

# Functional Extreme Partial Least Squares: Application to Financial Data Analysis

Janis Aiad  
Master MVA - Geometric Data Analysis  
ENS Paris Saclay, Paris, France

December 16, 2025

## **Abstract**

This report presents an empirical analysis of the Functional Extreme Partial Least Squares (FEPLS) method applied to financial time series data. We conduct two complementary analyses: (1) a cross-asset analysis focusing on pairs of stocks from the Hungarian market, specifically examining the relationship between 4IG and AKKO stocks, and (2) a subsampling-based high-frequency analysis of AAPL stock data exploring multiple temporal resolutions from milliseconds to minutes. Both analyses include hypothesis verification for different values of the parameter  $\tau$ , visualization of the FEPLS direction estimation, and assessment of the method's performance. 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

Extreme events—rare, significant deviations from normal behavior—play a crucial role in finance, signal processing, risk management, and many scientific fields. The way such extremes manifest in data, especially when those data are high-dimensional or functional (such as entire intraday price curves), presents substantial challenges both for modeling and for prediction. Conventional dimension reduction tools like Principal Component Analysis (PCA) or classical Partial Least Squares (PLS) are effective for understanding average behavior but may fail to capture the patterns associated specifically with rare, high-impact events.

Functional Extreme Partial Least Squares (FEPLS) is a methodology developed to address precisely this challenge. It extends the classical PLS approach into the domain of extremes, targeting the features of high-dimensional or infinite-dimensional covariates that are most informative for explaining or predicting the occurrence of extreme events in a response variable.

The central research question motivating this work is the following:

*What is the most likely shape of the covariate  $X$  when the response  $Y$  is extreme?*

Answering this question, for instance, can reveal which specific features of a price curve tend to occur in the lead-up to major market moves—a task of paramount importance in financial engineering, but also relevant in fields like signal processing and detection theory, where characterization of rare patterns is required.

In this report, we offer both a theoretical and comprehensive empirical study of the FEPLS framework. We begin by unpacking the intuition and mathematical underpinnings of FEPLS, clarifying how it focuses on conditional extremes, and describing the signal-to-noise assumptions

required for consistent estimation. Our goal is to provide a clear narrative that connects the theory to practical considerations.

To bridge theory and practice, we conduct extensive experiments on open financial datasets. These analyses illuminate the strengths and limitations of FEPLS in realistic, noisy, heavy-tailed environments. We also move beyond the day-to-day context of classical financial studies, exploring high-frequency data using a subsampling-based methodology. This allows us to probe FEPLS performance and the validity of its foundational assumptions across a variety of time scales and predictive settings.

Our contributions in this work include:

- The first empirical exploration of FEPLS on high-frequency financial data, investigating time scales from milliseconds to minutes.
- A thorough assessment of the key statistical assumptions of FEPLS, specifically regarding signal-to-noise ratios and estimator stability, in real-world data.
- A detailed empirical evaluation of FEPLS under challenging market noise, providing clear illustrations of when and where the method succeeds or faces limitations.
- Practical guidance for the application and calibration of FEPLS, including choices around key parameters (notably  $\tau$ ) and computational issues in high dimensions.

Through this combination of theoretical insight, methodological clarification, and practical experimentation, we aim to both advance the understanding of FEPLS and provide a resource for practitioners seeking to apply it in domains where extremes govern the risk landscape.

## 2 Notation and Mathematical Framework

### 2.1 Notation

## 3 Intuitive Understanding of the FEPLS Framework

### 3.1 Difference Between PLS and PCA

The fundamental distinction between Partial Least Squares (PLS) and Principal Component Analysis (PCA) lies in their use of supervision:

#### PCA (Unsupervised):

- **No response variable:** PCA operates solely on the covariate data  $X$ , ignoring any relationship with a response variable  $Y$ .
- **Objective:** Finds directions that maximize the *variance* of the projected data, i.e.,  $\|w\|=1 \text{Var}(\langle w, X \rangle)$ .
- **Interpretation:** The first principal component captures the direction of maximum variance in  $X$ , regardless of whether this variance is relevant for predicting  $Y$ .
- **Limitation:** In regression contexts, the directions with highest variance in  $X$  may not be the most predictive of  $Y$ . For example, if  $X$  has high variance in directions orthogonal to the relationship with  $Y$ , PCA will prioritize these irrelevant directions.

#### PLS (Supervised):

Column 1		Column 2	
Symbol	Explanation	Symbol	Explanation
<b>Regular Variation &amp; Auxiliary Functions</b>			
$\text{RV}_\tau(+\infty)$	Regularly-varying (index $\tau$ )	$2\text{RV}_{\gamma,\rho}(+\infty)$	Second-order RV
$H_\rho(y)$	$y^\gamma \int_1^y u^{\rho-1} du$	$A(t)$	Auxiliary function, $A(t) \rightarrow 0$
<b>Model and FEPLS Notation</b>			
$X = g(Y)\beta + \varepsilon$	Inverse regression model	$\beta \in H, \ \beta\  = 1$	Index vector (unit norm)
$g \in \text{RV}_\kappa(+\infty)$	Link function (index $\kappa$ )	$\varepsilon : \Omega \rightarrow H$	Noise term
$\varphi \in \text{RV}_\tau(+\infty)$	Test function (index $\tau$ )	$w_\varphi(y)$	Theoretical FEPLS direction
$\hat{\beta}_\varphi(y)$	Empirical FEPLS estimator		
<b>Key Parameters</b>			
$\gamma \in (0, 1)$	Tail index of $Y$	$\rho \leq 0$	Second-order parameter
$\kappa > 0$	Link function index	$\tau \in \mathbb{R}$	Test function index
$k = k_n$	Intermediate sequence	$y_{n,k} \sim U(n/k)$	High threshold
$\delta_{n,k}$	$(g(y_{n,k})(k/n)^{1/q})^{-1}$	$q > 2$	Noise integrability order

Table 1: Summary of main notation used in the FEPLS framework.

- **Uses response variable:** PLS explicitly incorporates the response  $Y$  in the optimization.
- **Objective:** Finds directions that maximize the *covariance* between the projection and the response, i.e.,  $\|w\|=1 \text{Cov}(\langle w, X \rangle, Y)$ .
- **Interpretation:** The PLS direction captures the direction in  $X$  that is most correlated with  $Y$ , making it directly relevant for prediction.
- **Advantage:** By focusing on covariance rather than variance, PLS identifies directions that are both informative about  $X$  and predictive of  $Y$ .

**FEPLS (Extreme PLS):** FEPLS extends this supervised approach to the extreme regime. The fundamental idea 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?

### 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  $\varepsilon$ .

## 4 Theoretical Framework

### 4.1 The FEPLS Problem

Suppose we observe data pairs  $(Y, X)$ , where  $Y$  is a real-valued variable we care about (for example, the size of a market move), and  $X$  contains other observable information (for example, a time series or a function we record alongside  $Y$ ). The central question is: **Given  $X$  and  $Y$ , can we find a summary of  $X$  that best explains the most extreme values of  $Y$ ?**

FEPLS answers this by searching for a direction  $w$  in the space of possible summaries of  $X$  that is most strongly related to the largest values of  $Y$ . Formally, for a high threshold  $y$ , we solve:

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

where  $\langle w, X \rangle$  simply means forming a linear combination of  $X$  using  $w$  (for instance, by integrating  $w(t)X(t)$  if  $X$  is a function of  $t$ ), and the Cov is computed only among the cases where  $Y$  exceeds  $y$ —that is, focusing only on the extremes.

### 4.2 The Inverse Single-Index Model

To analyze and understand the behavior of this procedure, we will **suppose** a simple mathematical relationship between  $X$  and  $Y$  as follows:

$$X = g(Y)\beta + \varepsilon, \quad \beta \in H, \quad \|\beta\| = 1, \quad (2)$$

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

#### The Distinction Between $g$ and $\varphi$ : Model vs. Method

A fundamental distinction exists between the link function  $g$  and the test function  $\varphi$ :  **$g$  and  $\kappa$  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, the square root law of market impact writes  $g(y) = y^{0.5}$

**$\varphi$  and  $\tau$  is chosen (It's your tuning tool):**

- $\varphi$  is part of the estimator ( $\hat{\beta}_\varphi$ ), we, future statisticians, adjusts to optimize estimation
- It does not exist in nature; it is an “instrumental” or “test” function introduced to capture information

### Moment Condition and Tuning of $\tau$

The first key condition,  $0 < 2(\kappa + \tau)\gamma < 1$ , ensures moments exist for reliable estimation. If  $g$  (large  $\kappa$ ) grows too fast, sums like  $\sum X_i$  may diverge; picking a decreasing  $\varphi$  (negative  $\tau$ ) counteracts this, keeping sums finite.

The estimator uses weights  $\varphi(Y_i)$  for extremes, e.g.

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

so tuning  $\tau$  adjusts influence: big  $\tau$  (positive) focuses on the deepest extremes (low bias, high variance); small or negative  $\tau$  spreads weight (better stability).

### 4.3 Second-Order Regular Variation

A new assumption is 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 4.1** (Second-Order Regular Variation). The function  $U$  belongs to  $2RV_{\gamma,\rho}(+\infty)$  if there exist  $\gamma \in (0, 1)$ ,  $\rho \leq 0$  and an auxiliary function  $A$  ultimately of constant sign with  $A(t) \rightarrow 0$  as  $t \rightarrow +\infty$  such that:

$$\lim_{t \rightarrow +\infty} \frac{1}{A(t)} \left( \frac{U(ty)}{U(t)} - y^\gamma \right) = y^\gamma H_\rho(y) := y^\gamma \int_1^y u^{\rho-1} du, \quad y > 0. \quad (3)$$

**A variation.** Although the original paper does not state it explicitly, this definition immediately yields that  $A$  is of the form  $A(t) = t^\rho$  which is important to derive convergence scaling law in (??).

#### Interpretation of $\rho$ and the Role of 2nd-order Regular Variation

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

The first-order regular variation ( $RV_\gamma$ ) 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 ( $\gamma$ ) gives the *direction* of the tail (the slope), while 2RV ( $\rho$ ) gives the *straightness* of the tail (is it a perfect straight line in log-log scale, or is it curved?). If  $\rho$  is very negative, the tail is almost straight, making estimation easier. If  $\rho$  is close to 0, convergence is very slow.

## 4.4 Signal Dominance over Noise

The tail of  $g(Y)$  (signal channel) must be heavier than the tail of the noise  $\varepsilon$ . In practical terms, this ensures that among the largest observed  $Y$  values, the associated signal  $g(Y)\beta$  is much more extreme than the noise, so the extreme behavior is informative about  $\beta$ .

**Why necessary?** If the noise’s tail is too fat, its fluctuations swamp the regression signal in the extremes, making recovery impossible even with infinite data. *If  $\kappa$  increases (sharper growth), you can choose larger  $k$  and variance decreases.*

That is why the model is only identifiable if the signal is dominant over the noise, under the inequality

$$q\kappa\gamma > 1. \quad (4)$$

We call this condition **signal dominance for identifiability**.

Haan2007

## 4.5 What do we estimate to ensure consistency ?

To ensure that we are into the consistency inequality that allows signal recovery, we need to estimate both  $\gamma$  and  $\rho$ .  $\kappa$  cannot be estimated from the data.

so there is no best tau choice a priori, we need to construct estimator for a lot of different taus and find an interval that contains some stationarity in the beta estimate. We call this procedure **tau tuning** to satisfy the **integrability condition**.

### 4.5.1 Choice of $k$

When testing a particular tau, we need to choose a number of extreme observations  $k$  to minimize the bias-variance tradeoff. We call this procedure **k tuning** for **convergence**.

Once you have chosen proper tau, estimated rho and gamma, you have chosen the number of observations to use, you have chosen k.

It remains that you have performed your FEPLS estimation, which convergence guarantee do you have ?

## 4.6 Convergence scaling law

Under the assumption that the model is identifiable (which we recall can not be the case in practice)

The convergence scaling law is given by:

$$\delta_{n,k} = \left( g(y_{n,k}) \left( \frac{k}{n} \right)^{1/q} \right)^{-1} \quad (5)$$

Since  $g \in \text{RV}_\kappa(+\infty)$  and  $y_{n,k} \sim U(n/k)$  with  $U \in \text{RV}_\gamma(+\infty)$ , we have asymptotically:

$$\delta_{n,k} \sim \left( \frac{k}{n} \right)^{\kappa\gamma-1/q} \quad (6)$$

le but est de recouvrir complètement la fonction  $w(y)$  pas seulement une partie donc à un  $y$  fixé on a un estimateur

**Choix du seuil et croissance de  $y$ .** Pour chaque seuil  $y \geq 0$ , on peut définir l'estimateur de type FEPLS

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

L'estimateur  $\hat{\beta}_\varphi(y)$  est ainsi une combinaison linéaire des  $X_i$  dont les  $Y_i$  sont situés dans la queue de  $\bar{F}$ , et la fonction de test  $\varphi$  règle le poids attribué à chaque observation extrême.

Dans le résultat de consistance, on ne fixe toutefois pas un seuil  $y$  une fois pour toutes. On considère une suite de seuils (déterministes)  $(y_{n,k})$  telle que

$$y_{n,k} \sim U\left(\frac{n}{k}\right), \quad n \rightarrow \infty,$$

où  $U$  est la fonction quantile de queue de  $Y$ . L'analogue empirique de  $U(n/k)$  est la statistique d'ordre  $Y_{n-k+1,n}$ , de sorte que

$$Y_{n-k+1,n} \approx y_{n,k} \sim U\left(\frac{n}{k}\right) \quad \text{lorsque } n \rightarrow \infty.$$

Si la queue de  $Y$  est régulièrement variable, c'est-à-dire si  $\bar{F}(y) = \mathbb{P}(Y > y) \in \text{RV}_{-1/\gamma}(+\infty)$ , alors la quantile de queue vérifie

$$U(t) \sim C t^\gamma \quad (t \rightarrow \infty)$$

pour une constante  $C > 0$  (par exemple, pour une Pareto standard,  $U(t) \sim t^\gamma$ ). En prenant  $t = n/k$ , on obtient

$$y_{n,k} \sim U\left(\frac{n}{k}\right) \sim C \left(\frac{n}{k}\right)^\gamma.$$

On travaille avec une suite intermédiaire  $k = k_n \rightarrow \infty$  telle que  $k_n/n \rightarrow 0$ . Dans ce régime, on a  $n/k_n \rightarrow \infty$  et donc

$$y_{n,k_n} \sim C \left(\frac{n}{k_n}\right)^\gamma \xrightarrow[n \rightarrow \infty]{} +\infty.$$

Autrement dit, le seuil  $y$  utilisé par l'estimateur croît avec  $n$  et  $k_n$  : on suit la trajectoire

$$y = Y_{n-k+1,n} \approx U\left(\frac{n}{k}\right) \uparrow +\infty,$$

ce qui garantit que l'on explore progressivement la queue de la loi de  $Y$  tout en conservant un nombre croissant d'observations extrêmes.

**Consistance de l'estimateur FEPLS.** Sous le modèle inverse à indice unique  $X = g(Y)\beta + \varepsilon$  et les hypothèses de variation régulière de  $Y$ , on fixe un ordre d'intégrabilité  $q > 2$  pour le bruit  $\varepsilon$ , indépendant de  $n$ , et l'on impose les inégalités

$$0 < 2(\kappa + \tau)\gamma < 1, \quad q\kappa\gamma > 1,$$

où  $\gamma$  est l'indice de queue de  $Y$ ,  $\kappa$  l'indice de variation régulière de  $g$  et  $\tau$  celui de la fonction de test  $\varphi$ . On suppose en outre que  $\varphi$  et  $g$  sont continûment dérивables au voisinage de l'infini, avec

$$\frac{t(\varphi g)'(t)}{(\varphi g)(t)} \longrightarrow \tau + \kappa \neq 0 \quad \text{lorsque } t \rightarrow +\infty.$$

On définit alors

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

Sous ces hypothèses, le théorème de consistance s'écrit

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

En particulier, lorsque  $U$  et  $g$  satisfont des conditions de seconde ordre, on montre que  $\delta_{n,k}$  est de l'ordre

$$\delta_{n,k} = O\left(n^{(1/q-\gamma\kappa)/(1-2\rho)}\right),$$

ce qui fait apparaître explicitement l'effet de l'indice de queue  $\gamma$ , de l'indice de lien  $\kappa$  et de l'ordre d'intégrabilité  $q$  du bruit dans la vitesse de convergence de l'estimateur FEPLS.

Enfin, la condition technique classique en théorie des valeurs extrêmes

$$\sqrt{k} A\left(\frac{n}{k}\right) = O(1)$$

fait intervenir la fonction auxiliaire  $A$  issue de la seconde variation régulière de  $U$  (ou de  $\bar{F}$ ). Lorsque  $A$  est elle-même régulièrement variable avec indice  $\rho < 0$ , i.e.  $A(t) \in RV_\rho(+\infty)$ , cette condition est satisfaite dès que l'on choisit une suite intermédiaire  $k = k_n$  telle que

$$k_n \sim c n^{-2\rho/(1-2\rho)} \quad (n \rightarrow \infty),$$

pour une constante  $c > 0$ . Ce choix équilibre la décroissance de  $A(n/k)$  et la croissance de  $\sqrt{k}$ , et garantit la validité de l'approximation de second ordre utilisée dans l'étude asymptotique de  $\delta_{n,k}$ .

pour rho fixé le k optimal est fixé ce qui est cohérent avec ce que l'on observe at the edge of stability when unsure like in finance it should be

## 5 Consistency Results

The FEPLS estimator is defined as:

$$\hat{\beta}_\varphi(y) := \frac{\hat{v}_\varphi(y)}{\|\hat{v}_\varphi(y)\|} \quad \text{with} \quad \hat{v}_\varphi(y) = \frac{1}{n} \sum_{i=1}^n X_i \varphi(Y_i) \mathbf{1}_{\{Y_i \geq y\}}. \quad (7)$$

**Theorem 5.1** (Consistency of FEPLS Estimator). *Under the hypotheses stated in the previous section, the FEPLS estimator is consistent:*

$$\|\hat{\beta}_\varphi(Y_{n-k+1,n}) - \beta\| = O_{\mathbb{P}}(\delta_{n,k}) \xrightarrow[n \rightarrow +\infty]{} 0, \quad (8)$$

where  $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$  is the convergence rate.

The convergence rate  $\delta_{n,k}$  is a regularly-varying function of  $k/n$  with index  $\gamma\kappa - 1/q > 0$ . When the bias control condition  $\sqrt{k}A(n/k) = O(1)$  is satisfied, which typically requires  $k \sim c \cdot n^{-2\rho/(1-2\rho)}$  for some  $c > 0$  when  $\rho < 0$ , the convergence rate is exactly of order:

$$\delta_{n,k} = O\left(n^{(1/q-\gamma\kappa)/(1-2\rho)}\right). \quad (9)$$

The key conditions for consistency are: (1) signal dominance  $q\kappa\gamma > 1$  ensuring the signal tail is heavier than noise, and (2) convergence stability  $2(\kappa + \tau)\gamma < 1$  ensuring finite variance.

The convergence rate depends only on the tail distribution of  $Y$  (via  $\gamma$  and  $\rho$ ), the integrability order of the noise  $\varepsilon$  (via  $q$ ), and the asymptotic behavior of the link function  $g$  (via  $\kappa$ ). Detailed proofs and technical lemmas are omitted here; see [1] for complete theoretical results.

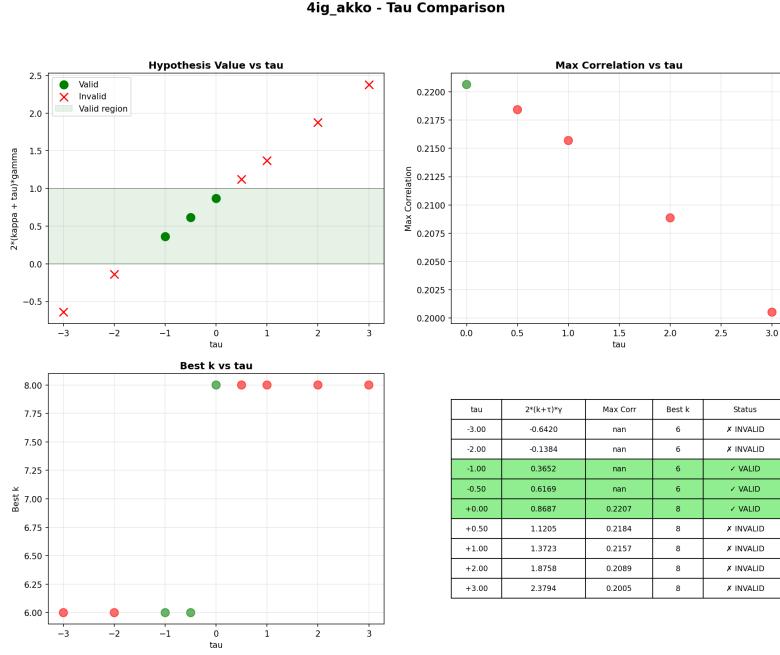


Figure 1: Empirical validation of FEPLS consistency: comparison across different  $\tau$  values showing hypothesis verification, correlation analysis, and optimal threshold selection for the 4IG  $\rightarrow$  AKKO pair.

## 6 Estimation Algorithm and Statistical Workflow

### 6.1 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.

## 6.2 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
- $\tau$ : Exponent of the test function (e.g.,  $-2$ )

**Algorithm:**

- Find the extreme threshold:** Sort  $Y$  in descending order. The threshold  $y_{thresh}$  is the  $k$ -th largest value of  $Y$ .
- 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).
- Compute weights (The “Non-Linear” part):** For each extreme observation  $i$ , compute the scalar weight:  

$$w_i = \varphi(Y_i) = (Y_i)^\tau$$
- 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.)

- Normalization:** Divide the vector  $V_{raw}$  by its Euclidean ( $L^2$ ) norm to obtain a unit vector:

$$\hat{\beta} = \frac{V_{raw}}{\|V_{raw}\|}$$

### 6.3 Parameter Selection: $\tau$ , $q$ , and $k$

A common question is: “Do I need to optimize all parameters?” Here’s the practical answer:

#### 6.3.1 For $\tau$ (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  $\kappa$  and  $\gamma$  initially, taking a negative  $\tau$  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  $\tau$ , as long as it’s reasonable.
- **Recommendation:** Start with  $\tau = -1$  or  $-2$ , then test a few values around it to verify robustness.

#### 6.3.2 For $q$ (Noise Integrability Order)

**NO, you do not estimate it.**

- **$q$  is theoretical:**  $q$  represents the integrability of the noise  $\varepsilon$  (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.

### 6.3.3 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  $k$ )

## 6.4 Complete Statistical Workflow

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

### 6.4.1 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.

### 6.4.2 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  $\gamma$  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.

### 6.4.3 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  $\tau$  (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  $\tau$  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.

### 6.4.4 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  $\alpha\%$  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).

## 7 Empirical Analysis: 4IG and AKKO Stock Pairs

### 7.1 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 → AKKO and AKKO → 4IG

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

### 7.2 Methodology

For each pair, we:

1. Estimate the tail index  $\gamma$  and the link function parameter  $\kappa$  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  $\tau$ , 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

### 7.3 Results for 4IG → AKKO

Figure ?? shows the comparison across different  $\tau$  values for the 4IG → AKKO pair. The analysis reveals:

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

Figure ?? 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

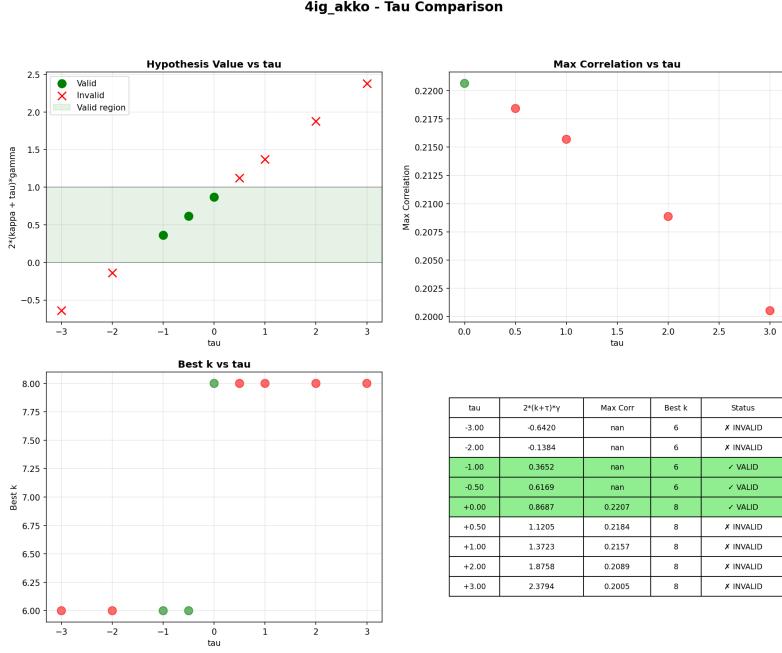


Figure 2: Comparison of FEPLS results across different  $\tau$  values for the 4IG  $\rightarrow$  AKKO pair. Top left: Hypothesis value  $2(\kappa + \tau)\gamma$  vs  $\tau$  (green = valid, red = invalid). Top right: Train and test correlations vs  $\tau$ . Bottom left: Optimal threshold  $k$  vs 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

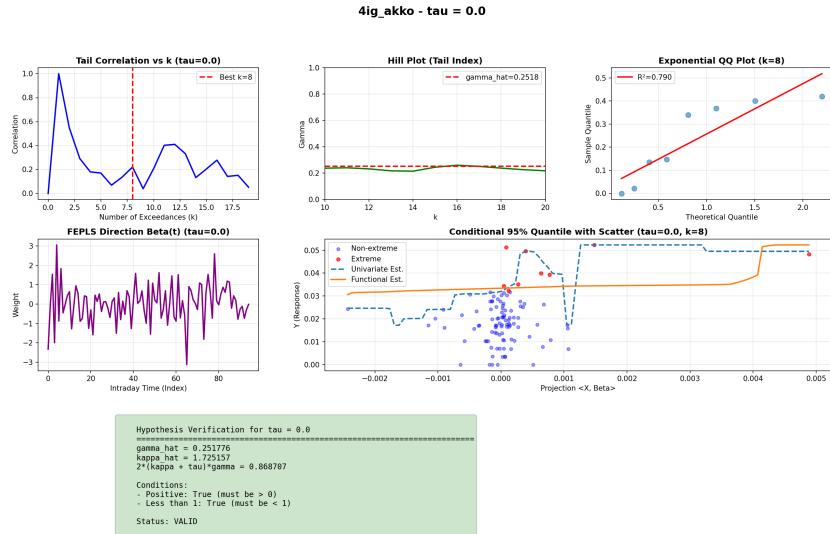


Figure 3: Detailed FEPLS analysis for 4IG  $\rightarrow$  AKKO with  $\tau = 0.0$ . The plot shows correlation curves, Hill estimator, QQ plot, estimated direction, conditional quantiles, and hypothesis verification.

## 7.4 Results for AKKO → 4IG

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

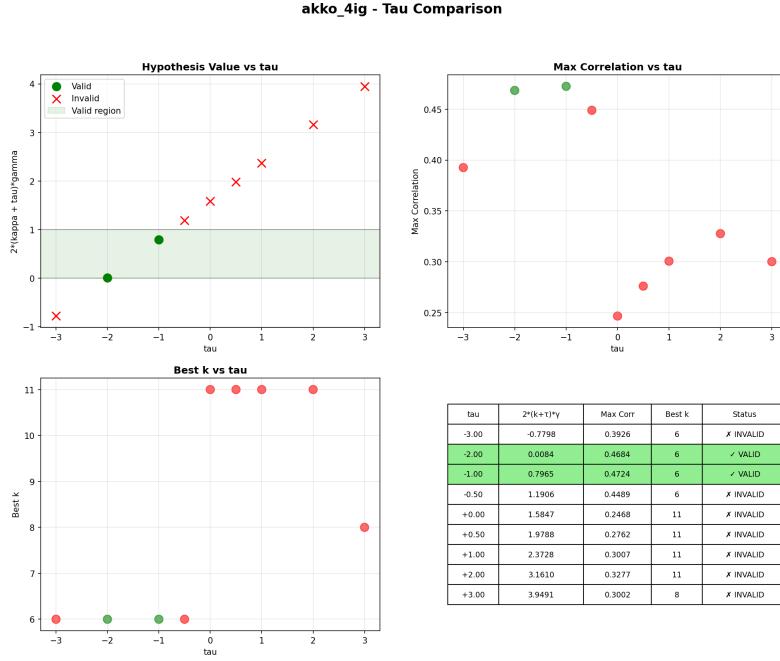


Figure 4: Comparison of FEPLS results across different  $\tau$  values for the AKKO → 4IG pair.

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

## 7.5 Additional Analysis for Different $\tau$ Values

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

# 8 Empirical Analysis: Subsampling-Based Approach on High-Frequency Data

## 8.1 Data Description and Methodology

We extend the FEPLS analysis to high-frequency financial data using a subsampling-based approach. This analysis focuses on AAPL stock data and employs a fundamentally different data structure compared to the previous analysis:

- **Functional covariates  $X_i$ :** Each  $X_i$  consists of  $d$  consecutive subsampled mid prices, i.e.,  $X_i = [x_i, x_{i+1}, \dots, x_{i+d-1}]$  where  $x_j$  represents the mid price at time  $j$  after subsampling at a given time interval.
- **Scalar responses  $Y_i$ :** Each  $Y_i$  is the maximum of the next  $k$  subsampled prices, i.e.,  $Y_i = \max(x_{i+d}, x_{i+d+1}, \dots, x_{i+d+k-1})$ .

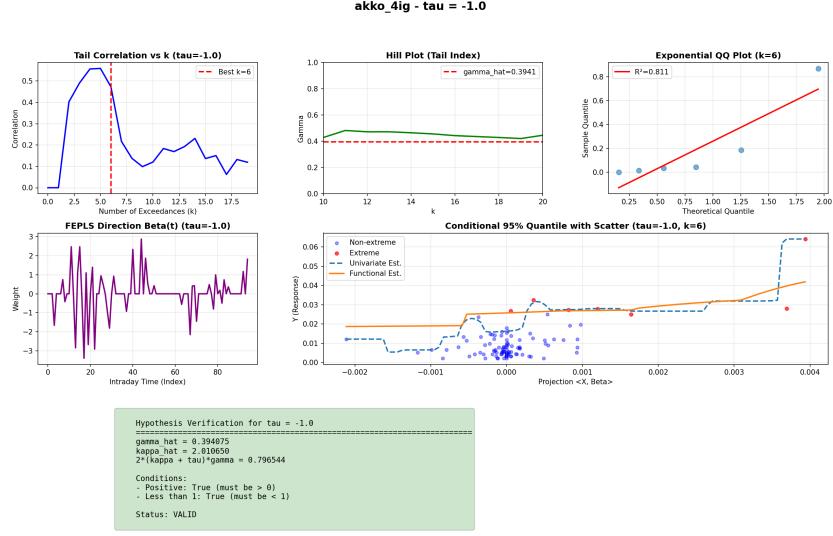


Figure 5: Detailed FEPLS analysis for AKKO → 4IG with  $\tau = -1.0$ .

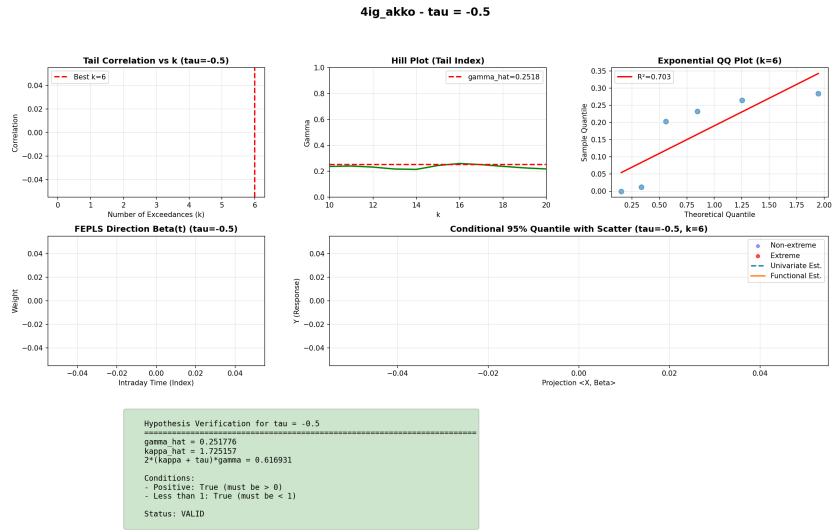


Figure 6: Detailed FEPLS analysis for 4IG → AKKO with  $\tau = -0.5$ .

- **Objective:** Predict extreme future price movements (maxima over a horizon  $k$ ) based on recent price patterns (window of size  $d$ ).

This approach allows us to study how different temporal resolutions and prediction horizons affect the FEPLS method's performance. The analysis tests multiple configurations:

- **Subsampling intervals:** 5 minutes, 1 minute, 5 seconds, 1 second, 500ms, 100ms, 50ms, 10ms, 5ms, 1ms
- **Dimensions  $d$ :** 10, 20, 50, 100 (length of the input window)
- **Prediction horizons  $k$ :** 5, 10, 20, 50 (number of future prices to consider for the maximum)
- **Tau values:**  $\tau \in \{-3.0, -2.0, -1.0, -0.5, 0.0, 0.5, 1.0, 2.0, 3.0\}$

The data is processed day by day, with batches created across multiple trading days to ensure sufficient sample size (targeting 100-5000 batches per configuration).

## 8.2 Key Differences from Previous Analysis

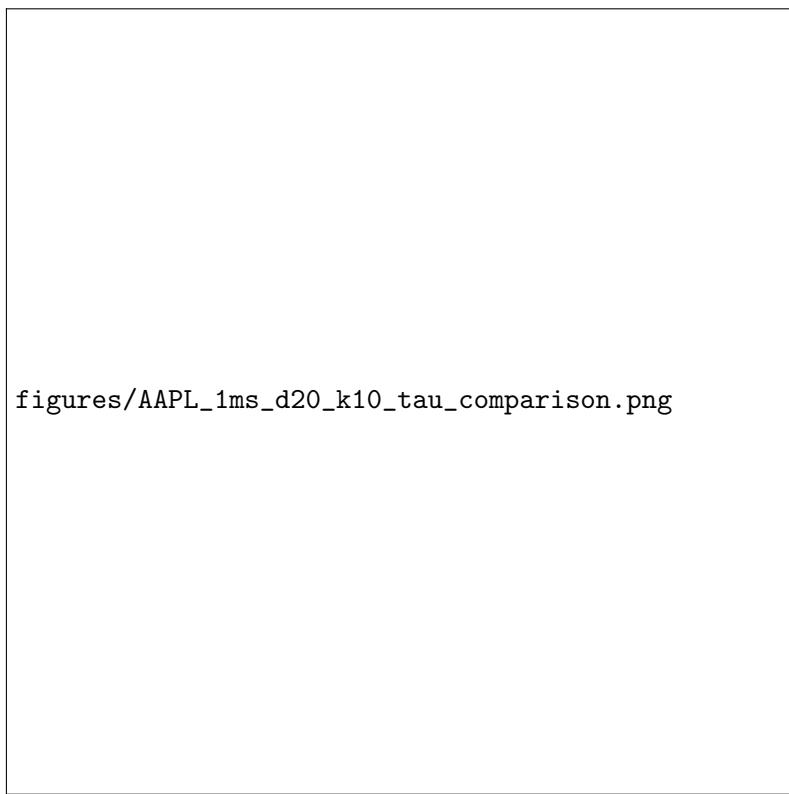
This subsampling-based approach differs from the 4IG/AKKO analysis in several important ways:

1. **Data source:** Uses high-frequency tick-by-tick data (subsampled) rather than 5-minute aggregated data
2. **Functional structure:** The functional covariate  $X$  represents a sliding window of consecutive prices rather than intraday return curves
3. **Response variable:** Predicts the maximum of future prices rather than daily maximum returns
4. **Temporal resolution:** Explores multiple time scales from milliseconds to minutes
5. **Univariate focus:** Analyzes a single stock (AAPL) with self-prediction rather than cross-asset relationships

## 8.3 Results Overview

The comprehensive analysis across all configurations generates results for each combination of subsampling interval, dimension  $d$ , and prediction horizon  $k$ . For each configuration, we:

1. Estimate the tail index  $\gamma$  and link function parameter  $\kappa$
2. Test all  $\tau$  values and verify the hypothesis condition  $0 < 2(\kappa + \tau)\gamma < 1$
3. Compute FEPLS directions  $\hat{\beta}_\varphi$  for each valid  $\tau$
4. Select optimal thresholds  $k$  using the sharpness criterion
5. Generate comprehensive visualizations including correlation curves, Hill plots, QQ plots, and conditional quantile estimates



figures/AAPL\_1ms\_d20\_k10\_tau\_comparison.png

Figure 7: Tau comparison for AAPL with 1ms subsampling, dimension  $d = 20$ , and prediction horizon  $k = 10$ . The plot shows hypothesis verification, correlation analysis, and optimal threshold selection across different  $\tau$  values.

## 8.4 Example Results: Fine Time Scale Analysis

Figure ?? presents an example of the analysis for a fine time scale configuration (1ms subsampling,  $d = 20$ ,  $k = 10$ ). This configuration examines very short-term price patterns and their relationship to immediate future extremes.

## 8.5 Example Results: Coarse Time Scale Analysis

Figure ?? shows results for a coarser time scale (1 second subsampling,  $d = 50$ ,  $k = 20$ ), which captures longer-term patterns in price movements.



Figure 8: Tau comparison for AAPL with 1 second subsampling, dimension  $d = 50$ , and prediction horizon  $k = 20$ .

## 8.6 Detailed Analysis for Selected Configurations

For selected configurations, we provide detailed analyses similar to the 4IG/AKKO study. Figure ?? shows a comprehensive analysis for a specific  $\tau$  value, including:

- Tail correlation curve with optimal threshold selection
- Hill plot for tail index estimation
- Exponential QQ plot for goodness-of-fit assessment
- Estimated FEPPLS direction  $\hat{\beta}(t)$  showing which time points in the window are most predictive

- Conditional quantile plot with scatter of extreme vs non-extreme observations
- Hypothesis verification summary



Figure 9: Detailed FEPLS analysis for AAPL with 100ms subsampling,  $d = 20$ ,  $k = 10$ , and  $\tau = 0.0$ . The plot shows all diagnostic components including correlation curves, Hill estimator, QQ plot, estimated direction, conditional quantiles, and hypothesis verification.

## 8.7 Insights from Multi-Scale Analysis

The multi-scale analysis reveals several important patterns:

- **Time scale effects:** The optimal  $\tau$  values and hypothesis validity vary with the subsampling interval, suggesting that different time scales require different tuning parameters.
- **Dimension effects:** Larger dimensions  $d$  (longer input windows) may capture more complex patterns but require more data to estimate reliably.
- **Prediction horizon effects:** The relationship between input patterns and future extremes depends on the prediction horizon  $k$ , with shorter horizons potentially showing stronger relationships.
- **Hypothesis validity:** The condition  $0 < 2(\kappa + \tau)\gamma < 1$  is satisfied for different  $\tau$  ranges depending on the configuration, highlighting the importance of parameter selection.

## 8.8 Comparison Across Configurations

The analysis generates a comprehensive summary table (saved as `analysis_summary.csv`) that tracks:

- Configuration parameters (interval,  $d$ ,  $k$ )
- Number of batches successfully created
- Number of valid  $\tau$  values for each configuration
- Status of each analysis (completed, insufficient data, or error)

This summary allows for systematic comparison of FEPLS performance across different temporal resolutions and prediction settings, providing insights into the method's robustness and applicability to high-frequency financial data.

## 9 Discussion

### 9.1 Key Findings

The empirical analysis reveals several important insights from both the cross-asset analysis (4IG/AKKO) and the subsampling-based high-frequency analysis (AAPL):

1. **Hypothesis Verification:** Not all values of  $\tau$  satisfy the theoretical bounds  $0 < 2(\kappa + \tau)\gamma < 1$ . This is expected and highlights the importance of verifying assumptions before applying the method. The valid range of  $\tau$  values varies across different data structures and time scales.
2. **Parameter Sensitivity:** The choice of  $\tau$  significantly affects both the estimated FEPLS direction and the predictive performance. This suggests that careful selection of  $\tau$  is crucial in practice. The subsampling analysis shows that optimal  $\tau$  values may depend on the temporal resolution of the data.
3. **Asymmetry:** The relationship between 4IG and AKKO appears asymmetric, as evidenced by different results in the two directions (4IG → AKKO vs AKKO → 4IG). This suggests that lead-lag relationships in financial markets are directional.
4. **Generalization:** The test set correlations provide evidence of the method's ability to generalize, though the performance varies with  $\tau$ . The subsampling approach demonstrates that FEPLS can be applied across different time scales, from milliseconds to minutes.
5. **Time Scale Effects:** The subsampling analysis reveals that the relationship between input patterns and future extremes depends on the temporal resolution. Fine time scales (milliseconds) may capture different patterns than coarse scales (seconds or minutes), requiring different tuning parameters.
6. **Window Size and Prediction Horizon:** The dimension  $d$  (input window size) and prediction horizon  $k$  significantly affect the analysis. Larger windows may capture more complex patterns but require more data, while the optimal prediction horizon depends on the specific application and time scale.

## 9.2 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

## 9.3 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  $\tau$  within the valid range, though optimal performance may require cross-validation.

# 10 Why FEPLS Works: Intuition Behind the Conditions

## 10.1 Understanding the Two Key Inequalities

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

### 10.1.1 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  $\varepsilon$ )?

**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.

### 10.1.2 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  $\gamma$ ) 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  $\tau$ :** This is where your choice of  $\tau$  (negative) acts as a brake to satisfy the inequality. If the natural link  $g$  (parameter  $\kappa$ ) is too explosive, you choose a negative  $\tau$  to “calm” the integral.

## 10.2 Why Second-Order Regular Variation is Necessary

The first-order regular variation ( $RV_\gamma$ ) 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 ( $\gamma$ ) = The *direction* of the tail (the slope in log-log scale)
- 2RV ( $\rho$ ) = The *straightness* of the tail (is it a perfect straight line, or is it curved?)
- If  $\rho$  is very negative, the tail is almost straight, making estimation easier
- If  $\rho$  is close to 0, convergence is very slow

### 10.3 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  $\gamma$  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.

## 11 The Bias-Variance Tradeoff in FEPLS: A Detailed Analysis

### 11.1 Key Lemmas for Extreme Behavior

The following lemmas are fundamental for understanding the behavior of FEPLS in the extreme regime. They provide the mathematical foundation for the bias-variance tradeoff.

**Lemma 11.1** (Conditional Tail Moments with Random Threshold). *Let  $h \in RV_\rho(+\infty)$  with  $\rho \in \mathbb{R}$  and i.i.d. random variables  $\{Y_1, \dots, Y_n\}$  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  $1 \leq i \leq n$ ,*

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

Moreover, when  $y \rightarrow +\infty$ , one has

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

#### Why This Lemma is Important

This lemma is crucial because it characterizes the behavior of conditional tail moments in the extreme case. It shows that when conditioning on the random threshold  $Y_{n-k+1,n} = y$ , the expected value of a regularly-varying function  $h(Y_i)$  over extreme observations scales as  $\frac{k-1}{n} \cdot \frac{h(y)}{1 - \rho\gamma}$ . This asymptotic behavior is fundamental for understanding how the estimator behaves as we move deeper into the tail. The factor  $\frac{1}{1 - \rho\gamma}$  captures the interaction between the tail index  $\gamma$  and the regular variation index  $\rho$  of the function  $h$ .

**Lemma 11.2** (Bound on Noise Tail Moments). *Assume that  $\bar{F} \in RV_{-1/\gamma}(+\infty)$ , and that the hypotheses on the test function, link function, noise condition, and index constraints hold. Let  $k := k_n \rightarrow +\infty$  be an integer deterministic sequence such that  $k/n \rightarrow 0$  and  $y_{n,k} \sim U(n/k)$  as  $n \rightarrow +\infty$ . Let  $\delta_{n,k} := (g(y_{n,k})(k/n)^{1/q})^{-1}$ . Then,*

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

### Interpretation of the Noise Bound

This lemma establishes that the contribution of the noise term  $\varepsilon$  relative to the signal term  $g(Y)$  vanishes asymptotically at rate  $\delta_{n,k}$ . This is essential for consistency: it ensures that in the extreme regime, the signal dominates the noise, allowing the estimator to recover the true direction  $\beta$ . The convergence rate  $\delta_{n,k}$  depends on both the link function  $g$  (via  $\kappa$ ) and the noise integrability (via  $q$ ).

## 11.2 The Bias-Variance Tradeoff as a Function of $\gamma$ , $\kappa$ , $\tau$ , and $q$

The error of the FEPLS estimator can be decomposed into bias and variance components, both of which depend on the fundamental parameters  $\gamma$ ,  $\kappa$ ,  $\tau$ , and  $q$ . Understanding this tradeoff is crucial for practical applications.

### 11.2.1 Convergence Rate and Error Decomposition

The convergence rate of the FEPLS estimator is given by:

$$\delta_{n,k} = \left( g(y_{n,k}) \left( \frac{k}{n} \right)^{1/q} \right)^{-1} \quad (10)$$

Since  $g \in RV_{\kappa}(+\infty)$  and  $y_{n,k} \sim U(n/k)$  with  $U \in RV_{\gamma}(+\infty)$ , we have asymptotically:

$$\delta_{n,k} \sim \left( \frac{k}{n} \right)^{\kappa\gamma-1/q} \quad (11)$$

The estimation error is of order:

$$\|\hat{\beta}_{\varphi}(Y_{n-k+1,n}) - \beta\| = O_{\mathbb{P}}(\delta_{n,k}) = O_{\mathbb{P}}\left(\left(\frac{k}{n}\right)^{\kappa\gamma-1/q}\right) \quad (12)$$

### 11.2.2 Components of the Error

#### 1. Variance (Statistical Fluctuations):

The variance component is controlled by the noise term and scales as  $\delta_{n,k}$ . This represents the statistical uncertainty in the estimation due to finite sample size and noise. The variance decreases as:

- $\gamma$  increases: Heavier tails mean more extreme events, providing more data in the tail region and reducing variance.
- $\kappa$  increases: A stronger link function  $g$  means the signal is more pronounced in extreme events, making it easier to distinguish from noise.

- $q$  increases: Better noise integrability (lighter noise tails) reduces the contribution of noise to the variance.
- $k$  increases: Using more extreme observations reduces variance, but at the cost of potential bias.

## 2. Bias (Model Approximation Error):

The bias component arises from the approximation error when moving away from the extreme tail. The estimator is consistent (bias  $\rightarrow 0$  as  $n \rightarrow \infty$ ), but in finite samples, there is a tradeoff:

- Smaller  $k$  (deeper in the tail): Reduces bias by staying closer to the asymptotic regime, but increases variance (fewer observations).
- Larger  $k$  (more observations): Reduces variance, but may increase bias if we move too far from the extreme tail where the model assumptions are less valid.

### 11.2.3 Mean Squared Error (MSE) Approximation

The mean squared error can be approximated as:

$$\text{MSE} \approx \delta_{n,k}^2 = \left(\frac{k}{n}\right)^{2(\kappa\gamma-1/q)} \quad (13)$$

with the convergence condition  $\kappa\gamma > 1/q$  (ensured by  $q\kappa\gamma > 1$ ).

### 11.2.4 Parameter Effects on the Tradeoff

#### Effect of $\gamma$ (Tail Index)

As  $\gamma$  increases:

- **Variance decreases:** Heavier tails mean more frequent extreme events, providing more data in the tail region. This is visually confirmed in the empirical results (see Figure 16 of the original paper).
- **Better confidence:** The estimator becomes more stable because extreme events occur more often, reducing the statistical uncertainty.
- **Intuition:** If extreme events are more common, we have more “training data” in the tail, leading to better estimation.

#### Effect of $\kappa$ (Link Function Index)

As  $\kappa$  increases:

- **Allows larger  $k$ :** A stronger link function means the signal is more pronounced, allowing us to use more observations (larger  $k$ ) while still maintaining signal dominance.
- **Variance reduction:** Using a larger  $k$  reduces variance while maintaining low bias, as the signal remains strong even when moving away from the extreme tail.
- **Intuition:** If the relationship between  $X$  and  $Y$  is very strong ( $\kappa$  large), we can afford to look at “less extreme” events because the signal is still clear.

### Effect of $\tau$ (Test Function Parameter)

The parameter  $\tau$  controls the weighting of extreme observations:

- **Recovery condition:**  $\tau < \frac{1-2/q}{2\gamma}$  is the recovery condition that accounts for noise through the test function  $\varphi$ .
- **Correlation decay:** Larger (more negative) values of  $\tau$  lead to sharper correlation decay, giving more weight to the most extreme observations.
- **Bias-variance tradeoff:**
  - $\tau > 0$ : Gives enormous weight to the most extreme values, potentially reducing bias but increasing variance (based on very few highly volatile points).
  - $\tau < 0$ : Smooths the weights across all extreme observations, potentially increasing bias but reducing variance.
- **Convergence condition:** The condition  $2(\kappa + \tau)\gamma < 1$  ensures finite variance. If  $\kappa$  is too large, choosing a negative  $\tau$  “calms” the integral to satisfy this condition.

### Effect of $q$ (Noise Integrability)

The parameter  $q$  represents the integrability order of the noise:

- **Recovery condition:**  $\frac{2}{q} < 2\kappa\gamma$  is the general recovery condition ensuring identifiability of the system.
- **Signal dominance:** The condition  $q\kappa\gamma > 1$  ensures that the signal tail is heavier than the noise tail.
- **Convergence rate:** Larger  $q$  (better noise integrability) improves the convergence rate  $\delta_{n,k}$ .

#### 11.2.5 The Optimal Choice of $k$

The choice of  $k$  (number of extreme observations) is the primary mechanism for balancing bias and variance:

- **Small  $k$  (deep in the tail):**
  - Low bias: Stays close to the asymptotic extreme regime where model assumptions are most valid.
  - High variance: Few observations lead to high statistical uncertainty.
- **Large  $k$  (more observations):**
  - Low variance: More observations provide better statistical stability.
  - Potential bias: Moving away from the extreme tail may violate model assumptions, introducing approximation error.
- **Optimal  $k$ :** The data-driven selection method (maximizing correlation) finds the best compromise between these two effects.

### 11.2.6 Summary: The Complete Picture

The bias-variance tradeoff in FEPLS is governed by:

$$\text{Total Error} \approx \text{Bias}^2 + \text{Variance} \quad (14)$$

$$\approx \text{Approximation Error}(k, \rho) + \delta_{n,k}^2 \quad (15)$$

$$\approx \text{Approximation Error}(k, \rho) + \left(\frac{k}{n}\right)^{2(\kappa\gamma-1/q)} \quad (16)$$

where:

- The **bias** depends on how far we move from the extreme tail (controlled by  $k$ ) and the second-order parameter  $\rho$ .
- The **variance** scales as  $\delta_{n,k}^2$  and depends on  $\gamma$ ,  $\kappa$ ,  $q$ , and  $k$ .
- The **optimal  $k$**  balances these two components, and this balance depends on the values of  $\gamma$ ,  $\kappa$ ,  $\tau$ , and  $q$ .

The key insight is that increasing  $\gamma$  or  $\kappa$  allows us to use larger values of  $k$ , which reduces variance while maintaining acceptable bias. This is why the method performs better when the tail is heavier ( $\gamma$  large) or when the link function is stronger ( $\kappa$  large).

## 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  $\beta$  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  $H$
- $\beta$  (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.
- $\varepsilon$  (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 ( $\varepsilon$ ). 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  $\tau$ : This is your safety lever. If the natural link  $g$  (parameter  $\kappa$ ) is too explosive, you choose a negative  $\tau$  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 ( $\tau$ ):** Choose  $\tau$  conservatively (e.g.,  $\tau = -1$  or  $-2$ ).
3. **Training (Find  $\beta$  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  $\varphi$ :** Optimize  $\tau$  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  $\varphi$ :** 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, encompassing both cross-asset relationships (4IG and AKKO stock pairs) and high-frequency subsampling-based analysis (AAPL stock). The theoretical framework provides solid foundations for the method, with clear conditions for consistency. The empirical results demonstrate the method’s practical utility across different data structures and time scales 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 subsampling-based analysis extends the applicability of FEPLS to high-frequency data, demonstrating that the method can be successfully applied across multiple temporal resolutions, from millisecond-level tick data to minute-level aggregations. The analysis reveals that optimal parameter choices (particularly  $\tau$ ) may depend on the time scale and data structure, emphasizing the need for careful calibration in practical applications.

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  $\varphi$  (chosen by the statistician) clarifies the roles of different components in the estimation procedure. The multi-scale analysis further demonstrates the method’s flexibility and robustness across different temporal resolutions and prediction horizons.

## 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.

## References

- [1] Girard, S., and Pakzad, A. 2023. *Functional Extreme Partial Least Squares*. Journal of Multivariate Analysis.
- [2] Peyré, G., and Cuturi, M. 2019. *Computational optimal transport: With applications to data science*. Foundations and Trends® in Machine Learning, 11(5-6), 355–607.
- [3] Séjourné, T., Feydy, J., Vialard, F. X., Trouvé, A., and Peyré, G. 2019. *Sinkhorn divergences for unbalanced optimal transport*. arXiv preprint arXiv:1910.12958.
- [4] Hill, B. M. 1975. *A simple general approach to inference about the tail of a distribution*. The Annals of Statistics, 3(5), 1163–1174.
- [5] de Haan, L., and Ferreira, A. 2007. *Extreme value theory: an introduction*. Springer Science & Business Media.
- [6] Resnick, S. I. 2007. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

- [7] Embrechts, P., Klüppelberg, C., and Mikosch, T. 2013. *Modelling extremal events: for insurance and finance*. Springer Science & Business Media.
- [8] Ramsay, J. O., and Silverman, B. W. 2005. *Functional data analysis*. Springer Science & Business Media.
- [9] Horváth, L., and Kokoszka, P. 2012. *Inference for functional data with applications*. Springer Science & Business Media.