No 565

December 2025

CefES

Center for European Studies

PAPER SERIES

The Risk Premia from the European Equity Market: An application of the Three-Pass Estimation Methodology

Elisa Ossola, Irina Trifan

DIPARTIMENTO DI ECONOMIA, METODI QUANTITATIVI E STRATEGIA DI IMPRESA

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

The Risk Premia from the European Equity Market: An application of the Three-Pass Estimation Methodology

Elisa Ossola[1*] and Irina Trifan[1]

[1]University of Milan-Bicocca

December 2025

Abstract

We develop an empirical application on a large dataset of European stock returns, in order to estimate the risk premia. We propose an application of the Three-Pass Estimation Method (3PEM) by Xiu and Giglio (2021) as a multipurpose tool in asset pricing. By assuming the Fama–French Five-Factor model (Fama and French (2015)) as baseline model, we show that the 3PEM yields risk premium estimates that are more economically plausible and statistically robust than those obtained using the traditional two-pass estimation method (2PEM). Moreover, we extend the results by Xiu and Giglio (2021) showing that the 3PEM is able to detect noise in tradable factors. Furthermore, the method is used to denoise the observed factors, providing purified versions that better capture the systematic components of risk. We also identify noisy factors, and yield denoised factor series that improve the estimation of stock-level exposures and expected returns.

Keywords: three-pass estimator, Empirical asset pricing, PCA, large panels.
JEL Codes: G12, C58, C55

Acknowledgements: Research paper developed within the PreDoc Project "The Three-Pass Estimation Methodology in Asset Pricing and Asset Allocation: An empirical application to the European Stock Market" - Dip. di Eccellenza 2023-2027, Department of Economics, Management and Statistics (DEMS), Cod. 25B037. *Corresponding Author: elisa.ossola@unimib.it

1 Introduction

Linear factor models lie at the heart of modern asset pricing. Their origins trace back to the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), which formalized the idea that expected returns compensate investors for exposure to a single source of systematic risk. The restrictive structure of the CAPM motivated the development of more flexible models, most notably the Arbitrage Pricing Theory (APT, Ross (1976)), which allows multiple factors — economic, financial, or statistical — to jointly explain the cross-section of returns.

Factor models are essential both for portfolio allocation and for evaluating investment performance because they provide structured estimates of expected returns and the covariance matrix, which many optimization procedures rely on. However, classical mean–variance optimization (Markowitz (1952)) is highly sensitive to estimation errors, making factor-structure assumptions a practical way to reduce dimensionality and stabilize inputs. At the same time, factor models allow practitioners to distinguish true managerial skill (alpha) from compensation for systematic risks (beta), but this requires specifying the model correctly: omitting relevant factors or including spurious ones can bias alpha estimates, distort inference, and ultimately lead to poor investment decisions (see, e.g., Chincarini and Kim (2006), Lazzari and Navone (2003)).

Despite their wide applicability, empirical implementation of factor models faces two enduring challenges: model specification and measurement error. First, selecting the number and nature of factors is non-trivial. Including irrelevant or weak factors introduces noise and biases risk-premium estimates (Bryzgalova (2015)), whereas omitting important factors leads to classic omitted-variable bias, affecting both factor loadings and premia. A rich econometric literature provides tools for estimating the number of latent factors (e.g., Bai and Ng (2002); Onatski (2010); Ahn and Horenstein (2013); Gagliardini et al. (2019)), but these methods do not identify which specific factors are missing nor correct for the resulting bias. Second, observable factors—tradable or otherwise—may suffer from measurement error. Tradable factors such as the Fama–French portfolios are straightforward to use, but they remain noisy proxies constructed from finite samples and sorting rules (Racicot et al. (2011)). Non-tradable macroeconomic factors introduce even larger identification issues, as their premia cannot be directly estimated from time-series averages. Traditional methods such as the Fama–MacBeth two-pass estimator (Fama and MacBeth (1973)) assume noise-free factors and are therefore sensitive to both omitted variables and measurement error.

Recent advances have addressed these limitations. The Three-Pass Estimation Method (3PEM), introduced by Xiu and Giglio (2021), offers a unified and robust procedure for estimating risk premia when observable factors are noisy or incomplete. Leveraging the rotation-invariance property of latent factor spaces, the 3PEM yields consistent estimates of the prices of risk even in the presence of omitted factors or errors-in-variables. While originally designed for non-tradable factors, the method naturally extends to tradable ones, where measurement error remains non-negligible. In this paper, we empirically show that, since the tradable factors contain a measurement error, the 3PEM is also a crucial tool involving tradable factors.

We show that the 3PEM is a multi-purpose tool in asset pricing. It is employed to perform several intermediary tasks, such as estimating risk premia, analyzing observable factors, and detecting weak factors to guide the selection of an appropriate model specification, with the final objective of computing the expected returns of stocks. In particular, the suggested methodology is integrated in the following algorithm:

(i) Estimate the risk premia of factors (in this contribution, factors are tradable) and conduct inference on them. It allows for accounting for the omitted variable and error-in-variables problems. Even when some priced factors are missing from the model, the estimated risk premia of the included observable factors remain consistent.

(ii) Given a model specification, one can use 3PEM to detect spurious or weak factors. Therefore, it can be used as a tool for model specification and factor selection by identifying and excluding factors that are not significantly priced.

(iii) 3PEM can be next used to denoise observable factors, including tradable ones, by filtering out idiosyncratic or measurement error components and, thus, providing a cleaner representation of the true systematic sources of risk. These factor observations are used to further compute the factor loadings at the stock level and the expected values of stock returns.

In this paper, we provide these results by developing an empirical application on the European equity market, using as a base model the well-known Fama and French Five (FF5) factors model (Fama and French (2015)). The FF5 factors serve as a practical benchmark, but the approach is fully general and applies to any linear factor model. We compare the results of the 3PEM with those of the traditional two-pass estimator and with the time-series averages of the tradable factors. The analysis shows that the 3PEM delivers more economically plausible and statistically robust risk-premium estimates, identifies noisy factors, and yields denoised factor series that improve the estimation of stock-level exposures and expected returns.

The remainder of the paper is organized as follows. Section 2 introduces the theoretical model. Section 3 outlines the Three-Pass Estimation Method. Section 4 presents the data and empirical findings, including risk-premium estimation, factor denoising, and stock-level implications. Section 5 concludes.

2 Model Setting

In this section, we introduce the theoretical framework describing the linear setting for excess returns. In particular, we follow the notation in Xiu and Giglio (2021). First, we introduce the setting that accounts for the restricted zero-beta rate, and then generalize it to allow for an unrestricted zero-beta rate.

Let us define r_t the $n \times 1$ vector of excess returns for n testing assets. At each time period t, r_t is described by the following linear model:

$$r_t = \beta\gamma + \beta v_t + u_t, \tag{1}$$

with

$$f_t = \mu + v_t$$

and $\mathbb{E}(v_t) = 0$, where f_t is a vector gathering the p true latent factors, defined as the sum of two components: the constant mean μ capturing the long-run average values of the true factors, and the time-varying innovation component v_t reflecting new, unexpected information that causes deviations from the mean. Thus, μ represents the mean around which the factors tend to fluctuate over time. Moreover, assuming that $\mathbb{E}(v_t) = 0$ implies that the factors, on average, revert to their mean μ over time, ensuring that the shocks are temporary, although they may have a significant impact on asset prices in the short run. Futhermore, in Eq. (1), the vector γ collects the risk premia of factors f_t. The $n \times p$ matrix β gathers the factor loadings, i.e., each of its elements β_{ij} measures the sensitivity of the return on asset i to factor j. The terms $\beta\gamma$ and βv_t are capturing the average expected risk premium and the time-varying impact of unexpected factor shocks on returns, respectively. Lastly, u_t is the vector of idiosyncratic errors, such that $\mathbb{E}(u_t) = 0$ and $\text{Cov}(u_t, v_t) = 0$.

In general, we do not observe the p factors f_t, but we usually observe their $d \leq p$ proxies. Thus, we define g_t as the set of d observable factors (e.g., macroeconomic variables, financial market indexes, portfolios,...) that relate to the true unobservable factors f_t through the following measurement error model:

$$g_t = \xi + \eta v_t + z_t, \tag{2}$$

where ξ is a vector of constants, and η is a $d \times p$ loading matrix, where each element η_{ij} measures the sensitivity of the i-th observed factor proxy to the j-th true factor. Finally, z_t is the vector of measurement errors, such that $\mathbb{E}(z_t) = 0$ and $\text{Cov}(z_t, v_t) = 0$. Since the observed factor proxies do not perfectly capture the true underlying factors, the measurement error arises and is capturing the noise or inaccuracies due to, for example, the data limitations, incorrect proxy selection, or other empirical measurement issues. Consequently, the vector of d risk premia of the observable factors, g_t, is defined as:

$$\gamma_g = \eta\gamma. \tag{3}$$

The estimation model from Eq. (1) and (2), can be rewritten in matrix notation as follows,

$$R = \beta\gamma\iota_T' + \beta V + U, \text{ and } G = \xi + \eta V + Z, \tag{4}$$

where $R \in \mathbb{R}^{n\times T}$ is the matrix of excess returns, $\beta \in \mathbb{R}^{n\times p}$ is the matrix of factor loadings, ι_T is the T vector of ones, and risk premia and innovations are collected in $\gamma \in R^{p\times 1}$ and $V \in \mathbb{R}^{p\times T}$, respectively. Furthermore, the matrix of observable factors is $G \in \mathbb{R}^{d\times T}$, $\xi \in \mathbb{R}^{d\times T}$ refers to the constant terms, and matrix $\eta \in \mathbb{R}^{d\times p}$ collects the loadings to the observed proxies factors. Finally, error terms are gathered in matrices $U \in \mathbb{R}^{n\times T}$ and $Z \in \mathbb{R}^{p\times T}$ with $\mathbb{E}[V] = \mathbb{E}[U] = \mathbb{E}[Z] = 0$. Moreover, we also assume that $\mathbb{E}[UV'] = 0$ guarantees the ability to separate common factors from idiosyncratic errors, and $\mathbb{E}[ZV'] = 0$ ensures that the model can distinguish the true source of risk from noise in observable factors. The above orthogonality condition is fundamentally satisfied because applying PCA on a large panel of asset returns recovers the factor space precisely by exploiting the common variation across returns, filtering out the idiosyncratic part. We also note that this theoretical framework allows for heteroskedasticity and autocorrelation of v_t, u_t, and z_t. Finally, the setting proposed in Eq. (4) can be easily extended to the unrestriced zero-beta rate as follows,

$$R = \gamma_0\iota_n\iota_T' + \beta\gamma\iota_T' + \beta V + U, \tag{5}$$

where $\iota_n \in \mathbb{R}^{n\times 1}$ is a vector of ones.

In this setting, Xiu and Giglio (2021) show that the risk premia γ_g defined in Eq. (3) can be estimated due to the rotation invariance property. In fact, from Eq. (2), the observed factors are expressed as a projection onto true latent factors, which span the true factor space. If we omit some of the true factors, standard estimators will suffer from omitted variable bias. Nevertheless, we can apply PCA on a large panel of returns to get some rotated version of the factor space, not the exact v_t, i.e., we can recover a rotation of the latent factors $\tilde{v}_t = Hv_t$, where $H \in \mathbb{R}^{p\times p}$ is an invertible matrix such that $H'H = HH' = I_p$, $H^{-1} = H'$, and $det(H) = 1$. Thus, Eq. (1) and (2) becomes

$$\begin{aligned} r_t &= \hat{\beta}\hat{\gamma} + \hat{\beta}\hat{v}_t + u_t, \\ g_t &= \xi + \hat{\eta}\hat{v}_t + z_t, \end{aligned}$$

where $\hat{\beta} = \beta H^{-1}$, $\hat{\gamma} = H\gamma$, $\hat{v}_t = Hv_t$, and $\hat{\eta} = \eta H^{-1}$. Noting that $\hat{v}_t$ can be easily extracted by applying Principal Component Analysis (PCA) on the large panel of returns and, since it is a rotated version of the true factors, i.e., it spans the same risk space, the estimator for the risk premia of observable factors is given by:

$$\hat{\gamma}_g = \hat{\eta}\hat{\gamma} = \eta H^{-1} H\gamma = \eta\gamma.$$

Thus, even if we do not know the true H, we are able to find an estimate of the risk premia of observable factors. However, the true matrix of factor loadings η and the risk premia of latent factors, γ cannot be identified separately.

3 The Three-Pass Estimation Methodology

In this section, we review the 3PEM presented in Xiu and Giglio (2021). This methodology is based on three steps and aims to estimate the risk premia vector of g_t. Hereafter, we provide a description of the three passes.

1. Estimation of V and β. We apply PCA on $(nT)^{-1}\bar{R}'\bar{R}$, where $\bar{R}$ is the matrix of demeaned returns, such that the element $[\bar{r}_{i,t}]$ is defined as $\bar{r}_{i,t} = r_{i,t} - \frac{1}{T}\sum_{t=1}^{T} r_{i,t}$. This matrix captures the variance and covariance structure of the asset returns, reflecting the true structure of data without being distorted by its scale. We then obtain the following eigenvalue decomposition:

$$(nT)^{-1}\bar{R}'\bar{R} = Q\Lambda Q',$$

where Q is an $T \times T$ matrix whose columns are the orthogonal eigenvectors of $(nT)^{-1}\bar{R}'\bar{R}$, and Λ is an $T \times T$ diagonal matrix whose diagonal elements are the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_T$. Each eigenvector points in a direction where the data varies the most when projected onto this direction. Similarly, eigenvalues indicate the magnitude of variance captured by each eigenvector, such that larger eigenvalues correspond to eigenvectors that explain more of the total variance in the data. By decomposing $(nT)^{-1}\bar{R}'\bar{R}$ into its eigenvectors and eigenvalues, it is possible to identify the components that capture the most significant patterns in the asset returns.

In this setting, the true number p of latent factors is an unknown parameter. In line with Xiu and Giglio (2021), the following consistent estimator for p is introduced:

$$\hat{p} = \arg \min_{1 \leq j \leq p_{\max}} \left((nT)^{-1} \lambda_j (\bar{R}'\bar{R}) + j \times \phi(n,T) \right) - 1, \tag{6}$$

where the term $\lambda_j(\bar{R}'\bar{R})$ corresponds to the j-th eigenvalue of the matrix $\bar{R}'\bar{R}$, and $\phi(n,T)$ is a penalty term depending on the number of assets and the time periods considered.

Then, the following approximation holds:

$$(nT)^{-1}\bar{R}'\bar{R} \approx Q_{\hat{p}}\Lambda_{\hat{p}}Q'_{\hat{p}} = \sum_{i=1}^{\hat{p}} \lambda_i \xi_i \xi_i', \tag{7}$$

where $\Lambda_{\hat{p}}$ is the diagonal matrix containing the first $\hat{p}$ sorted eigenvalues, and $Q_{\hat{p}}$ is the matrix of their corresponding orthogonal eigenvectors. Thus, each term $\lambda_i \xi_i \xi_i'$ represents the contribution of the i-th principal component to the total covariance matrix, scaled by its eigenvalue λ_i. Finally, the estimated matrix of latent factors is defined as:

$$\hat{V} = T^{1/2}(\xi_1 : \xi_2 : \ldots : \xi_{\hat{p}})',$$

and the corresponding estimated matrix of factor loadings:

$$\hat{\beta} = T^{-1}\bar{R}\hat{V}'.$$

2. Estimation of γ. In order to estimate the risk premia of latent factors, a cross-sectional regression (CSR) technique is employed. In particular, we regress the average returns of the n testing assets on the estimated factor loadings, $\hat{\beta}$:

$$\bar{r} = \hat{\beta}\gamma + u_t,$$

where $\bar{r} = \begin{bmatrix} \bar{r}_1 & \bar{r}_2 & \cdots & \bar{r}_n \end{bmatrix}'$, and the risk premia of estimated latent factors are given by [1]:

$$\hat{\gamma} = \left(\hat{\beta}'\hat{\beta} \right)^{-1} \hat{\beta}'\bar{r}.$$

3. Estimation of γ_t. From Eq.(4), we have

$$\bar{G} = G - E[G] = \xi + \eta V + Z - \varepsilon = \eta V + Z, \tag{8}$$

where $\bar{G}$ is the matrix $(d \times T)$ of the demeaned observable factors. The true factors V are not directly observable and we can deduce a rotated version of them, $\hat{V}$ through PCA, and

$$\bar{G} = \eta \hat{V} + Z. \tag{9}$$

Thus, the time-series regression allows to project g_t onto the space spanned by the estimated factors $\hat{V}$ and we get:

$$\hat{\eta} = \bar{G}\hat{V}'(\hat{V}\hat{V}')^{-1}.$$

[1]The full derivation is available in Appendix B.

Note that if the factors are tradable, then

$$G = \eta \hat{V} + Z,$$

and $\hat{\eta} = G\hat{V}'(\hat{V}\hat{V}')^{-1}$. Finally, we estimate the risk premia of observable factors as follows:

$$\hat{\gamma}_g = \hat{\eta}\tilde{\gamma}. \tag{10}$$

Moreover, we also get the denoised version of the observable factors as the fitted value of the model (3):

$$\hat{G} = \hat{\eta}\hat{V}. \tag{11}$$

From Eq. (3) is also possible to decompose the variance of the observable factors in the variance due to the exposure to the true risk factors and the residual variance which can be attributable to the measurement error and noise:

$$Var(G) = Var(\xi + \eta V + Z) = Var(\eta V + Z) = Var(\eta V) + Var(Z)$$

Substituting for the estimated parameters $\hat{\eta}$ and $\hat{V}$:

$$Var(G) = Var(\hat{\eta}\hat{V} + Var(Z) = Var(\hat{G}) + Var(Z) \tag{12}$$

Therefore, in the case of tradable factors, rather then using the noisy version of observable factors, it is possible to use their denoised version, $\hat{G}$, as represented in 11.

For the unrestricted zero-beta rate in model (5), we get

$$\begin{aligned}
\hat{\gamma}_0 &= \left(\iota_n' \mathbb{M}_{\hat{\beta}}\, \iota_n\right)^{-1} \iota_n' \mathbb{M}_{\hat{\beta}}\, \bar{r} \\
\hat{\gamma}_g &= \bar{G}\hat{V}'(\hat{V}\hat{V}')^{-1}\left(\hat{\beta}'\mathbb{M}_{\iota_n}\hat{\beta}\right)^{-1}\hat{\beta}'\mathbb{M}_{\iota_n}\bar{r},
\end{aligned}$$

where $\mathbb{M}_{\hat{\beta}} = I_n - \hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'$ and $\mathbb{M}_{\iota_n} = I_n - \iota_n(\iota_n'\iota_n)^{-1}\iota_n'$.

4 Data and Empirical Analysis

This section mainly presents empirical results. First, we describe the datasets employed in the analysis. Next, we present the results of our estimation of European risk premia. Furthermore, we analyse the denoised Fama-French factors, and finally investigate the risk exposures and expected returns.

4.1 Data Description

The empirical analysis conducted in this study draws on two complementary datasets: (i) a dataset of observable factors and portfolio returns constructed in accordance with the Fama–French methodology, and (ii) a dataset of individual European stock returns. Both datasets are sampled at a monthly frequency.

The first dataset is employed to estimate factor risk premia and to implement the denoising procedure. It is sourced from Kenneth R. French's Data Library French (2023) and pertains to the European market. This portfolio-level dataset comprises returns for 132 characteristic-sorted portfolios, including 25 size–book-to-market portfolios, 25 size-profitability portfolios, 25 size–investment portfolios, 25 momentum portfolios, and 32 additional style-based portfolios. In addition, observations on five tradable factors —namely, the market excess return ($Mkt - Rf$), size (SMB), value (HML), profitability (RMW), and investment (CMA) — are collected for the same sample period.

The second dataset represents the European investment equity universe. The analysis covers the period from January 2005 to May 2025. The dataset consists of time series of adjusted closing prices for 2,327 European equities sourced from the Europe FactSet Market Index. The index is designed to capture a broad, investable cross-section of European stocks by including firms listed on major European exchanges that satisfy specific liquidity and capitalization requirements. Only actively traded, primary equity listings are retained, while inactive securities, secondary listings, and non-equity instruments are excluded. Furthermore, only stocks with at least five years of historical price data are included[2].

4.2 Estimation Results and Risk Premia

We collect results on the implementation of the 3PEM and the conventional 2PEM framework on the European equity market to assess its empirical performance. Following the procedure outlined in Section 3, the first step is to recover the latent factor space. Thus, we estimate the number of common latent factors. Figure 1 reports the eigenvalue structure and provides clear evidence of a low-dimensional factor structure in European portfolio returns. Panel A displays the distribution of the first ten eigenvalues of the sample covariance matrix. The first eigenvalue dominates the spectrum, accounting for the largest share of common variation in returns. Upon removing the first eigenvalue (Panel B), the remaining spectrum exhibits a smooth, monotonic decline. The cumulative explained variance shown in Panel C rises sharply with the inclusion of the first two principal components—surpassing 90% of total variance—and reaches roughly 95% when the first ten components are included. This pattern supports the presence of a small number of economically meaningful latent factors. Panel D compares the normalised information criteria proposed by Bai and Ng (2002). In particular, we compute the PCP1, ICP1, and BIC3.[3] All three criteria experience a pronounced drop between one and two factors, and then flatten substantially. The minima occur in the range between five and seven factors, with the Bayesian criterion (BIC3) reaching its lowest value at $\hat{p} = 6$. This result motivates the choice of six latent factors as the empirical benchmark for the subsequent estimation of risk premia. The convergence of different criteria around this dimension reinforces the robustness of the selected factor space. It provides a balanced trade-off between parsimony and explanatory power, in line with the asymptotic guidance in Bai and Ng (2002) and the empirical framework of Xiu and Giglio (2021).

A complementary perspective comes from analysing how well the latent factor structure extracted in the first step is able to reconstruct the demeaned returns. In this regard, Figure 2 demonstrates that the latent space is able to capture a substantial share of the common variation in returns.

[2] Newly listed firms represent a distinct analytical category and are therefore excluded from the present analysis. We also allow for prices below one unit , as the analysis relies on simple rather than logarithmic returns. All prices are denominated in U.S. dollars to ensure consistency with the Fama–French factors, which are reported as simple returns based on dollar-denominated prices.

[3] Based on the minimization approach in Eq. (6), Bai and Ng (2002) proposed two classes of PC and IC criteria w.r.t the definition of the variance and penalty functions.

Moreover, a further diagnostic concerns the economic interpretability of the recovered latent space. To this end, we compute the correlation matrix between the estimated latent factors and the observable Fama-French factors, as shown in Table 1. We observe that the first latent factor correlates almost perfectly with the market risk factor and it shows that the market component is largely recovered by the latent representation. A similarly strong correlation is observed between the second latent factor and SMB, which reflects the independence of the size premium from the remaining factor dimensions. However, HML, RMW, and CMA exhibit a different pattern. The third latent factor loads strongly on HML and only moderately on RMW and CMA. Therefore, it suggests that these factors do not correspond to clean, isolated directions in the latent space. Instead, they share common variation and occupy a more diffuse and weaker subspace of the pricing kernel. This behavior is thus consistent with their known lack of orthogonality in empirical datasets.

Using the five-factor model of Fama and French (2015) as the baseline linear specification, Table 2 reports the estimated risk premia. Specifically, we present the estimates obtained from the traditional two-pass estimation method (2PEM) and from the Three-Pass Estimation Method (3PEM), and compare them with the time-series averages of the corresponding factor returns. It is important to notice that the choice of Fama-French five-factor specification serves purely as an empirical illustration of the 3PEM framework rather than a structural requirement of the method. The 3PEM can be applied to any linear factor model, including specifications based on alternative tradable or non-tradable factors. In fact, the observable factors simply act as proxies for the underlying sources of systematic risk, and the first step of the 3PEM procedure recovers the latent factor space independently of the chosen economic or fundamental model. Hence, the FF5 model is one possible application among many, selected in this work due to its extensive use in the literature and its relevance for explaining equity returns. Additionally, as we show in the subsequent analysis, the 3PEM also helps to evaluate the actual relevance of the factors included in the specification.

Before analysing the estimated risk premia obtained from the 2PEM and 3PEM, it is useful to examine some intermediary results from the first two steps of the traditional two-pass procedure. In particular, we inspect the cross-sectional distribution of the estimated factor loadings and their relation to average excess returns, as represented in Figures 3 and 4. Although the estimated betas show reasonably and economically plausible distributions, the beta-return scatter plots reveal that the cross-sectional pricing relationship is weak for most factors. The absence of a clear slope and the large dispersion of points is an indication of the fact that differences in factor exposures do not translate into systematic differences in average returns. As a result, the second-pass regression in the 2PEM might not perform well in determining a stable pricing signal, producing noisy and unreliable risk-premia estimates.

Regarding the estimation of risk premia, and in order to illustrate the differences in detail, we examine two model specifications: one imposing the zero-beta rate to equal the T-bill rate, following Fama and French (2015), and one allowing the zero-beta rate to be freely estimated. These correspond to the restricted and unrestricted models shown in Table 2. In order to study the significance of the estimates, we apply the inference results from Xiu and Giglio (2021) for the 3PEM estimates and Shanken (1992) for the 2PEM. Xiu and Giglio (2021) shows that $\hat{\gamma}_g$ converge to a normal distribution when the cross-sectional and time-series dimensions converge to infinity simultaneously, $n, T \to \infty$ and such that $T1/2n^{-1} \to 0$.[4] We note that the asymptotic variance-covariance matrix of estimator of the risk premia of the observable factors g_t does not depend on the covariance matrix of the residual u_t or the estimation error of β. This is the main difference with the classical two pass methodology and the Shanken's correction, in which both β and Σ^u impact on the accuracy of the risk premia.[5]. The results indicate that the 3PEM yields risk premium estimates that are more economically plausible and statistically robust than those obtained using the traditional 2PEM. Notably, the market risk premium becomes positive and statistically significant in the unrestricted specification under the 3PEM, whereas it is estimated to be negative under the 2PEM. This shift reinforces the view that the 3PEM produces estimates more consistent with asset pricing theory, which predicts that the market factor should command a positive price of risk. Moreover, the 3PEM estimates closely align with the simple time-series averages of the factor returns, suggesting that the method effectively captures the systematic component of risk premia rather than noise arising from idiosyncratic variation. The presence of a positive market premium, together with the close correspondence between the 3PEM estimates and the time-series factor means, mirrors the findings of Xiu and Giglio (2021) for the U.S. market and is likewise confirmed here for the European equity market.

It is important to note that the alignment between the estimated and model-free risk premia is less pronounced for the profitability factor (RMW). This factor appears particularly noisy, a point that will be examined in greater detail later. Interestingly, RMW displays strong statistical significance under the 2PEM but not under the 3PEM. As demonstrated in the following section, despite the apparent significance produced by the 2PEM, the estimated factor loadings (betas) on RMW are largely insignificant across individual assets. Once these exposures are corrected using the 3PEM, the number of stocks exhibiting statistically significant loadings on the RMW factor increases, indicating that the de-noising step embedded in the 3PEM facilitates a clearer and more reliable identification of true factor exposures.

To assess the robustness of the estimated risk premia to potential model misspecification, we perform additional estimations under three alternative specifications: the CAPM, the Fama–French three-factor model (FF3F), and the five-factor model (FF5). Table 3 reports the estimated premia obtained using both the traditional two-pass estimation method and the 3PEM under the restricted zero-beta setting. The results indicate that the estimates of the risk premia from the 3PEM remain stable across all model specifications. Introducing or omitting observable factors does not materially alter the premia, suggesting that the 3PEM is robust to model misspecification and insulated from omitted-variable bias. This stability arises from the structure of the methodology. Indeed in the third pass of estimation approach (see Section 3) each observable factor is projected onto the latent space extracted in the first step, ensuring that the contribution of unobserved components is already accounted for. In contrast, the 2PEM estimates exhibit substantial variation across the different model specifications.

However, while the 3PEM yields stable estimates with respect to the observable model specification, it remains sensitive to the underlying latent structure. Table 4 and Figure 5 illustrate how the estimated risk premia vary with the number of latent factors, $\hat{p}$. The

[4]See Theorem 1 in Xiu and Giglio (2021)

[5]See, for example, Cochrane (2005), Shanken (1992) and Gagliardini et al. (2020)

market premium remains positive and stable across all configurations, confirming its robustness and strong identification. By contrast, the estimated premia and their statistical significance for the remaining factors depend on the chosen latent dimensionality.

Moreover, Figures 6 and 7 examine the stability of the estimated risk premia under changes in sample composition, both across portfolios and over time. Specifically, Figure 6 assesses the sensitivity of the estimates to cross-sectional resampling, while Figure 7 presents results from a rolling-window analysis along the time-series dimension. In both cases, the distributions of the estimated premia show that the traditional two-pass approach yields wider and more dispersed estimates, indicating greater variability and weaker robustness. This effect is particularly pronounced for the factors identified as noisy under the 3PEM, such as RMW and CMA. Despite the presence of noise in certain factors, the premia estimated using the 3PEM remain substantially more stable under both cross-sectional and time-series perturbations.

4.3 Denoised Fama-French Factors

As previously discussed, the 3PEM was originally developed to estimate the risk premia of non-tradable factors, but it is equally valid for tradable factors. This empirical work shows that 3PEM can also be extended to detect noise in observable factors, in line with the findings of Xiu and Giglio (2021). Moreover, it goes further by demonstrating that 3PEM can be used to identify both weak and noisy factors in a chosen model specification. In particular, this paper proposes obtaining a denoised version of tradable factors, which can then be employed in subsequent empirical applications. The empirical exercise carried out in this section focuses on the Fama and French Five-Factor model, but the approach can be applied to a wide range of linear factor models. The literature has already raised concerns regarding measurement error in the Fama and French factors. For instance, Racicot et al. (2011) propose incorporating correction terms for factor exposures to address the errors-in-variables problem. In contrast, this work advocates using the denoised version of observable factors—an approach applicable to any tradable factor, provided that the factor is priced.

Indeed, given any specification of a linear factor model, 3PEM can be used: (i) to detect whether the factor is weak, and if it is, to choose a model specification that excludes the weak factors; (ii) if the factor is not weak, but it is measured with errors, that is, if it is noisy, to obtain a denoised version of it.

In the 3PEM setting, a factor g_t is weak if $\eta \to 0$, or $\eta = 0$, or equivalently, if its explanatory power with respect to a rotated version of latent factors is almost equal to noise. Xiu and Giglio (2021) propose to use the signal-to-noise ratio of each observable factor, defined as follows:

$$R^2_{g,i} = \frac{[\eta \Sigma^v \eta']_{ii}}{[\Sigma^v \eta' + \Sigma^z]_{ii}}.$$

Its counterpart estimator is given by:

$$\hat{R}^2_{g,i} = \frac{[\hat{\eta}\hat{V}\hat{V}'\hat{\eta}']_{ii}}{[\bar{G}\bar{G}']_{ii}}. \tag{13}$$

R^2_g gives information about how noisy g_d is. In particular, if R^2_g is close to 1, then most of the variation in g is due to latent factors, meaning that g is strong: the factor is pervasive and measured with little noise. Conversely, a value close to zero signals that most of the variation in g_t is idiosyncratic and not linked to the underlying factor space.

Besides the signal-to-noise ratio, Xiu and Giglio (2021) also propose a formal statistical test of factor strength. The Wald statistic is constructed to test the null hypothesis that the observable factor is weak,

$$H_{0,i} : \eta_{i1} = \cdots = \eta_{i\hat{p}} = 0, \text{with } i = 1, ..., d, \tag{14}$$

against the alternative that it loads on at least one latent factor. The test is therefore designed to detect whether g_t has any systematic exposure to the latent factor space recovered in the first step of the 3PEM.

Thus, we compute the signal-to-noise ratios and formally test hypotheses on η. The evidence reported in Tables 5 and 6 consistently points to the strength of the Fama–French observable factors. The estimated signal-to-noise ratios $\hat{R}_g^2$ are generally high, indicating that most of the variation in these factors is captured by the latent components extracted in the model. The market, SMB, and HML factors clearly behave as strong factors, while RMW shows a comparatively lower value, suggesting a greater role for idiosyncratic noise. The Wald test results reinforce this interpretation: for all five factors, the null hypothesis of weakness is decisively rejected. This provides formal evidence that each observable factor significantly loads on at least one latent component. Overall, the findings imply that all Fama–French factors are statistically and economically non-weak (that is, they are priced) in the European stock market, supporting the use of this model in subsequent empirical analysis.

However, several factors appear to be affected by measurement error. As shown in Figure 8, the variance decomposition indicates that although most of the variation in the observable factors is captured by the latent components, some factors retain a non-negligible proportion of idiosyncratic noise. In particular, the HML, RMW, and CMA European factors exhibit measurement errors, with the noise component especially pronounced for RMW and CMA. This suggests that these factors are less tightly connected to the latent structure identified by the 3PEM. Similar results can be observed in Figures 9 and 10, where the actual and denoised factor observations are plotted against each other. Factors such as Mkt-RF, SMB, and display points that lie very close to the 45-degree line, indicating that their variation is largely captured by the latent factor structure and that measurement noise is limited. In contrast, the wider dispersion visible for RMW and, to a lesser extent, for CMA signals a weaker relationship with the latent factors. Their observations deviate more substantially from the 45-degree line and suggests a higher noise component and lower signal strength within the 3PEM framework.

Moreover, Figures 6 and 7 also confirm that RMW is the noisiest factor in the set since it displays greater estimation uncertainty in its corresponding risk premium. Notably, when comparing the 2PEM and 3PEM estimates, the 3PEM produces a more stable, lower-variance estimate of the RMW risk premium (which is also statistically significant under 2PEM). This contrast indicates that the 3PEM is more effective at absorbing spurious variation generated by measurement error.

To further illustrate the ability of 3PEM to distinguish economically meaningful factors from irrelevant ones, we introduce a purely spurious factor generated as random noise and subject it to the same variance-decomposition and correlation analysis as the observable Fama-French factors. The results in Figure 11 and Table 7 clearly show that this artificial factor is almost entirely orthogonal to the latent factor space, with correlations that remain close to zero across all latent components. Consistent with this, the variance decomposition assigns nearly all of its variation to the idiosyncratic noise component. This behaviour demonstrates that the 3PEM is able to correctly classify the spurious factor as non-informative since it is not aligned with any latent source of systematic risk. Therefore, these findings reinforce the suggestion to use the 3PEM to filter out irrelevant or noisy factors and confirm its theoretical advantage in settings where observable characteristics may include weak or non-priced components.

Hence, in the next part of analysis, we compute the denoised version of the observable factors according to Eq.(11) which allow us to separate the systematic component captured by the latent structure from the idiosyncratic noise associated with each factor.

4.4 Risk Exposures and Expected Values at Stock-Level

Building on prior results, which confirmed that the five Fama-French factors are strong in the European stock market and that the optimal latent dimension is six, the analysis now extends to individual stocks. The objective is to investigate how the estimation of factor loadings and expected returns varies when moving from portfolios to single assets, and to compare the performance of raw versus denoised factors within the 2PEM and 3PEM frameworks.

In this section, the analysis applies the Fama-French five-factor (FF5) model, under the restricted zero-beta rate assumption, to estimate factor loadings and expected returns for 2,327 European stocks using monthly data from January 2005 to May 2025. The study aims to compare results under two distinct setups: (i) Raw factors: Stock-level betas are estimated using the raw FF5 factors through time-series regressions, while the risk premia are those computed at the portfolio level using the 2PEM framework; (ii) Denoised factors: Stock-level betas are estimated using the denoised FF5 factors obtained via the 3PEM framework, together with risk premia computed at the portfolio level through the 3PEM. In both setups, stock-level betas are estimated using ordinary least squares (OLS) regressions corrected for heteroskedasticity and autocorrelation, with standard errors adjusted for up to six lags. Expected stock returns are subsequently derived by multiplying the estimated betas by the corresponding vector of risk premia. In the time-series regressions used to estimate stock-level factor loadings, employing denoised factors leads to a modest but consistent improvement in the model's explanatory power. The maximum adjusted R^2 rises from 0.777 to 0.812, while both the median and upper quartile values also increase, indicating that the improvement in fit is widespread across stocks rather than driven by a few outliers. Although the gain is moderate, it suggests that denoising the factors reduces idiosyncratic noise in the time-series estimation of betas, resulting in more stable and better-specified factor exposures.

Figure 12 shows the distributions of stock-level betas estimated using raw and denoised factor observations. The results indicate that, for the market and size factors, the distributions remain nearly identical, suggesting that these factors are already well captured by their observable counterparts. In contrast, for HML, RMW, and CMA, the distributions widen after denoising, exhibiting greater cross-sectional dispersion in stock exposures. This increased dispersion highlights that the denoising process uncovers meaningful heterogeneity in how individual stocks load on these factors—heterogeneity that was previously masked by noise. The effect is particularly pronounced for the profitability and investment factors, confirming that denoising identifies the true underlying exposures rather than merely increasing estimation error. Overall, this evidence supports the conclusion that denoising enhances the precision and interpretability of factor loadings, especially for factors subject to measurement error.

Figure 13 illustrates the differences in the distribution of expected monthly stock returns across estimation procedures. The 2PEM approach yields a wider dispersion with heavier tails, resulting in both overestimation and underestimation of expected returns compared 3PEM approach. These deviations are likely driven by noise contaminating the tradable factors, which raises spurious covariances for some stocks while reducing true exposures for others. Furthermore, the average Euclidean distance between the expected returns estimated using the 2PEM approach and the returns obtained using denoised factors and 3PEM risk premia is approximately 1.18% per stock. This implies that, on average, the expected monthly returns differ by about one percentage point between the two procedures. While this difference may seem modest, it can be significant in portfolio allocation decisions and may accumulate over time, affecting investment performance and the interpretation of risk–return trade-offs.

In certain cases, the discrepancies are substantially larger. For example, Figure 14 highlights individual stocks to illustrate how expected values change when denoised factors are used instead of raw factors. For stocks such as ATO–FR and IGR–GB, expected returns estimated with raw factors are biased and substantially underestimated, particularly when the raw specification yields large negative values. After denoising, expected returns adjust to more plausible levels, indicating that the 3PEM correction effectively mitigates distortions caused by measurement error. For other stocks, such as ONE–AU and HTRO–SE, the differences are smaller, reflecting that the impact of denoising depends on the extent to which each asset's exposure is affected by noise in the factor structure.

5 Conclusions

This paper provides an empirical application of the Three-Pass Estimation Method introduced by Xiu and Giglio (2021) to the European equity market, with the objective of estimating factor risk premia in the presence of omitted factors and measurement error. Using a large panel of European portfolio and stock returns, and adopting the Fama–French five-factor model as a benchmark specification, we show that traditional two-pass procedures are highly sensitive to noise in observable factors and to model misspecification. In contrast, the 3PEM delivers risk-premium estimates that are both economically plausible and statistically robust, even when the observable factor set is incomplete or contaminated by measurement error.

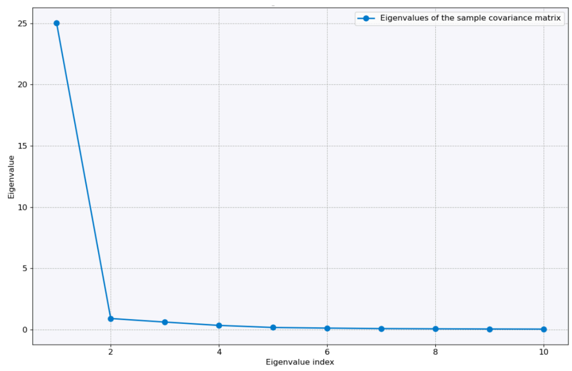
Our empirical results highlight several key advantages of the 3PEM. First, the estimated risk premia are stable across alternative observable model specifications and align closely with the time-series averages of tradable factor returns, particularly for the market factor. This stability reflects the ability of the method to account for omitted-variable bias by conditioning on a latent factor space extracted from the cross-section of returns. Second, the 3PEM provides a natural diagnostic framework to assess factor relevance. Through signal-to-noise ratios and formal Wald tests, we show that all Fama–French factors are priced in the European market, while simultaneously identifying substantial measurement error in some of them—most notably in the profitability and investment factors.

Furthermore, we demonstrate how the 3PEM can be employed as a practical tool for denoising tradable factors. By projecting observable factors onto the estimated latent space, we obtain purified factor series that more accurately capture systematic sources of risk. These denoised factors result in more precise and economically meaningful estimates of stock-level factor loadings and expected returns. Notably, using denoised factors reduces the dispersion and instability observed in the traditional two-pass framework, revealing heterogeneity in factor exposures that would otherwise be obscured by noise. The differences in expected returns across estimation procedures are economically significant, suggesting that denoising factors can have important implications for portfolio construction and asset allocation.

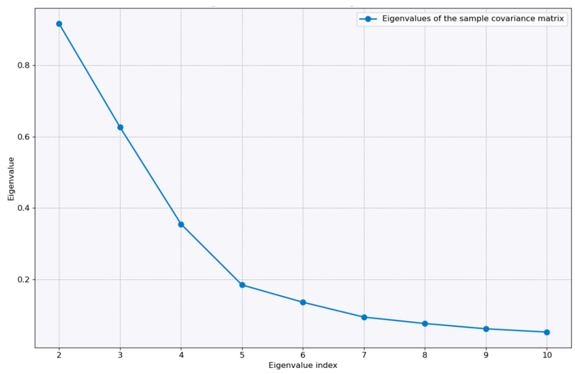
Overall, the 3PEM offers a consistent approach to factor selection, noise identification, and the development of cleaner factor inputs for subsequent applications. Although the analysis focuses on the Fama–French model and the European equity market, the methodology is entirely flexible and can be applied to different factor structures, asset classes and markets.

Figures and Tables

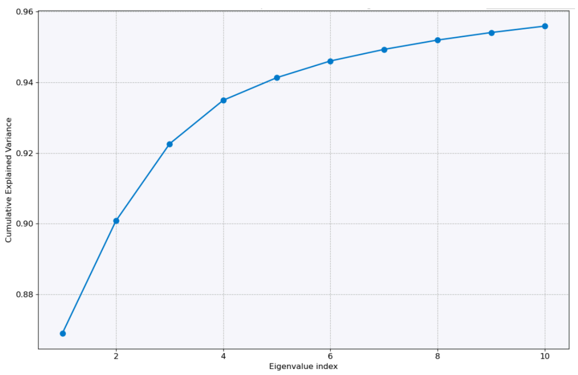
Figure 1: Eigenvalue Structure and Information Criteria for Determining the Number of Latent Factors



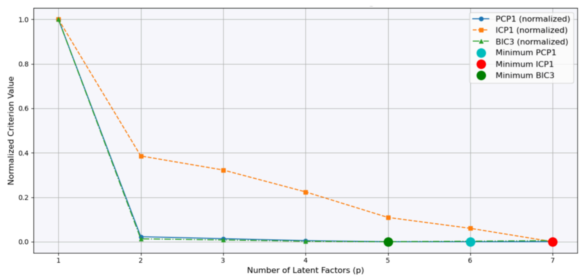
Panel A. First 10 eigenvalues distribution



Panel B. First 10 eigenvalues distribution (1st eigenvalue excluded)



Panel C. Cumulative explained variance vs. eigenvalues



Panel D. Normalized information criteria for determining the number of factors

Figure 2: Reconstruction of Demeaned Excess Returns Using PCA-Estimated Latent Factors

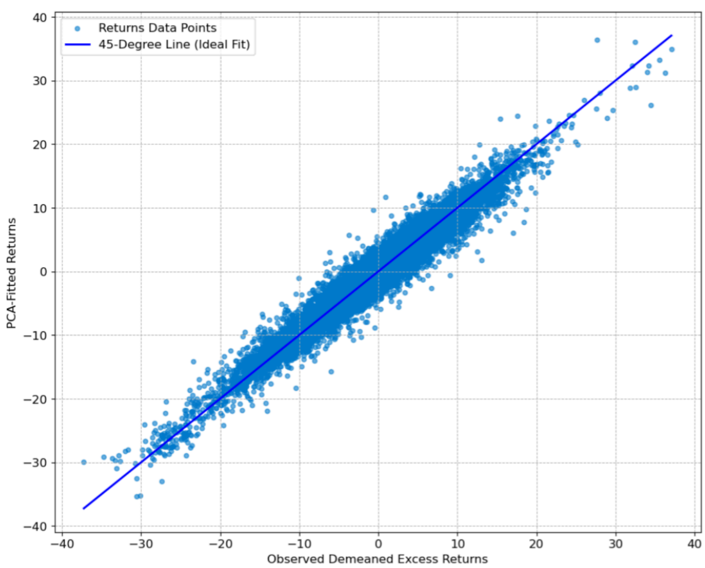


Figure 3: Cross-Sectional Distribution of First-Pass Estimated Betas within 2PEM Framework

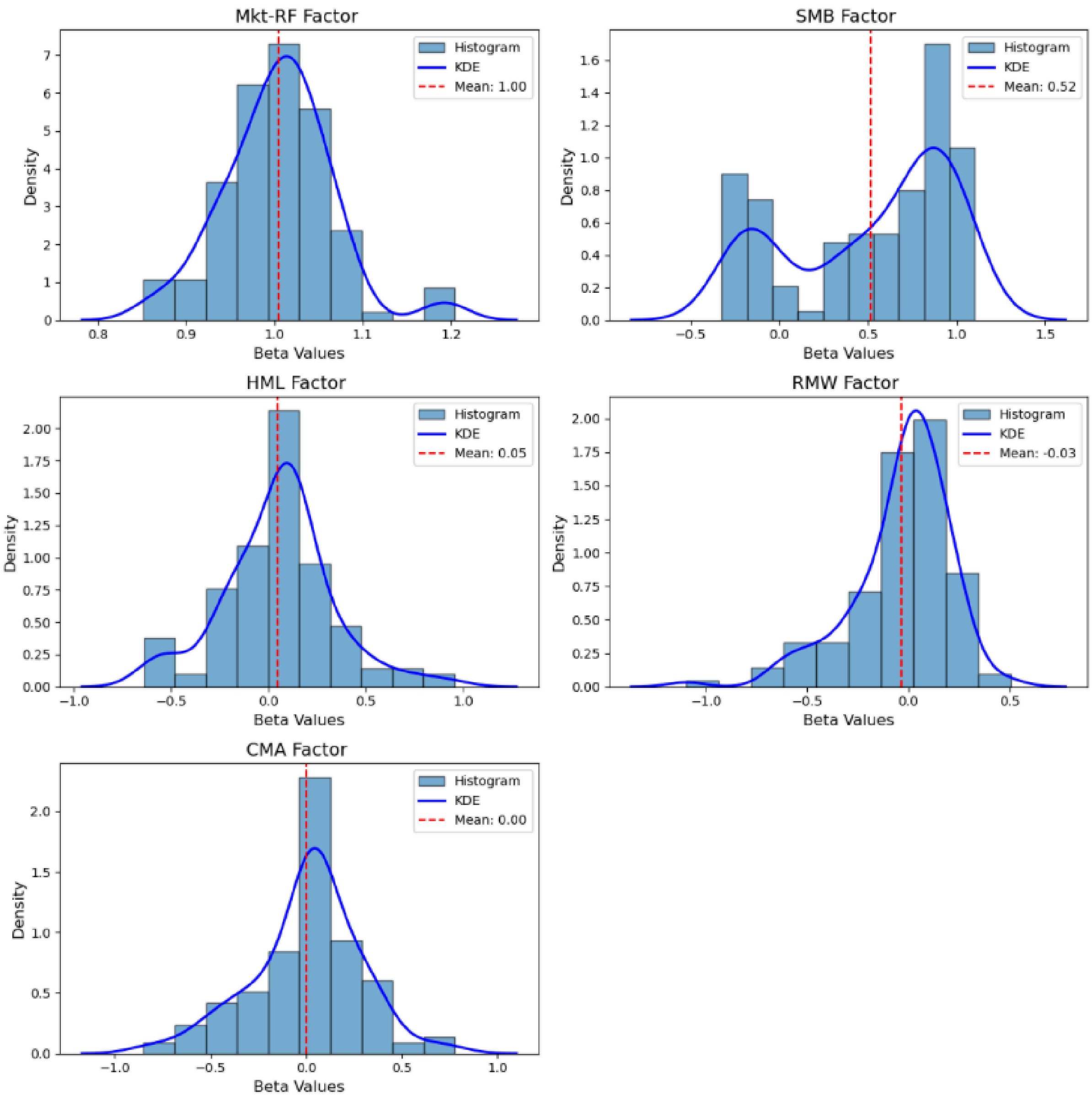


Figure 4: Scatter Plots of Average Excess Returns Against First-Pass Betas for the Five Fama–French Factors within 2PEM Framework

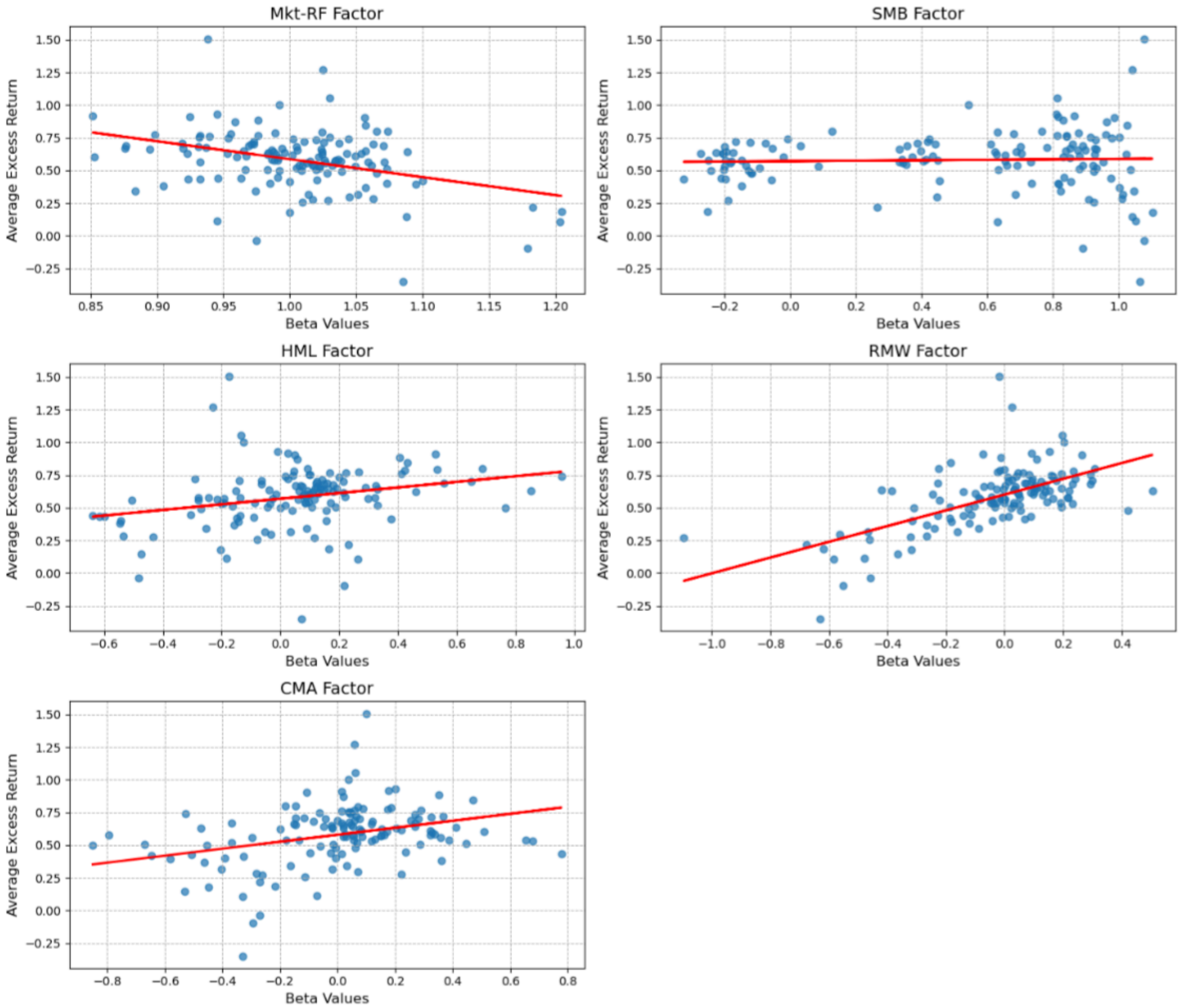


Figure 5: Dynamics of Risk Premia under Shifts in the Latent Factor Space (Restricted Case, $\hat{p} = 6$.)

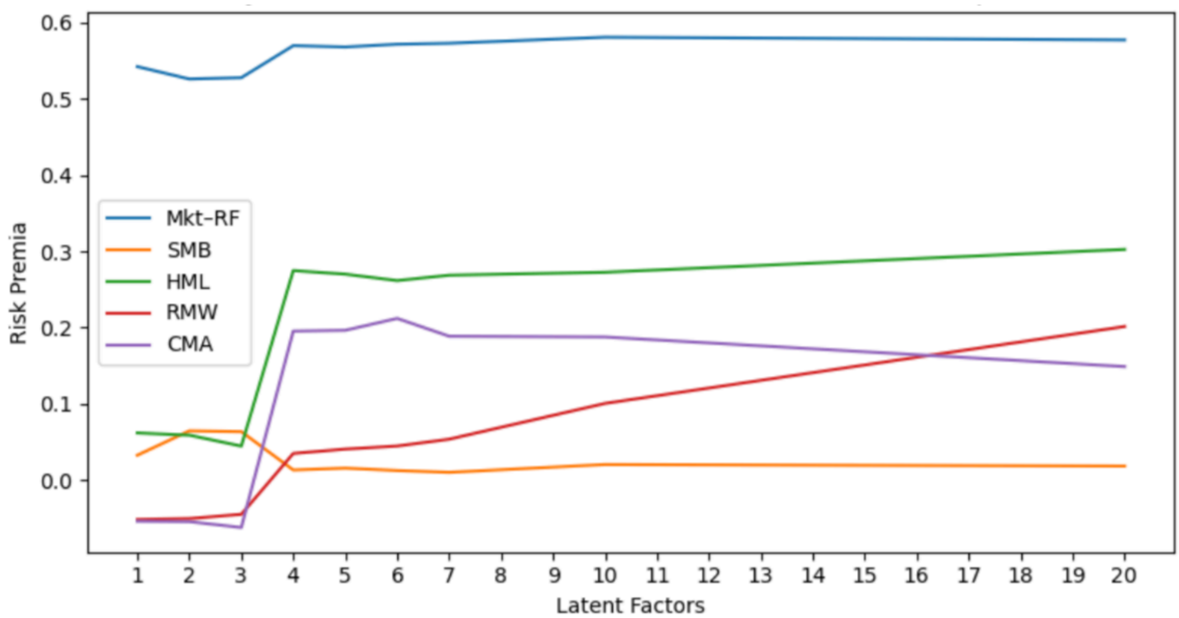
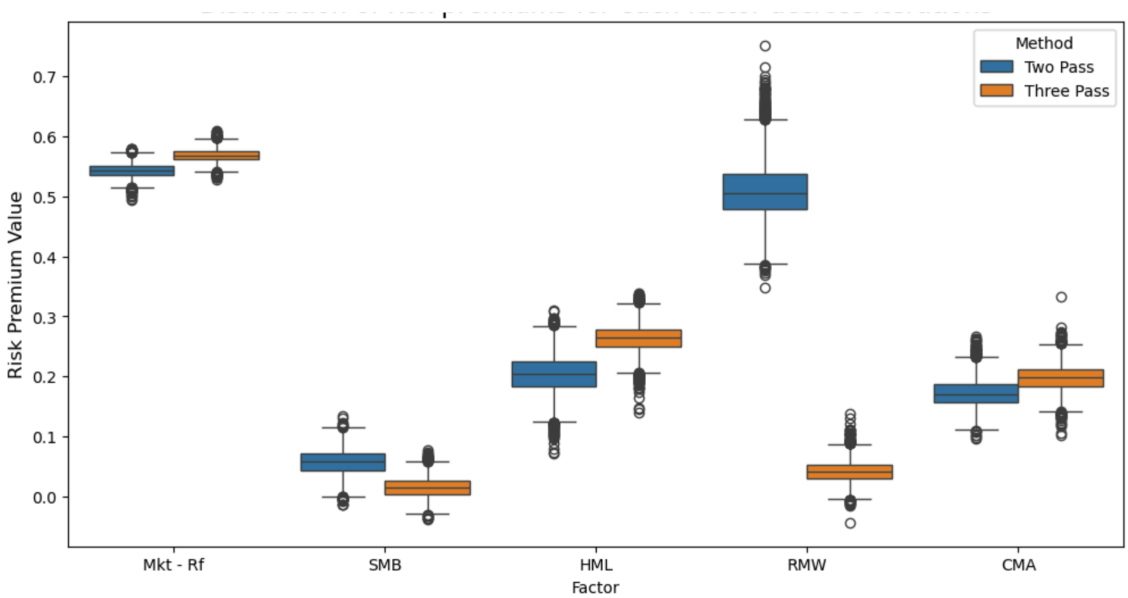
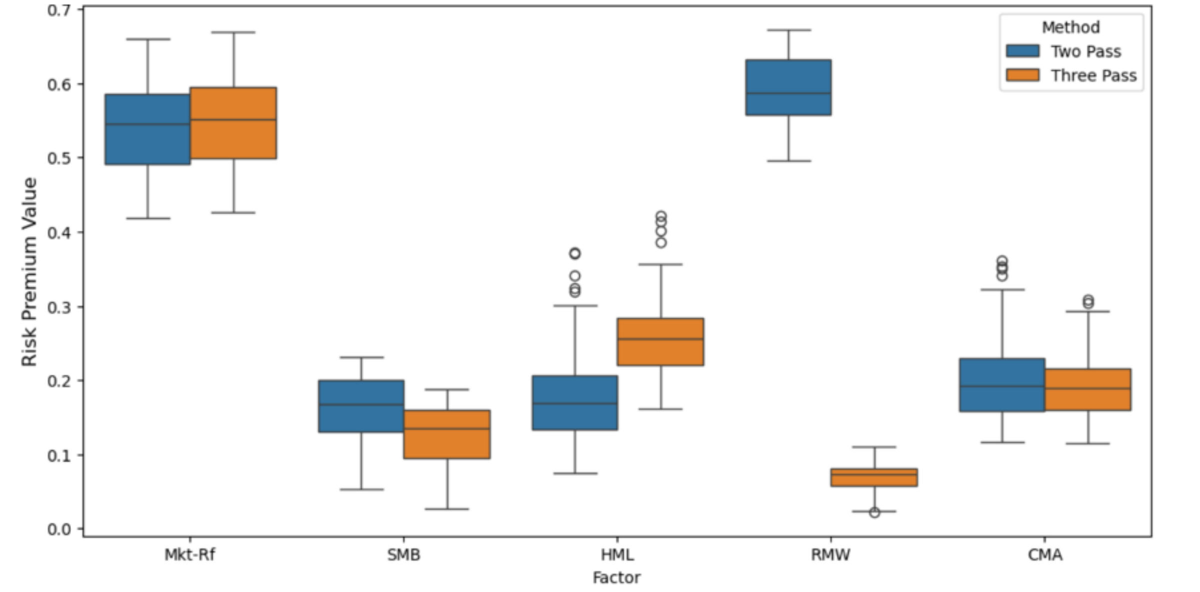


Figure 6: Risk Premia Sensitivity to Cross-Sectional Changes



Note: Each boxplot shows the empirical distribution of factor risk premia obtained via cross-sectional resampling, for the restricted zero-beta rate specification of the model. At each iteration $n = 1, \ldots, 5000$, a random subset of $N = 100$ portfolios is drawn, the two-pass and three-pass estimators are computed. For the 3PEM, the latent dimension $\hat{p}$ is case-specific. The boxplots illustrate the sensitivity of estimated risk premia to changes in the cross-sectional composition of portfolios.

Figure 7: Distribution of Factor Risk Premia: Rolling Window over the Time-Series Dimension



Note: Each boxplot shows the time-series distribution of factor risk premia obtained through a rolling-window estimation procedure. The model is estimated over overlapping windows of 300 months, shifting by one period at each step, for the restricted zero-beta rate specification. For the 3PEM, $\hat{p}$ is case-specific. This approach captures the temporal dynamics of risk premia and their stability across time.

Figure 8: Variance Decomposition for Observable Factors (Restricted case, $\hat{p} = 6$)

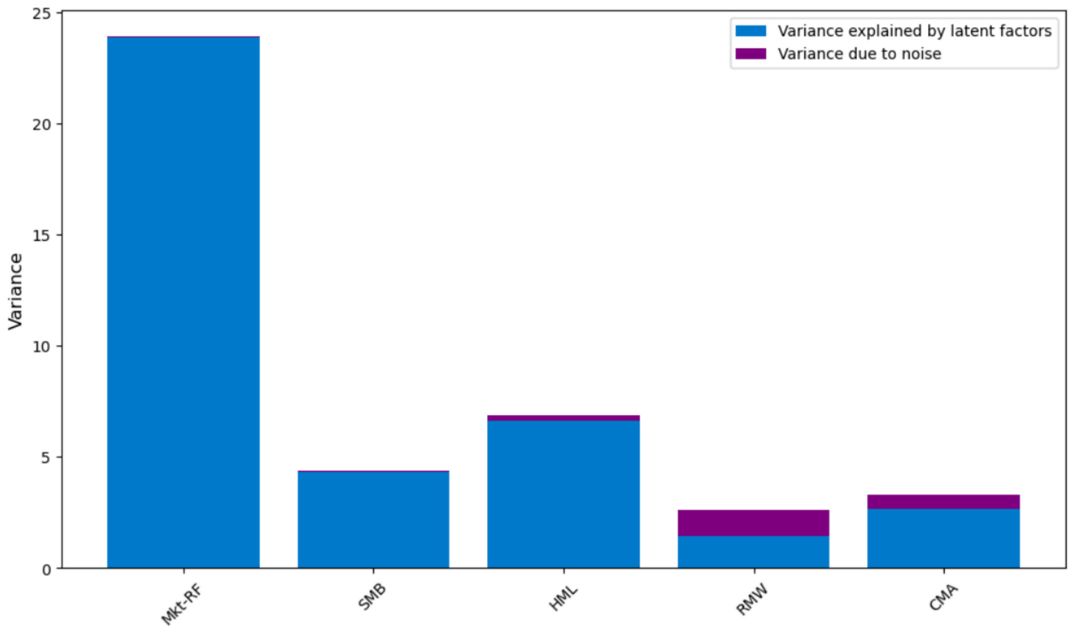


Figure 9: Actual versus Denoised Factor Observations

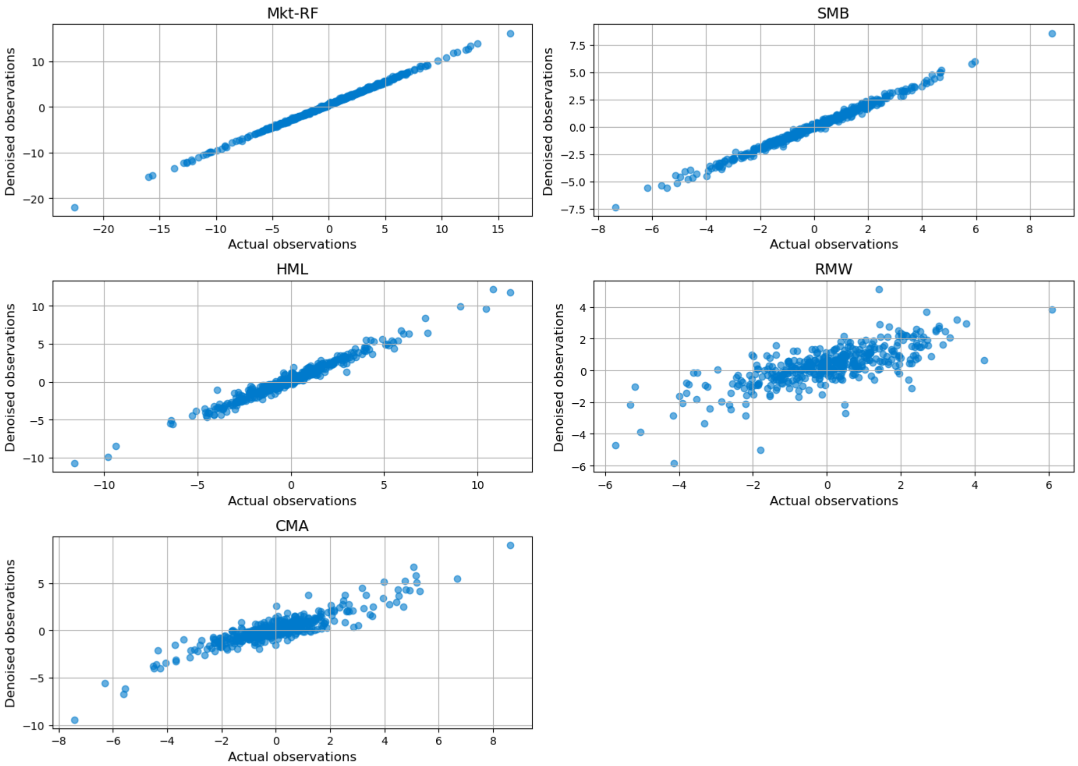


Figure 10: Actual vs. Denoised Time-Series of Observable Factors

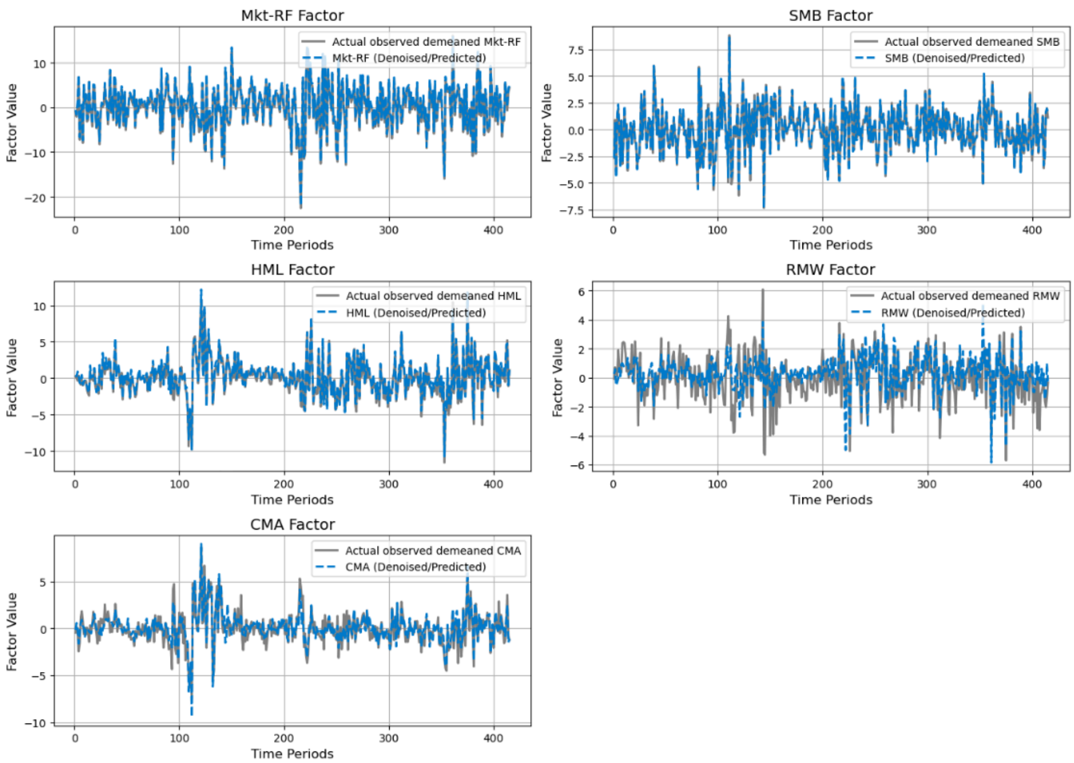


Figure 11: Detection of a Spurious Factor via Variance Decomposition

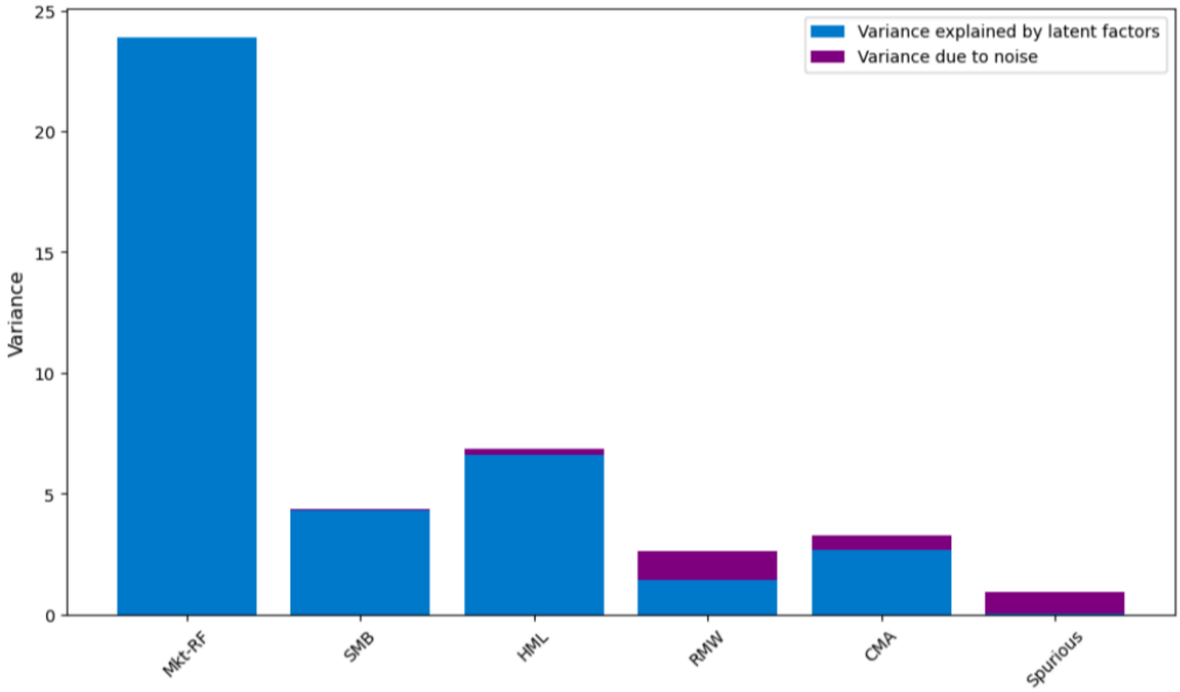


Figure 12: Beta Distributions across Stocks – Raw vs Denoised Factors

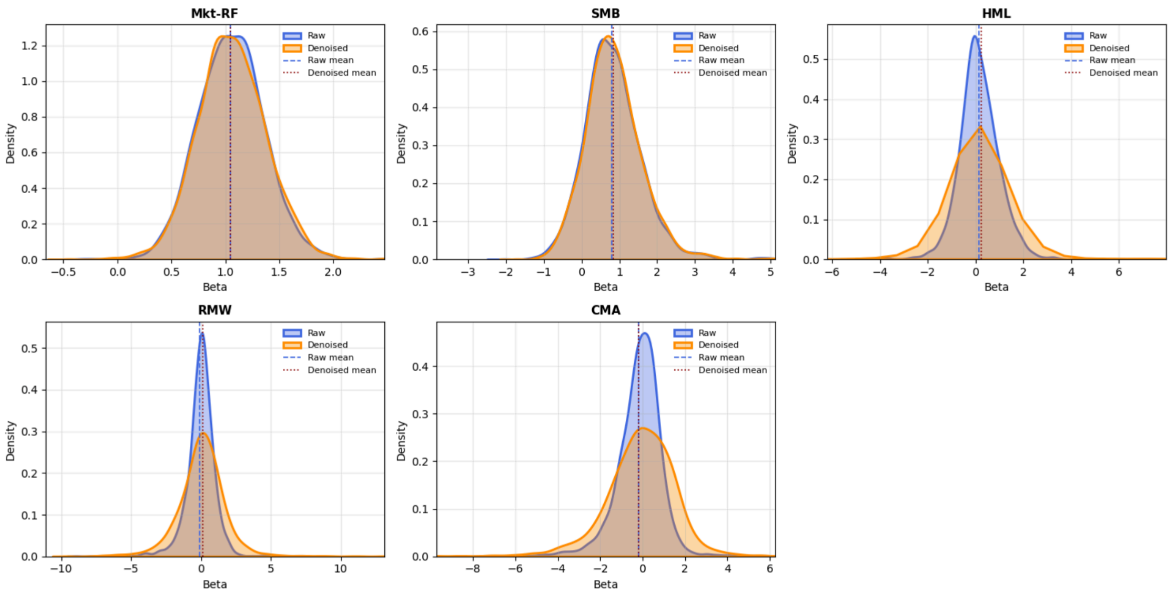


Figure 13: Distribution of Expected Stock Returns under 2PEM and 3PEM (FF5F Framework)

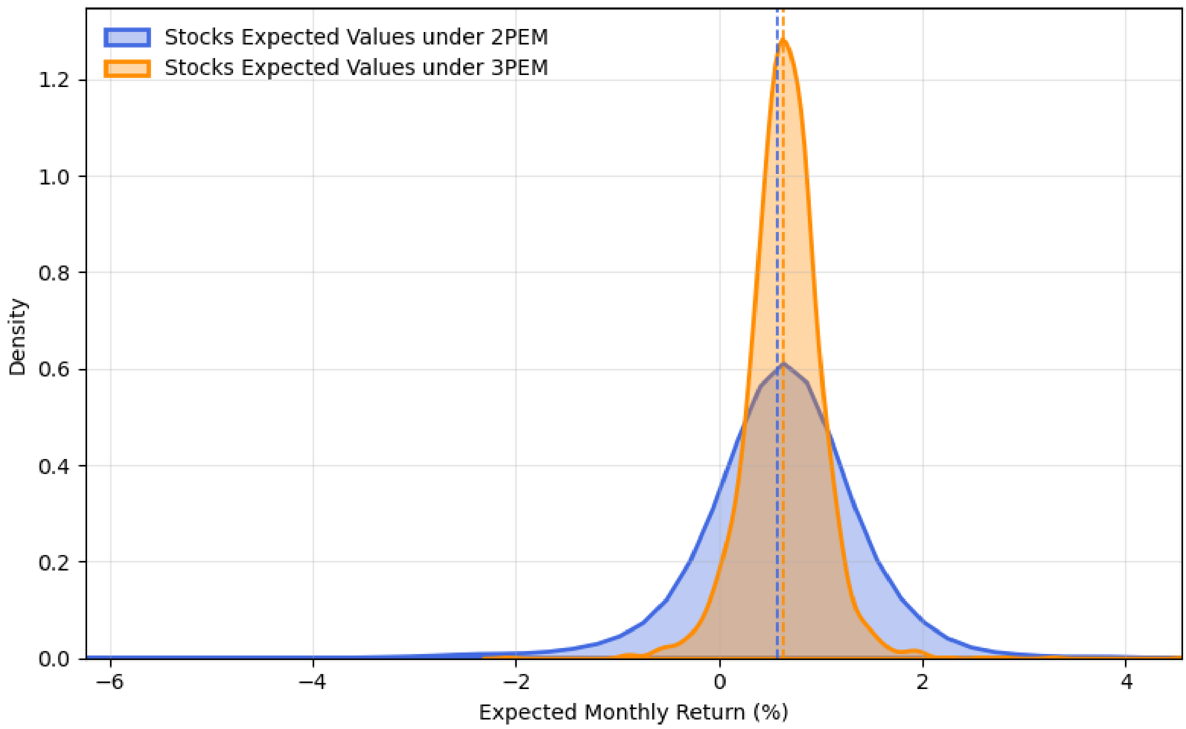


Figure 14: Illustrative Cases of Stock Excess Returns and Expected Values under Raw and Denoised Factors

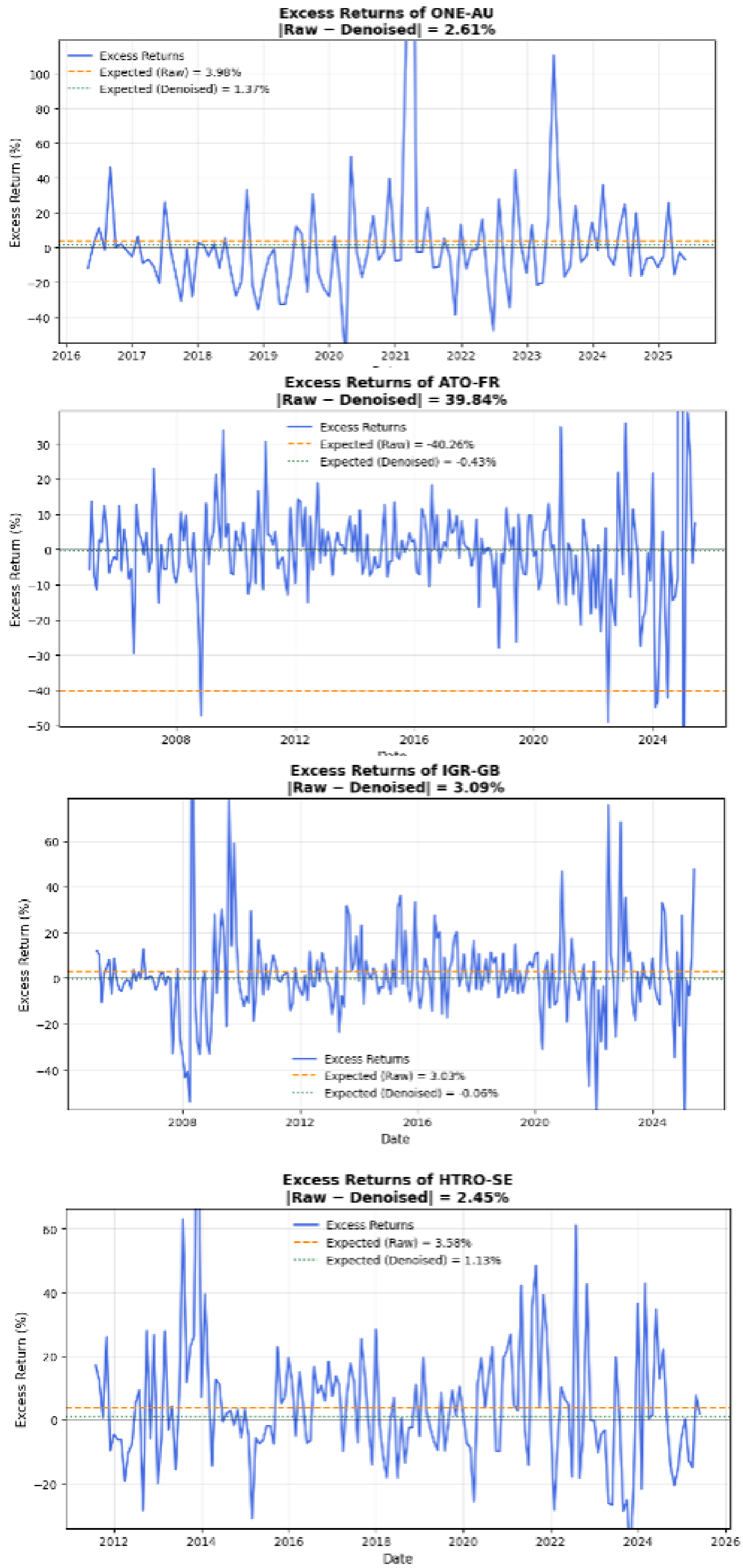


Table 1: Correlation between Latent Factors and Fama–French Observable Factors

Latent factor	Mkt–RF	SMB	HML	RMW	CMA
1st	0.9738	0.1378	0.2078	-0.2794	-0.2607
2nd	-0.2116	0.9726	-0.0700	0.0484	-0.0208
3rd	0.0591	-0.0803	-0.9038	0.5720	-0.6950
4th	-0.0240	0.0673	-0.2461	-0.1379	-0.3982
5th	0.0342	-0.1089	0.1692	-0.3348	-0.0551
6th	0.0277	-0.0549	-0.1197	0.0927	0.3206

Table 2: Risk Premia Results

	Time-Series Avg	2PEM				3PEM			
Factor	Estimate	Restricted		Unrestricted		Restricted		Unrestricted	
Mkt–RF	0.541	0.557	**	−0.411		0.571	***	0.371	*
SMB	0.018	0.054		0.050		0.013		0.015	
HML	0.319	0.205	*	0.250	*	0.262	*	0.264	*
RMW	0.314	0.594	***	0.529	***	0.045		0.048	
CMA	0.129	0.182	**	0.174	*	0.212	**	0.215	**

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All estimates refer to the restricted zero-beta rate specification. Standard errors for the 2PEM estimates are computed using the Fama–MacBeth procedure with Shanken correction. For the 3PEM, the latent structure is estimated with $\hat{p} = 6$. All risk premia are expressed in percentage terms and are computed using simple monthly returns.

Table 3: Risk Premia Estimates under 3PEM and 2PEM for Different Model Specifications

	CAPM			FF3F Model			FF5F Model		
Factor	Estimate	SE		Estimate	SE		Estimate	SE	
Model: 3PEM, restricted ($\hat{p} = 6$)									
Mkt–RF	0.571	0.260	**	0.571	0.260	**	0.571	0.260	**
SMB				0.013	0.103		0.013	0.103	
HML				0.262	0.165	*	0.262	0.165	*
RMW							0.045	0.082	
CMA							0.212	0.116	**
Model: 2PEM, restricted									
Mkt–RF	0.571	0.248	**	0.529	0.243	**	0.557	0.258	**
SMB				0.044	0.105		0.054	0.111	
HML				0.184	0.134	*	0.205	0.142	*
RMW							0.594	0.115	***
CMA							0.182	0.099	**

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All model setups refer to the restricted zero-beta rate case with $\hat{p} = 6$ latent factors. 2PEM standard errors are Fama–MacBeth estimates adjusted using the Shanken correction. All risk premia are expressed in percentage terms and are computed using simple monthly returns.

Table 4: 3PEM Risk Premia at Changing the Number of Latent Factors

	$\hat{p}=1$			$\hat{p}=2$			$\hat{p}=3$			$\hat{p}=4$			$\hat{p}=5$			$\hat{p}=6$			$\hat{p}=7$			$\hat{p}=10$			$\hat{p}=20$		
Factor	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.	Est.	SE	Sig.
Mkt–RF	0.542	0.274	**	0.526	0.258	**	0.528	0.259	**	0.570	0.259	**	0.568	0.260	**	0.571	0.260	**	0.573	0.261	**	0.581	0.262	**	0.577	0.262	**
SMB	0.033	0.021	*	0.065	0.104		0.064	0.104		0.013	0.105		0.016	0.104		0.013	0.103		0.010	0.102		0.021	0.102		0.019	0.102	
HML	0.062	0.043	*	0.059	0.040	*	0.045	0.156		0.275	0.170	*	0.270	0.168	*	0.262	0.165	*	0.269	0.165	*	0.273	0.165	**	0.303	0.171	**
RMW	-0.052	0.030		-0.050	0.029		-0.045	0.070		0.035	0.085		0.041	0.083		0.045	0.082		0.054	0.080		0.101	0.079		0.201	0.083	***
CMA	-0.054	0.025		-0.054	0.026		-0.062	0.087		0.195	0.111	**	0.196	0.113	**	0.212	0.116	**	0.189	0.124	*	0.188	0.125	*	0.149	0.123	

Notes: Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All estimates are based on the restricted zero-beta rate specification. All risk premia are expressed in percentage terms and are computed using simple monthly returns.

Table 5: Estimated Signal-to-Noise Ratios ($\hat{R}_g^2$) for Fama and French Factors

Factor	$\hat{R}_g^2$	Interpretation
Mkt–Rf	0.999	Strong
SMB	0.991	Strong
HML	0.968	Strong
RMW	0.547	Moderate
CMA	0.816	Strong

Notes: The table reports the estimated signal-to-noise ratios $\hat{R}_g^2$ for the Fama and French five observable factors. A value close to one indicates that the variation in the factor is largely explained by latent components, implying a strong factor, whereas lower values suggest higher idiosyncratic noise.

Table 6: Row-wise Wald Tests for Weak Observable Factors (Fama and French Factors)

Factor	$W_{i,T}$
MKT–RF	408,143.84
SMB	33,407.41
HML	7,880.41
RMW	303.48
CMA	1,159.24

Notes: The table reports the Wald test statistics for the null hypotheses $H_{0,i} : \eta_{i1} = \eta_{i2} = \ldots = \eta_{i\hat{p}} = 0$, testing whether each observable factor i is weak. η_{il} are the coefficients linking observable factor i to the $\hat{p} = 6$ latent factors. The 5% critical value corresponds to the 95% quantile of the χ_6^2 distribution, equal to 12.5916. For all factors, the associated p-values are extremely small (numerically equal to zero due to machine precision), implying a strong rejection of the null hypothesis and indicating that all observable factors are statistically non-weak.

Table 7: Correlation Between Latent Factors and Observable Factors, Including a Spurious Factor

Latent factor	Mkt–RF	SMB	HML	RMW	CMA	Spurious
1st	0.9738	0.1378	0.2078	-0.2794	-0.2607	-0.0489
2nd	-0.2116	0.9726	-0.0700	0.0484	-0.0208	-0.1318
3rd	0.0591	-0.0803	-0.9038	0.5720	-0.6950	-0.0062
4th	-0.0240	0.0673	-0.2461	-0.1379	-0.3982	0.0893
5th	0.0342	-0.1089	0.1692	-0.3348	-0.0551	0.0581
6th	0.0277	-0.0549	-0.1197	0.0927	0.3206	0.0882

References

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. Econometrica, 81(3), 1203–1227.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70(1), 191–221.

Bryzgalova, S. (2015). Spurious factors in linear asset pricing models.

Chincarini, L., & Kim, D. (2006). Quantitative equity portfolio management: An active approach to portfolio construction and management. McGraw-Hill.

Cochrane, J. H. (2005). Asset pricing (Revised Edition). Princeton University Press.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), 1–22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. Journal of Political Economy, 81(3), 607–636.

French, K. R. (2023). Data library [Accessed: 2024-09-25]. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Gagliardini, P., Ossola, E., & Scaillet, O. (2019). A diagnostic criterion for approximate factor structure. Journal of Econometrics, 212(2), 503–521.

Gagliardini, P., Ossola, E., & Scaillet, O. (2020). Estimation of large dimensional conditional factor models in finance. In S. N. Durlauf, L. P. Hansen, J. J. Heckman, & R. L. Matzkin (Eds.), Handbook of econometrics (pp. 219–282, Vol. 7A). Elsevier. https://doi.org/10.1016/bs.hoe.2020.10.001

Lazzari, V., & Navone, M. (2003). The selection ability of italian mutual fund.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics, 47(1), 13–37.

Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91. Retrieved August 7, 2024, from http://www.jstor.org/stable/2975974

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. Review of Economics and Statistics, 92(4), 1004–1016.

Racicot, F.-É., Thèoret, R., & Coën, A. (2011). A new empirical version of the fama and french model based on the hausman specification test: An application to hedge funds. Journal of Derivatives Hedge Funds, 16(4), 278–302. https://doi.org/10.1057/jdhf.2010.22

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), 341–360.

Shanken, J. (1992). On the estimation of beta-pricing models. The Review of Financial Studies, 5(1), 1–33.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425–442.

Xiu, D., & Giglio, S. (2021). Asset pricing with omitted factors. Journal of Political Economy, 129(7), 1947–1990.