



Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Long-History Principal Component Analysis in a Dynamic Factor Model with Weak Loadings

Robert M. Anderson, Baeho Kim, Dean Ryu

To cite this article:

Robert M. Anderson, Baeho Kim, Dean Ryu (2026) Long-History Principal Component Analysis in a Dynamic Factor Model with Weak Loadings. *Operations Research*

Published online in Articles in Advance 11 Mar 2026

<https://doi.org/10.1287/opre.2024.1134>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. You are free to download this work and share with others for any purpose, except commercially, and you must attribute this work as “*Operations Research*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/opre.2024.1134>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc/4.0/>.”

Copyright © 2026 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes. For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Contextual Areas

Long-History Principal Component Analysis in a Dynamic Factor Model with Weak Loadings

Robert M. Anderson,^{a,b} Baeho Kim,^{c,*} Dean Ryu^d

^aHarbin Institute of Technology, Harbin 150001, China; ^bUniversity of California Berkeley, Berkeley, California 94720; ^cKorea University Business School, Seoul 02841, Republic of Korea; ^dDepartment of Business Administration at Instituto Tecnológico Autónomo de México (ITAM), Mexico City 01080, Mexico

*Corresponding author

Contact: robert.anderson@berkeley.edu,  <https://orcid.org/0000-0002-4674-1088> (RMA); baehokim@korea.ac.kr,  <https://orcid.org/0000-0001-5069-5996> (BK); dean.ryu@sbs.ox.ac.uk,  <https://orcid.org/0009-0005-0347-4951> (DR)

Received: June 23, 2024

Revised: May 15, 2025; December 8, 2025;
January 6, 2026

Accepted: January 13, 2026


Published Online in Articles in Advance:
March 11, 2026

Area of Review: Financial Engineering

<https://doi.org/10.1287/opre.2024.1134>

Copyright: © 2026 The Author(s)

Abstract. Estimated covariance and precision matrices of asset returns significantly influence the set of portfolios compliant with risk budgets and their potential losses. Statistical risk modeling approaches often assume temporal stability for consistency with a static factor structure, typically estimated within $T_M \sim 250$ days of data history, resulting in finite-sample estimation error when the dimension of the population exceeds the number of observations. Our study investigates the application of Principal Component Analysis (PCA) over extended data histories (e.g., $T_L \sim 1,500$ trading days), an approach we term Long-History PCA, to forecast the daily risk profile based on dynamic factor structures with heterogeneous factor strengths. The use of a longer data history mitigates excess dispersion bias in the estimated factor loadings, particularly in the presence of weak factors. As shown in simulations and empirical data from the United States and European stock markets, our approach substantially mitigates second-order risk bias compared with traditional methods using medium horizons (T_M), both with and without the augmentation of Responsive Covariance Adjustment using a short half-life (T_S) of 40 days.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. You are free to download this work and share with others for any purpose, except commercially, and you must attribute this work as “Operations Research. Copyright © 2026 The Author(s). <https://doi.org/10.1287/opre.2024.1134>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by-nc/4.0/>.”

Funding: R. M. Anderson gratefully acknowledges the financial support of Swiss Re through the Consortium for Data Analytics in Risk. B. Kim’s research is partially supported by a Korea University Business School Research Grant.

Supplemental Material: All supplemental materials, including the code, data, and files required to reproduce the results, are available at <https://doi.org/10.1287/opre.2024.1134>.

Keywords: dynamic factor model • long history • principal component analysis • second order risk • weak loadings

1. Introduction

In the classical mean-variance framework, second-order risk (SOR) arises from the uncertain nature of the joint distribution of asset returns, particularly due to inaccuracies in estimating the population covariance matrix (Σ) or its inverse, the precision matrix (Σ^{-1}). Small errors in these estimates can significantly distort efficient allocations in portfolio optimization,¹ potentially biasing the optimized portfolio toward the deficiencies of the estimated model (Shepard 2009, Bernardi et al. 2019). Consequently, understanding and mitigating SOR bias is crucial for investors seeking admissible portfolio compositions that align with a predefined risk budget.

Nevertheless, quantifying and managing SOR remains challenging due to the unobservable and rapidly evolving

nature of the true covariance structure of asset returns. Statistical factor-based risk modeling approaches, such as Principal Component Analysis (PCA), often rely on assumptions of temporal stability and uniform factor strengths² to ensure consistent factor identification. These assumptions frequently fail to capture the dynamic complexities of real-world financial data. As a result, practitioners tend to limit their analysis to medium-length data histories (e.g., $T_M \sim 250$ or 500 trading days)³ to estimate factor exposures, while using short half-life techniques, such as the Exponentially Weighted Moving Average (EWMA) with a half-life of $T_S \sim 40$ days, to estimate current factor variances.⁴ For portfolios of practical relevance, which often include $N \sim 2,000$ or more stocks, this mismatch leads to an

inaccurate representation of the underlying covariance structure and exacerbates the SOR bias in portfolio optimization.

To address this challenge, this paper introduces the Long-History PCA (LH-PCA), a method designed to accurately forecast the precision matrix on a daily basis using an extended historical window (e.g., $T_L \sim 1,500$ trading days). It is important to clarify that LH-PCA is not a new statistical method or estimator. Rather, it represents the application of conventional PCA to a significantly longer history of data than is typical in financial risk modeling. Our primary contribution is the theoretical and empirical insight that this application yields useful and robust factor estimates, even when the underlying factor model exhibits time variation and weak factor loadings. We use the label “LH-PCA” as a convenient shorthand to refer to this specific application and its associated findings.

Building on the seminal contributions of Bates et al. (2013) and Bai and Ng (2023), we propose a dynamic factor model that accounts for variable factor volatilities and temporal instability in factor loadings, allowing for variations in their strength. We prove that PCA consistently estimates the underlying factor structure in the “large- N and large- T ” framework, enabling precise risk forecasts for portfolios of practical and reasonable size. Beyond theoretical advancements, our study makes practical contributions by developing a robust framework for detecting and quantifying second-order risk bias in high-dimensional portfolios. Because the true return covariance structure is unobservable in practice, we propose an observable proxy of the SOR bias by constructing the Global Minimum Variance Portfolio (GMVP) using the empirically estimated precision matrix and comparing its out-of-sample realized volatility to its predicted volatility. This approach enables the development of reliable bias statistics that capture discrepancies between estimated and actual portfolio risk, providing an effective tool for risk-aware investment decisions.

Empirically, we validate the effectiveness of LH-PCA through Monte Carlo simulations and real-world stock market data, including U.S. and European equities. Compared with traditional benchmarks, such as variants of PCA with medium-length historical windows and shrinkage techniques (Ledoit and Wolf 2004b), we demonstrate that LH-PCA significantly reduces second-order risk bias by extending the data history from T_M to T_L . This improvement remains robust across diverse scenarios, including the use of responsive covariance adjustments (RCA) with short data histories (e.g., EWMA with a half-life of $T_S \sim 40$ days). Our findings confirm that LH-PCA enhances risk estimation accuracy in fast-evolving financial markets, establishing it as a valuable tool for high-dimensional portfolio construction and risk management.

Overall, the proposed LH-PCA approach provides a robust and intuitive framework for analyzing high-dimensional financial data sets, particularly for portfolio optimization in the context of modern big data analytics. By effectively reducing dimensionality while retaining key factors, it facilitates accurate estimation of covariance and precision matrices, which are critical for constructing resilient and well-diversified portfolios. The adaptability of the proposed approach to both strong and weak factors, as well as dynamic and static loadings, renders it well-suited for the modeling of pragmatic dynamic factor structures, thereby addressing the limitations of traditional approaches.

2. Related Literature

Our study contributes to the literature on portfolio risk optimization, particularly in improving covariance matrix estimation for large portfolios. Traditional research on estimation errors in the Markowitz framework (Markowitz 1952) has largely focused on portfolio allocation strategies, such as constructing minimum-variance portfolios. Among various asset allocation models (see DeMiguel et al. 2009b, table 1), GMVP has been widely studied due to its ability to minimize risk while avoiding reliance on noisy mean return estimates (Jagannathan and Ma 2003). Moreover, GMVPs serve as foundational building blocks for other portfolio strategies (DeMiguel et al. 2009a) and are favored by investors for their attractive risk-adjusted returns. Recent refinements to GMVP have been proposed by Basak et al. (2009), Shi et al. (2020), and Ding et al. (2021).

Our approach diverges from this literature by focusing on daily forecasting of the precision matrix for large portfolios, a critical component in managing SOR. To achieve this, we emphasize identifying all relevant risk factors, regardless of their risk premiums. Including factors with negative or zero risk premiums can inadvertently increase portfolio volatility without enhancing expected returns. Although we do not assess whether factors are priced—because a factor can influence volatility without carrying a risk premium—our findings on SOR contribute to recent asset pricing discussions (Anderson et al. 2009, Bekaert et al. 2009a, Brenner and Izhakian 2018, Ait-Sahalia et al. 2025). Given the need for robust risk estimation, we leverage PCA-based methods, which are computationally efficient in capturing market dynamics.

We also build upon research improving PCA methodologies. Risk Premium-PCA (Lettau and Pelger 2020) extracts weak factors with high Sharpe ratios, Projected-PCA (Kim et al. 2021) constructs portfolios that hedge systematic risks, and Instrumental-PCA (Kelly et al. 2019) identifies latent factors explaining cross-sectional stock returns. These methods estimate

low-dimensional factor structures under Arbitrage Pricing Theory (Ross 1976), prioritizing priced risks but often overlooking nonbenchmark assets' time-series information (Pastor 2000), which is central to our study. Although fundamental risk factors such as 'small minus big' and 'high minus low' (Fama and French 1993) are interpretable, they often exhibit strong correlations. In contrast, PCA extracts orthogonal latent factors that efficiently capture market trends, despite limited interpretability.⁵ Given that capturing short-term trends—whether they persist long enough to be integrated into a fundamental factor model—is crucial for dynamic investment strategies, PCA-based models play a key role in adapting to evolving market conditions (Fan et al. 2016, Lettau and Pelger 2020).

The literature on Dynamic Factor Models (DFMs) has evolved to address challenges in modeling time-varying dynamics in high-dimensional data. Early contributions, such as Del Negro and Otrok (2008), introduced models with time-varying factor loadings and stochastic volatility, shedding light on international business cycle fluctuations. Su and Wang (2017) extended this work with a nonparametric approach to estimate DFMs with smoothly varying loadings, enabling tests for structural changes. More recently, Pelger and Xiong (2022b) developed an inferential framework for state-varying factor models, improving the estimation of time-dependent factor structures, while Barigozzi et al. (2021) advanced estimation methods for locally stationary DFMs, emphasizing financial connectedness and heterogeneous shock responses. Complementing these studies, our work leverages DFMs to estimate covariance and precision matrices for asset returns, a key component of portfolio risk management. Our proposed LH-PCA utilizes extended data histories to mitigate estimation errors from dispersion bias. This approach enhances traditional medium-horizon approaches by reducing second-order risk bias, as we elaborate on in subsequent sections.

Finally, our work complements research on estimation error in covariance matrices, including model-free shrinkage methods (Ledoit and Wolf 2003, 2004a, 2004b), eigenvalue corrections (Ledoit and P ech e 2011, Wang and Fan 2017), and eigenvector adjustments (Goldberg et al. 2022). By addressing the challenges of SOR in high-dimensional portfolios, our study provides a robust framework for improving precision matrix estimation and enhancing portfolio optimization.

3. Problem Formulation

Throughout the paper, we define the matrix norm of a matrix A as the Frobenius norm, represented by $\|A\|_F = \sqrt{\text{tr}(A^T A)}$. Given $b \in \mathbb{R}^m$, we define $\text{diag}(b)$ to be the $m \times m$ matrix with the elements of b on the diagonal and zeros off the diagonal, while $\text{diag}(B)$ for some

$B \in \mathbb{R}^{m \times m}$ represents the vector in \mathbb{R}^m whose elements are the diagonal elements of B . We define the volatility function $\sigma(\gamma, C) = \sqrt{\gamma^T C \gamma} \geq 0$ as the volatility of a given portfolio $\gamma \in \mathbb{R}^{N \times 1}$ calculated under a covariance matrix $C \in \mathbb{R}^{N \times N}$. As usual, $I_N \in \mathbb{R}^{N \times N}$ represents an identity matrix, and $\mathbf{1} \in \mathbb{R}^{N \times 1}$ denotes the vector of ones, where all N entries are identical and equal to one.

3.1. Preliminaries

Over a finite time window of length $T > 0$ days, we observe the dynamics of N tradable security returns in the market, where $X_t \in \mathbb{R}^{N \times 1}$ represents the column vector of daily security returns on each date $t \in \{1, \dots, T\}$. The true population covariance matrix and its estimate of

$$\mathbf{X} = (X_1, \dots, X_T)^T \in \mathbb{R}^{T \times N},$$

are indicated by $\Sigma \in \mathbb{R}^{N \times N}$ and $\hat{\Sigma} \in \mathbb{R}^{N \times N}$, respectively.

Estimating the precision matrix (the inverse of the covariance matrix) of stock returns is crucial in optimizing asset allocation to identify assets that reduce overall portfolio volatility and improve the effectiveness of hedging strategies. Although the precision matrix enables more informed decision making by balancing risk and return in a portfolio, misestimating it can lead to flawed conclusions, suboptimal portfolio construction, and increased exposure to unintended risks. Recognizing the critical role of the precision matrix managing portfolio-level risk, we define the SOR bias matrix as $\epsilon = \hat{\Sigma}^{-1} - \Sigma^{-1} \in \mathbb{R}^{N \times N}$ and set the SOR bias measure as

$$(\text{SOR Bias}) = \left\| \frac{1}{N} (\hat{\Sigma}^{-1} - \Sigma^{-1}) \right\|_F := \|\epsilon\|, \quad (1)$$

where the average Frobenius norm serves as a measure of the average error within the entries of the estimated precision matrix. Our objective is to minimize the SOR bias measured by $\|\epsilon\|$ in (1).

A fundamental challenge in detecting and addressing SOR bias lies in its measurement because the true precision matrix Σ^{-1} is unobservable in practice. We therefore need an observable proxy to quantify the extent of SOR bias resulting from the misrepresentation in $\hat{\Sigma}^{-1}$. A reasonable approach is to construct optimized portfolios (such as the GMVP) based on the estimated $\hat{\Sigma}^{-1}$ and evaluate their out-of-sample performance by comparing it to the in-sample prediction.⁶ In the absence of the short-selling constraint within the classical mean-variance framework, the true GMVP (ω) and the estimated GMVP ($\hat{\omega}$) can be expressed as $\omega = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}$ and $\hat{\omega} = \frac{\hat{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T \hat{\Sigma}^{-1} \mathbf{1}}$, which are fully characterized by Σ^{-1} and $\hat{\Sigma}^{-1}$, respectively.

Then, the true GMVP variance of ω is given by

$$\begin{aligned} (\text{True variance of } \omega) &= \sigma^2(\omega, \Sigma) \\ &= \omega^\top \Sigma \omega = \frac{1}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}}. \end{aligned}$$

In addition, the (in-sample) predicted variance of the estimated $\hat{\omega}$ takes the form of

$$\begin{aligned} (\text{Predicted variance of } \hat{\omega}) &= \sigma^2(\hat{\omega}, \hat{\Sigma}) \\ &= \hat{\omega}^\top \hat{\Sigma} \hat{\omega} = \frac{1}{\mathbf{1}^\top \hat{\Sigma}^{-1} \mathbf{1}}, \end{aligned}$$

whereas the (out-of-sample) actual variance of $\hat{\omega}$ to be realized from Σ is derived as

$$\begin{aligned} (\text{Actual variance of } \hat{\omega}) &= \sigma^2(\hat{\omega}, \Sigma) \\ &= \hat{\omega}^\top \Sigma \hat{\omega} = \frac{\mathbf{1}^\top \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \mathbf{1}}{(\mathbf{1}^\top \hat{\Sigma}^{-1} \mathbf{1})^2}. \end{aligned}$$

3.2. Excess Dispersion Bias from Finite Sample Error

Statistical risk modeling approaches typically assume temporal stability to ensure consistency with a static risk structure over the estimation horizon. However, when the number of observations (T) is significantly less than the dimension (N) of $\mathbf{X} \in \mathbb{R}^{T \times N}$, finite-sample estimation error arises, and the estimated model is subject to the SOR bias. When $N \gg T$, for example, the sample covariance matrix $\frac{\mathbf{X}^\top \mathbf{X}}{T}$ is singular,⁷ so its inverse matrix is not well-defined, and the GMVP is not unique and has estimated variance of zero. One of the most widely used methods to mitigate this bias is shrinkage estimation of the sample covariance matrix, as proposed by Ledoit and Wolf (2004b), which serves as one of our benchmark methods.

In practice, variants of PCA estimation for linear factor models are typically applied using a *medium* horizon of one or two years of historical data to ensure the temporal stability of the factor structure. Assuming $r < \min\{N, T\}$ where r represents the true number of factors, the factor structure is often represented by a *static* linear factor model in the form of

$$\underbrace{\mathbf{X}}_{(T \times N)} = \underbrace{\mathbf{F}}_{(T \times r)} \underbrace{\boldsymbol{\Lambda}^\top}_{(r \times N)} + \underbrace{\mathbf{e}}_{(T \times N)},$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times r}$ represents the static loadings of the factor returns $\mathbf{F} = [F_1, \dots, F_T]^\top \in \mathbb{R}^{T \times r}$, and $\mathbf{e} \in \mathbb{R}^{T \times N}$ denotes the idiosyncratic return matrix.

Let $\text{rank}(\mathbf{X})$ be the full-rank of \mathbf{X} and the time-normalized data $\mathbf{Z} = \frac{\mathbf{X}}{\sqrt{T}}$ admit the singular value decomposition expressed as

$$\mathbf{Z} = \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r^\top + \sum_{j=r+1}^{\text{rank}(\mathbf{X})} \mathbf{d}_j \mathbf{u}_j \mathbf{v}_j^\top,$$

where $\mathbf{D}_r = \text{diag}(\mathbf{d}_1, \dots, \mathbf{d}_r) \in \mathbb{R}^{r \times r}$ contains the top r

singular values of \mathbf{Z} in descending order, and $\mathbf{U}_r = [\mathbf{u}_1, \dots, \mathbf{u}_r] \in \mathbb{R}^{T \times r}$ and $\mathbf{V}_r = [\mathbf{v}_1, \dots, \mathbf{v}_r] \in \mathbb{R}^{N \times r}$ form the corresponding left and right singular vectors, respectively.⁸ The principal component (PC) estimator with K (that may be different from r) eigenfactors is obtained by solving the optimization problem given by

$$\underset{\hat{\mathbf{F}}_K \in \mathbb{R}^{T \times K}, \hat{\boldsymbol{\Lambda}}_K \in \mathbb{R}^{N \times K}}{\text{minimize}} \quad \|\mathbf{X} - \hat{\mathbf{F}}_K \hat{\boldsymbol{\Lambda}}_K^\top\|_F^2,$$

subject to $\frac{\hat{\mathbf{F}}_K^\top \hat{\mathbf{F}}_K}{T} = \mathbf{I}_K$ and $\hat{\boldsymbol{\Lambda}}_K^\top \hat{\boldsymbol{\Lambda}}_K$ being diagonal, where the solution is given by

$$(\tilde{\mathbf{F}}_K, \tilde{\boldsymbol{\Lambda}}_K) = (\sqrt{T} \mathbf{U}_K, \mathbf{V}_K \mathbf{D}_K).$$

Although the factor-based approach is effective in mitigating finite-sample error problems when the number of factors (r) is effectively smaller than the number of observations, it still yields inaccurate estimated factor loadings when this medium horizon (T_M) is smaller than the number of securities. This discrepancy leads to what is known as excess dispersion bias in the estimated factor loadings, which remains one of the primary sources of second-order risk in factor-based models. More specifically, estimated factor loadings are contaminated by idiosyncratic returns when using static PCA with a medium data history $T_M \ll N$, as we have with $\mathbf{Y} = \frac{1}{T} (\mathbf{D}_K^{-1} \mathbf{V}_K^\top)$

$$\begin{aligned} \tilde{\boldsymbol{\Lambda}}_K^\top &= \mathbf{D}_K \mathbf{V}_K^\top = \frac{1}{T} (\mathbf{D}_K^{-1} \mathbf{V}_K^\top) (\underbrace{T \mathbf{V}_K \mathbf{D}_K^2 \mathbf{V}_K^\top}_{\approx \mathbf{X}^\top \mathbf{X}}) \\ &\approx \mathbf{Y} (\mathbf{F} \boldsymbol{\Lambda}^\top + \mathbf{e})^\top (\mathbf{F} \boldsymbol{\Lambda}^\top + \mathbf{e}) \\ &= \mathbf{Y} \left(\underbrace{\boldsymbol{\Lambda} \mathbf{F}^\top \mathbf{F} \boldsymbol{\Lambda}^\top}_{\text{primary source of SOR bias}} + \underbrace{\boldsymbol{\Lambda} \mathbf{F}^\top \mathbf{e} + (\boldsymbol{\Lambda} \mathbf{F}^\top \mathbf{e})^\top}_{\text{primary source of SOR bias}} + \mathbf{e}^\top \mathbf{e} \right). \end{aligned}$$

In finite sample, with $K \ll T (= T_M) \ll N$, the idiosyncratic returns of some of the stocks will appear to be very significantly correlated with the returns of some of the factors. The realized correlations between the factor returns and the idiosyncratic returns appear explicitly in the terms $\boldsymbol{\Lambda} \mathbf{F}^\top \mathbf{e} + (\boldsymbol{\Lambda} \mathbf{F}^\top \mathbf{e})^\top$, adding dispersion to the estimated factor loadings.

3.3. Biased Estimation of Weak Factor Loadings

We next turn to an examination of weak latent factors. Standard factor models typically assume that the factors are *strong*, which implies that $\frac{\boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}}{N}$ is asymptotically positive definite as $N \rightarrow \infty$. Recent literature has made efforts to relax this condition, allowing $\frac{\boldsymbol{\Lambda}^\top \boldsymbol{\Lambda}}{N^\alpha}$, where $\alpha \in (0, 1]$, to be asymptotically positive definite in the limit; for example, see Freyaldenhoven (2022), Bai and Ng (2023), and the references therein. Factors with $\alpha < 1$, in particular, are commonly referred to as *weak*.

As noted by Freyaldenhoven (2022), a risk factor can be weak if its influence is limited to a subset of stocks in

the market. We classify such weak factors in this context as *narrow*, whereas Freyaldenhoven (2022) refers to them as *local*. The crucial point is that PCA cannot accurately capture narrow factors because the extracted eigenfactors are supposed to be orthogonal to one another. That is, each eigenfactor may be an amalgam of several narrow factors with broad factors and idiosyncratic returns, while each true narrow factor may be partially captured in each of multiple (orthogonal) eigenfactors.

Although a narrow factor does not appear as an *explicit* factor in the output of PCA, it nonetheless is a source of risk. This risk will be captured in the estimated covariance matrix produced by PCA, as long as we extract enough eigenvectors to ensure that “the best rank K approximation” in the form of $\mathbf{X}/\sqrt{T} \sim \mathbf{U}_K \mathbf{D}_K \mathbf{V}_K^T$ is sufficiently accurate.⁹ A portfolio that is concentrated in a single industry or country will, all else being equal, be riskier than a portfolio that is diversified over industries and countries. In order to properly evaluate the risk of a portfolio, and especially to produce optimized portfolios, we need our estimated covariance matrix to accurately reflect the narrow factors.

When $K \ll T (= T_M) \ll N$, the idiosyncratic returns of some stocks will appear to be very significantly correlated with the returns of some narrow factors, just as is the case with the broad factors. For example, all European stocks are exposed to the market factor, with true factor loadings that are roughly equal. By contrast, only some stocks are exposed to each narrow factor, with true factor loadings that are roughly equal and comparable in magnitude to the true loadings on the market portfolio. The idiosyncratic return contaminates the narrow and market factor loadings essentially equally, but the relative contamination of the narrow factor is much greater than that of the market factor. This contamination assigns nonzero estimated factor loadings on the narrow factors, when the stocks are, in fact, not exposed to them. One can think of this as excess dispersion of the estimated factor loadings that are, in fact, zero.¹⁰

3.4. The SOR Bias Mechanism for GMVP

Consider how the excess dispersion affects the construction of the GMVP. The factor loadings of some stocks appear to be smaller than they actually are, and the optimizer will choose to increase their weights in the portfolio. Conversely, the factor loadings of some stocks appear to be larger than they actually are, and the optimizer will choose to decrease their weights in the portfolio.

Remark 1 (A Consequence of the SOR Bias). A stylized fact stemming from the excess dispersion bias, as detailed in Sections 3.2 and 3.3, is that as the error in

the finite-sample estimate of $\hat{\Sigma}^{-1}$ increases, it tends to underestimate the volatility of the estimated GMVP. This indicates that finite-sample estimation errors, in the presence of weak factors, cause an apparent inflation of $\mathbf{1}^T \hat{\Sigma}^{-1} \mathbf{1}$ relative to its true value, $\mathbf{1}^T \Sigma^{-1} \mathbf{1}$, as illustrated by

$$\sigma^2(\hat{\omega}, \hat{\Sigma}) = \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1} + \underbrace{\mathbf{1}^T \boldsymbol{\epsilon} \mathbf{1}}_{\uparrow \text{ as } \|\boldsymbol{\epsilon}\| \uparrow}} \downarrow \text{ as } \|\boldsymbol{\epsilon}\| \uparrow, \quad (2)$$

which shows the mechanism behind the underprediction of the estimated GMVP’s predicted volatility.¹¹

As stated in Remark 1, the excess dispersion resulting from contamination by idiosyncratic volatility results in estimated GMVPs that have significantly higher variance than the true GMVP. Moreover, the estimated GMVP is substantially more volatile than predicted. In the presence of narrow factors, such as those driven by industry- or country-specific influences, an optimizer subject to excess dispersion may attempt to offset a narrow factor’s exposure. It does so by taking long positions in stocks that appear to be negatively correlated with the factor and short positions in stocks that appear to be positively correlated. However, if these stocks have no true exposure to the narrow factor, the positions fail to provide any actual offset to the factor exposure.¹²

To tackle the issue of excess dispersion bias, for example, Goldberg et al. (2022) proposed the Goldberg-Papanicolaou-Shkolnik (GPS) correction method that corrects the leading eigenvector based on PCA using the medium horizon T_M . They show that the GPS correction can asymptotically mitigate this bias under the “large- N and fixed- T ” framework. However, our preliminary simulation study indicates that the marginal benefit of using GPS correction is limited with a single layer of narrow factors (e.g., industries) and deteriorates further with multiple layers (e.g., countries and industries). Detailed results are available upon request.

Instead, our approach employs a longer history ($T_L > T_M$) of sample data to estimate the dynamic factor structure, with the goal of mitigating the SOR bias at the portfolio level under the “large- N and large- T ” framework. This involves extending the history of the sample data to a longer horizon (T_L) in the hope that we can reduce the contamination of the estimated factor loadings. Of course, several challenges arise when extending covariance estimation into the temporal domain, presenting both theoretical and empirical hurdles. As the factor loadings of the securities change over longer horizons, extending the sample period creates challenges in ensuring consistent estimates of principal components in a dynamic factor model. In addition, detecting and extracting weaker (eigen-)factors, which typically contain multiple narrow factor

components, poses a challenge due to inconsistencies in principal component estimates when weaker factors predominate. From this perspective, Bai and Ng (2023) argue that accurately estimating weaker loadings requires using a longer T .

3.5. Observable Measures of the SOR Bias

The volatility ratio (VR) is defined as the actual volatility, $\sigma(\hat{\omega}, \Sigma)$, divided by the predicted volatility of the estimated GMVP, $\sigma(\hat{\omega}, \hat{\Sigma})$. Simply put, the VR of $\hat{\omega}$ captures the discrepancy between the in-sample $\sigma(\hat{\omega}, \hat{\Sigma})$ and the out-of-sample $\sigma(\hat{\omega}, \Sigma)$. The VR of $\hat{\omega}$ is an observable measure in a statistical sense, and its ideal value would be one when the estimation of the precision matrix is perfectly accurate. According to Proposition 1, the VR can be approximated by the square root of a ratio expression where the error term in the numerator is twice that in the denominator. This suggests that the VR is monotonically increasing in the magnitude of $\|\epsilon\|$, subject to the SOR bias, as noted in Remark 1.

Proposition 1 (A Definitive SOR Bias Measure). *Under the condition specified in (2) of Remark 1, the SOR bias measured by $\|\epsilon\|$ monotonically increases the volatility ratio of the estimated $\hat{\omega}$; that is, VR can be approximated as*

$$(VR) := \frac{\sigma(\hat{\omega}, \Sigma)}{\sigma(\hat{\omega}, \hat{\Sigma})} = \frac{(\text{Actual volatility of } \hat{\omega})}{(\text{Predicted volatility of } \hat{\omega})} \\ \approx \sqrt{\frac{\mathbf{1}^\top \Sigma^{-1} \mathbf{1} + 2 \cdot \mathbf{1}^\top \epsilon \mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1} + \mathbf{1}^\top \epsilon \mathbf{1}}} \geq 1,$$

which is monotone increasing as $\|\epsilon\|$ gets larger.

(Proof of Proposition 1). Refer to Section A of the Electronic Companions. \square

One might argue that comparing the realized risk of the estimated GMVP would be the primary criterion for assessing the degrees of SOR bias across different methodologies, rather than focusing on the “risk-prediction error” that the VR attempts to capture. The intuition is that a more accurately estimated precision matrix would result in lower realized volatilities of the estimated GMVP. However, we have discovered that this intuition may not always hold true and can be misleading in many realistic cases. Specifically, recall that the actual volatility is the product of the predicted volatility and the volatility ratio by definition. As a result, the actual volatility of $\hat{\omega}$ (i.e., $\sigma(\hat{\omega}, \Sigma)$) may not have a monotone relationship with $\|\epsilon\|$ as

$$\sigma(\hat{\omega}, \Sigma) = \underbrace{\sigma(\hat{\omega}, \hat{\Sigma})}_{\downarrow \text{ as } \|\epsilon\| \uparrow} \cdot \underbrace{(VR)}_{\uparrow \text{ as } \|\epsilon\| \uparrow},$$

where $\sigma(\hat{\omega}, \hat{\Sigma})$ decreases and the VR increases with the estimation error under the condition in (2). As a result,

the overall behavior may not be a monotonic function of the error, which suggests that VR is a more reliable proxy for SOR than $\sigma(\hat{\omega}, \Sigma)$ for measuring the estimation error in $\hat{\Sigma}^{-1}$ across different approaches.

We conduct a numerical experiment to quantify the relationship between different SOR bias metrics and the excess dispersion bias in the estimated factor exposures specific to the GMVP. The data-generating process involves the contaminated mean-variance stock return model with $K = 5$ mutually uncorrelated factors, where the idiosyncratic returns of $N = 2,000$ stocks are uncorrelated with each other, as well as with the factor returns. For simplicity, we assume homogeneous factor volatilities of 10% and idiosyncratic volatilities of 5%. The true factor loadings are drawn from a normal distribution with mean one, where the true dispersion is represented by $\theta_0 = 1/200$. The estimated covariance matrix, $\hat{\Sigma}$, is subject to contamination by an excess dispersion parameter, $\theta \geq 0$, where the total dispersion is given by $\theta_0 + \theta$. Sample paths are generated from 1,000 different simulation seeds.

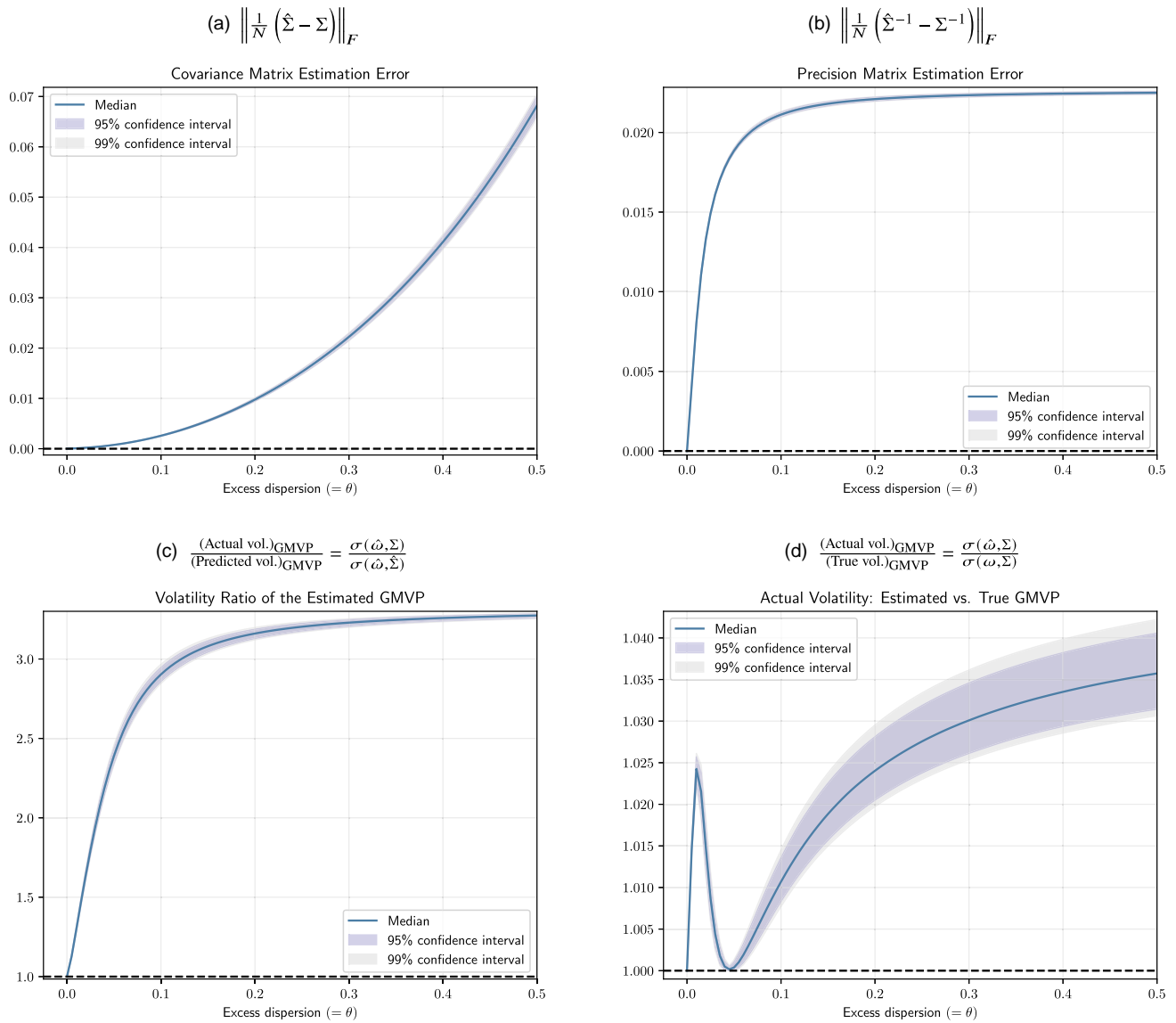
Figure 1 illustrates the results of our numerical experiment. As shown in panels (a) and (b), the ideal (or unobservable) SOR bias measure, computed using the average Frobenius norm of the estimation error in both the covariance and precision matrices, exhibits a monotonically increasing association with the excess dispersion parameter, θ . In particular, the estimation error in $\hat{\Sigma}^{-1}$ appears to be more sensitive to relatively small deviations from the true Σ^{-1} . Panels (c) and (d) confirm that VR serves as a more definitive measure of the accuracy of $\hat{\Sigma}^{-1}$ compared with the actual volatility of the estimated GMVP. The latter displays a nonmonotonic pattern influenced by the excess dispersion of the estimated factor loadings, along with its noisier reflection of the SOR bias, as indicated by relatively wider confidence intervals. Related to this observation, Proposition 2 draws attention to a potential SOR measurement problem concerning the actual volatility of $\hat{\omega}$, revealing that the true precision matrix may not be the unique minimizer of the GMVP’s actual volatility. This illustrates that a naive comparison of realized GMVP risks may result in an inaccurate assessment of the extent of the SOR bias across different methods employing the same data set.

Proposition 2 (An Illusive SOR Bias Measure). *For $\hat{\Sigma} \neq \Sigma$, suppose that $\Sigma \hat{\Sigma}^{-1}$ has an eigenvector close to the vector of ones—that is, $\Sigma \hat{\Sigma}^{-1} \mathbf{1} \approx \lambda \mathbf{1}$ for some $\lambda > 0$. Then, we observe*

$$(\text{Actual variance of } \hat{\omega}) \approx (\text{True variance of } \omega),$$

and the relationship between the SOR bias ($\|\epsilon\|$) and the actual (i.e., to-be-realized) volatility of the estimated GMVP may not be monotonic.

Figure 1. (Color online) The Metrics for SOR Bias as Functions of the Excess Dispersion Imposed in the Factor Loadings



Notes. This figure illustrates the relationship between various SOR bias metrics and the excess dispersion bias in estimated factor exposures for the GMVP, using a contaminated mean-variance stock return model with $K = 5$ mutually uncorrelated factors and $N = 2,000$ stocks with uncorrelated idiosyncratic returns and factor returns. Assuming homogeneous factor volatilities of 10% and idiosyncratic volatilities of 5%, the true factor loadings are drawn from a normal distribution with a mean of one, with true dispersion represented by $\theta_0 = 1/200$. The estimated covariance matrix, $\hat{\Sigma}$, is contaminated by an excess dispersion parameter, $\theta \geq 0$, resulting in total dispersion $\theta_0 + \theta$. Panel (a) shows the average estimation error in the covariance matrix across different levels of excess dispersion. Panel (b) measures the error in the precision matrix, demonstrating that the inverse of the covariance matrix is significantly more sensitive to relatively small deviations from the true values. Panel (c) displays the volatility ratio, which captures the discrepancy between out-of-sample actual risk and in-sample predicted risk. Panel (d) compares the realized volatility of the estimated minimum-variance portfolio to the true minimum risk. Sample paths are generated from 1,000 different simulation seeds.

(Proof of Proposition 2) Refer to Section A of the Electronic Companions. \square

Long-History PCA in an asymptotic framework, where both N and T are large.

4. A Dynamic Factor Model with Weak Loadings

This section introduces our dynamic factor model with weak loadings for stock returns and outlines the key assumptions. Furthermore, it explores the identification approaches and the theoretical justification of

4.1. Econometric Framework and Assumptions

The DFM of stock returns considered in this paper generalizes the standard definition by incorporating time-varying security sensitivities to factors and heterogeneous strengths in factor loadings. This setup builds on the frameworks of Bates et al. (2013), who propose a decomposition of factor loadings into a static term and

a purely innovative component, and Bai and Ng (2023), who extend the assumption of homogeneous factor loadings strengths to allow heterogeneity across factors. It also aligns with the local factor setup of Freyaldenhoven (2022), making it a versatile framework for capturing the empirical dynamics of observables with heterogeneous factor strengths.

Definition 1 (A Dynamic Factor Model). The log returns of the securities are generated by the linear factor model specified as

$$\underbrace{X_t}_{(N \times 1)} = \underbrace{\Lambda_t}_{(N \times r)} \underbrace{F_t}_{(r \times 1)} + \underbrace{e_t}_{(N \times 1)} \quad \text{for } t = 1, \dots, T, \quad (3)$$

where Λ_t , F_t , and e_t represent the time-varying factor loadings, latent factor returns, and idiosyncratic returns at time t , respectively.

Definition 1 establishes the econometric framework for our DFM. We presume that the daily log returns of N stocks are generated by a fixed number ($r > 0$) of unobserved common factors ($F_t \in \mathbb{R}^{r \times 1}$) with the possibly time-varying factor loadings ($\Lambda_t \in \mathbb{R}^{N \times r}$) and idiosyncratic errors $e_t \in \mathbb{R}^{N \times 1}$ on each date $t = 1, \dots, T$. Assumptions 1 and 2, which are central to our framework, rely on the existence of a finite $M < \infty$ to ensure that the DFM remains well-posed and robust across varying empirical contexts.

Assumption 1. Latent factor and idiosyncratic returns

- A. (Latent factor returns): $F_t = (F_{1t}, \dots, F_{rt})^\top \in \mathbb{R}^{r \times 1}$
- $\mathbf{E}(F_t) = \mathbf{0}$ and $\mathbf{E}(\|F_t\|_F^4) \leq M$.
 - For $k = 1, \dots, r$, $\mathbf{E}(F_{kt}^2) = \mu_{kt}^2$, where $\mu_{kt} \in [0, M]$ is the nonrandom (but possibly time-varying) factor volatility of F_{kt} .
 - Define $\mu_t = \text{diag}(\mu_{1t}, \dots, \mu_{rt}) \in \mathbb{R}^{r \times r}$ and the diagonal elements of

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mu_t^2 := \Sigma_F,$$

are strictly positive.

- d. We have

$$\frac{1}{T} \sum_{t=1}^T F_t F_t^\top \xrightarrow{p} \Sigma_F \quad \text{as } T \rightarrow \infty.$$

- B. (Idiosyncratic returns): $e_t = (e_{1t}, \dots, e_{Nt})^\top \in \mathbb{R}^{N \times 1}$
- For $n = 1, \dots, N$, we have $\mathbf{E}(e_{nt}) = 0$, $\mathbf{E}(|e_{nt}|^8) \leq M$ and $\mathbf{E}(e_{nt}^2) = v_{nt}^2$, where $v_{nt} \in [0, M]$ is the nonrandom idiosyncratic volatility of e_{nt} . For notational consistency, define $v_t = \text{diag}(v_{1t}, \dots, v_{Nt}) \in \mathbb{R}^{N \times N}$ and

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_t^2 := \Sigma_\varepsilon.$$

- Both $\gamma_N(s, t) = \frac{1}{N} \mathbf{E}(e_s^\top e_t)$ and $\tau_{ij, st} = \mathbf{E}(e_{it} e_{js})$ exist for all $i, j \in \{1, \dots, N\}$ and $s, t \in \{1, \dots, T\}$.

- We have $|\gamma_N(s, s)| \leq M$ for all $s \in \{1, \dots, T\}$, and

$$\frac{1}{T} \sum_{s, t=1}^T |\gamma_N(s, t)| \leq M.$$

- For all $t \in \{1, \dots, T\}$, $|\tau_{tt, ij}| \leq |\tau_{ij}|$ holds for some τ_{ij} , and we have

$$\frac{1}{N} \sum_{i, j=1}^N |\tau_{ij}| \leq M \quad \text{and} \quad \frac{1}{NT} \sum_{s, t=1}^T \sum_{i, j=1}^N |\tau_{ts, ij}| \leq M.$$

- For all $s, t \in \{1, \dots, T\}$, we have

$$\mathbf{E} \left(\left| \frac{1}{\sqrt{N}} \sum_{n=1}^N e_{sn} e_{nt} - \mathbf{E}(e_{sn} e_{nt}) \right|^4 \right) \leq M.$$

- For all $i \in \{1, \dots, N\}$ and $s, t \in \{1, \dots, T\}$, e_{it} is independent of F_s .

The conditions outlined in Assumption 1 align closely with those in Bai and Ng (2002), accommodating weak dependence between factors and idiosyncratic returns, as well as limited cross-sectional and temporal dependence among the idiosyncratic errors. As such, the dynamic nature of the factor model arises from two sources: (i) variable volatilities of both factor and idiosyncratic returns, and (ii) the time-varying factor loadings, with their conditions detailed in Assumption 2.

Assumption 2. Assumptions on the factor loading dynamics

- A. (Factor loading dynamics)
- The dynamics of Λ_t for $t = 1, \dots, T$ is expressed as

$$\Lambda_t = \Lambda_0 + h_{NT} \xi_t, \quad (4)$$

where $\Lambda_0 \in \mathbb{R}^{N \times r}$ specifies the (static) baseline factor loadings, $h_{NT} \in \mathbb{R}$ is a deterministic scalar that may depend on the pair of (N, T) , and $\xi_t = (\xi_{1t}, \dots, \xi_{Nt})^\top \in \mathbb{R}^{N \times r}$ is a (possibly degenerate) stochastic innovation process.

- For all $i, j \in \{1, \dots, N\}$ and $s, t \in \{1, \dots, T\}$, e_{it} is independent of ξ_{js} .
- B. (Static baseline component): Λ_0
- Λ_0 has rank r , the columns of Λ_0 are orthogonal, and $\|\Lambda_0\|_F \leq M$ holds.
 - Allowing the strengths of baseline loadings to vary across factors, let

$$1 \geq \alpha_1 \geq \dots \geq \alpha_r > 0,$$

so that the weakest baseline loading has strength $\alpha_r \in (0, 1]$. Letting the $r \times r$ normalization matrix

$$B_N = \text{diag}(N^{\alpha_1}, \dots, N^{\alpha_r}),$$

there exists some diagonal matrix $D \in \mathbb{R}^{r \times r}$ such that

$$B_N^{-1} \Lambda_0^\top \Lambda_0 B_N^{-1} \xrightarrow{p} D \quad \text{as } N \rightarrow \infty.$$

- C. (Stochastic innovation component): $h_{NT}\xi_t$
 a. There exist envelope functions $Q_1(N, T)$, $Q_2(N, T)$, and $Q_3(N, T)$ such that the following conditions hold for all N, T and factor indices $p, q, k, \ell = 1, \dots, r$:

$$\begin{aligned} \sup_{s, t \leq T} \sum_{i, j=1}^N |\mathbf{E}((\xi_{is})_p (\xi_{jt})_q F_{ps} F_{qt})| &\leq Q_1(N, T) \\ \sum_{s, t=1}^T \sum_{i, j=1}^N |\mathbf{E}((\xi_{is})_p (\xi_{js})_q F_{ps} F_{qs} F_{kt} F_{\ell t})| &\leq Q_2(N, T) \\ \sum_{s, t=1}^T \sum_{i, j=1}^N |\mathbf{E}((\xi_{is})_p (\xi_{js})_q (\xi_{it})_k (\xi_{jt})_\ell F_{ps} F_{qs} F_{kt} F_{\ell t})| &\leq Q_3(N, T) \end{aligned}$$

- b. The following conditions hold:

$$\begin{aligned} h_{NT}^2 Q_1(N, T) &= \mathcal{O}(N) \\ h_{NT}^2 Q_2(N, T) &= \mathcal{O}(NT^2) \\ \min\{N, T\} h_{NT}^4 Q_3(N, T) &= \mathcal{O}(N^2 T^2). \end{aligned}$$

Assumption 2 demonstrates that our approach significantly advances the conventional static framework by extending the setup of Bates et al. (2013) to account for time-varying security sensitivities to factors with heterogeneous strengths. The time-varying loadings (Λ_t) are decomposed into a static component (Λ_0) and a term driven by pure innovations ($h_{NT}\xi_t$). A central implication of this structure, governed by the conditions in Assumptions 2 and 3, is that the noise term originating from the loading innovations, $h_{NT}\xi_t F_t$, can be treated analogously to an idiosyncratic error.¹³ Thus, the principal component estimator can tolerate significant instability in the factor loadings precisely because the effects of these shifts are mitigated by averaging across a large number of series (N) and a long time horizon (T). As inspired by Bates et al. (2013), the envelope conditions in Assumptions 2 and 3 formalize this by accommodating several empirically relevant forms of instability, such as “white noise” deviations, “random walk” innovations, and even discrete “single large break” events affecting a subset of series, under the key condition that these instabilities are sufficiently idiosyncratic to be diversified away. This structure, which facilitates the consistent estimation of factor models through PCA under mild regularity conditions, accommodates stylized features including mean-reverting¹⁴ stationary dynamics of factor exposures, occasionally interrupted by structural breaks, a feature commonly observed in empirical data sets. This enhancement builds on the contributions of Bai and Ng (2023), further expanding the local factor structure proposed by Freyaldenhoven (2022).¹⁵

Whereas the *pure innovation* component is structured to be diversified away, the *static* component of the factor loadings, Λ_0 , is the primary focus of our analysis, as it approximates the baseline factor sensitivities that

cannot be diversified away. PCA-based estimation of such models typically assumes homogeneous factor loadings ($\alpha = 1$), a framework extended by Bai and Ng (2023) to allow heterogeneity using a parameter $\alpha \in (0, 1]$. We incorporate this extension in our normalization matrix B_N , enabling the model to accommodate variations in factor strength across different securities and factors. Accordingly, we adhere to the assumptions of Bai and Ng (2023) for addressing heterogeneity in the strengths of the static component. These assumptions ensure consistency in estimation while allowing for flexibility in modeling empirical data. For example, when Λ_t, μ_t , and v_t are independent of t , the model reduces to the standard definition of a static factor model with constant factor sensitivities, volatilities, and idiosyncratic terms over time. Overall, the proposed model combines the strengths of existing literature while providing greater flexibility to capture time-varying dynamics and heterogeneous strengths in factor loadings.

4.2. Identification Approaches

Our DFM specification can also be reformulated as a static factor model under additional restrictive conditions, accommodating the stochastic variation in the loadings by attributing some of this variability to the factors. Specifically, assume there exists $\xi_0 \in \mathbb{R}^{N \times r}$ with full column rank r . Let $\xi_0^+ = (\xi_0^\top \xi_0)^{-1} \xi_0^\top \in \mathbb{R}^{r \times N}$ denote the Moore-Penrose pseudo-inverse of ξ_0 . Under these conditions, our DFM model can be interpreted as a static factor model with time-invariant loadings in the form of

$$X_t = (\Lambda_0 + h_{NT}\xi_t)F_t + e_t = \Lambda_0 F_t + \tilde{\Lambda}_0 \tilde{F}_t + e_t,$$

where $\tilde{\Lambda}_0 = h_{NT}\xi_0 \in \mathbb{R}^{N \times r}$ and $\tilde{F}_t = \xi_0^+(\xi_t F_t) \in \mathbb{R}^{r \times 1}$. Observe that this formulation highlights that our DFM with time-varying loadings is both more general and parsimonious, utilizing fewer factors and avoiding reliance on the construction of ξ_0^+ that must be invertible.

Our approach also introduces a simplified representation of a linear factor model in the sense that the observed data process can be expressed as

$$X_t = \Lambda_0 F_t + \omega_t + e_t, \quad (5)$$

where $\omega_t = h_{NT}\xi_t F_t$ represents additional time-varying noise term. Such representation is comparable to Pelger and Xiong (2022b), where the time-varying factor loadings are assumed to be $\Lambda_t = \Lambda(S_t)$ as a deterministic function of the observable state process S_t . The state-dependent factor structure is given by

$$X_t = (\Lambda(S_t) + \mathcal{E}_t)F_t + e_t,$$

where the noise term \mathcal{E}_t captures variations in loading coefficients that cannot be fully explained by the state process (S_t), potentially arising from measurement errors or omitted variables. Although Pelger and Xiong (2022b) assumes the existence of an observable state

process that accounts for the factor loading dynamics in real time, if this is not feasible, the state-dependent dynamic factor structure can be reformulated as given in Equation (5) by imposing¹⁶

$$\Lambda(S_0) = \Lambda_0 \quad \text{and} \quad \Lambda(S_t) - \Lambda(S_0) + \mathcal{E}_t = h_{NT}\xi_t.$$

In short, the representation in Equation (5) simplifies our DFM by assuming time-invariant loadings, while treating the composite noise as a combination of idiosyncratic and measurement errors that can be diversified away under mild assumptions, providing a robust, yet interpretable, formulation.

4.3. Theoretical Basis of Long-History PCA

To mitigate the realized correlation between idiosyncratic returns and factor returns, we consider a longer data history of length $T_L (> T_M)$ —for example, daily returns over six years. As a result, we cannot presume that the factor loadings remain constant over the entire (extended) sample period T_L . Nevertheless, Theorem 1 and Corollary 1 confirm that, under the “large- N and large- T ” framework, PCA can consistently estimate the covariance and precision matrices in a dynamic factor model under mild regularity conditions outlined in Assumptions 1 and 2.

Theorem 1 (Convergence of a Population Covariance and Precision Matrices). *Consider the population covariance matrix within the observation time window T given by*

$$\Sigma_N(T) = \frac{1}{T} \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{N \times N},$$

where $\mathbf{X} \in \mathbb{R}^{T \times N}$ is the observed daily return data of N securities. In addition, we introduce $\bar{\Sigma}_N(T)$ as a static version of $\Sigma_N(T)$ given by

$$\bar{\Sigma}_N(T) = \Lambda_0 \left(\frac{1}{T} \sum_{t=1}^T \mu_t^2 \right) \Lambda_0^\top + \left(\frac{1}{T} \sum_{t=1}^T v_t^2 \right),$$

where both $\Sigma_N(T)$ and $\bar{\Sigma}_N(T)$ are assumed to be invertible. By letting $N, T \rightarrow \infty$, we have

$$\begin{aligned} \left\| \frac{1}{N} (\Sigma_N(T) - \bar{\Sigma}_N(T)) \right\|_F^2 &\asymp \left\| \frac{1}{N} (\Sigma_N^{-1}(T) - \bar{\Sigma}_N^{-1}(T)) \right\|_F^2 \\ &= O_p \left(\frac{1}{\min\{N, T\}} \right), \end{aligned}$$

under the conditions stated in Assumptions 1 and 2, where $a_N(T) \asymp b_N(T)$ indicates $a_N(T) = O(b_N(T))$ and $b_N(T) = O(a_N(T))$ as $N, T \rightarrow \infty$.¹⁷

(Proof of Theorem 1) Refer to Section A of the Electronic Companions. \square

Theorem 1 asymptotically justifies that the variation introduced by $h_{NT}\xi_t$ remains sufficiently small, ensuring that it does not dominate the factor structure in the limiting representations of the covariance and precision matrices. Note that the speed of convergence is

governed by the minimum of N and T , emphasizing the importance of extending T for large N , while ensuring that T is not excessively large relative to N .

Corollary 1 (Consistent Estimation of LH-PCA). *Let $(\tilde{F}_r, \tilde{\Lambda}_r)$ denote the principal component estimator of $\mathbf{X} \in \mathbb{R}^{N \times T}$, where r represents the true number of factors. Define $\tilde{e}_r := \mathbf{X} - \tilde{F}_r \tilde{\Lambda}_r^\top \in \mathbb{R}^{N \times T}$ and*

$$\tilde{\Sigma}_N(T) := \tilde{\Lambda}_r \left(\frac{1}{T} \tilde{F}_r^\top \tilde{F}_r \right) \tilde{\Lambda}_r^\top + \frac{1}{T} \tilde{e}_r \tilde{e}_r^\top.$$

Then, as $N, T \rightarrow \infty$, we have

$$\begin{aligned} \left\| \frac{1}{N} (\Sigma_N(T) - \tilde{\Sigma}_N(T)) \right\|_F^2 &\asymp \left\| \frac{1}{N} (\Sigma_N^{-1}(T) - \tilde{\Sigma}_N^{-1}(T)) \right\|_F^2 \\ &= O_p \left(\frac{1}{\min\{N, T\}} \right), \end{aligned}$$

under the conditions stated in Assumptions 1 and 2.

(Proof of Corollary 1) Refer to Section A of the Electronic Companions. \square

In particular, Corollary 1 justifies the use of LH-PCA in the setting of variable factor and idiosyncratic volatility, along with time-varying factor loadings with heterogeneous strengths under our assumptions. This implies that, accounting for time variation in both factor and idiosyncratic volatilities, changes in the values of the factor-loading matrix under our assumption do not pose an additional obstacle to the consistent estimation of the factor-driven covariance matrix. Furthermore, the estimation remains consistent under the dynamic factor model specification, even with weaker loadings in their static components.

Empirically, factor volatility is highly variable, even over short horizons. Factor loadings are also variable over long, and possibly medium, horizons. The presence of weak factors complicates the process of portfolio volatility prediction. Note that even if factor loadings were constant, and there were no weak factors, variation in factor volatility would alter the return correlation matrix over short horizons. PCA estimation over medium horizons makes sense only if factor loadings can be consistently estimated in the presence of rapidly varying factor volatility. PCA estimation over long horizons requires, in addition, that factor loadings can be consistently estimated in the presence of variable factor loadings and weak factors. For this reason, commercial latent factor models have avoided using long histories of daily return data. Our theoretical results justify estimation over longer histories, even when factor volatilities and factor loadings are varying and weak factors are present. Our empirical and simulation results show that the use of longer histories improves the prediction of portfolio volatility.¹⁸

4.4. Number of Eigenfactors

Implementing the PCA estimator requires selecting the appropriate number of eigenfactors (K) to extract. Although data-driven criteria such as the Bai and Ng (2002) (hereafter “BN”) estimator have the theoretical advantage of adapting to the evolving structure of the stock market and reducing the misspecification risk associated with arbitrary fixed choices, they can exhibit practical instability in finite samples. To ensure the credibility and real-world applicability of our findings, our main analysis presents results using a fixed K , complemented by results where K is diagnosed dynamically.¹⁹

To validate that the performance of LH-PCA is not an artifact of the specific choice of K , we compare it against standard one-year PCA and GPS across both a wide fixed range ($K \in [3, 100]$) and varying deviations from the BN estimator ($K = BN + \Delta$ for $\Delta \in [-3, 100]$), as shown in Figure 2. The results illustrate that although performance is similar at very low K (where the model is misspecified), the outperformance of LH-PCA widens significantly as K increases. This confirms that the primary source of LH-PCA’s advantage is the use of a long history (T_L), which enables the model to effectively extract signal from the weaker factors found at higher K levels, rather than the specific choice of K itself.

For our main empirical analysis, we set the representative fixed K to the median of the BN estimator calculated over the long-history window ($T_L = 1,500$). This choice is supported by simulation evidence confirming that T_L enables the detection of weak, but structurally important, volatility-driving factors (consistent with Bai and Ng 2023), while filtering out nonsystematic signals that are often conflated with noise in medium windows ($T_M = 250$). As such, using T_L mitigates the finite-sample instability typically observed under T_M , whose BN estimates tend to have higher dispersion

and a lower median, as shown in Figure EC.1 of the Electronic Companions. The extended history improves the signal-to-noise ratio, allowing the estimator to suppress transient idiosyncratic shocks and converge toward the market’s effective rank.²⁰

5. Simulation Study

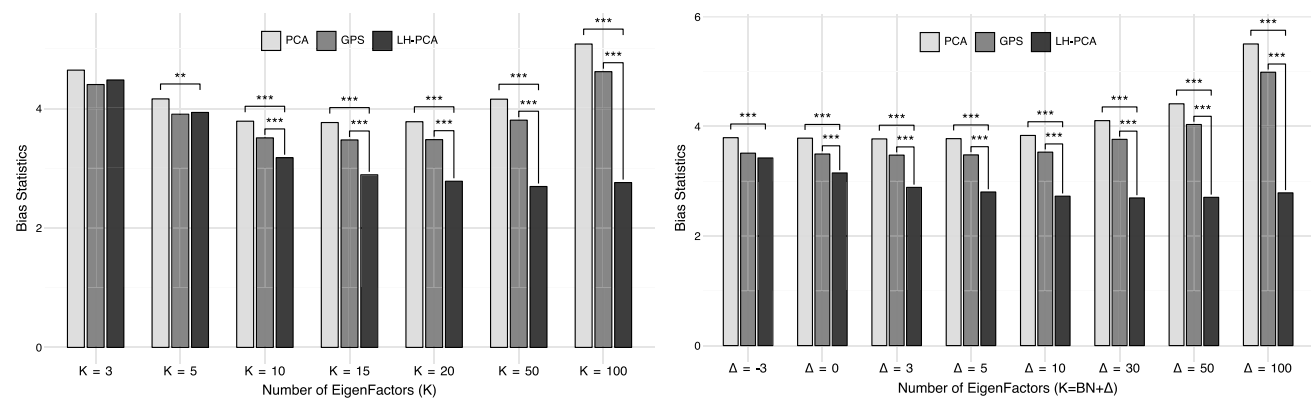
In this section, we employ Monte Carlo simulations to assess the performance of the SOR bias. Having access to the true risk factor structure in the simulations, we can directly compute the difference between the true and estimated precision matrices. This not only supports the validation of our observable SOR bias metrics, but also provides a solid basis for our subsequent empirical analysis, as detailed in the following section.

5.1. Simulation Setup

In our simulation, we assume the presence of 4 *broad* factors, consisting of 1 market factor and 3 style factors, along with 27 narrow factors, which include 16 country factors plus 11 industry factors, to mimic the narrow factor structure of the European stock market. We set the annualized daily volatility of the market factor at 16% and the daily volatility of the three style factors at 8%, 8%, and 4%, respectively (Clarke et al. 2011, Morozov et al. 2012). The country and industry factor volatilities are calibrated from the European stock market data. We draw idiosyncratic volatilities uniformly from [32%, 64%] per annum. We randomly assign $N = 2,000$ stocks to countries and industries; see Table 1 for details. The true factors are normally distributed with mean zero, and they are assumed to be pairwise uncorrelated. The sample stock returns are generated across 1,000 different seeds.

5.1.1. Variable Volatility Structure. We assume that the return-generating process follows a Markov regime-

Figure 2. Comparison of Bias Statistics Across Different Factor Specifications (K)



Notes. This figure illustrates the Bias Statistics across different choices of the number of extracted factors (K). In the right panel, “BN” represents the number of eigenfactors determined using the Bai and Ng (2002) estimator, and the variants are constructed by adding or subtracting a fixed increment (Δ). The data consist of CRSP daily returns from 2001 to 2021. The brackets denote the statistical significance of the difference in BS between LH-PCA and the benchmark schemes. **Significant at the 5% level; ***significant at the 1% level.

Table 1. True Factor Structure in Simulation

Factors	Volatility (%)	No. of stocks
Broad factors		
Market	16.00	2,000
Style 1	8.00	2,000
Style 2	4.00	2,000
Style 3	4.00	2,000
Industry factors		
Industry 1	11.51	143
Industry 2	12.74	291
Industry 3	13.00	131
Industry 4	15.08	144
Industry 5	13.99	115
Industry 6	17.80	251
Industry 7	14.07	258
Industry 8	12.22	53
Industry 9	9.80	109
Industry 10	13.11	171
Industry 11	21.23	334
Industry sum		2,000
Country factors		
Country 1	13.00	154
Country 2	12.38	69
Country 3	18.53	159
Country 4	15.91	76
Country 5	24.93	178
Country 6	18.14	99
Country 7	11.86	177
Country 8	17.57	19
Country 9	16.43	203
Country 10	24.41	73
Country 11	14.31	56
Country 12	12.03	247
Country 13	13.33	171
Country 14	13.79	75
Country 15	18.29	135
Country 16	22.59	109
Country sum		2,000

Notes. This table reports the true structure of broad, country, and industry factors in our simulation study. Factor volatilities are annualized, in percent. The true factors are assumed to be normally distributed with a mean of zero within the same regime and are considered pairwise uncorrelated.

switching model (MRS). There is a latent Markov chain of the state process $s_t \in \{0, 1\}$, in which $s_t = 0$ denotes a normal factor-volatility structure, and $s_t = 1$ denotes a crisis structure. This Markov chain is described by its transition matrix $\mathbf{\Pi}$ with the probabilities $p_{ij} = \mathbb{P}(s_t = i | s_{t-1} = j)$ of switching from regime j at time $t - 1$ to regime i at time t . In particular, we have

$$\mathbf{\Pi} = \begin{bmatrix} p_{0|0} & p_{1|0} \\ p_{0|1} & p_{1|1} \end{bmatrix} = \begin{bmatrix} 0.998 & 0.002 \\ 0.008 & 0.992 \end{bmatrix}.$$

Furthermore, we assume that the initial distribution is given by $\mathbb{P}(s_t = 0) = 0.8$. During the normal period ($s_t = 0$), the factor volatilities are drawn from the same distribution as the baseline setup. For the crisis regime ($s_t = 1$), we assume that the volatilities of the (market, style, country, industry, idiosyncratic) factors are multiplied

by (2.0, 1.5, 1.5, 1.5, 1.25), respectively. The true factors are normally distributed with mean zero within the same regime, and they are assumed to be pairwise uncorrelated.

5.1.2. Time-Varying Factor Loadings. We further extend the simulation setup by allowing variable factor loadings over time, following Adrian and Franzoni (2009) and the references therein. We first assume that the *broad* factor exposures (i.e., the market beta and the style factor loadings) are time-varying based on a discretized version of the mean-reverting Ornstein-Uhlenbeck process given by the following stochastic differential equation:

$$d\Lambda_t^i = \kappa(\Lambda_0^i - \Lambda_t^i)dt + D_\sigma dW_t^i,$$

where $\Lambda_t^i = [\lambda_t^{i0}, \lambda_t^{i1}, \lambda_t^{i2}, \lambda_t^{i3}]^\top$ represents the *broad* factor loadings specific to the i -th stock, $\kappa = 1.0$ governs the (common) mean-reversion speed, $\Lambda_0^i = [\lambda_0^{i0}, \lambda_0^{i1}, \lambda_0^{i2}, \lambda_0^{i3}]^\top$ specifies the long-run mean level of the factor loadings, D_σ is a 4×4 diagonal matrix with its diagonal entries equal to [10%, 5%, 5%, 5%] per annum, and W_t^i is an independent four-dimensional standard Brownian motion. We set the initial broad factor loadings as the long-run mean level of the factor loadings. For simplicity, we assume that the narrow factor loadings are binary constants.

5.2. Main Findings

Table 2 summarizes the simulation results within the setup characterized by the MRS volatility structure and the mean-reverting factor loading dynamics. We compare the performance of four distinct approaches—that is, LW denotes the shrinkage estimation of the sample

Table 2. The 99% Confidence Intervals for the Expected Performance Metrics (without RCA)

	Expected performance statistics (bootstrapped)		
	Ideal SOR bias $\mathbb{E}\ \frac{1}{N}(\hat{\Sigma}^{-1} - \Sigma^{-1})\ _F$	Volatility ratio $\mathbb{E}\left(\frac{\sigma(\hat{\omega}, \Sigma)}{\sigma(\hat{\omega}, \hat{\Sigma})}\right)$	Actual volatility $\mathbb{E}(\sigma(\hat{\omega}, \Sigma))$ (% , p.a.)
LW	[51.5739, 56.4899]	[22.0352, 24.4072]	[9.5471, 9.8187]
PCA	[5.5964, 5.8933]	[2.9599, 3.0431]	[7.1718, 7.3970]
GPS	[5.5829, 5.8793]	[2.7943, 2.8709]	[7.1195, 7.3338]
LH-PCA	[4.6756, 4.9940]	[2.6009, 2.6842]	[7.3720, 7.6024]

Notes. This table shows the 99% confidence intervals of the estimated GMVP’s expected performance metrics obtained from 1,000 different seeds under the simulation setup with the MRS volatility structure and the mean-reverting factor loading dynamics. The performance statistics are calculated from a sample of $N = 2,000$ stocks. LW denotes the shrinkage estimation of the sample covariance matrix proposed by Ledoit and Wolf (2004b) with $T_M = 250$. PCA stands for the plain PCA method with T_M ; GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M . We employ $T_L = 1,500$ for LH-PCA. These results are obtained from 1,000 simulation runs with varying seeds, acquired through bootstrapping with replacements for cross-validation.

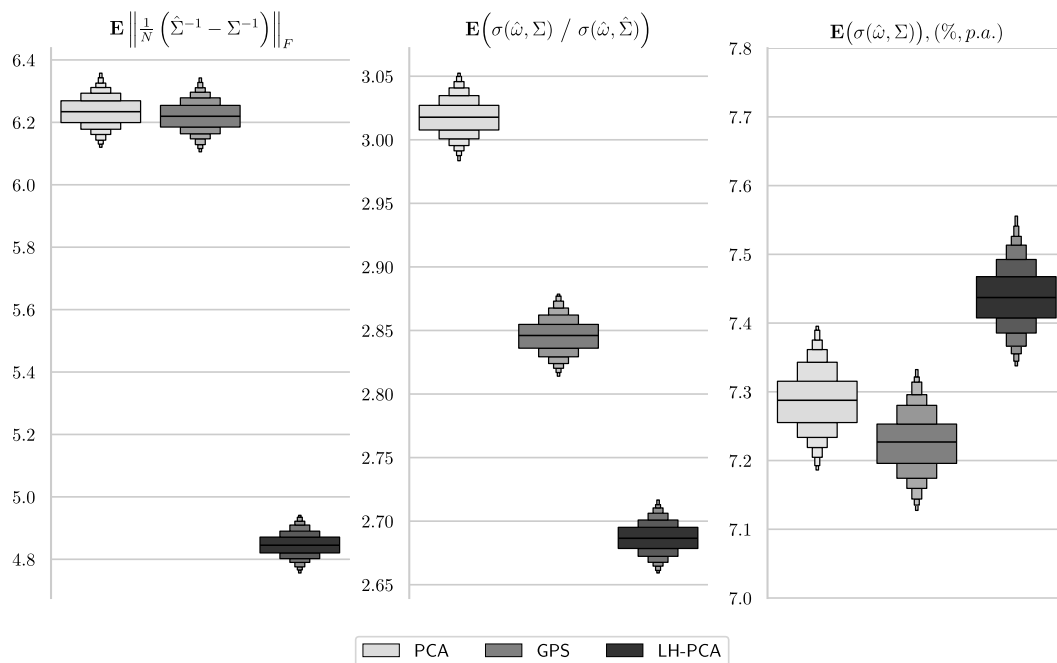
covariance matrix, proposed by Ledoit and Wolf (2004b) using one year of historical data;²¹ PCA represents the plain PCA estimation based on one year ($T_M = 250$) of data history; GPS refers to the PCA method adjusted by the GPS correction in accordance with the guidelines of Goldberg et al. (2022) to correct the leading eigenvector; and LH-PCA represents the PCA estimation utilizing a six-year history ($T_L = 1,500$) of sample data. The simulation results are obtained from 1,000 simulation runs with varying seeds, acquired through bootstrapping with replacements for cross-validation.

As indicated by the ideal SOR bias defined in Equation (1), which is quantified by the average Frobenius norm of the estimation error in the precision matrix and is only available in simulation, the factor-based approach demonstrates approximately 10 times superior performance to the use of the shrunk sample covariance matrix. This observation is corroborated by both metrics for expected volatility ratio and actual volatility, though the performance differentials in terms of actual volatility become much less pronounced. Moreover, the actual volatility measure exhibits a contrasting direction compared with the observations from the other two metrics within the factor-based approaches. This is consistent with our observations in Proposition 2 and Figure 1, in that the actual volatility may lose its monotone relationship with the estimation error when the estimated $\hat{\Sigma}^{-1}$ is reasonably close to the true Σ^{-1} .

In a dynamic factor model, the covariance matrix estimated by PCA (and its variant) is distinct from the covariance matrix encountered on the following day of estimation. Such discrepancies have the potential to result in empirically inaccurate out-of-sample predictions of the true factor structure on a daily basis. Understanding these discrepancies underscores the potential necessity for responsive covariance adjustments in empirical analysis; see Section C of the Electronic Companions for more details of its implementation under the factor model approaches. Nevertheless, similar patterns can be observed in the enhanced box plots in Figure 3, where the estimated covariance matrices are adjusted by RCA with $T_S = 40$ days for PCA, GPS, and LH-PCA. Specifically, the enhanced box plots consistently illustrate the distributions of the expected metrics within each panel, specific to the estimated GMVPs across different estimation schemes.

The fact that expected VR statistics are not adversely affected by the introduction of variable factor volatility and mean-reverting stock factor exposures is a direct consequence of Corollary 1, which asserts that (LH-)PCA consistently estimates the baseline factor exposures in the presence of (potentially) unstable factor structure dynamics under mild regularity conditions. In addition, the outcomes demonstrate that the use of a longer data history leads to a significant improvement in the estimation of factor loadings. This is evidenced

Figure 3. Expected Performance Statistics of the Factor-Based Approaches (with RCA)



Notes. This figure shows the distributions of the estimated GMVP's expected metrics obtained from 1,000 different seeds under the extended setup with the MRS volatility structure and the mean-reverting factor loading dynamics. The performance statistics are calculated from a sample of $N = 2,000$ stocks. PCA stands for the plain PCA method with $T_M = 250$; GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M . We employ $T_L = 1,500$ for LH-PCA. The estimated covariance matrices are adjusted by RCA with $T_S = 40$ days. These results are obtained from 1,000 simulation runs with varying seeds, acquired through bootstrapping with replacements for cross-validation.

by the fact that LH-PCA outperforms one-year PCA and GPS in the presence of narrow factors under both the ideal SOR bias and the volatility ratio metrics.²² Overall, our numerical results confirm that the use of long histories such as $T_L = 1,500$ trading days, or six years, substantially improves the performance of the unconstrained GMVPs, compared with using a medium history of $T_M = 250$ days, indicating superior SOR bias mitigation.

6. Empirical Analysis

6.1. Data and Samples

In our empirical study, we investigate two financial markets: the United States and Europe. Daily stock return data in the United States are obtained from the Center for Research in Security Prices (CRSP), and European stock return data from 16 countries²³ are obtained from Compustat-Capital IQ's Global Daily (henceforth EURO). Our database spans from 2001 to 2021. We selected this period for several reasons: (i) it provides a substantial timeframe for estimating the plane spanned by factor returns over a long history (T_L); (ii) it encompasses major market shocks, including the global financial crisis and COVID-19; and (iii) most European countries transitioned from their national currencies to the Euro in the late 1990s and early 2000s. One particular feature of our empirical setting is that we consider the entire universe of stocks at a given point in time. This is crucial because small stocks, which tend to be highly volatile, can make the estimation problem more challenging than focusing on more established samples, such as the S&P 500 constituents.

As in the previous section, we fix $T_L = 250 \times 6 = 1,500$ and $T_M = 250$ trading days. Our decision to use $T_L = 1,500$ trading days (approximately six trading years) is based on observed stock market patterns in the United States and Europe. For instance, the U.S. stock market experienced a bull market from late 2002 to 2007, followed by the Global Financial Crisis in 2008–2009 and a recovery afterward. Similarly, from 2013 to 2019, markets underwent a prolonged period of sustained growth, which was disrupted by the COVID-19 crash in 2020. In Europe, the Eurozone debt crisis (2010–2012) was preceded by a postcrisis recovery and followed by a stabilization phase from 2013 to 2018. Taken together, our choice of six trading years as the long-history window is grounded in the stock market cycles observed during our main sample period, 2001–2021. In Section D.3 of the Electronic Companions, we cross-check our empirical results using alternative lengths for the T_L moving window to ensure that our main findings are not sensitive to this parameter choice.

Table 3 provides a comprehensive overview of our sample, presenting the number of stocks and the yearly realized volatility across countries and industries. On

average, a moving window in the CRSP data set contains approximately 2,260 stocks, with the total number of firms in the sample being $N = 6,516$. At the start of our moving window analysis, around 2,800 securities are available, but this number gradually declines to approximately 1,800 by the end of the sample period, consistent with the *listing gap* argument in Doidge et al. (2017). In the EURO data set, a total of $N = 7,350$ firms are included in our sample, with an average of about 2,500 stocks available in each moving window. Unlike in CRSP, the number of available securities in EURO has increased over the sample period, likely due to the broader coverage of the database and the expanding stock market.

6.2. Empirical Assessment of SOR Bias

In Section 3.5, we introduced the Volatility Ratio, the ratio of the GMVP's actual (out-of-sample) volatility to its predicted (in-sample) volatility, as a definitive, albeit unobservable, measure of SOR bias. As shown in Proposition 1, the VR is an excellent theoretical metric that is monotonically increasing with the magnitude of the SOR bias. However, the VR cannot be calculated directly in empirical applications, when the true population covariance matrix, Σ , is unobservable.

To overcome this limitation, we require an observable and statistically robust proxy for the VR that can be computed from realized returns. As proposed by Shepard (2009), the building block of our performance statistics is *Z-score* of $\hat{\omega}$, defined as

$$z_t = \frac{(\text{Realized GMVP return})_{t+1}}{(\text{Predicted GMVP volatility})_t} = \frac{R_{t+1} \hat{\omega}_t}{\sigma(\hat{\omega}_t, \hat{\Sigma}_t)}, \quad (6)$$

where the variance of the Z-scores should be equal to one, provided that the predicted volatilities are perfectly precise. We consider *Bias Statistic* (BS), introduced in Menchero et al. (2013), as our main performance measure to evaluate the SOR bias for the GMVP, expressed as

$$BS_t(\tau) = \sqrt{\frac{1}{\tau} \sum_{k=t-\tau+1}^t z_k^2}, \quad (7)$$

where τ denotes the number of trading days in the testing (out-of-sample) period. The BS can be interpreted as the sample standard deviation of the realized Z-scores, irrespective of their distribution. Ideally, in the absence of SOR bias, the BS value should be equal to one. Thus, the Z-score-based BS is not a conceptually different metric but, rather, the empirical counterpart to the VR measure. The VR provides the theoretical foundation for why we are interested in the ratio of actual to predicted volatility, whereas the BS provides the statistical methodology to measure this ratio using observable market data. This allows us to assess the

Table 3. Empirical Data Description

Sources:	United States			Europe		
	CRSP			Compustat-Capital IQ Global		
Dimensions:	$T = 5,284, N = 6,516$			$T = 5,359, N = 7,350$		
Category	Classification	Volatility (%)	No. of stocks	Classification	Volatility (%)	No. of stocks
Industry	1. Consumer Nondurables	19.32	346	1. Consumer Nondurables	12.08	679
	2. Consumer Durables	26.65	155	2. Consumer Durables	13.12	216
	3. Manufacturing	26.30	614	3. Manufacturing	12.21	1,082
	4. Energy	39.02	302	4. Energy	13.00	344
	5. Chemicals	23.67	134	5. Chemicals	13.94	218
	6. Business Equipment	21.82	1,400	6. Business Equipment	12.37	1,764
	7. Telecom	23.75	278	7. Telecom	12.92	250
	8. Utilities	18.94	157	8. Utilities	13.94	151
	9. Shops	23.28	719	9. Shops	11.83	767
	10. Health	20.80	812	10. Health	12.30	746
	11. Money (Finance)	23.91	1,599	11. Money (Finance)	11.98	1,347
Country	United States		6,516	1. Austria (AUT)	14.46	98
				2. Belgium (BEL)	13.01	151
				3. Switzerland (CHE)	12.16	337
				4. Germany (DEU)	11.14	997
				5. Denmark (DNK)	13.76	231
				6. Spain (ESP)	15.25	201
				7. Finland (FIN)	11.92	203
				8. France (FRA)	12.70	909
				9. United Kingdom (GBR)	11.70	2,148
				10. Greece (GRC)	14.75	239
				11. Ireland (IRL)	15.87	78
				12. Italy (ITA)	16.47	487
				13. Netherland (NLD)	16.04	226
				14. Norway (NOR)	15.33	356
				15. Portugal (PRT)	17.88	48
			16. Sweden (SWE)	15.35	855	

Notes. In each dataset, the Category reflects a narrow factor classification (industry or country), with the Classification column identifying the specific industries or countries within each category. Annualized realized volatility is reported for each factor in percentage terms, along with the number of stocks represented in each classification. Industry classifications follow the definitions provided on Kenneth French’s website.

degree of SOR bias across different models in a practical setting. The BS captures whether the risk forecasts were accurate, on average, for the testing portfolio over time. To assess the statistical reliability of BS, confidence intervals can be approximated by bootstrapping the realized Z-scores with replacement for cross-validation purposes.

A potential pitfall of the BS measure lies in the time-resolution problem, which arises because the statistic may appear overly optimistic, nearing one, due to cyclical prediction errors. These errors may alternatively overestimate or underestimate the predicted risk, but offset each other over the long term. The *Mean Rolling Absolute Deviation* (MRAD) is an alternative way to mitigate this problem because it computes the bias statistic over the u -day horizon subperiods and then calculates the average of those rolling windows until we exhaust the sample. The MRAD metric is defined as

$$\text{MRAD}_t(u; \tau) = \frac{1}{\tau - u + 1} \sum_{k=t-\tau+u}^t |\text{BS}_k(u) - 1|, \quad (8)$$

which is the mean absolute deviation of $\text{BS}_k(u)$ from

the overall BS, defined in Equation (7). We set $u = 12$ following the approach of Menchero et al. (2013).²⁴

We further employ Q-statistics, which is a QLIKE loss function proposed in Patton (2011). As mentioned in Menchero et al. (2013), the Q-statistics address both time-resolution and portfolio-resolution issues by penalizing both underprediction and overprediction of risks, defined as

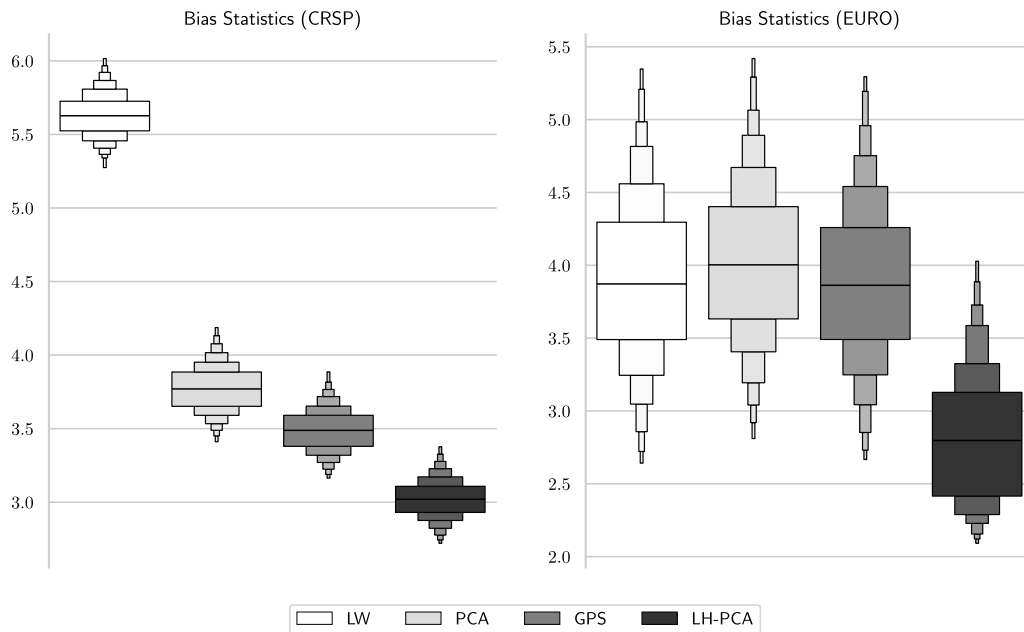
$$\text{Q - statistic} = z_t^2 - \log(z_t^2), \quad (9)$$

which is minimized in expectation when the estimation is perfectly accurate.²⁵

6.3. Main Findings

Similar to the simulation section, we assess the predictive performance of the four estimation methods. Our main empirical measure is the Bias Statistics, defined in Equation (7). For our benchmark specification, we use the median of Bai and Ng (2002) estimator computed over T_L as the fixed number of PCA eigenfactors to extract, without applying RCA.²⁶

Figure 4. Bias Statistic Results



Notes. This figure shows the distributions of the estimated GMVP's Bias Statistics obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. LW stands for Ledoit-Wolf shrinkage method; PCA stands for the plain PCA method with $T_M = 250$; GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M . We employ $T_L = 1,500$ for LH-PCA. These results are obtained from bootstrapping with replacement for cross-validation, using 1,000 different random seeds. For the European data, we cap Z-scores at 100 and -100 to mitigate the impact of a small number of extreme days that cause factor models based on medium-length windows (such as LW, PCA, and GPS) to become overly sensitive to outliers. This conservative treatment favors competing methods over LH-PCA.

Figure 4 presents enhanced box plots²⁷ using 1,000 bootstrapped Z-score samples with replacement. In the CRSP sample, the median Bias Statistic for the Ledoit-Wolf method (white) is 5.63, which is significantly higher than the median Bias Statistic of the standard PCA estimation of 3.77 (light gray). We see that eigenvector correction of GPS (dark gray) shows a modest improvement to 3.49. Our proposed method, LH-PCA (black), shows the best performance at the median Bias Statistic of 3.02. Our analysis now turns to the empirical results from the EURO data set. In Figure 4, the right panel conveys the same overall message as in the CRSP case: despite practitioners' reluctance to apply PCA to data histories longer than one or two years, using a long-history approach (LH-PCA) improves portfolio volatility prediction.²⁸

The realized (or actual) volatility $\sigma(\hat{\omega}, \Sigma)$ results are depicted in Figure 5. Here, we observe a slight deterioration in LH-PCA's realized volatility performance in both CRSP and EURO. In the CRSP sample, the median realized volatility of LH-PCA is 7.33% per annum, which is comparable to the 7.36% achieved by LW. It is slightly higher than the medians for PCA and GPS, although these differences are empirically indistinguishable. In the case of EURO, LH-PCA achieves lower median realized volatility (2.91%) than other benchmark schemes. However, other methods exhibit a wider distribution of realized volatility, which at

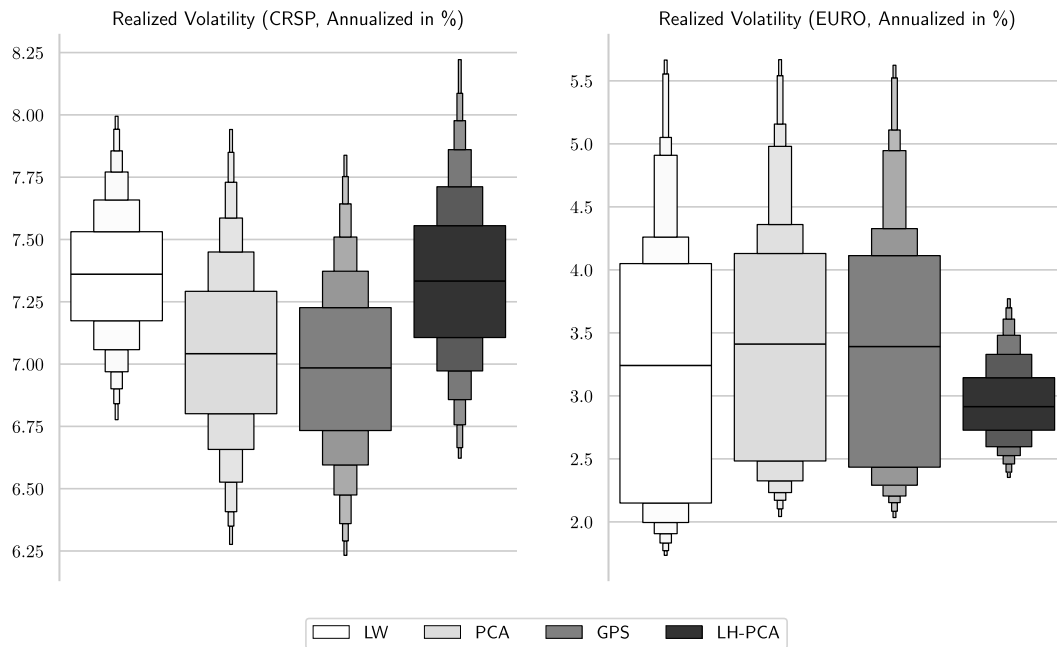
times results in lower realized volatility in the lower quantiles. This mixed result is exactly what we documented in Section 3.5, where the level of actual volatility $\sigma(\hat{\omega}, \Sigma)$ is not a monotonic function of SOR bias $\|\epsilon\|$. Putting this together, LH-PCA demonstrates powerful performance in mitigating SOR bias in both the U.S. and European stock markets, as indicated by Figure 4. However, because of the nonmonotonic nature between the degree of SOR bias and realized volatility, Figure 5 produces less-clear-cut results.

Figures EC.3 and EC.4 in Section D.4 of the Electronic Companions present the empirical comparison using an alternative performance metrics, MRAD and Q -statistics, demonstrating that LH-PCA consistently outperforms competing methods in estimating factor models. Overall, the results in this section provide strong evidence that adopting a long-history approach such as LH-PCA leads to more accurate GMVP risk forecasts and substantially reduces second-order risk.

6.4. Robustness Checks

As a robustness check, we verify that our results remain consistent when using the extended CRSP sample (1961–2021). Table 4 reports subsample estimates based on this extended data set, which provides 61 years of daily stock returns. We partition the sample into three periods: 1961–1980, 1981–2000, and our main CRSP sample of 2001–2021. Although this is not a strict

Figure 5. Realized Volatility Results



Notes. This figure shows the distributions of the estimated GMVP's Realized Volatility obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. LW stands for Ledoit-Wolf shrinkage method; PCA stands for the plain PCA method with $T_M = 250$; GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M . We employ $T_L = 1,500$ for LH-PCA. These results are obtained from bootstrapping with replacement for cross-validation, using 1,000 different random seeds. For the European data, we cap Z-scores at 100 and -100 to mitigate the impact of a small number of extreme days that cause factor models based on medium-length windows (such as LW, PCA, and GPS) to become overly sensitive to outliers. This conservative treatment favors competing methods over LH-PCA.

Table 4. CRSP Longer Sample Results

Periods	PCA	LH-PCA	Difference
Panel A: Excess bias statistics			
CRSP 1961–1980	1.39	0.70	0.69*** (21.10)
CRSP 1981–2000	2.73	1.36	1.38*** (5.22)
CRSP 2001–2021	2.77	2.02	0.75*** (8.71)
CRSP 1961–2021	2.31	1.32	0.99*** (12.10)
Panel B: Excess MRAD			
CRSP 1961–1980	0.98	0.43	0.55*** (14.73)
CRSP 1981–2000	1.89	0.88	1.01*** (11.34)
CRSP 2001–2021	1.92	1.31	0.61*** (8.20)
CRSP 1961–2021	1.61	0.81	0.80*** (22.98)
Panel C: Excess Q-statistics			
CRSP 1961–1980	3.34	1.17	2.17*** (12.79)
CRSP 1981–2000	10.98	3.29	7.69*** (3.81)
CRSP 2001–2021	11.15	6.55	4.60*** (6.76)
CRSP 1961–2021	8.14	3.22	4.92*** (8.55)

Notes. This table presents robustness checks using an extended CRSP sample, covering 61 years of daily data (1961–2021). The data set is divided into three periods: 1961–1980, 1981–2000, and 2001–2021, with the 2001–2021 results corresponding to our main sample period. Additionally, we analyze results for the entire 1961–2021 period. PCA refers to the standard PCA method using $T_M = 250$, whereas LH-PCA employs $T_L = 1,500$. For each panel, we report the deviation of the corresponding test statistics from their ideal values under the assumption of normally distributed daily returns, adopting a conservative perspective. Specifically, the reported figures include Excess Bias Statistics (Panel A), Excess MRAD (Mean Rolling Absolute Deviation) (Panel B), and the Excess mean of Q-statistics (Panel C), all computed using Global Minimum Variance Portfolios across different subsamples and estimation methods. The last column (Difference) reports the difference in the test statistic, along with the associated t -statistics shown in parentheses.

***Significant at the 1% level.

three-fold validation, it serves a similar purpose for robustness analysis. This extension is particularly informative because the earlier periods encompass major market disruptions, including the 1973 oil crisis, 1970s stagflation, and the 1987 Black Monday crash.

Each panel employs a different test statistic to ensure that our findings are not driven by any single performance measure. Table 4, Panel A reports the Excess Bias Statistic (Excess BS), defined as the deviation of the empirical BS from its ideal value of one under normally distributed daily returns and perfectly accurate volatility forecasts. Table 4, Panel B presents the Excess MRAD, which subtracts the MRAD implied by the ideal normal benchmark (0.17), and Table 4, Panel C provides the Excess Q-statistic, obtained by removing the theoretical Q-stat value associated with a standard normal distribution (2.27). Across all three subsamples, as well as the full 61-year data set, we consistently find that LH-PCA outperforms standard PCA because its volatility predictions are consistently closer to the ideal values suggested by each test statistic. Therefore, we conclude that the performance advantage of LH-PCA is neither sample-specific nor method-specific.

In Section D.3 of the Electronic Companions, we detail our choices of tuning parameters. We describe how we select the benchmark number of factors (K) and demonstrate that the outperformance of LH-PCA is robust to variations in the long-history window (T_L).

7. Conclusion

This paper introduces Long-History Principal Component Analysis, a robust framework for estimating precision matrices in high-dimensional portfolios under dynamic factor structures with weak loadings. By extending the data history used for PCA estimation from traditional short-to-medium horizons (e.g., $T_M = 250$ days) to a long history ($T_L = 1,500$ days), LH-PCA mitigates second-order risk bias more effectively than competing methodologies, including Ledoit-Wolf shrinkage, standard PCA, and the GPS eigenvector correction (Goldberg et al. 2022). Theoretically, we show that LH-PCA maintains consistency in estimating the factor-driven covariance and precision matrices, even under variable factor volatilities and time-varying loadings. Empirically, both simulation results and real-world stock return data from the U.S. (CRSP) and European markets demonstrate that LH-PCA produces more accurate forecasts of GMVP volatility, as measured by bias statistics, volatility ratios, MRAD, and Q-statistics.

Our findings have important implications for portfolio risk management in the presence of high-dimensional and dynamically evolving return structures. The superior performance of LH-PCA is particularly evident in environments where narrow factors are prominent and

traditional PCA-based estimators are vulnerable to dispersion bias. Moreover, the method remains effective, even when responsive covariance adjustments are incorporated, further validating its applicability in fast-changing financial markets. In light of increasing investor reliance on data-driven risk models, LH-PCA offers a scalable, theoretically grounded, and empirically validated approach to improving portfolio construction and volatility forecasting under realistic and noisy conditions.

Acknowledgments

The authors thank Jeffrey Bohn, Young Ho Eom, Lisa Goldberg, Markus Pelger, Raman Uppal, and seminar participants at 2023 Financial Management Association Annual Meeting, the INFORMS Annual Meeting 2024, Korea University Business School, Southwestern University of Finance and Economics, the University of California Berkeley, University of Oxford Saïd Business School, and Yonsei University Business School for many valuable comments. B. Kim gratefully acknowledges that a significant portion of this work was completed while he was a visiting scholar at the University of California, Berkeley.

Endnotes

¹ Neglecting the misestimation of expected returns can have a significant impact, similar to the mismeasurement of the covariance matrix. However, this study focuses solely on covariance matrix estimation, leaving expected return estimation errors for future research. Notably, the significance of expected return estimation errors may diminish in short-horizon optimization scenarios.

² We use the terms *strong* and *weak* factor loadings within the framework of Bai and Ng (2023), which extends and generalizes the concepts introduced in Lettau and Pelger (2020), Freyaldenhoven (2022), and Uematsu and Yamagata (2023), among others.

³ Throughout the paper, we use T_L , T_M , and T_S to represent Long, Medium, and Short historical windows, respectively.

⁴ An exception is Northfield, which employs a hybrid model using 60 months of exponentially weighted regressions on fundamental factors and PCA factors derived from residuals. However, it still contends with SOR bias due to high dimensionality—that is, typically $N \gg T \sim 60$.

⁵ Pelger and Xiong (2022a) discuss interpretation of latent PCA factors.

⁶ As illustrated by Proposition 2, achieving the minimum realized variance of the constructed GMVP, which is a practically important goal, may be a conceptually different objective from measuring SOR in our context.

⁷ Because mean daily returns of stocks are small, and the sample mean is a very noisy estimate of the true mean, it is customary not to demean, and therefore to divide by T rather than $T - 1$.

⁸ The Eckart and Young (1936) theorem implies that $U_r D_r V_r^T$ is the best rank- r approximation of Z .

⁹ If the approximation is poor, then the risk of the narrow factor will not be captured by PCA. As a consequence, an optimizer will not attempt to diversify away the risk of that narrow factor and potentially expose the investor to unnecessary and unexpected risk.

¹⁰ Goldberg et al. (2022) (GPS) propose a correction to the estimated eigenvector corresponding to the largest eigenvalue (i.e., market factor). But they do not address this excess dispersion in the narrow (or weak) factors.

¹¹ When $\hat{\Sigma}$ is estimated by the sample covariance matrix with $T > N$ and the true covariance matrix Σ is time-invariant, the estimated precision matrix $\hat{\Sigma}^{-1}$ is characterized by the *inverse-Wishart* distribution.

¹² When over 90% of the stocks have zero true exposure to a given narrow factor, the number of false positives—stocks that falsely appear to have nonzero exposure at the 5% confidence level—will be about half the number of true positives. As a consequence, about one-third of the stocks in our estimate (embedded in the PCA output) of the narrow factor will be false positives, and we cannot capture the narrow factor as cleanly as we would wish. Because the longer history (T_L) reduces the correlation between factor and idiosyncratic return, it reduces the estimated loadings of the false positives, but does not reduce the number of false positives. This suggests that the factor return model is misspecified in the presence of two layers of narrow factors, most likely because it does not allow for interaction terms between narrow factors in the two layers.

¹³ The envelope conditions outlined in Assumptions 2 and 3 align with those in Bates et al. (2013, corollary 1), serving as sufficient criteria for the envelope functions specified in Bates et al. (2013, assumption 4).

¹⁴ Empirical evidence suggests that the factor loadings of individual stocks exhibit mean-reverting behavior over time (Blume 1971, Alexander and Chervany 1980).

¹⁵ More specifically, Freyaldenhoven (2022) states that a factor can be considered weak if it influences only a subset of the observables, where such factors are referred to as *local*.

¹⁶ The positive recurrence of S_t and mild conditions of \mathcal{E}_t ensure that the term $\Lambda(S_t) - \Lambda(S_0) + \mathcal{E}_t$ does not drift indefinitely, provided that the deterministic function $\Lambda(\cdot)$ is well-behaved.

¹⁷ The average Frobenius norm serves as a measure of the average error within the entries of the approximated population covariance and precision matrices.

¹⁸ In Bai and Ng (2023), proposition 6 proves that the average error in estimating Λ_0 asymptotically vanishes provided that $\alpha_r > 0$ and $\frac{N^{1-\alpha_r}}{T} \rightarrow 0$, where α_r is the weakest factor strength out of the r factor loadings in Λ_0 .

¹⁹ We acknowledge and appreciate the advice of the anonymous Associate Editor on this matter.

²⁰ Robustness checks for different BN penalty functions are provided in Section D.3 of the Electronic Companions.

²¹ See Section B of the Electronic Companions for details.

²² It is noteworthy that the performance of one-year GPS is fairly decent when only broad factors are present. However, the performance of the method deteriorates with the inclusion of a single layer of narrow factors and deteriorates further with the addition of a second layer of narrow factors. Details are available upon request.

²³ These are Austria (AUT), Belgium (BEL), Switzerland (CHE), Germany (DEU), Denmark (DNK), Spain (ESP), Finland (FIN), France (FRA), United Kingdom (GBR), Greece (GRC), Ireland (IRL), Italy (ITA), Netherlands (NLD), Norway (NOR), Portugal (PRT), and Sweden (SWE). These 16 countries are commonly studied in analyses of European stock markets (see Fama and French 1998 and Bekaert et al. 2009b).

²⁴ Under the assumption of normally distributed daily returns, the ideal value of MRAD is approximately 0.17. In the presence of leptokurtic return distributions, this value tends to increase, implying that our evaluation based on the normal benchmark adopts a conservative perspective.

²⁵ Patton (2011) shows that the Q -statistic is the unique loss function that solely relies on the Z -scores. The ideal value of the expected Q -statistic is approximately 2.27 assuming the normal distribution of daily returns.

²⁶ Refer to Section D.4 of the Electronic Companions for the empirical analyses with RCA.

²⁷ We use the Python library Seaborn to produce enhanced box plots. This enhanced version plots more quantiles than the standard, providing more information about the shape of the distribution, particularly in the tails.

²⁸ The gains in the EURO case appear less pronounced than those in the CRSP case, but this is largely attributable to our treatment of outliers. Unlike the CRSP data set, the European stock market data are less comprehensive and require additional filtering, which likely removes many small and volatile stocks. Even after cleaning, short-window methods (LW, PCA, and GPS) occasionally produced highly inaccurate daily risk forecasts, resulting in extremely large Z -scores (e.g., a Z -score of 400 under the standard PCA factor model). Such distortions undermine meaningful comparisons across forecasting methods. To address this, we capped Z -scores at 100 and -100 for the affected observations—an adjustment that effectively biases the evaluation in favor of short-history algorithms. Despite this conservative treatment, LH-PCA consistently outperforms the alternative approaches; without these caps, its relative advantage would appear even more substantial.

References

- Adrian T, Franzoni F (2009) Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *J. Empirical Finance* 16(4):537–556.
- Ait-Sahalia Y, Matthys F, Osambela E, Sircar R (2025) When uncertainty and volatility are disconnected: Implications for asset pricing and portfolio performance. *J. Econom.* 248:105654.
- Alexander GJ, Chervany NL (1980) On the estimation and stability of beta. *J. Financial Quant. Anal.* 15(1):123–137.
- Anderson EW, Ghysels E, Juergens JL (2009) The impact of risk and uncertainty on expected returns. *J. Financial Econom.* 94(2):233–263.
- Bai J, Ng S (2002) Determining the number of factors in approximate factor models. *Econometrica* 70(1):191–221.
- Bai J, Ng S (2023) Approximate factor models with weaker loadings. *J. Econom.* 235:1893–1916.
- Barigozzi M, Hallin M, Soccorsi S, von Sachs R (2021) Time-varying general dynamic factor models and the measurement of financial connectedness. *J. Econom.* 222(1):324–343.
- Basak GK, Jagannathan R, Ma T (2009) Jackknife estimator for tracking error variance of optimal portfolios. *Management Sci.* 55(6):990–1002.
- Bates BJ, Plagborg-Møller M, Stock JH, Watson MW (2013) Consistent factor estimation in dynamic factor models with structural instability. *J. Econom.* 177:289–304.
- Bekaert G, Engstrom E, Xing Y (2009a) Risk, uncertainty, and asset prices. *J. Financial Econom.* 91(1):59–81.
- Bekaert G, Hodrick RJ, Zhang X (2009b) International stock return comovements. *J. Finance* 64(6):2591–2626.
- Bernardi S, Leippold M, Lohre H (2019) Second-order risk of alternative risk parity strategies. *Risk* 21(3):1–25.
- Blume ME (1971) On the assessment of risk. *J. Finance* 26(1):1–10.
- Brenner M, Izhakian Y (2018) Asset pricing and ambiguity: Empirical evidence. *J. Financial Econom.* 130(3):503–531.
- Clarke R, Silva RD, Thorley S (2011) Minimum-variance portfolio composition. *J. Portfolio Management* 37(2):31–45.
- Del Negro M, Otrok C (2008) Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. Federal Reserve Board of New York Staff Report 326, Federal Reserve Board of New York, New York.
- DeMiguel V, Garlappi L, Uppal R (2009b) Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *Rev. Financial Stud.* 22(5):1915–1953.
- DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009a) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Sci.* 55(5):798–812.

- Ding Y, Li Y, Zheng X (2021) High dimensional minimum variance portfolio estimation under statistical factor models. *J. Econom.* 222(1):502–515.
- Doidge C, Karolyi AG, Stulz RM (2017) The U.S. listing gap. *J. Financial Econom.* 123(3):464–487.
- Eckart C, Young G (1936) The approximation of one matrix by another of lower rank. *Psychometrika* 1(3):211–218.
- Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* 33(1):3–56.
- Fama EF, French KR (1998) Value versus growth: The international evidence. *J. Finance* 53(6):1975–1999.
- Fan J, Liao Y, Wang W (2016) Projected principal component analysis in factor models. *Ann. Statist.* 44(1):219–254.
- Freyaldenhoven S (2022) Factor models with local factors: Determining the number of relevant factors. *J. Econom.* 229(1):80–102.
- Goldberg L, Papanicolaou A, Shkolnik A (2022) The dispersion bias. *SIAM J. Financial Math.* 13(2):521–550.
- Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance* 58(4):1651–1684.
- Kelly B, Pruitt S, Su Y (2019) Characteristics are covariances: A unified model of risk and return. *J. Financial Econom.* 134(3):501–524.
- Kim S, Korajczyk RA, Neuhierl A (2021) Arbitrage portfolios. *Rev. Financial Stud.* 34(6):2813–2856.
- Ledoit O, P ech e S (2011) Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* 151:233–264.
- Ledoit O, Wolf M (2003) Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance* 10(5):603–621.
- Ledoit O, Wolf M (2004a) A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* 88(2):365–411.
- Ledoit O, Wolf M (2004b) Honey, I shrunk the sample covariance matrix. *J. Portfolio Management* 30(4):110–119.
- Lettau M, Pelger M (2020) Estimating latent asset-pricing factors. *J. Econom.* 218(1):1–31.
- Markowitz H (1952) Portfolio selection. *J. Finance* 7(1):77–91.
- Menchero J, Morozov A, Pasqua A (2013) Predicting risk at short horizons: A case study for the USE4D Model. MSCI Model Insight, MSCI Barra documentation, MSCI Barra, New York.
- Morozov A, Wang J, Borda L, Mencehro J (2012) The Barra global equity model (GEM3). MSCI Barra model documentation, MSCI Barra, New York.
- Pastor L (2000) Portfolio selection and asset pricing models. *J. Finance* 55(1):179–223.
- Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. *J. Econom.* 160(1):246–256.
- Pelger M, Xiong R (2022a) Interpretable sparse proximate factors for large dimensions. *J. Bus. Econom. Statist.* 40(4):1642–1664.
- Pelger M, Xiong R (2022b) State-varying factor models of large dimensions. *J. Bus. Econom. Statist.* 40(3):1315–1333.
- Ross SA (1976) The arbitrage theory of capital asset pricing. *J. Econom. Theory* 13(3):341–360.
- Shepard P (2009) Second order risk. Working paper, MSCI Barra, New York.
- Shi F, Shu L, Yang A, He F (2020) Improving minimum-variance portfolios by alleviating overdispersion of eigenvalues. *J. Financial Quant. Anal.* 55(8):2700–2731.
- Su L, Wang X (2017) On time-varying factor models: Estimation and testing. *J. Econom.* 198(1):84–101.
- Uematsu Y, Yamagata T (2023) Estimation of sparsity induced weak factor models. *J. Bus. Econom. Statist.* 41(1):213–227.
- Wang W, Fan J (2017) Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Statist.* 45(3):1342–1374.

Robert M. Anderson is a Long-Term Visiting Professor at the Institute for Advanced Study in Mathematics at Harbin Institute of Technology; and the Coleman Fung Professor Emeritus of Risk Management and a Distinguished Professor Emeritus of Economics and Mathematics at the University of California, Berkeley. His research bridges advanced mathematics and economic theory, applying rigorous mathematical modeling to the fields of mathematical economics and quantitative finance.

Baeho Kim is a professor of finance and business analytics at Korea University Business School. His research integrates financial analytics, risk management, and derivatives pricing and hedging through quantitative modeling of financial risk.

Dean Ryu is an assistant professor of finance at ITAM Business School. His research focuses on asset pricing, with particular emphasis on understanding volatility dynamics using statistical methods and machine learning techniques.