:

- If k is very small (deep in the tail), the approximation error is small (close to the limit), but the variance is huge (few points).
- If k is larger (to stabilize variance), we move away from the extreme tail where the distribution deviates from pure Pareto. The approximation error grows.

The 2RV tells us *how fast* this error grows. The condition $\sqrt{k}A(n/k) = O(1)$ balances:

- $\sqrt{k}$: The statistical variance (noise)
- $A(n/k)$: The model bias (error quantified by 2RV)

This condition says: "You may increase k as long as your model error (A) remains smaller than your statistical noise ($\sqrt{k}$)."
In summary: 1RV (γ) gives the *direction* of the tail (the slope), while 2RV (ρ) gives the *straightness* of the tail (is it a perfect straight line in log-log scale, or is it curved?). If ρ is very negative, the tail is almost straight, making estimation easier. If ρ is close to 0, convergence is very slow.

5 Hypotheses in Extreme Value Theory Language

Before stating the consistency results, we reformulate the key hypotheses using the terminology and intuition of Extreme Value Theory (EVT):

5.1 Membership in the Fréchet Domain of Attraction (Heavy Tails)

The hypothesis $U \in \mathrm{RV}_\gamma(+\infty)$ means that the distribution of the response Y belongs to the **Fréchet Maximum Domain of Attraction**. Concretely, this means the distribution is heavy-tailed (Pareto-type): the probability of observing extreme values decays polynomially, not exponentially. This is the fundamental assumption that places us in the extreme value framework.

5.2 Intermediate Order Statistics Regime

The hypothesis on the sequence k ($k \to \infty$ and $k/n \to 0$) places inference in an **intermediate regime**. We look neither at the absolute maximum (too volatile) nor at the center of the distribution (biased), but at a portion of the tail that contains increasingly more observations ($k \to \infty$) while remaining a negligible fraction of the total sample to stay in the extreme asymptotic regime.

5.3 Classical Bias Control Condition

The condition $\sqrt{k}A(n/k) = O(1)$ is the standard condition for bias control in extreme value theory. It ensures that the error made by approximating the tail distribution by its limit model (generalized Pareto) does not dominate the variance of the estimator when moving away in the tail.

5.4 Signal Tail Dominance Over Noise

The inequality $q\kappa\gamma > 1$ is interpreted as a **tail dominance condition**. It guarantees that the tail index of the "signal" part ($g(Y)$) is higher (heavier tail) than that of the "noise" part (ε). In other words, in extreme events, it is the signal structure that drives the phenomenon, with noise becoming negligible compared to the explosion of Y.

5.5 Preservation of Tail Regularity

The hypotheses on g and φ (regular variation) ensure stability under transformation. Since Y is regularly varying, applying power functions (or similar) g and φ preserves the "Fréchet" nature of the data: the transformed variables remain heavy-tailed, allowing the application of standard limit theorems.

5.6 Existence of Tail-Moments

The condition $2(\kappa+\tau)\gamma < 1$ simply ensures that the tail moments necessary for constructing the covariance exist (the integral converges). This is the extreme equivalent of saying "the variance exists," but adapted to power laws where higher-order moments often explode.

6 Consistency Results

6.1 The Empirical FEPLS Estimator

Given a sample $(X_i, Y_i)_{1\le i\le n}$ of independent copies of (X, Y), we define the empirical tail-moment:

$$\hat{m}_W(y) := \frac{1}{n}\sum_{i=1}^{n} W_i \mathbf{1}_{\{Y_i \ge y\}}.$$

The FEPLS estimator is then:

$$\hat{\beta}_\varphi(y) := \frac{\hat{v}_\varphi(y)}{\|\hat{v}_\varphi(y)\|} \quad \text{with} \quad \hat{v}_\varphi(y) = \hat{m}_{X\varphi(Y)}(y) = \frac{1}{n}\sum_{i=1}^{n} X_i\varphi(Y_i)\mathbf{1}_{\{Y_i \ge y\}}. \tag{6}$$

6.2 Main Consistency Theorem

Theorem 6.1 (Consistency of FEPLS Estimator)**.** *Assume the following conditions:*

1. *The test function $\varphi \in RV_\tau(+\infty)$ for some $\tau \in \mathbb{R}$*
2. *The link function $g \in RV_\kappa(+\infty)$ for some $\kappa > 0$*
3. *The tail quantile function $U \in 2RV_{\gamma,\rho}(+\infty)$ with $\gamma \in (0, 1)$ and $\rho \le 0$*
4. *The noise condition: there exists $q > 2$ such that*

$$\limsup_{n\to+\infty} \sup_{y\ge 0} \mathbb{E}\left(\|\varepsilon_1\|^q \mid Y_{n-k+1,n} = y\right) < +\infty \tag{7}$$

5. *The threshold condition: $\sqrt{k}A(n/k) = O(1)$*
6. *The bounds:*

$$0 < 2(\kappa+\tau)\gamma < 1, \tag{8}$$
$$q\kappa\gamma > 1 \tag{9}$$

7. *Regularity: φ and g are continuously differentiable in a neighbourhood of infinity such that $t(\varphi g)'(t)/(\varphi g)(t) \to \tau + \kappa \neq 0$ as $t \to +\infty$*

Let $k := k_n \to +\infty$ be an intermediate sequence (i.e., $k/n \to 0$ as $n \to +\infty$), and let $y_{n,k} \sim U(n/k)$ as $n \to +\infty$. Define the convergence rate:

$$\delta_{n,k} := \left(g(y_{n,k}) \left(\frac{k}{n}\right)^{1/q} \right)^{-1}.$$

Then, the FEPLS estimator satisfies:

$$\|\hat{\beta}_\varphi(Y_{n-k+1,n}) - \beta\| = O_{\mathbb{P}}(\delta_{n,k}) \xrightarrow[n\to+\infty]{} 0.$$

Interpretation of the Bounds

The consistency theorem requires two key bounds that ensure the estimator converges:
Bound 1: $0 < 2(\kappa + \tau)\gamma < 1$
This condition ensures the existence of the second moment of $\varphi \cdot g(Y)\mathbf{1}_{\{Y \geq y\}}$ for all $y \geq 0$.

- The product $(\kappa + \tau)\gamma$ represents the effective tail index of the weighted response $\varphi(Y) \cdot g(Y)$
- The factor of 2 comes from requiring the second moment to exist
- The upper bound < 1 ensures that the variance remains finite in the tail
- This implies that $\tau < \frac{1}{2\gamma} - \kappa$, which can be either positive or negative depending on γ and κ

Bound 2: $q\kappa\gamma > 1$
This condition ensures that the convergence rate $\delta_{n,k} \to 0$ as $n \to +\infty$, regardless of the slowly-varying parts of g and $\bar{F}$.

- The product $\kappa\gamma$ is the tail index of $g(Y)$
- The factor q (where $q > 2$ from the noise condition) ensures sufficient integrability
- This condition implies that $g(Y)$ has a heavier tail than $\|\varepsilon\|$, which is necessary for the signal to dominate the noise in the extreme regime
- The convergence rate $\delta_{n,k}$ is of order $n^{(1/q-\gamma\kappa)/(1-2\rho)}$, so this bound ensures the exponent is negative

Together, these bounds ensure that the FEPLS estimator is both well-defined and consistent.
Summary:

- $q\kappa\gamma > 1$: Extract signal from noise (signal tail heavier than noise tail)
- $2(\kappa + \tau)\gamma < 1$: Ensure $\hat{\beta}$ is consistent and convergent (variance remains finite)

6.3 Technical Lemmas

The consistency proof relies on several technical lemmas that we state here for completeness. These lemmas handle the technical aspects of working with random thresholds and controlling the noise terms.

Lemma 6.2 (Bochner Expectation and Inner Product Commutation)**.** *For any independent Bochner-integrable random variables W_1, W_2 having values in $(H, \langle \cdot, \cdot \rangle)$ a separable Hilbert space and being independent conditionally to $\mathcal{F} \in \mathcal{B}(\mathbb{R})$,*

$$\mathbb{E}_{\mathcal{F}}(\langle W_1, W_2 \rangle) = \langle \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_1), \mathbb{E}_{\mathcal{F},\mathcal{B}}(W_2) \rangle.$$

Remark 6.3. *This lemma is crucial for manipulating conditional expectations involving inner products in Hilbert spaces. It essentially says that when two random variables are conditionally independent, the expectation of their inner product equals the inner product of their expectations. This is the infinite-dimensional analog of the fact that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ for independent random variables.*

Lemma 6.4 (Joint Density of Order Statistics). *Let $\{Y_1, \ldots, Y_n\}$ be i.i.d. random variables with common cdf F. Let k be an integer such that $2 \leq k < n$ and let any $1 \leq i \leq n$. The joint density $(t, y) \mapsto f_{n,k}(t, y)$ of $(Y_i, Y_{n-k+1,n})$ is given for all $y \leq t$ by:*

$$f_{n,k}(t, y) = f(t)f(y)\frac{(n-1)!}{(n-k)!(k-2)!}F^{n-k}(y)\bar{F}^{k-2}(y).$$

Lemma 6.5 (Conditional Expectation with Random Threshold). *Let $h \in RV_\rho(+\infty)$ with $\rho \in \mathbb{R}$ and i.i.d. random variables $\{Y_1, \ldots, Y_n\}$ with common density $f \in RV_{-1/\gamma-1}(+\infty)$, $\gamma > 0$. Assume that $\rho\gamma < 1$. Let $2 \leq k \leq n$ be some integer. Then, for any $y \geq 0$ and any $1 \leq i \leq n$:*

$$\mathbb{E}\left(h(Y_i)\mathbf{1}_{\{Y_i \geq Y_{n-k+1,n}\}} \mid Y_{n-k+1,n} = y\right) = \frac{k-1}{n} \cdot \frac{1}{\bar{F}(y)} \int_y^{+\infty} h(t)f(t)\mathrm{d}t.$$

Moreover, when $y \to +\infty$:

$$\mathbb{E}\left(h(Y_i)\mathbf{1}_{\{Y_i \geq Y_{n-k+1,n}\}} \mid Y_{n-k+1,n} = y\right) \sim \frac{k-1}{n} \cdot \frac{h(y)}{1-\rho\gamma}.$$

Lemma 6.6 (Bound on Empirical Tail-Moments of Noise). *Assume that $\bar{F} \in RV_{-1/\gamma}(+\infty)$, the test function condition, link function condition, noise condition, and the bounds hold. Let $k := k_n \to +\infty$ be an integer deterministic sequence such that $k/n \to 0$ and $y_{n,k} \sim U(n/k)$ as $n \to +\infty$. Let $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$. Then:*

$$\frac{\|\hat{m}_{\varphi(Y)\varepsilon}(Y_{n-k+1,n})\|}{m_{\varphi \cdot g(Y)}(y_{n,k})} = O_{\mathbb{P}}(\delta_{n,k}) \xrightarrow[n \to +\infty]{} 0.$$

Lemma 6.7 (Bound on Inner Product with Noise). *Let $\beta \in H$ with $\|\beta\| = 1$ and suppose $\bar{F} \in RV_{-1/\gamma}(+\infty)$. Assume the test function, link function, noise condition, and bounds hold. Let $k := k_n \to +\infty$ be an integer deterministic sequence such that $k/n \to 0$ and $y_{n,k} \sim U(n/k)$ as $n \to +\infty$. Let $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$. Then:*

$$\frac{\hat{m}_{\langle \beta, \varphi(Y)\varepsilon \rangle}(Y_{n-k+1,n})}{m_{\varphi \cdot g(Y)}(y_{n,k})} = O_{\mathbb{P}}(\delta_{n,k}) \xrightarrow[n \to +\infty]{} 0.$$

7 Complete List of Hypotheses

For reference, we provide here the complete list of all hypotheses required for the consistency theorem, organized by category:

7.1 Hypotheses on the Response Variable Y (Distribution and Tail)

1. **Integrability:** Y is integrable (necessary to define the covariance).
2. **Second-Order Regular Variation:** The tail quantile function U belongs to the class $2\mathrm{RV}_{\gamma,\rho}(+\infty)$. This implies:
   - $U \in \mathrm{RV}_\gamma(+\infty)$ with $\gamma \in (0, 1)$. The condition $\gamma < 1$ is required for the existence of the first-order moment.
   - The survival function $\bar{F} \in \mathrm{RV}_{-1/\gamma}(+\infty)$.
   - The second-order parameter $\rho \leq 0$.
3. **Density:** The density $f = F'$ exists and is regularly varying with index $1/\gamma - 1$.

7.2 Hypotheses on the Regression Model

The assumed model is $X = g(Y)\beta + \varepsilon$:

1. **Covariate X:** X is Bochner-integrable.
2. **Parameter β:** $\beta \in H$ and is unitary ($\|\beta\| = 1$).
3. **Link Function g:**
    - g is a deterministic function such that $g \in \mathrm{RV}_\kappa(+\infty)$ with $\kappa > 0$.
    - g is continuously differentiable in a neighbourhood of infinity.
4. **Noise ε:**
    - There exists $q > 2$ (independent of n) such that the q-th moment of the norm of the noise, conditionally on the extreme threshold, is uniformly bounded:

$$\limsup_{n\to+\infty} \sup_{y\geq 0} \mathbb{E}\left(\|\varepsilon_1\|^q \mid Y_{n-k+1,n} = y\right) < +\infty$$

    - This implies $\mathbb{E}(\|\varepsilon\|^q) < +\infty$.

7.3 Hypotheses on the Inference Method (Test Function and Parameters)

1. **Test Function φ:**
    - $\varphi \in \mathrm{RV}_\tau(+\infty)$ with $\tau \in \mathbb{R}$.
    - φ is continuously differentiable in a neighbourhood of infinity.
2. **von Mises Condition (Asymptotic Regularity):**

$$\frac{t(\varphi g)'(t)}{(\varphi g)(t)} \to \tau + \kappa \neq 0 \quad \text{as } t \to +\infty$$

3. **Index Constraints:**
    - $0 < 2(\kappa + \tau)\gamma < 1$ (ensures convergence of the involved tail moments).
    - $q\kappa\gamma > 1$ (ensures that the tail of $g(Y)$ is heavier than that of $\|\varepsilon\|$).

7.4 Asymptotic Hypotheses (for $n \to \infty$)

1. **Intermediate Sequence k:** $k = k_n \to +\infty$ and $k/n \to 0$.
2. **Bias Control:**

$$\sqrt{k}A(n/k) = O(1)$$

   This classical condition in extreme value theory controls the approximation error of the quantile by the power law limit.

8 Estimation Algorithm and Statistical Workflow

8.1 The FEPLS Estimation Algorithm: Step-by-Step Intuition

The algorithm for computing $\hat{\beta}_\varphi$ is straightforward: it is a weighted average of the vectors X_i, where only the "extreme" days count. Here's the intuitive reasoning behind each step:

Step 1: Find the Extreme Threshold

- **What we're doing:** Identifying which observations are "extreme."
- **Why:** We only want to use the tail data, where the signal-to-noise ratio is highest.
- **How:** Sort Y and take the k-th largest value as the threshold.
- **Intuition:** This is like saying "we'll focus on the worst k days."

Step 2: Identify Extreme Indices

- **What we're doing:** Finding all days where Y exceeds the threshold.
- **Why:** These are the days where extreme events occurred.
- **Intuition:** We're filtering the dataset to keep only "crisis days."

Step 3: Compute Weights

- **What we're doing:** Assigning weights $\varphi(Y_i) = Y_i^\tau$ to each extreme observation.
- **Why:** This allows us to give more or less importance to the most extreme values.
- **Intuition:** If $\tau > 0$, the worst days get enormous weight. If $\tau < 0$, we smooth the weights across all extreme days.
- **Mathematical reason:** This ensures the moments exist (convergence condition).

Step 4: Weighted Sum

- **What we're doing:** Computing $\sum w_i X_i$ over extreme observations.
- **Why:** This aggregates the functional patterns from all extreme days.
- **Intuition:** We're finding the "average pattern" of intraday returns on crisis days, weighted by severity.
- **Mathematical reason:** This is the empirical version of the tail-moment $m_{X\varphi(Y)}(y)$.

Step 5: Normalization

- **What we're doing:** Dividing by the norm to get a unit vector.
- **Why:** We only care about the direction, not the magnitude.
- **Intuition:** The direction tells us which parts of the day matter most; the magnitude is just a scaling factor.

8.2 The FEPLS Estimation Algorithm: Formal Description

Input:

- X: Matrix $(n \times d)$ of functional curves
- Y: Vector (n) of scalar responses
- k: Number of extremes to consider
- τ: Exponent of the test function (e.g., -2)

Algorithm:

1. **Find the extreme threshold:** Sort Y in descending order. The threshold y_{thresh} is the k-th largest value of Y.
2. **Identify extreme indices:** Find all indices i where $Y_i \geq y_{thresh}$. There will be exactly k indices (or slightly more if there are ties).
3. **Compute weights (The "Non-Linear" part):** For each extreme observation i, compute the scalar weight:
$$w_i = \varphi(Y_i) = (Y_i)^\tau$$
4. **Weighted sum (The "Projection" part):** Compute the weighted sum of the retained vectors X_i:
$$V_{raw} = \sum_{i \in \text{extremes}} w_i \cdot X_i$$
(Note: The factor $1/n$ from the formula disappears during normalization, so it can be ignored for computing the direction.)
5. **Normalization:** Divide the vector V_{raw} by its Euclidean (L^2) norm to obtain a unit vector:
$$\hat{\beta} = \frac{V_{raw}}{\|V_{raw}\|}$$

8.3 Parameter Selection: τ, q, and k

A common question is: "Do I need to optimize all parameters?" Here's the practical answer:

8.3.1 For τ (Test Function Parameter)

Yes, you test multiple values, but often you take a "safe" value.

- **In practice:** Choose often negative values (e.g., $\tau = -1, -2, -3$).
- **Why:** As explained, we want to ensure $2(\kappa + \tau)\gamma < 1$. Since we don't know the true parameters κ and γ initially, taking a negative τ is "insurance" to calm the integral and guarantee that moments exist.
- **Sensitivity:** The paper notes that results are "not very sensitive" to the precise choice of τ, as long as it's reasonable.
- **Recommendation:** Start with $\tau = -1$ or -2, then test a few values around it to verify robustness.

8.3.2 For q (Noise Integrability Order)

NO, you do not estimate it.

- **q is theoretical:** q represents the integrability of the noise ε (the noise has a finite q-th moment).
- **Not in the algorithm:** Look at the formula for $\hat{\beta}_\varphi(y)$: q does not appear anywhere.
- **The hidden assumption:** By applying the method, you simply make the implicit "bet" that the noise is lighter than the signal (the hypothesis $q\kappa\gamma > 1$). You don't need to calculate q.
- **In practice:** You assume the model is valid. If the method fails (e.g., $\hat{\beta}$ is unstable or points in random directions), it might indicate that this assumption is violated.

8.3.3 For k (Number of Extremes)

YES, this is the crucial step.

Since k (the number of extreme values considered) is the most critical parameter (the bias-variance tradeoff), we use a "data-driven" method:

1. Define a search range (e.g., between 5 and $n/5$).
2. For each k:
   - Compute the estimator $\hat{\beta}$.
   - Project the data onto this $\hat{\beta}$.
3. **Selection criterion:** Compute the correlation between Y and this projection (restricted to extreme events).
4. Keep the k that maximizes this correlation.

Why this works: The correlation measures how well the projection captures the extreme behavior. Maximizing it finds the best compromise between:

- Using enough data to be stable (larger k)
- Staying in the extreme regime where the model is valid (smaller k)

8.4 Complete Statistical Workflow

For a rigorous statistical workflow with financial data (e.g., Stooq 5-minute data), the following procedure should be followed:

8.4.1 Phase 1: Data Engineering (Construction of Functional Objects)

- **Cleaning & Log-returns:** Never work on raw prices (non-stationary). Compute log-returns: $R_t = \ln(P_t/P_{t-1})$. Handle missing data (trading gaps) by linear interpolation or "last observation carried forward" to have identical time grids.
- **Definition of X (The curve) and Y (The scalar):**
  - X_i (Covariate): The vector of 5-minute log-returns for day i (e.g., a vector of dimension $d = 78$ if the market is open for 6.5 hours).

- Y_i (Response): An extreme scalar measure for day $i+1$ (or the same day for concomitant analysis). The paper uses the maximum of log-returns for the day.

- **Temporal Dependence (Critical Point):** Financial markets have "volatility clusters" (GARCH). If day i is extreme, day $i+1$ will likely be extreme too, violating the i.i.d. assumption. **Solution:** Space out observations. For example, take every 3rd or 5th day to reduce serial dependence and approach the theoretical assumptions.

8.4.2 Phase 2: Hypothesis Verification (Sanity Check)

Before running the algorithm, verify that you are in the application framework (Heavy Tail):

- **Hill Plot on Y:** Plot the Hill estimator of your Y_i. If the tail index γ is close to 0 (or < 0), stop: FEPLS is not designed for this (it's for Pareto/Fréchet-type laws). In finance, we often find $\gamma \in [0.2, 0.4]$, which is perfect.
- **Canonical Correlation Test (Optional):** Verify that there is a minimal link between X and Y. If correlation is zero everywhere, dimension reduction will only find noise.

8.4.3 Phase 3: Calibration and Estimation (The Core of FEPLS)

- **Train/Test Split:** Split your dataset chronologically (e.g., 2010-2018 for Train, 2019-2023 for Test). Never optimize k on the Test set.
- **Choice of τ (Test Function):** Fix a conservative value, e.g., $\tau = -1$ or $\tau = -2$. Reason: This stabilizes the integral if the relationship g is explosive. No need for a complex "grid search" on τ initially.
- **k Optimization Loop (on Train Set):**
  1. Define a grid of k (e.g., from $k = 10$ to $k = 150$).
  2. For each k:
     - Compute $\hat{\beta}_{\tau,k}$.
     - Project X_{train} onto it: $z_i = \langle X_i, \hat{\beta} \rangle$.
     - Compute the correlation between z_i and Y_i only for extreme Y_i (the k largest).
  3. Select the k^* that maximizes this correlation.

8.4.4 Phase 4: Validation and Use (Backtest)

Once $\hat{\beta}$ and k^* are obtained, apply them to the Test Set:

- **Dimension Reduction:** For each day in the Test Set, transform your complex curve X_{new} into a simple number: $x_{new}^{proj} = \langle X_{new}, \hat{\beta} \rangle$.
- **Risk Estimation (Conditional VaR):** Use this univariate score x_{new}^{proj} to estimate tomorrow's Value-at-Risk (as shown in Figure 5d of the paper). You can use quantile regression on the projected score, or the Nadaraya-Watson method described in the paper.
- **Success Metrics:**
  - If your VaR curve is exceeded exactly α% of the time (e.g., 1% for a 99% VaR), then the model has correctly captured the extreme dependence structure.
  - **Kupiec Test:** Statistically verify if the exceedance rate is significantly different from the target.
  - **Independence of Exceedances:** Verify that VaR failures are not clustered (Cluster test).

9 Empirical Analysis: 4IG and AKKO Stock Pairs

9.1 Data Description

We analyze the relationship between pairs of stocks from the Hungarian market, focusing on the 4IG and AKKO stocks. The data consists of:

- Functional covariates X_i: intraday log-return curves for one stock
- Scalar responses Y_i: daily maximum log-returns for another stock
- The analysis is performed in both directions: 4IG $\to$ AKKO and AKKO $\to$ 4IG

The dataset is split into training (80%) and testing (20%) sets to assess the predictive performance of the FEPLS method.

9.2 Methodology

For each pair, we:

1. Estimate the tail index γ and the link function parameter κ on the training set
2. Test multiple values of $\tau \in \{-3.0, -2.0, -1.0, -0.5, 0.0, 0.5, 1.0, 2.0, 3.0\}$
3. For each τ, compute the FEPLS direction $\hat{\beta}_\varphi$ and verify the hypothesis condition: $0 < 2(\kappa + \tau)\gamma < 1$
4. Select the optimal threshold k using the sharpness criterion on the correlation curve
5. Evaluate the correlation between projections and responses on both training and test sets
6. Generate comprehensive visualizations

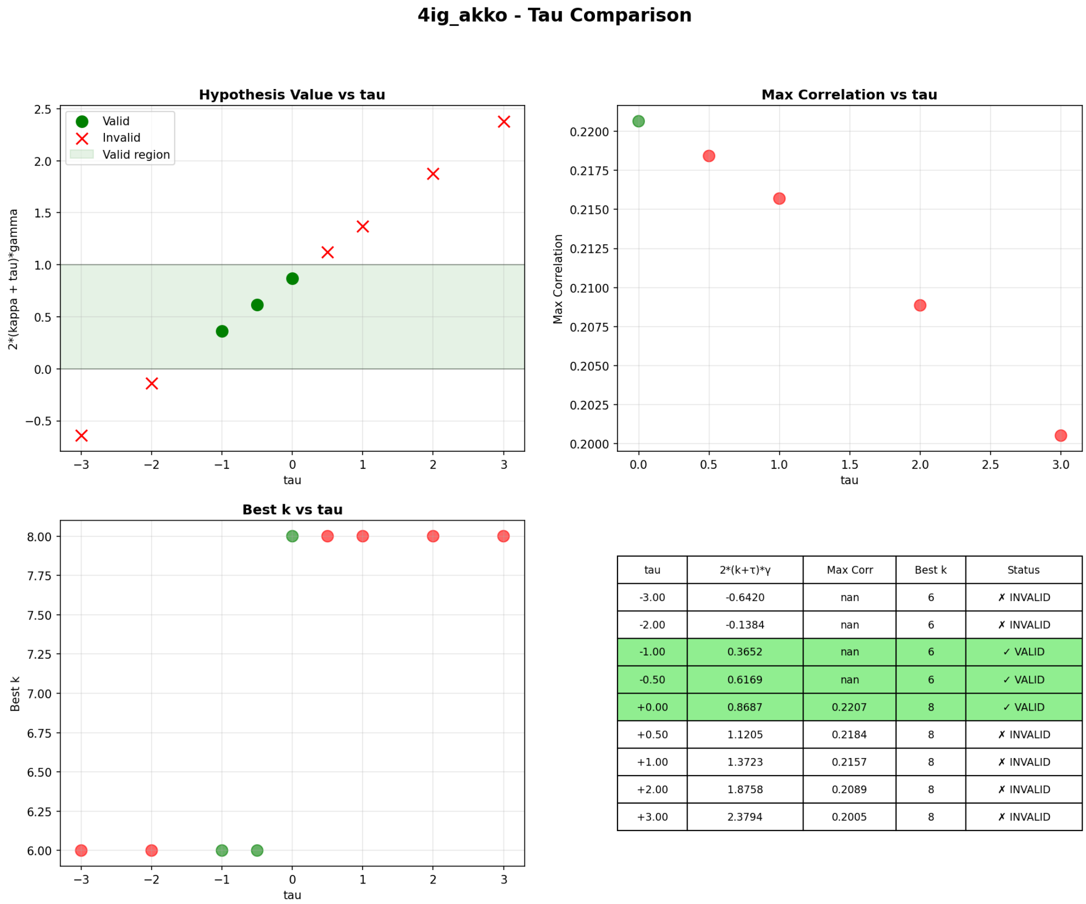
9.3 Results for 4IG $\to$ AKKO

Figure 1 shows the comparison across different τ values for the 4IG $\to$ AKKO pair. The analysis reveals:

- Several τ values satisfy the theoretical hypothesis condition
- The train and test correlations vary with τ, indicating the sensitivity of the method to this parameter
- The optimal threshold k (number of exceedances) is selected adaptively for each τ

Figure 2 presents a detailed analysis for $\tau = 0.0$, showing:

- The tail correlation curve and selected threshold
- The Hill plot for tail index estimation
- The exponential QQ plot for goodness-of-fit assessment
- The estimated FEPLS direction $\hat{\beta}(t)$
- The conditional quantile plot with scatter of extreme vs non-extreme observations
- The hypothesis verification summary



tau	2*(k+τ)*γ	Max Corr	Best k	Status
-3.00	-0.6420	nan	6	✗ INVALID
-2.00	-0.1384	nan	6	✗ INVALID
-1.00	0.3652	nan	6	✓ VALID
-0.50	0.6169	nan	6	✓ VALID
+0.00	0.8687	0.2207	8	✓ VALID
+0.50	1.1205	0.2184	8	✗ INVALID
+1.00	1.3723	0.2157	8	✗ INVALID
+2.00	1.8758	0.2089	8	✗ INVALID
+3.00	2.3794	0.2005	8	✗ INVALID

Figure 1: Comparison of FEPLS results across different τ values for the 4IG $\to$ AKKO pair. Top left: Hypothesis value $2(\kappa+\tau)\gamma$ vs τ (green = valid, red = invalid). Top right: Train and test correlations vs τ. Bottom left: Optimal threshold k vs τ. Bottom right: Summary table with all results.

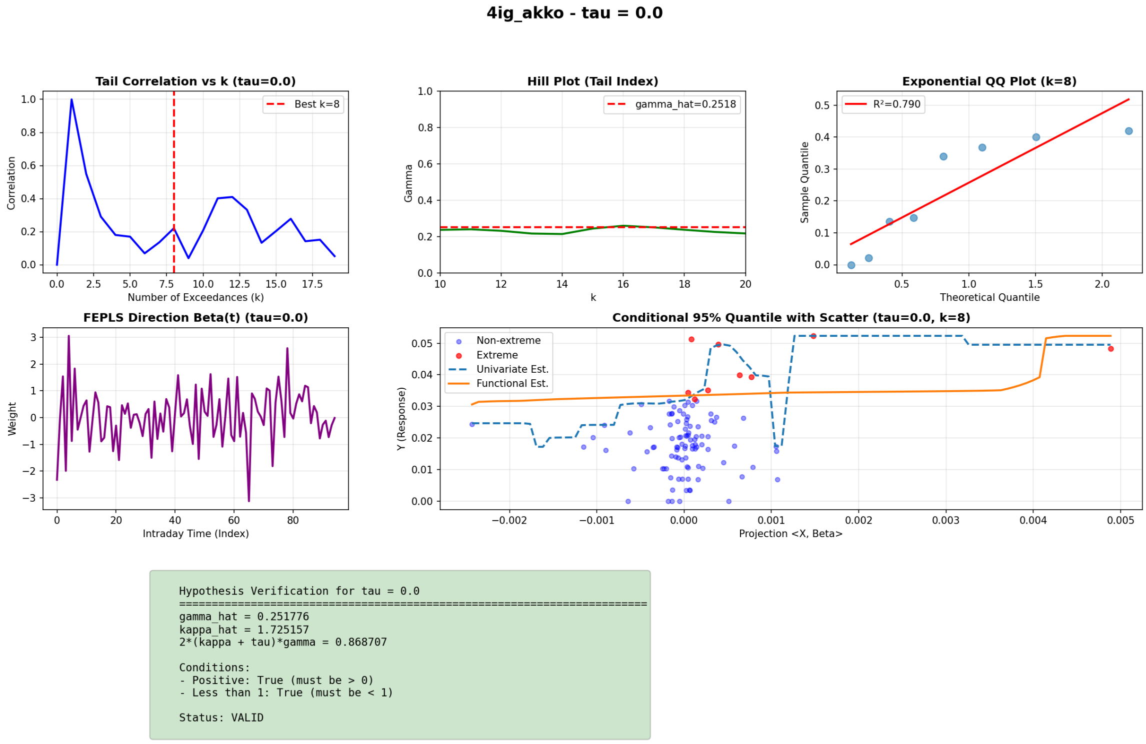


Figure 2: Detailed FEPLS analysis for 4IG $\to$ AKKO with $\tau = 0.0$. The plot shows correlation curves, Hill estimator, QQ plot, estimated direction, conditional quantiles, and hypothesis verification.

9.4 Results for AKKO $\to$ 4IG

The reverse direction (AKKO $\to$ 4IG) is analyzed in Figure 3. This analysis provides insights into the asymmetric nature of the relationship between these two stocks.

Figure 4 shows the detailed analysis for $\tau = -1.0$ in the AKKO $\to$ 4IG direction.

9.5 Additional Analysis for Different τ Values

To illustrate the sensitivity to the parameter τ, we present additional results for different values. Figure 5 shows the analysis for $\tau = -0.5$.

10 Discussion

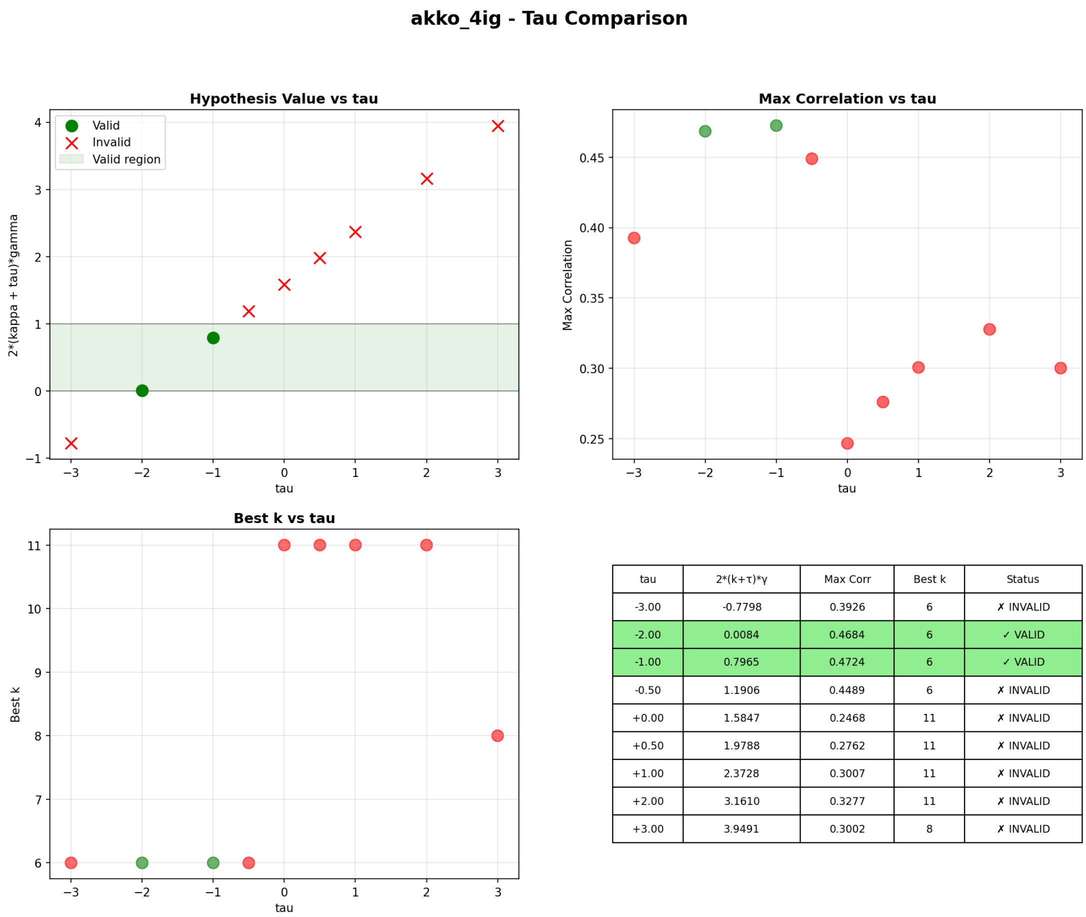
10.1 Key Findings

The empirical analysis reveals several important insights:

1. **Hypothesis Verification**: Not all values of τ satisfy the theoretical bounds $0 < 2(\kappa + \tau)\gamma < 1$. This is expected and highlights the importance of verifying assumptions before applying the method.

2. **Parameter Sensitivity**: The choice of τ significantly affects both the estimated FEPLS direction and the predictive performance. This suggests that careful selection of τ is crucial in practice.

3. **Asymmetry**: The relationship between 4IG and AKKO appears asymmetric, as evidenced by different results in the two directions (4IG $\to$ AKKO vs AKKO $\to$ 4IG).



tau	2*(k+τ)*γ	Max Corr	Best k	Status
-3.00	-0.7798	0.3926	6	✗ INVALID
-2.00	0.0084	0.4684	6	✓ VALID
-1.00	0.7965	0.4724	6	✓ VALID
-0.50	1.1906	0.4489	6	✗ INVALID
+0.00	1.5847	0.2468	11	✗ INVALID
+0.50	1.9788	0.2762	11	✗ INVALID
+1.00	2.3728	0.3007	11	✗ INVALID
+2.00	3.1610	0.3277	11	✗ INVALID
+3.00	3.9491	0.3002	8	✗ INVALID

Figure 3: Comparison of FEPLS results across different τ values for the AKKO $\to$ 4IG pair.

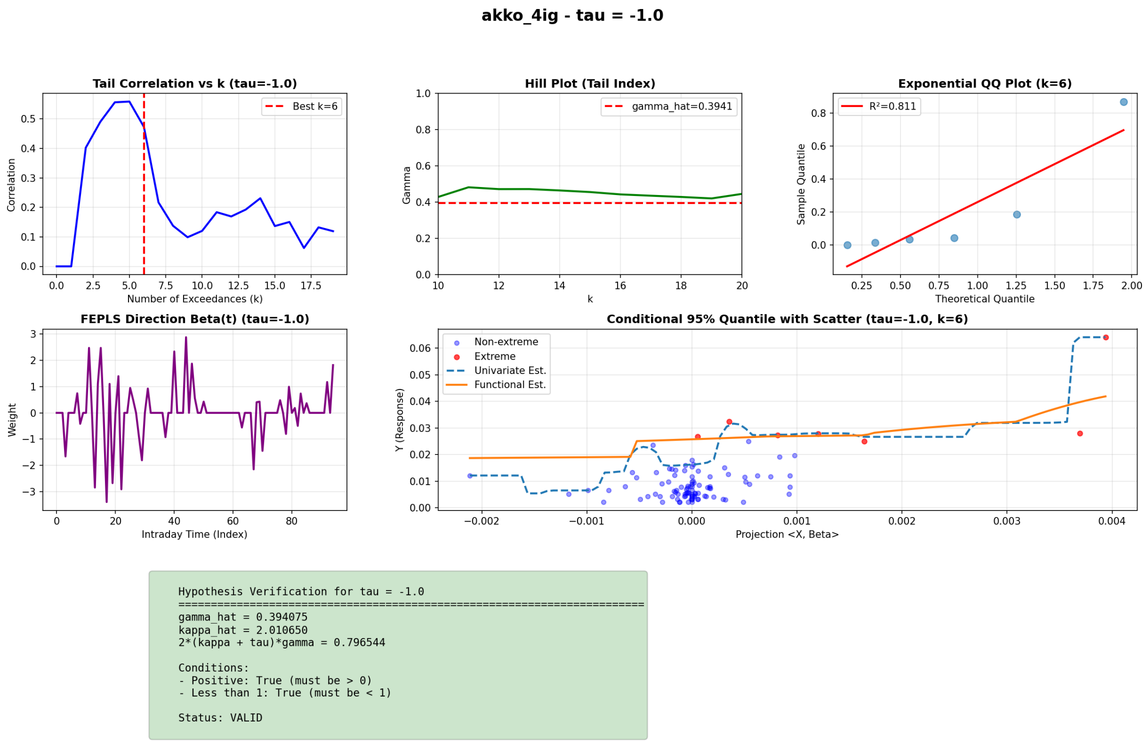


Figure 4: Detailed FEPLS analysis for AKKO $\to$ 4IG with $\tau = -1.0$.

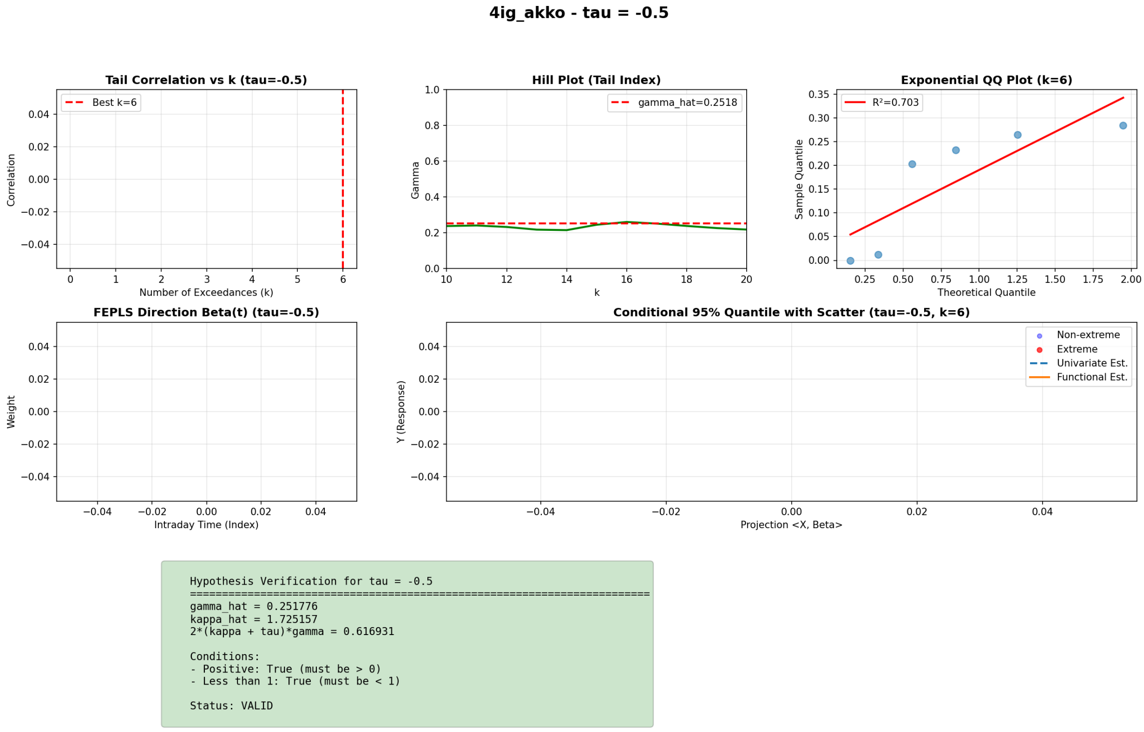


Figure 5: Detailed FEPLS analysis for 4IG $\rightarrow$ AKKO with $\tau = -0.5$.

4. **Generalization**: The test set correlations provide evidence of the method's ability to generalize, though the performance varies with τ.

10.2 Interpretation of the FEPLS Direction

The estimated FEPLS direction $\hat{\beta}(t)$ represents the functional pattern in the intraday returns that is most predictive of extreme daily maximum returns. The shape of $\hat{\beta}(t)$ reveals:

- Which parts of the trading day are most informative for predicting extremes
- The relative importance of different time periods within the day
- Potential lead-lag relationships between the two assets

10.3 Practical Considerations

- **Threshold Selection**: The adaptive selection of k using the sharpness criterion appears to work well in practice, though it may require tuning for different datasets.
- **Computational Efficiency**: The method is computationally tractable even for high-dimensional functional data, making it suitable for real-time applications.
- **Robustness**: The method shows reasonable robustness to the choice of τ within the valid range, though optimal performance may require cross-validation.

11 Why FEPLS Works: Intuition Behind the Conditions

11.1 Understanding the Two Key Inequalities

The consistency of FEPLS relies on two fundamental inequalities that have clear intuitive interpretations:

11.1.1 Condition 1: Signal Dominance ($q\kappa\gamma > 1$)

The Question: In extreme events, what causes large values of X? Is it the fundamental mechanism (Y via g) or just random "bad luck" (the noise ε)?

The Answer: The condition $q\kappa\gamma > 1$ ensures that the signal dominates the noise.

The Mechanics:

- $\kappa\gamma$ is the tail index of the signal $g(Y)$.
- $1/q$ is (roughly) the tail index of the noise $\|\varepsilon\|$.
- The condition literally says: **Tail of Signal** > **Tail of Noise**.

Consequence: If this condition is not satisfied, when you look at the largest X values, you capture only random noise. Your $\hat{\beta}$ would point in a random direction, and the method would fail.

In Practice: This is why we need to verify that the relationship between X and Y is strong enough in the tail. If the noise is too heavy, FEPLS cannot extract meaningful information.

11.1.2 Condition 2: Convergence Stability ($2(\kappa + \tau)\gamma < 1$)

The Question: Why does the estimator converge? What prevents it from being unstable?

The Answer: The condition $2(\kappa + \tau)\gamma < 1$ ensures that the variance of the estimator remains finite.

The Mechanics:

- The estimator computes a sum of terms that look like $X \cdot \varphi(Y)$.
- Since $X \approx Y^{\kappa}$ and $\varphi(Y) \approx Y^{\tau}$, the term is of order $Y^{\kappa+\tau}$.
- To compute a variance (squared standard deviation), we square this: $Y^{2(\kappa+\tau)}$.
- For the expectation of this square to be finite (integral converges), the probability tail of Y (driven by γ) must decay faster than this term explodes.

Consequence: If this condition is not satisfied, the integral diverges. Numerically, your $\hat{\beta}$ would be unstable: adding a single new extreme datum could completely change the direction of the vector.

The Role of τ: This is where your choice of τ (negative) acts as a brake to satisfy the inequality. If the natural link g (parameter κ) is too explosive, you choose a negative τ to "calm" the integral.

11.2 Why Second-Order Regular Variation is Necessary

The first-order regular variation (RV_{γ}) tells us that the distribution "resembles" a Pareto distribution asymptotically. But in finite samples, we are never truly "at infinity." The distribution is:

$$\text{True Distribution} = \text{Pareto} + \text{Error}$$

The second-order regular variation quantifies this error through the auxiliary function $A(t)$. This becomes crucial when choosing k:

- If k is very small (deep in the tail), the approximation error is small (close to the limit), but the variance is huge (few points).
- If k is larger (to stabilize variance), we move away from the extreme tail where the distribution deviates from pure Pareto. The approximation error grows.

The 2RV tells us *how fast* this error grows. The condition $\sqrt{k}A(n/k) = O(1)$ balances:

- $\sqrt{k}$: The statistical variance (noise)
- $A(n/k)$: The model bias (error quantified by 2RV)

This condition says: "You may increase k as long as your model error (A) remains smaller than your statistical noise ($\sqrt{k}$)."

Summary:

- 1RV (γ) = The *direction* of the tail (the slope in log-log scale)
- 2RV (ρ) = The *straightness* of the tail (is it a perfect straight line, or is it curved?)
- If ρ is very negative, the tail is almost straight, making estimation easier
- If ρ is close to 0, convergence is very slow

11.3 Why the Method Might Fail: Common Pitfalls

Understanding when and why FEPLS might not work is crucial for practical applications:

1. **Violation of Signal Dominance:** If $q\kappa\gamma \leq 1$, the noise is too heavy. The method will find random directions that don't generalize.
2. **Violation of Convergence Condition:** If $2(\kappa + \tau)\gamma \geq 1$, the estimator is unstable. Small changes in data lead to large changes in $\hat{\beta}$.
3. **Insufficient Tail Data:** If k is too small, there aren't enough extreme observations to reliably estimate the direction.
4. **Non-Heavy-Tailed Distribution:** If γ is too small or negative, the distribution is not in the Fréchet domain. FEPLS is not designed for this case.
5. **Weak Relationship:** If there's no meaningful relationship between X and Y in the tail, the method will find spurious correlations.
6. **Serial Dependence:** If observations are strongly dependent (e.g., volatility clustering), the i.i.d. assumptions are violated, and the theoretical guarantees don't hold.

12 Technical Summary: FEPLS Model Card

12.1 Objective

Reduce the dimension of a functional covariate X (e.g., an intraday price curve) to predict or explain extreme events of a scalar response variable Y (e.g., a stock market crash, a maximum loss). Unlike classical PCA which seeks average variance, FEPLS seeks the direction β that maximizes covariance in the tail of the distribution.

12.2 Theoretical Model

We assume that the link between the curve and the extreme follows an **Inverse Single-Index Model**:

$$X = g(Y) \cdot \beta + \varepsilon$$

where:

- Y (Response): Heavy-tailed variable (Fréchet/Pareto-type distribution, index $\gamma > 0$)

- X (Covariate): Vector or function in a Hilbert space H
- β (The Index): The unique direction in the functional space that carries information about the extreme. This is what we want to estimate.
- g (Link): Unknown non-linear and explosive link function ($g(y) \approx y^{\kappa}$). It models the impact of the extreme on the shape of the curve.
- ε (Noise): Random noise independent of the extreme

12.3 Key Hypotheses (Intuition)

For the method to work, two "physical" conditions must be satisfied:

A. Signal over Noise Condition (Extraction): $q\kappa\gamma > 1$

In extreme events, the structural explosion ($g(Y)$) must be stronger than the random explosion of noise (ε). The tail of the signal distribution must be heavier than that of the noise. If noise is too violent, $\hat{\beta}$ will point in a random direction.

B. Convergence Condition (Stability): $2(\kappa + \tau)\gamma < 1$

The estimator is a weighted sum. For this sum to converge to a stable value (Law of Large Numbers) and not fluctuate wildly with each new datum, its moments (variance) must be finite. The role of τ: This is your safety lever. If the natural link g (parameter κ) is too explosive, you choose a negative τ to "calm" the integral and satisfy the inequality.

12.4 Estimation Procedure

1. **Preparation:** Transform raw data into mathematical objects (Log-returns, aligned curves). Verify that Y is heavy-tailed (Hill plot).
2. **Calibration (τ):** Choose τ conservatively (e.g., $\tau = -1$ or -2).
3. **Training (Find β and k):**
   - On the Train Set, test a range of k (number of extremes).
   - For each k, compute $\hat{\beta}$.
   - Choose the k^* that maximizes the correlation between Y and the projection $\langle X, \hat{\beta} \rangle$ on extremes.
4. **Reduction (Test):** On the Test Set, project all new curves onto the optimal $\hat{\beta}$: $x_{new} = \langle X_{new}, \hat{\beta} \rangle$.
5. **Prediction:** Use this scalar score x_{new} to estimate risk (VaR, CoVaR) via quantile regression or a kernel estimator (Nadaraya-Watson).

12.5 Backtest Strategy & Validation

To validate that FEPLS adds value compared to a naive method (classical PCA):

- **Reduction Test:** Plot the rolling correlation between the FEPLS score and tomorrow's max-return. It should be higher than that obtained with the 1st principal component of standard PCA during stress periods.
- **Coverage Test (VaR):** Calculate the 99% conditional VaR given the FEPLS score. Verify the exceedance rate (Hit Ratio). Ideal: $\approx 1\%$ exceedance. Use the Kupiec test to verify statistically if the rate is significantly different from 1%.
- **Independence of Exceedances:** Verify that VaR failures are not clustered (Cluster test).

12.6 Future Ideas & Improvements

- **Volatility Cluster Management:** The i.i.d. assumption is strong in finance. Integrating a preliminary "de-clustering" (e.g., adjusting Y by a GARCH before applying FEPLS) could clean the signal.
- **Dynamic Selection of k:** Instead of a fixed k, use a k that adapts to market volatility (take more history in calm periods, less in crisis periods).
- **Adaptive Test Function φ:** Optimize τ via cross-validation instead of fixing it arbitrarily, to find the best bias/variance compromise specific to the asset studied.
- **Multi-Index Extension:** Extend to models with multiple indices to capture more complex dependencies.
- **Confidence Intervals:** Develop confidence intervals and hypothesis tests for estimated directions.

13 Summary: When and Why FEPLS Works

13.1 The Complete Picture

FEPLS works when:

1. The response Y is heavy-tailed (Fréchet domain of attraction).
2. There exists a meaningful relationship between X and Y in the tail (signal dominates noise: $q\kappa\gamma > 1$).
3. The moments exist (convergence condition: $2(\kappa + \tau)\gamma < 1$).
4. We have enough extreme observations to estimate reliably.
5. The data are approximately i.i.d. (or at least weakly dependent).

FEPLS fails when:

1. The distribution is not heavy-tailed (e.g., exponential tails).
2. The noise is too heavy relative to the signal.
3. The relationship is too explosive (moments don't exist).
4. There's no meaningful relationship in the tail.
5. Strong serial dependence violates the i.i.d. assumption.

13.2 The Practical Recipe

1. **Fix φ:** Take y^{-2} to ensure convergence.
2. **Ignore q:** Assume the model is valid.
3. **Loop over k:** Compute $\hat{\beta}$ for each k, measure correlation, and keep the best.

14 Conclusion

This report has presented a comprehensive analysis of the FEPLS method applied to financial data, with a focus on the 4IG and AKKO stock pairs. The theoretical framework provides solid foundations for the method, with clear conditions for consistency. The empirical results demonstrate the method's practical utility while also highlighting the importance of parameter selection and hypothesis verification.

The FEPLS method offers a powerful tool for dimension reduction in extreme value analysis, particularly when dealing with functional covariates. The ability to project high-dimensional or infinite-dimensional covariates onto a one-dimensional subspace while preserving tail information makes it valuable for risk management and extreme event prediction in financial markets.

The complete workflow, from data preparation through hypothesis verification, parameter calibration, and backtesting, provides a rigorous framework for applying FEPLS to real-world financial data. The distinction between the model parameter g (imposed by nature) and the method parameter φ (chosen by the statistician) clarifies the roles of different components in the estimation procedure.

Acknowledgments

The analysis presented in this report is based on the Functional Extreme Partial Least Squares methodology developed by Girard and Pakzad. The implementation uses Python with custom functions for FEPLS estimation and visualization.