Analysis of James-Stein for the Leading Eigenvector

Simon Ribas

Florida State University

April 18, 2023

Table of Contents

Markowitz Optimization

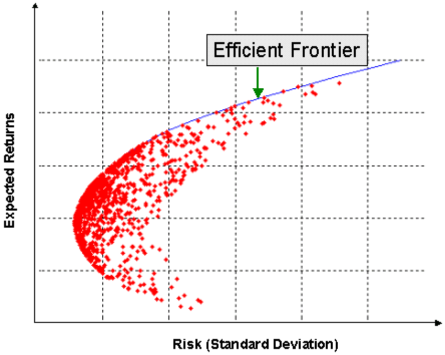

Figure: The Efficient Frontier

Since Markowitz 1952, quantitative investors have constructed portfolios with mean-variance optimization.

The Minimum Variance Portfolio

We look for the fully invested minimum variance portfolio w^* which is the solution to:

$$\min_{w \in \mathbb{R}^p} w^\top \Sigma w$$

subject to the following constraint

$$w^\top \mathbf{1}_p = 1$$

However, we don't know Σ. Instead we estimate $\hat{\Sigma}$ and get $\hat{w}$

Estimating the Covariance Matrix

The Problem: We consider estimating Σ in the high dimension low sample size regime (HL)

Estimating the Covariance Matrix

The Problem: We consider estimating Σ in the high dimension low sample size regime (HL)

Another Problem: The sample covariance matrix $\mathbf{S}$ is singular or "ill-conditioned" in HL

Estimating the Covariance Matrix

The Problem: We consider estimating Σ in the high dimension low sample size regime (HL)

Another Problem: The sample covariance matrix $\mathbf{S}$ is singular or "ill-conditioned" in HL

- Singular or "ill conditioned" matrices imply the sample covariance matrix is not of full rank
- Could result in unstable estimates of portfolio weights and potentially misleading portfolio optimization results

Factor Models

- Statistical models used to explain the variation of asset returns based on set of underlying factors
- Real world factor models include dozens of factors
- Industry, volatility, earnings, and dividend yield are just some of the factors
- With each factor there is an associated return

The One Factor Estimate of Σ

Suppose returns to p securities follow a one-factor model:

$$r = \beta x + \epsilon$$

where r, β, $\epsilon \in \mathbb{R}^p$ and $x \in \mathbb{R}$. Assume $m(\beta) > 0$.
With common assumptions,

$$\Sigma = \sigma^2 \beta\beta^\top + \delta^2 I$$

where $\sigma^2 = \mathrm{var}(x)$, $\delta^2 = \mathrm{var}(\epsilon_i)$ and I is the identity matrix
OR:

$$\Sigma = \eta^2 bb^\top + \delta^2 I$$

where $\eta^2 = \sigma^2|\beta|^2$ and $b = \frac{\beta}{|\beta|}$

Principal Component Analysis

- Technique for reducing dimension of large data sets.
- Looks for the direction of greatest variance which is always an eigenvector
- PCA uses the leading eigenvector of $\mathbf{S}$ as the factor
- Preserve's as much information possible from $\mathbf{S}$ while simplifying the model

PCA estimate of Σ

So estimating Σ amounts to estimating its parameters η^2, b, and δ^2:

$$\hat{\Sigma} = \hat{\eta}^2 \hat{b}\hat{b}^\top + \hat{\delta}^2 I$$

In a PCA model, we take $\hat{b}$ to be the leading sample eigenvector of $\mathbf{S}$. The eigenvalues of $\mathbf{S}$ can be used to determine $\hat{\eta}^2$ and $\hat{\delta}^2$.

Three Estimators

We consider three data-driven estimators:

$$\Sigma_{\mathrm{raw}} = (\lambda^2 - \frac{n-1}{p-1}\ell^2)hh^\top + \frac{n-1}{p-1}\ell^2 I$$
$$\Sigma_{\mathrm{PCA}} = (\lambda^2 - \ell^2)hh^\top + (n/p)\ell^2 I$$
$$\Sigma_{\mathrm{JSE}} = (\lambda^2 - \ell^2)h^{\mathrm{JSE}}(h^{\mathrm{JSE}})^\top + (n/p)\ell^2 I$$

where h^{JSE} is defined by shrinking the entries of h toward their average $m(h)$

$$h^{\mathrm{JSE}} = m(h)\mathbf{1} + c^{\mathrm{JSE}}(h - m(h)\mathbf{1})$$

with shrinkage constant c^{JSE}

Numerical Experiment Set Up

We consider the problem of estimating a covariance matrix with a year's worth of daily observations for stocks in an index like the S&P-500

- Let $p = 500$ and $n = 252$
- Draw independent returns from the one factor model
- Compute $\mathbf{S}$
- Derive the estimators and solve for w_{raw}, w_{pca}, and w_{jse}
- Measure the results
- Repeat experiment 400 times

Errors in portfolio weights

Squared tracking error of an optimized portfolio $\hat{w}$ measures its distance from the optimal portfolio w^*:

$$TE^2 = (\hat{w} - w^*)^\top \Sigma (\hat{w} - w^*)$$

Tracking error is the width of the distribution of return differences between $\hat{w}$ and w^*.

Error in the variance forecast

Variance forecast ratio measures the error in the risk forecast:

$$VFR = \frac{\hat{w}^\top \hat{\Sigma} \hat{w}}{\hat{w}^\top \Sigma \hat{w}}$$

and calculates the estimated portfolio risk over the actual risk of the estimated portfolio.

True variance ratio measures excess variance in the estimated portfolio:

$$TVR = \frac{{w^*}^\top \Sigma w^*}{\hat{w}^\top \Sigma \hat{w}}$$

by dividing the true variance of the true minimum variance portfolio by the true variance of the estimated portfolio.

One Factor Tracking Error

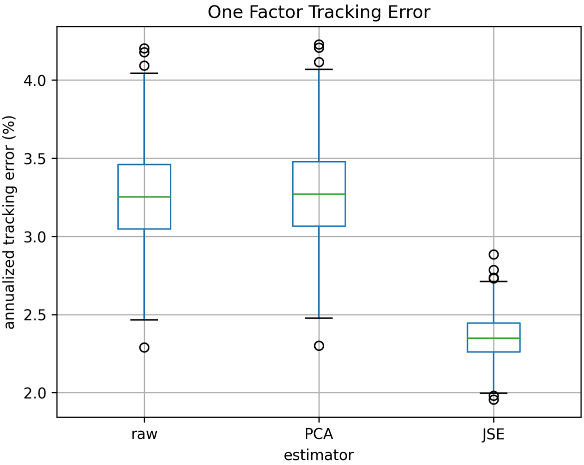

Figure: Tracking Error for 400 experiments under the one factor model of returns. A perfect tracking error is equal to 0

One Factor Variance Forecast Ratio

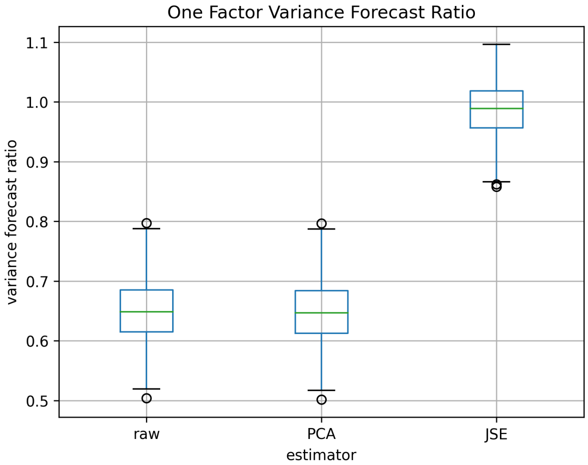

Figure: Variance Forecast Ratio for 400 experiments under the one factor model of returns. A perfect variance forecast ratio is equal to 1

One Factor True Variance Ratio

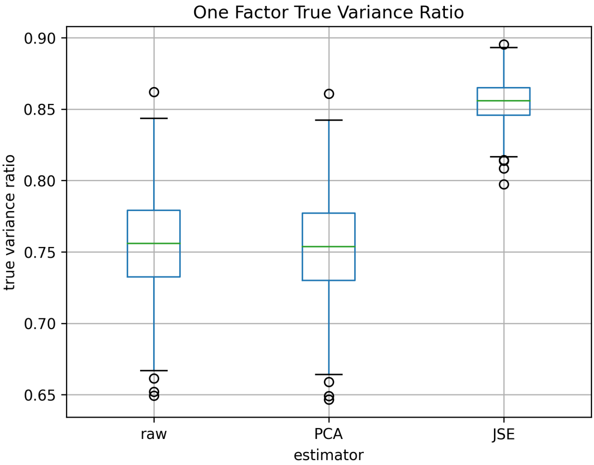

Figure: True Variance Ratio for 400 experiments under the one factor model of returns. A perfect TVR is equal to 1

The Four Factor Model

Suppose now that returns to p securities follow a four-factor model:

$$R = \mathbf{B}_* \psi + \epsilon$$

- ψ is a 4×1 vector of returns to the factors
- $\mathbf{B}_*$ is a $p \times 4$ matrix of exposures to the factors
- ϵ is a $p \times 1$ vector of heterogeneous specific returns

Four Factor Covariance Estimation

Then the covariance matrix of R takes the form

$$\Sigma_* = \mathbf{B}_* \Omega \mathbf{B}_*^\top + \Delta$$

assuming all ψ_k and ϵ_i are mean zero and pairwise uncorrelated.

- Ω and Δ denote the covariance matrices of ψ and ϵ respectively
- Under assumptions both diagonal matrices

Four Factor Covariance Estimation

Construct matrix $\mathbf{Z}$ which are the residuals to the regression of the security returns onto the factors giving:

$$\delta_p^2 = \frac{1}{T}\sum_{t=1}^{T}\mathbf{Z}_{pt}^2$$

- Specific variance given by $\delta^2 = m(\delta_p^2)$
- Market variance given by $\sigma^2 = \lambda^2 - \delta^2$

We can also make use of the Marchenko-Pastur correction such that δ^2 and σ^2 become

$$\delta_{mp}^2 = \frac{tr(\mathbf{S}) - (\lambda^2 + ... + \lambda_4^2)}{p - 4(1 - p/T)}$$

$$\sigma_{mp}^2 = \lambda^2 - \delta_{mp}^2(1 + p/T)$$

Estimators

Three distinct estimators for the four factor simulation:

- Σ_{raw^*} will use δ^2 and σ^2 as the diagonal elements of Δ and Ω respectively.
- Σ_{PCA^*} will improve upon this by using the Marchenko-Pastur Correction
- Σ_{JSE^*} will go one step further by replacing the first column of $\mathbf{B}_*$ with h^{JSE}

Four Factor Tracking Error

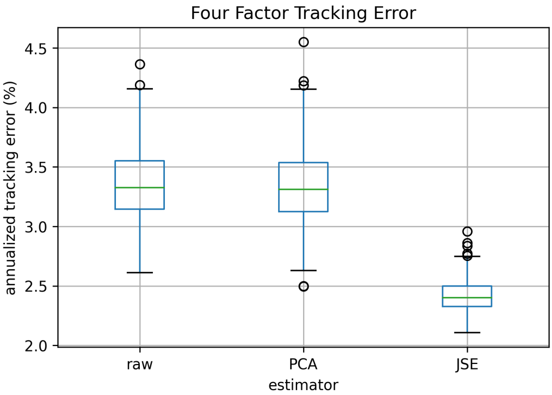

Figure: Tracking Error for 400 experiments under the four factor model of returns. A perfect tracking error is equal to 0

Four Factor Variance Forecast Ratio

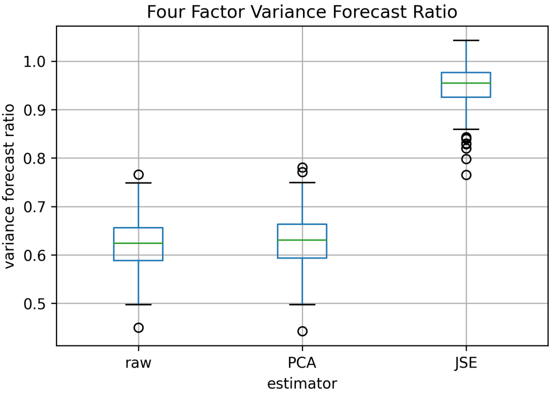

Figure: Variance Forecast Ratio for 400 experiments under the four factor model of returns. A perfect variance forecast ratio is equal to 1

Four Factor True Variance Ratio

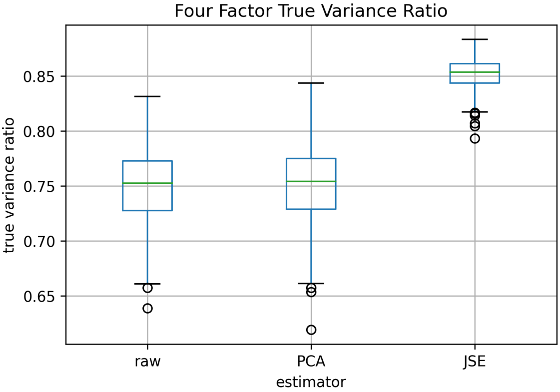

Figure: True Variance Ratio for 400 experiments under the four factor model of returns. A perfect TVR is equal to 1

Empirical Study

- Market viewed through one-factor lens for empirical study
- Conduct tests (in-sample, out-of-sample, bias) to compare JSE's performance to PCA and RAW
- Consider $p = 100$ assets and $n = 50$ observations

In Sample Tests

We want to see whether the JSE estimator does well in correcting for optimization bias. So we follow the following procedure:

- Collect a large time series of stock returns
- Start at the beginning of the data set and construct **S** from our first 50 observations
- Solve the mean variance optimization problem giving us w_{raw}, w_{pca}, and w_{jse}
- Calculate the in-sample variance: $\hat{w}^\top \mathbf{S} \hat{w}$
- Shift up one index in the data set and repeat

In Sample Results

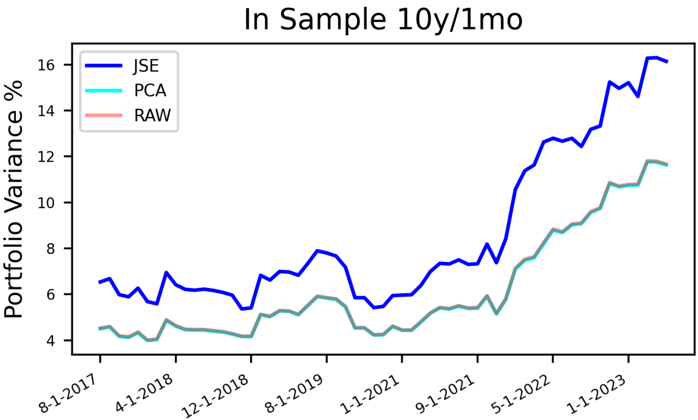

Figure: In-Sample portfolio variance for JSE, PCA, and RAW estimates looking back 10 years using monthly returns.

Out of Sample Tests

Analyze how the estimators perform out of sample. We do this with the following procedure:

- Collect returns for the 100 assets
- Start at the beginning of the data set and construct **S** from out first 50 observations
- Solve the mean variance optimization problem
- Let the portfolios $w_{\mathrm{raw}}, w_{\mathrm{pca}}$, and w_{jse} perform for 6 iterations (months) out of sample
- Take the variance of the returns out of sample
- Shift up one index in the data set and repeat

Out of Sample Results

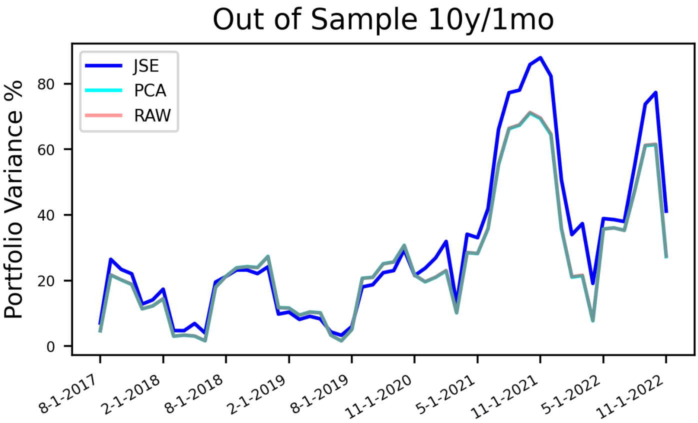

Figure: Out-of-Sample portfolio variance using 6 iterations for jse, PCA, and RAW estimates looking back 10 years using monthly returns

Bias Test Overview

Bias test aims to measure the quality of our covariance matrix estimate. Suppose we have a sequence of demeaned portfolio returns $x^{(1)}, x^{(1)}, ..., x^{(n)}$. If we had the correct estimator for the standard deviation in the forecast, i.e. we solved:

$$w_{\text{true}} = \arg\min_{w} \sqrt{w^T \Sigma w}$$

subject to $w^T \mathbf{e} = 1$. If we considered the following sequence of random variables:

$$\frac{x^{(1)}}{\sigma_{\text{true}}}, \frac{x^{(2)}}{\sigma_{\text{true}}}, ..., \frac{x^{(n)}}{\sigma_{\text{true}}}$$

where σ_{true} is the true standard deviation of the minimum variance portfolio, and took the standard deviation of a large enough subset, it would be approximately close to 1.

Bias Test Implementation

Although we don't have Σ, and thus can't solve for w_{true}, we can use this normalization technique to decide which $\hat{\Sigma}$ is closest to the true covariance matrix. Our sequence now becomes:

$$\frac{x_{\text{jse}}^{(1)}}{\sigma_{\text{jse}}^{(1)}}, \frac{x_{\text{jse}}^{(2)}}{\sigma_{\text{jse}}^{(2)}}, \ldots, \frac{x_{\text{jse}}^{(n)}}{\sigma_{\text{jse}}^{(n)}}$$

where now each $x_{\text{jse}}^{(i)}$ is the next months out of sample return divided by the in-sample standard deviation of the minimum variance portfolio. To get a time-series of this data we consider a large enough sequence of normalized random variables and continuously take standard deviations similar in a similar manner to how a moving average is calculated

Bias Test Results

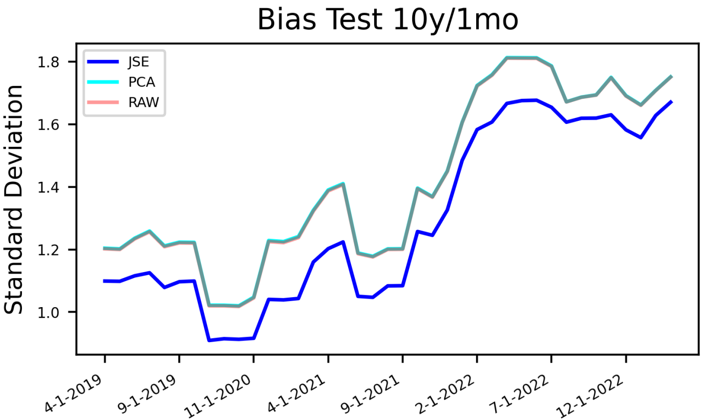

Figure: Bias test using 10 years of history and monthly returns for the JSE, PCA, and RAW estimators. A standard deviation closest to 1 is optimal

Empirical Conclusions

- Eigenvalue correction in the PCA-estimator had little to no effect supporting the simulation experiments
- JSE helps reduce optimization bias (in-sample)
- Bias tests hints that JSE is giving better covariance matrix estimations
- Mixed results out of sample suggest further empirical study is necessary

Some Limitations

- A one factor representation of the market is an over simplification
- Returns in the market are serially correlated, simulation assumes independence of returns
- The time-series encounters various volatility regimes, whereas in simulation, we assume a constant level of volatility.

Next Steps

Although the out-of-sample results in our case did not prove too effective for the JSE estimator, there is much more to analyze to name a few:

- Empirical implementation of the JSE estimator in a Barra-Style factor model
- Simulation accounting for serial correlation of returns
- Simulation with dynamic volatility regimes

References

- Goldberg, L.R. & Kercheval, A. (2023), 'James-stein for the leading eigenvector', *Proceedings of the National Academy of Sciences* **120**(2).
- Goldberg, L.R., Papanicalaou, A. & Shkolnik, A. (2022), 'The dispersion bias', SIAM *Journal of Financial Mathematics* **13**(2), 521-550.
- Goldberg, L.R., Papanicolaou, A., Shkolnik, A. & Ulucam, S. (2020), 'Better Betas', *The Journal of Portfolio Management* **47**(1), 119-136.
- Markowitz, H. (1952), 'Portfolio Selection', Journal of Finance **7**(1), 77-91.

Thank you!

$$h^{\mathrm{JSE}} = m(h)\mathbf{1} + c^{\mathrm{JSE}}(h - m(h)\mathbf{1}) \qquad c^{\mathrm{JSE}} = 1 - \frac{v^2}{s^2(h)}$$

where

$$s^2(h) = \frac{1}{p}\sum_{i=1}^{p}(\lambda h_i - \lambda m(h))^2$$

is a measure of the variation of the entries of λh around their average $\lambda m(h)$ and v^2 is equal to the average of the nonzero smaller eigenvalues of $\mathbf{S}$, scaled by $\frac{1}{p}$.

$$v^2 = \frac{tr(\mathbf{S}) - \lambda^2}{p(n-1)}$$

Simply: it calls for a lot of shrinkage when the average of the nonzero smaller eigenvalues dominates the variation of the entries of λh and only a little shrinkage when the reverse it true.

Solving for Minimum Variance Portfolio

Portfolio variance is given by $\sigma_p^2 = w^\top \Sigma w$ subject to $w^\top \mathbf{1} = 1$.

We can then express the Lagrangian as follows:

$$L = w^\top \Sigma w + \lambda(w^\top \mathbf{1} - 1)$$

Differentiate with respect to w to get:

$$\frac{dL}{dw} = 2\Sigma w + \lambda \mathbf{1} = 0$$
$$w = -\frac{1}{2}\lambda \Sigma^{-1}\mathbf{1}$$

After some rearrangement and simplification we see that:

$$w = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1}\mathbf{1}}$$

ℓ^2

$$\ell^2 = \frac{tr(\mathbf{S}) - \lambda^2}{n-1}$$

Regression for $\mathbf{Z}$

We want to minimize the following by solving for ψ:

$$\min||R - \mathbf{B}_*\psi||_2^2$$

where $\epsilon = R - \mathbf{B}_*\psi$