Long-History PCA in a Dynamic Factor Model with Weak Loadings

Electronic Companions

Robert M. Anderson, Baeho Kim and Dean Ryu

A. Proofs

(Proof of Proposition 1) Define $\tilde{\Sigma} := \hat{\Sigma}\Sigma^{-1}\hat{\Sigma}$ as the *sandwich* covariance matrix. Then, the actual variance of the estimated GMVP is given by

$$(\text{Actual variance of } \hat{\omega}) = \sigma^2(\hat{\omega}, \Sigma) = \frac{\mathbf{1}^\top \tilde{\Sigma}^{-1}\mathbf{1}}{(\mathbf{1}^\top \hat{\Sigma}^{-1}\mathbf{1})^2} = \frac{\sigma^4(\hat{\omega}, \hat{\Sigma})}{\sigma^2(\tilde{\omega}, \tilde{\Sigma})} = \frac{(\text{Predicted variance of } \hat{\omega})^2}{(\text{Predicted variance of } \tilde{\omega})} ,$$

where $\tilde{\omega}$ is the weight vector of the estimated GMVP based on $\tilde{\Sigma}$. This implies that $\sigma(\hat{\omega}, \hat{\Sigma})$ is the geometric mean of $\sigma(\hat{\omega}, \Sigma)$ and $\sigma(\tilde{\omega}, \tilde{\Sigma})$; i.e.,

$$(\text{VR}) = \frac{\sigma(\hat{\omega}, \Sigma)}{\sigma(\hat{\omega}, \hat{\Sigma})} = \frac{\sigma(\hat{\omega}, \hat{\Sigma})}{\sigma(\tilde{\omega}, \tilde{\Sigma})} . \tag{EC.1}$$

Next, by the definitions of ϵ and $\tilde{\Sigma}$, we have

$$\tilde{\Sigma}^{-1} = \left(\Sigma^{-1} + \epsilon\right) \Sigma \left(\Sigma^{-1} + \epsilon\right) = \Sigma^{-1} + 2\epsilon + \epsilon\Sigma\epsilon \approx \Sigma^{-1} + 2\epsilon ,$$

where the approximation is justified by small $\|\epsilon\|$. It follows that

$$\sigma^2(\tilde{\omega}, \tilde{\Sigma}) = \frac{1}{\mathbf{1}^\top \tilde{\Sigma}^{-1}\mathbf{1}} \approx \frac{1}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1} + 2 \cdot \mathbf{1}^\top \epsilon \mathbf{1}}$$

and combining (2) and (EC.1) completes the proof. □

(Proof of Proposition 2) Under the hypothesis of $\Sigma\hat{\Sigma}^{-1}1 \approx \lambda 1$ with some $\lambda > 0$, we have

$$\mathbf{1}^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mathbf{1} \approx \frac{\left(\mathbf{1}^\top \hat{\Sigma}^{-1}\mathbf{1}\right)^2}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}}$$

This implies that

$$(\text{Actual variance of } \hat{\omega}) = \frac{\mathbf{1}^\top \hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}\mathbf{1}}{\left(\mathbf{1}^\top \hat{\Sigma}^{-1}\mathbf{1}\right)^2} \approx \frac{\left(\mathbf{1}^\top \hat{\Sigma}^{-1}\mathbf{1}\right)^2}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}} \frac{1}{\left(\mathbf{1}^\top \hat{\Sigma}^{-1}\mathbf{1}\right)^2} = (\text{True variance of } \omega) ,$$

where the approximation becomes exact when $\Sigma\hat{\Sigma}^{-1}\mathbf{1} = \lambda\mathbf{1}$. □

(Proof of Theorem 1) Recall $\Lambda_t = \Lambda_0 + h_{NT}\xi_t \in \mathbb{R}^{N\times r}$. For a fixed pair of (N,T), we have

$$\begin{aligned}
\Sigma_N(T) &= \frac{1}{T}\sum_{t=1}^{T} \left(\Lambda_t F_t + e_t\right)\left(\Lambda_t F_t + e_t\right)^\top \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\left(\Lambda_0 + h_{NT}\xi_t\right)F_t + e_t\right)\left(\left(\Lambda_0 + h_{NT}\xi_t\right)F_t + e_t\right)^\top \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(a_t + b_t + b_t^\top + c_t\right) \ ,
\end{aligned}$$

where a_t, b_t and c_t are given by

$$\begin{aligned}
a_t &= \Lambda_0 F_t F_t^\top \Lambda_0^\top \\
b_t &= \Lambda_0 F_t \left(h_{NT}\xi_t F_t + e_t\right)^\top \\
c_t &= \left(h_{NT}\xi_t F_t + e_t\right)\left(h_{NT}\xi_t F_t + e_t\right)^\top \ .
\end{aligned}$$

As $\left(\sum_{j=1}^m x_j\right)^2 \le m\sum_{j=1}^m x_j^2$ holds for all integers $m \ge 1$, we have

$$\frac{1}{N^2}\left\|\Sigma_N(T) - \overline{\Sigma}_N(T)\right\|_F^2 \le 6\sum_{j=1}^{6}\Psi_j \ ,$$

where letting $N,T\to\infty$ justifies with our assumptions that

$$\begin{aligned}
\Psi_1 &= \frac{1}{N^2}\Big\|\Lambda_0 \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left(F_t F_t^\top - \mu_t^2\right)}_{=o_p(1)} \Lambda_0^\top\Big\|_F^2 \le \frac{1}{N^2}\underbrace{\left\|\Lambda_0\Lambda_0^\top\right\|_F^2}_{=O_p(N)} O_p(1) = O_p\left(\frac{1}{N}\right) \\
\Psi_2 &= \frac{1}{N^2}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(e_t e_t^\top - \nu_t^2\right)\right\|_F^2 \le \frac{1}{N^2}\left(O_p(N) + O_p\left(\frac{N(N-1)}{T}\right)\right) = O_p\left(\frac{1}{\min\{N,T\}}\right) \\
\Psi_3 &= \frac{1}{N^2}\left\|\frac{2}{T}\sum_{t=1}^{T}\Lambda_0 F_t e_t^\top\right\|_F^2 \le \frac{1}{N^2}O_p(N) \le O_p\left(\frac{1}{N}\right) \\
\Psi_4 &= \frac{h_{NT}^2}{N^2}\left\|\frac{2}{T}\sum_{t=1}^{T}\Lambda_0 F_t F_t^\top \xi_t^\top\right\|_F^2 \le O_p\left(\frac{h_{NT}^2}{N^2T^2}Q_2(N,T)\right) = O_p\left(\frac{1}{N}\right) \\
\Psi_5 &= \frac{h_{NT}^2}{N^2}\left\|\frac{2}{T}\sum_{t=1}^{T} e_t F_t^\top \xi_t^\top\right\|_F^2 \le O_p\left(\frac{h_{NT}^2}{N^2}Q_1(N,T)\right) = O_p\left(\frac{1}{N}\right) \\
\Psi_6 &= \frac{h_{NT}^4}{N^2}\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_t F_t F_t^\top \xi_t^\top\right\|_F^2 = O_p\left(\frac{h_{NT}^4}{N^2T^2}Q_3(N,T)\right) = O_p\left(\frac{1}{\min\{N,T\}}\right) \ ,
\end{aligned}$$

which completes the proof

$$\left\| \frac{1}{N} \left(\Sigma_N(T) - \overline{\Sigma}_N(T) \right) \right\|_F^2 = O_p \left(\frac{1}{\min\{N,T\}} \right) .$$

In addition, our assumptions yield that there exists some $c > 0$ such that

$$\min \left\{ \lambda_{\min}\big(\Sigma_N(T)\big),\ \lambda_{\min}\big(\overline{\Sigma}_N(T)\big) \right\} > c ,$$

where $\lambda_{\min}(A)$ represents the smallest eigenvalue of A. Note that this condition regarding the uniform boundedness (away from zero) of the minimum eigenvalue is consistent with a fixed $r < \infty$. This implies that we have

$$\begin{aligned}
\left\| \frac{1}{N} \left(\overline{\Sigma}_N^{-1}(T) - \Sigma_N^{-1}(T) \right) \right\|_F &= \left\| \frac{1}{N} \left(\Sigma_N^{-1}(T) \left(\Sigma_N(T) - \overline{\Sigma}_N(T) \right) \overline{\Sigma}_N^{-1}(T) \right) \right\|_F \\
&\leq \underbrace{\frac{1}{\underline{\lambda}(\Sigma_N(T))}}_{<1/c} \cdot \left\| \frac{1}{N} \left(\Sigma_N(T) - \overline{\Sigma}_N(T) \right) \right\|_F \cdot \underbrace{\frac{1}{\underline{\lambda}(\overline{\Sigma}_N(T))}}_{<1/c} \\
&= O \left(\left\| \frac{1}{N} \left(\Sigma_N(T) - \overline{\Sigma}_N(T) \right) \right\|_F \right) ,
\end{aligned}$$

which directly completes the proof. □

(Proof of Corollary 1) Let $\overline{X}_t = \Lambda_0 F_t + e_t \in \mathbb{R}^{N \times 1}$ and $\overline{\mathbf{X}} = [\overline{X}_1, \ldots, \overline{X}_T]^\top \in \mathbb{R}^{T \times N}$, where we have

$$\frac{1}{T} \overline{\mathbf{X}}^\top \overline{\mathbf{X}} = \Lambda_0 \left(\frac{1}{T} \mathbf{F}^\top \mathbf{F} \right) \Lambda_0^\top + \frac{1}{T} \mathbf{e}^\top \mathbf{e} .$$

Using the same notations as in the proof of Theorem 1, we obtain that the perturbation matrix defined as

$$\Delta_N(T) := \frac{1}{N} \left(\frac{1}{T} \mathbf{X}^\top \mathbf{X} - \frac{1}{T} \overline{\mathbf{X}}^\top \overline{\mathbf{X}} \right)$$

satisfies the condition expressed as

$$\|\Delta_N(T)\|_F^2 \leq 2 \left(\Psi_4 + \Psi_6 \right) = O_p \left(\frac{1}{\min\{N,T\}} \right) .$$

Since the leading eigenvalues of $\frac{1}{T} \overline{\mathbf{X}}^\top \overline{\mathbf{X}}$ are of order $O_p(N^{\alpha_k})$ for $k = 1, \ldots, r$ and of order $O_p(1)$ for the remaining components, the Davis-Kahan theorem ensures that the PC estimators based on $\mathbf{X}$ and $\overline{\mathbf{X}}$ must be asymptotically equivalent. As Λ_0 is time-invariant, the singular value decomposition of $\frac{\overline{\mathbf{X}}}{\sqrt{T}}$ yields $\overline{\mathbf{X}} = \sqrt{T} \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r^\top + \tilde{e}_r$, which leads to

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

within the PCA framework specified in Section 3.2, where the PC estimators of $\overline{\mathbf{X}}$ are expressed as $\left(\tilde{F}_r, \tilde{\Lambda}_r\right) = \left(\sqrt{T}\mathbf{U}_r, \mathbf{V}_r\mathbf{D}_r\right)$. It follows that we have

$$\left\| \frac{1}{N}\left(\overline{\Sigma}_N(T) - \widetilde{\Sigma}_N(T)\right)\right\|_F^2 \leq 2\left(\Psi_1 + \Psi_2\right) = O_p\left(\frac{1}{\min\{N,T\}}\right) .$$

Since our assumptions ensure that the smallest eigenvalue of $\overline{\mathbf{X}}$ is uniformly bounded away from zero, the convergence in probability holds for their inverses as well. Applying Theorem 1 along with the triangle inequality for the Frobenius norm completes the proof. □

B. Ledoit and Wolf model constructions

One of our benchmark methods in assessing the risk forecasting performance is Ledoit and Wolf (2004) shrinkage approach. This shrinkage estimator is computed as

$$\widehat{\Sigma}_{Shrink} = \widehat{\delta^*}\mathbf{F} + (1 - \widehat{\delta^*})\mathbf{S} , \tag{EC.2}$$

where $\mathbf{F}$ is the shrinkage target, $\mathbf{S}$ is the sample covariance matrix. The shrinkage constant, $\widehat{\delta^*} \in [0, 1]$, allows users to form a convex linear combination of the two matrices. We follow Appendix B in the Ledoit and Wolf (2004) in constructing the optimal shrinkage parameter, and use constant-correlation model as suggested in the same paper in constructing the shrinkage target matrix $\mathbf{F}$. Consistent with our simulation and empirical approach, we compute the shrinkage estimator at each moving window in obtaining the risk forecasts.

C. Responsive Covariance Adjustment

In real-world financial markets, volatility is constantly changing and the distribution of stock returns is not stationary over time. Accordingly, practitioners often assign more weight to recently realized returns than to distant observations to capture the contemporaneous covariance structure with a *responsive covariance adjustment* (RCA) based on the Exponentially Weighted Moving Averages (EWMA) model; see Harvey et al. (2018) and Bollerslev et al. (2018), for instance. The generalization to variable volatility is implicitly used in commercially available fundamental and latent factor models, which estimate factor and idiosyncratic volatilities using a short history, such as the EWMA with a half-life T_S of around 40 days (Menchero et al. 2013).

In what follows, we provide a detailed description of our approach to the implementation of the RCA framework. Let $\Sigma_N^{\psi}(t)$ denote the $N \times N$ covariance matrix of the $N \times 1$ factor-based return vector $\psi_t := \Lambda_t F_t$. Within each trailing window of T days, a decay factor $\eta \in (0, 1)$ yields the adjusted factor-based return covariance as a weighted average of $\psi_{t-m}\psi_{t-m}^{\top}$ with $m \in \{0, \ldots, T-1\}$.

Simply put, the daily forecast of factor-based return covariance matrix estimated on day t (for day $t+1$) is given by

$$\widehat{\Sigma}_t^{\psi}(T) = \sum_{m=0}^{T-1} \omega_m \psi_{t-m} \psi_{t-m}^{\top} , \tag{EC.3}$$

where the weights satisfy $\sum_{m=0}^{T-1} \omega_m = 1$ and decrease by fixed proportion as $\omega_{m+1} = \eta\omega_m$ for $m \geq 0$. It is also worth noting that our RCA formulation in Formula (EC.3) does not rely on the estimated factor model. We set $T_S = 40$ days, which yields a decay factor of $\eta = \exp(\frac{log0.5}{40}) \approx 0.9828$. This choice is motivated by Menchero et al. (2013) that use 42-day EWMA volatility half-life. Harvey et al. (2018) do volatility scaling using a half-life of 20 days. Notice that EWMA has been applied in similar contexts; e.g., see Harvey et al. (2018) and Bollerslev et al. (2018).

We apply the RCA schemes independently to both the factor returns and the idiosyncratic factor returns to correct for changing factor volatilities. Note that we treat the idiosyncratic returns as uncorrelated across stocks by zeroing out the non-diagonal elements of the idiosyncratic part of the covariance matrix to estimate the $N \times N$ idiosyncratic return covariance matrix on day t with window width T, which is denoted by $\widehat{\Sigma}_t^{\varepsilon}(T)$. This is consistent with the purpose of dimensional reduction by using a factor model for reducing the tendency of the minimum variance optimizer to overfit the data. We finally obtain the daily forecast of covariance matrix estimated on day t as $\widehat{\Sigma}_t(T) = \widehat{\Sigma}_t^{\psi}(T) + \widehat{\Sigma}_t^{\varepsilon}(T)$ by assuming that the idiosyncratic return is uncorrelated with the factor return in the population.

D. Additional Information on the Empirical Analyses

This section details the empirical analyses discussed in the main body of the paper.

D.1. Data Construction

In constructing the dataset, we apply filters consistent with standard practice in the empirical finance literature. For both the U.S. and European samples, we focus on common stocks listed on major exchanges and retain only those with valid country and industry classifications, ensuring exposure to both country- and industry-specific risk factors. We also exclude microcap stocks priced below \$5.

For the European sample, we identify 16 countries as European (with their corresponding code): 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). This selection aligns with existing studies of international stock returns, such as Fama and French (2017).

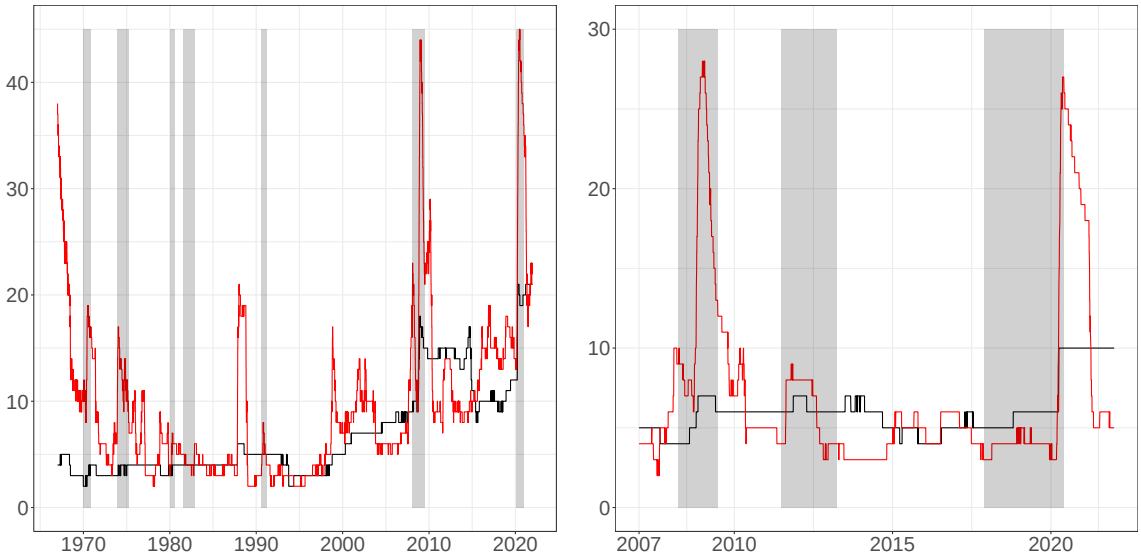
As noted in Section 6, our main CRSP and EURO samples begin in 2001 to facilitate a direct comparison between the U.S. and European stock markets. Euro adoption dates vary across countries—for example, Greece adopted the euro at the start of 2001. Although we could extend the Greek data back to 1999 by applying the fixed conversion rate (1€ = 340.75 GRD) to prices recorded prior to 2000, we choose to focus on daily prices from 2001 onward to ensure data consistency and reliability. Additional details regarding the Greek conversion are available at: https://ec.europa.eu/info/business-economy-euro/euro-area/euro/eu-countries-and-euro/greece-and-euro_en [Accessed 21 Dec. 2025].

D.2. Moving window analysis

We provide more details on our moving window approach that calculates the risk forecasts at a daily frequency. To maintain uniformity of the stock universe across all methods, we generate a single-day forecast on day $t+1$ only for the stocks that are present throughout the longer T_L moving window spanning from day $t-T_L+1$ to t, as well as on the forecast day $t+1$. If a stock is delisted on the prediction day $t+1$, we impute a loss on the prediction day as described in the previous section. Moreover, if a stock declines in value over several days leading up its delisting, then each of those days will have been counted as a prediction day in a previous rolling window. Thus, all of the losses leading up to the demise of a given stock are included in our analysis for the appropriate day and contribute to the bias statistics we compute; thus, we argue that our analysis is free of survivorship bias. Note, however, that our analysis in any given window excludes stocks that were first listed after the first day of the window. Thus, our analysis omits newly created firms, until there is a full T_L window's worth of data from which to make predictions.

Moreover, using the same period, one day, for both data and predictions obviates the need for serial correlation adjustments that would arise if the horizons were different; refer to the related discussion in Menchero et al. (2013). Schwert (1989) show that the estimates of volatility from daily data have much less error than the estimates from low frequency data. In this study, we argue that the benefits of the additional observations from a six-year history of daily returns outweigh the cost of changing factor loadings over that long history; note that six years of weekly data contain only 20% more observations than one year of daily data. We could increase the number of observations by using high-frequency intraday data, but this would require us to extrapolate overnight risk from trade during the day, when the market is open.

Figure EC.1 Number of PCA Eigenfactors using Bai and Ng (2002) Estimator



Note. The left (right) panel reports the number of PCA eigenfactors extracted at each moving window in the CRSP (EURO) daily stock return sample over 1961-2021 (2001-2021). The red solid line shows the number of factors selected using the Bai and Ng (2002) criterion with a medium window ($T_M = 250$), while the black solid line corresponds to the long-history window ($T_L = 1500$). The longer U.S. sample permits estimation back to 1961. The x-axis marks the ending date of each moving window, and the y-axis indicates the extracted number of eigenfactors. Shaded regions denote U.S. NBER recessions (left panel) and OECD-based recessions for the euro area (right panel).

D.3. Tuning Parameters

In implementing LH-PCA, we require two tuning parameters: the choice of the number of factors (K) extracted in each moving window and the choice of the long-history window (T_L). This section details our selection of these parameters and reaffirms that alternative reasonable choices do not alter the main implications presented in the primary analysis.

There are several factor selection criteria that provide consistent estimates even in the presence of weak factors, as supported by recent empirical studies (e.g., Uematsu and Yamagata 2023; Bai and Ng 2023). Freyaldenhoven (2022) explores the selection of K in the presence of 'local' factors, which aligns with our concept of 'narrow' factors and the broader framework of 'weak' factors proposed by Bai and Ng (2023). As noted by Han and Caner (2017) and Freyaldenhoven (2022), the Bai and Ng (2002) (BN) estimator remains consistent in identifying both strong and weak factors when $\alpha > 0$, where $\alpha \in (0, 1]$ represents the strength of the factors and the associated eigenvalues scale as $O_p(N^\alpha)$, with N denoting the population size.

Figure EC.1 plots the number of PCA eigenfactors selected by the BN estimator, where the red (black) line corresponds to its application to 250-day (1500-day) moving windows. Following the

broader literature, we implement the $PC_{p1}(k)$ criterion at each moving window to determine the number of extracted eigenfactors. The BN estimator applied to the medium-history window (red line) exhibits pronounced time variation, with notable increases during major downturns such as the Global Financial Crisis in 2008 and the onset of the COVID-19 crisis in early 2020. These episodes suggest that transient, crisis-related factors become more prominent during periods of economic stress. Consistent with the higher structural complexity of the U.S. equity market, the CRSP sample requires a larger number of factors than the EURO sample. Similar patterns also arise when using Minka (2000)'s PCA estimator (results omitted for brevity), indicating that this behavior is not specific to the BN method.

Given the substantial variability of the BN estimators, relying on its window-by-window output may lead to instability in downstream risk forecasts. For this reason, we adopt a fixed dimension K for all factor-based approaches. Specifically, we use the median number of factors implied by the BN estimator over the long-history window T_L, which yields $K = 23$ for simulation, $K = 12$ for CRSP, and $K = 6$ for EURO. This fixed-K choice provides a more stable benchmark while still reflecting the information contained in the BN procedure. That said, we note that in certain instances the BN estimator's time-varying selection can produce Bias Statistics that outperform those obtained under a fixed-K specification, suggesting that adaptive factor selection may be beneficial in some environments.

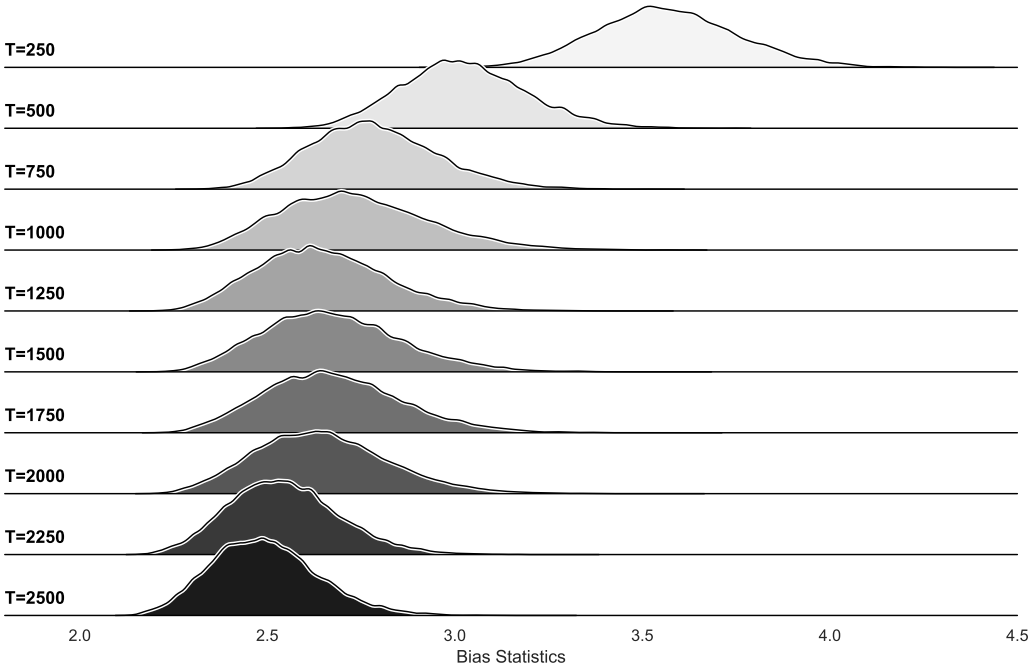
Next, we demonstrate that choosing a moving-window length different from our benchmark of $T_L = 1500$ trading days does not materially affect our empirical findings. As shown in Figure EC.2, the performance gap in Bias Statistics between PCA using $T_M = 250$ days and LH-PCA using $T_L = 1250$ days or more remains substantial. At the same time, the differences between $T_L = 1500$ and the longest window considered, $T_L = 2500$, are small and statistically insignificant, as reflected in the overlapping confidence intervals.

D.4. Additional Empirical Results

As an additional analysis, we present enhanced boxplots, which are analogous to Figure 4, using two alternative performance measures: MRAD and the Q-statistic, as shown in Figures EC.3 and EC.4. They further corroborate our conclusion that, across performance metrics, LH-PCA consistently delivers superior performance relative to competing estimators.

Finally, we verify that applying Responsive Covariance Adjustment (RCA) with a short half-life of $T_S = 40$ days (equivalent to a daily decay factor of 0.9828; see Section C) does not alter our

Figure EC.2 Bias Statistic Across Different Moving Window Lengths



Note. This figure reports the Bias Statistic (BS) across varying moving window lengths, using CRSP data from 2001 to 2021. A window length of 250 corresponds to the standard PCA method with T_M, while a window length of 1,500 represents the LH-PCA approach with T_L. To eliminate cross-sectional variation, we fix the starting point across all specifications so that any differences in BS can be attributed solely to the length of the moving window. Distributions for each BS estimate are derived from bootstrapping with replacement for cross-validation, using 1,000 different random seeds.

Table EC.1 Comparison of Results With and Without Responsive Covariance Adjustment (RCA)

Panel A. RCA Not Applied	Benchmark	LH-PCA	Difference
Benchmark: PCA	3.77	3.02	0.75*** (8.71)
Benchmark: GPS	3.49	3.02	0.46*** (6.04)

Panel B. RCA Applied ($T_s = 40$)	Benchmark	LH-PCA	Difference
Benchmark: PCA	3.63	2.91	0.72*** (9.16)
Benchmark: GPS	3.33	2.91	0.43*** (6.18)

This table reports Bias Statistics computed from daily CRSP returns over the period 2001–2021. 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 $T_L = 1500$ for LH-PCA. These results are obtained from bootstrapping with replacement for cross-validation, using 1,000 different random seeds. Panel A presents results without RCA, corresponding to the main findings in the paper, while Panel B reports results after applying RCA with a half-life of $T_s = 40$ days. For each competing estimator (LW, PCA, and GPS), we report its Bias Statistic alongside the corresponding value for LH-PCA. The final column reports the difference relative to LH-PCA together with the associated t-statistic in parentheses, with *** and ** indicating significance at the 1% and 5% levels, respectively.

main findings in that LH-PCA continues to outperform all competing estimators. This robustness is illustrated in Table EC.1. Figures EC.5, EC.6 and EC.7 further demonstrate that the superior

performance of LH-PCA is attributable to the long-history window itself, rather than the application of RCA. Consistent with our theoretical and simulation findings, LH-PCA reduces SOR bias more effectively than standard PCA regardless of RCA implementation, confirming that the observed improvements are intrinsic to the method rather than an artifact of the RCA technique.

References

Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235:1893–1916.

Bollerslev, T., Hood, B., Huss, J., and Pedersen, L. H. (2018). Risk Everywhere: Modeling and Managing Volatility. *Review of Financial Studies*, 31(7):2729–2773.

Fama, E. F. and French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, 123(3):441–463.

Freyaldenhoven, S. (2022). Factor models with local factors: Determining the number of relevant factors. *Journal of Econometrics*, 229(1):80–102.

Goldberg, L., Papanicolaou, A., and Shkolnik, A. (2022). The Dispersion Bias. *SIAM Journal of Financial Mathematics*, 13(2):521–550.

Han, X. and Caner, M. (2017). Determining the number of factors with potentially strong within-block correlation in error terms. *Econometric Reviews*, 36(6-9):946–969.

Harvey, C. R., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M., and Hemert, O. V. (2018). The Impact of Volatility Targeting. *Journal of Portfolio Management*, 45(1):14–33.

Ledoit, O. and Wolf, M. (2004). Honey, I Shrunk the Sample Covariance Matrix. *Journal of Portfolio Management*, 30(4):110–119.

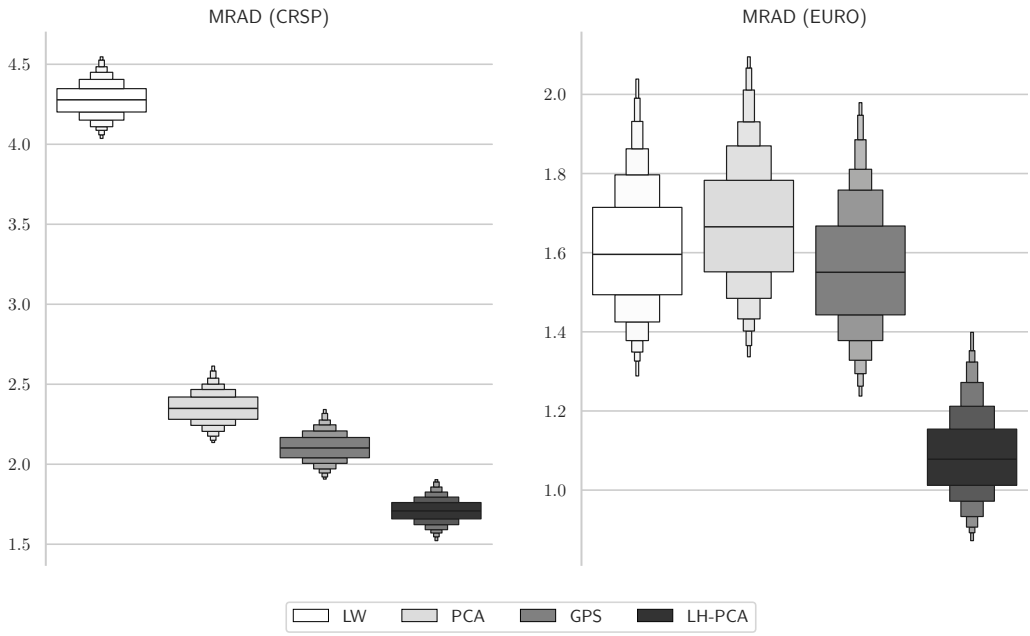
Menchero, J., Morozov, A., and Pasqua, A. (2013). Predicting Risk at Short Horizons: A Case Study for the USE4D Model. MSCI Model Insight, MSCI Barra documentation.

Minka, T. P. (2000). Automatic choice of dimensionality for PCA. *Advances in Neural Information Processing Systems*, 13:598–604.

Schwert, G. W. (1989). Why Does Stock Market Volatility Change Over Time? *Journal of Finance*, 44(5):1115–1153.

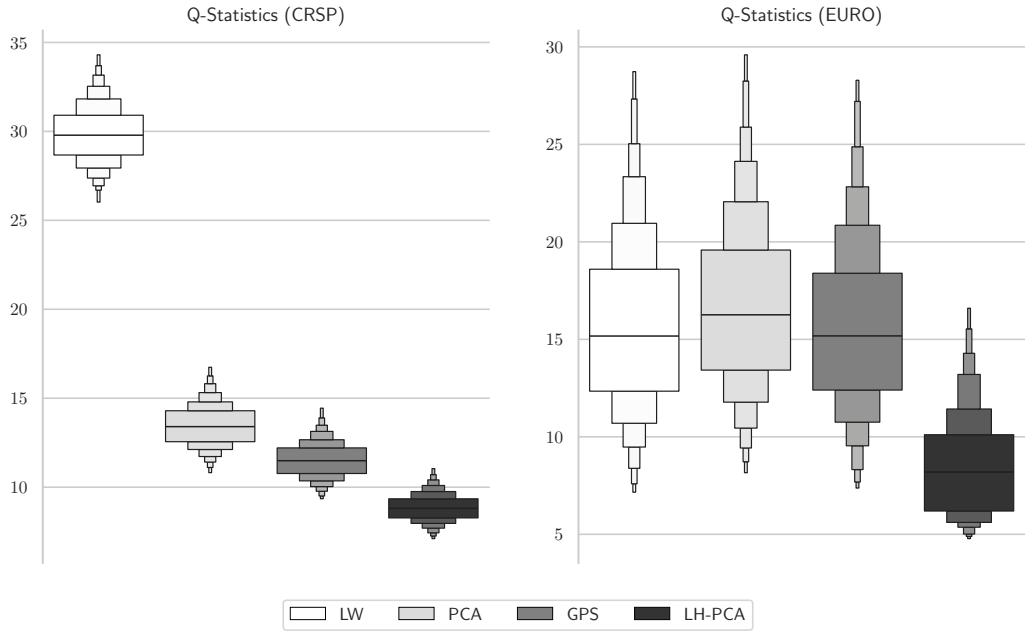
Uematsu, Y. and Yamagata, T. (2023). Estimation of sparsity induced weak factor models. *Journal of Business & Economic Statistics*, 41(1):213–227.

Figure EC.3 MRAD Results (w/o RCA)



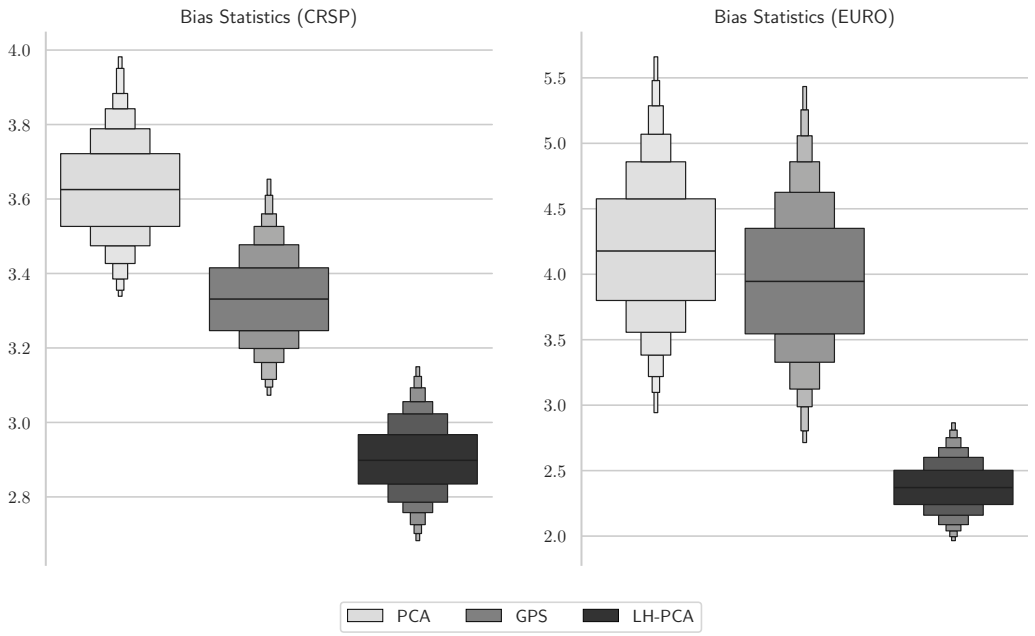
Note. This figure shows the distributions of the estimated GMVP's MRAD obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. LW stands for Ledoit-Wolf shrinkage method, PCA stands for the plain PCA method with T_M, GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M. We employ T_L 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 EC.4 Q-statistic Results (w/o RCA)



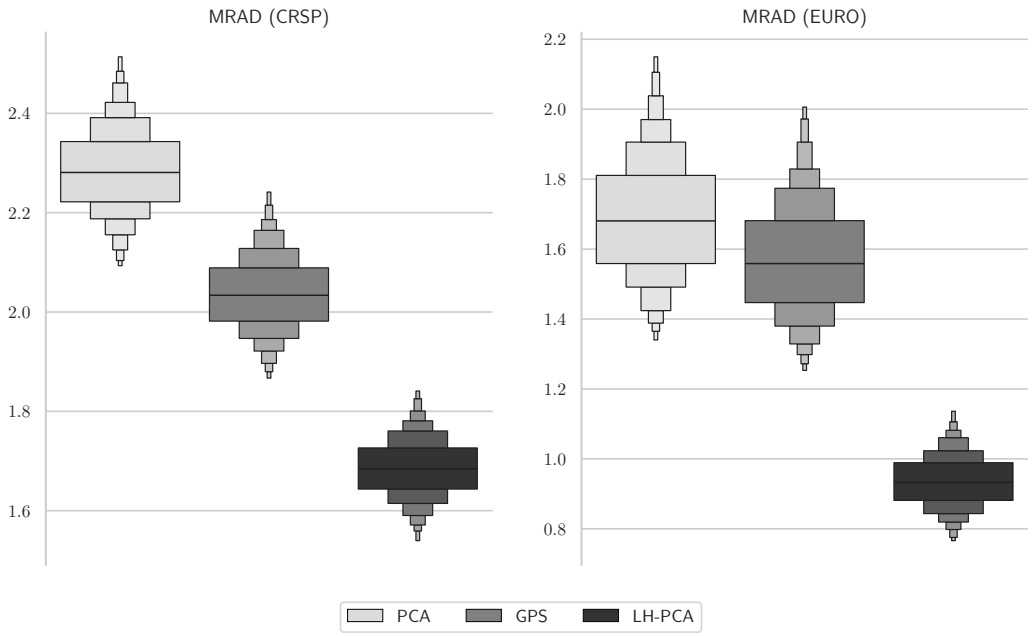
Note. This figure shows the distributions of the estimated GMVP's Q-statistics obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. LW stands for Ledoit-Wolf shrinkage method, PCA stands for the plain PCA method with T_M, GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M. We employ T_L 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 EC.5 Bias Statistic Results (w/ RCA)



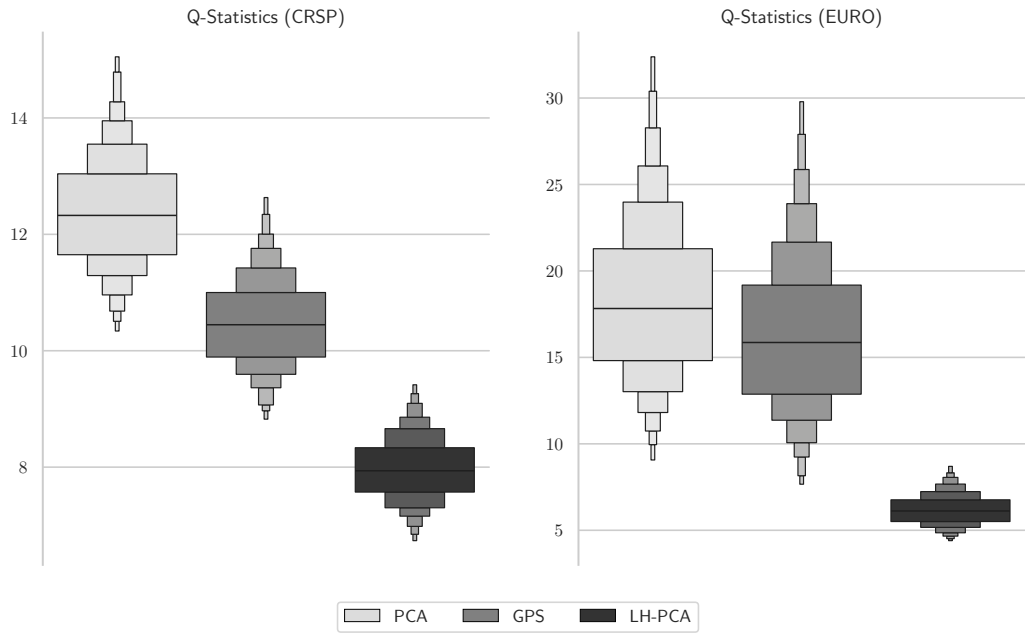
Note. This figure shows the distributions of the estimated GMVP's Bias Statistics obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. The estimated covariance matrices are adjusted by RCA with $T_S = 40$ days. PCA stands for the plain PCA method with T_M, GPS indicates the methodology proposed by Goldberg et al. (2022) with T_M. We employ T_L for LH-PCA. These results are obtained from bootstrapping with replacement for cross-validation, using 1000 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 EC.6 MRAD Results (w/ RCA)



Note. This figure shows the distributions of the estimated GMVP's MRAD obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. The estimated covariance matrices are adjusted by RCA with $T_S = 40$ days. 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 for LH-PCA. These results are obtained from bootstrapping with replacement for cross-validation, using 1000 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 EC.7 Q-statistic Results (w/ RCA)



Note. This figure shows the distributions of the estimated GMVP's Q-statistics obtained from daily CRSP (left panel) and EURO (right panel) stock returns data. The estimated covariance matrices are adjusted by RCA with $T_S = 40$ days. 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 for LH-PCA. These results are obtained from bootstrapping with replacement for cross-validation, using 1000 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.