James-Stein Shrinkage for High Dimensional Eigenvectors

Alec Kercheval[1], Lisa R. Goldberg[2], Hubeyb Gurdogan[3]

[1]Department of Mathematics, Florida State University, Tallahassee, FL 32306
[2]Departments of Economics and Mathematics, University of California, Berkeley 94720, and BlackRock, Inc, NY, NY
[3]Department of Mathematics, UCLA, Los Angeles, CA 90024

Abstract

For high dimensional covariance estimation, the phenomenon of concentration of measure causes a bias in the direction of the estimated leading eigenvector when the number of observations is limited. This can be mitigated by means of James-Stein shrinkage. We describe that method, including rigorous results on the degree of improvement it provides, and discuss applications to quadratic optimization with estimated covariance matrices.

1. Introduction

Markowitz [9] introduced mean-variance optimization to quantitative portfolio construction. A case of interest is the minimum variance portfolio, which is the solution to the optimization problem

$$\min w^\top \Sigma w \tag{1}$$

$$w^\top \mathbf{1} = 1 \tag{2}$$

where we are minimizing over all $w \in R^p$, representing a vector of portfolio weights of p securities, $w^\top$ denotes transpose, and $\mathbf{1}$ represents the p-vector with every entry equal to 1. The covariance matrix Σ is not observed but must be estimated as $\widehat{\Sigma}$, for example from a time series of observed security returns.

However, variance optimization sheds light on covariance matrix estimation error, because variance optimizers tend to be estimation error maximizers. Portfolio solutions will tend to overweight variances and covariances that have been underestimated, and and conversely. The result is a bias, sometimes severe, toward underestimating the variance of the estimated minimum variance portfolio. This has sparked a vast body of research aimed a mitigating estimation error in a covariance matrix.

Motivated by the financial applications mentioned above, we consider the typical setting in which the number of securities p is substantially larger than the number of observations n available for estimation. In the asymptotic limit when p tends to infinity and n is fixed, we call this the high dimension, low sample size (HL) asymptotic regime – the reverse of the classical regime in statistics where p is fixed and n tends to infinity.

Factor models of security returns are typically used to highlight structure and reduce the number of parameters that need to be estimated. In this paper we consider the simple one-factor model:

$$r = \beta x + \varepsilon, \tag{3}$$

where $r \in R^p$ represents a vector of returns, $\beta \in R^p$ is an unknown vector of (non-random) factor loading, $x \in R$ is the random factor return, and $\varepsilon \in R^p$ is a random vector of idiosyncratic returns not explained by the factor. We assume that x and ε are uncorrelated with mean zero, and the components ε_i of ε are mutually independent with bounded variances $var(\varepsilon_i) = \delta_i^2$ such that the limiting average variance is positive:

$$(1/p) \sum \delta_i^2 \to \delta^2 > 0 \tag{4}$$

as $p \to \infty$. (The assumption that the ε_i are independent may be relaxed – see [3].)

We do not need to assume that the variables are normal, subnormal, or belong to any particular parametric family. We do assume they have bounded fourth moments. We emphasize here that only r is observed in each of n time periods.

With this factor model, the $p \times p$ covariance matrix for r can be expressed as a sum:

$$\Sigma = \sigma^2 \beta \beta^\top + \Omega, \tag{5}$$

where $\sigma^2 = var(x)$ and $\Omega = var(\varepsilon)$ is a diagonal matrix. Since the scales of β and σ are not separately identifiable from observation, we write

$$\Sigma = \eta^2 b b^\top + \Omega, \tag{6}$$

where $\eta^2 = \sigma^2 |\beta|^2$ and $b = \beta / |\beta|$. We further assume that the components of β asymptotically do not cluster at zero or infinity, to the extent that $|\beta|^2/p$ converges to a positive finite limit as $p \to \infty$. This means that η^2 tends to infinity at rate p, while Ω has bounded eigenvalues, so we can describe this as a single spiked covariance model.

Since the number of securities p is greater than the number of observations n, the sample covariance matrix is singular, and hence, a poor estimate of Σ for the purpose of optimization. This is remedied by (6), given appropriate estimates of the scalar η^2, the unit vector b, and the diagonal matrix Ω, yielding:

$$\widehat{\Sigma} = \hat{\eta}^2 \hat{b} \hat{b}^\top + \widehat{\Omega}. \tag{7}$$

If Ω were the scalar matrix $\delta^2 I$, then the leading eigenvalue of Σ would be $\eta^2 + \delta^2$ with eigenvector b, and this remains approximately true for general Ω. We can therefore make use of the leading (principal component) eigenvalue and eigenvector of the sample covariance matrix. For the HL regime, we can consistency (as $p \to \infty$) estimate η^2 and δ^2, but not b. As described in the next section, the sample leading eigenvector is not a consistent estimator of the population leading eigenvector, but has a limiting non-zero bias.

In the next section, we describe this bias, discuss the James-Stein shrinkage method to mitigate it, and examine how that mitigation impacts the optimized portfolio when optimized with the estimated covariance matrix.

The results described here summarize a part of the results of a series of papers, primarily [3] and [4], by several researchers working over the past several years. For simplicity of exposition we restrict our discussion to the simple case of a one-factor model and optimization with the single equality constraint in (1). The reader may find more general results and details, full proofs, further background and generalizations, and fuller discussion of the related literature by consulting that paper and the other referenced works [4], [5], [6], [7], [10],[11].

2. James-Stein for Eigenvectors

With assumptions as above, denote by Y the $p \times n$ data matrix of n observations of p samples of the return r modeled in (3). The sample covariance matrix is $S = YY^\top/n$, which is singular when $n < p$, as in our HL regime.

Denote by λ the leading eigenvalue of S, h^{PCA} the (unit) leading eigenvector (principal component) of S, and

$$\ell^2 = \frac{trace(S) - \lambda^2}{n-1}, \tag{8}$$

the average of the remaining non-zero eigenvalues. By convention we choose the sign of h^{PCA} so that the inner product $\langle h^{\text{PCA}}, \mathbf{1} \rangle$ is non-negative.

The eigenvector h^{PCA} can be thought of as an estimator of the population spike b, but with significant bias. A better estimator is the JSE estimator h^{JSE}, defined as follows.

For a p-vector h, let $m(h)$ denote the average of its entries: $m(h) = (1/p)\sum h_i$. The estimator h^{JSE} is a normalized convex combination of h^{PCA} and $m(h^{\text{PCA}})\mathbf{1}$ given by

$$h^{\text{JSE}} = H^{\text{JSE}}/|H^{\text{JSE}}|, \tag{9}$$

where

$$H^{\text{JSE}} = m(h^{\text{PCA}})\mathbf{1} + c^{\text{JSE}}(h^{\text{PCA}} - m(h^{\text{PCA}})\mathbf{1}), \tag{10}$$

the shrinkage constant c^{JSE} is given by

$$c^{\text{JSE}} = 1 - \frac{\ell^2/p}{s^2(h^{\text{PCA}})} \tag{11}$$

and where, for any p-vector h, we define

$$s^2(h) = \frac{1}{p}\sum_{i=1}^{p}(\lambda h_i - \lambda m(h))^2. \tag{12}$$

The quantity $s^2(h)$ measures the variation of the entries of λh around their average $\lambda m(h)$.

The JSE (James-Stein for eigenvectors) estimator h^{JSE} is so-called because it bears a very close relationship to the classical James-Stein shrinkage estimator developed in [12] and [8] for a collection of averages.

Theorem 1 ([3]) *Assume, as $p \to \infty$, $\angle(\beta, \mathbf{1})$ has a positive limit $\Theta < \pi/2$.*

Then,

$$\lim_{p\to\infty} |h^{\text{JSE}} - b| < \lim_{p\to\infty} |h^{\text{PCA}} - b| \quad a.s. \tag{13}$$

Asymptotically,

$$\cos^2(\angle(h^{\text{JSE}}, b)) - \cos^2(\angle(h^{\text{PCA}}, b)) = \frac{(\ell^2/\lambda^2)^2 \cos^2\Theta}{\sin^2\Theta + (\ell^2/\lambda^2)} > 0 \tag{14}$$

Theorem 1 says that h^{JSE} is asymptotically a better estimate of the population eigenvector b than h^{PCA}. This can be intuitively understood in the following way. Consider the unit vector $u = \mathbf{1}/\sqrt{p}$ as the north pole on the unit sphere S, with equator $E = \{v \in S : \langle v, u \rangle = 0\}$. In high dimensions, the concentration of measure phenomenon (see [13] and [1]) means that the volume of S is concentrated near E, so a random unit vector will be approximately orthogonal to u with high probability. This means that additive noise will tend to push estimates of b (like h^{PCA}) toward E and away from u. Shrinking back toward u by the proper amount (such as h^{JSE}) improves the estimate. Theorem 1 quantifies the improvement. See [3], [4] and [2] for further explanation.

When the JSE estimator is used in favor of PCA in constructing an estimated covariance matrix, we observe the significance of the improvement. We compare the following two covariance estimators:

$$\begin{aligned} \Sigma^{\text{PCA}} &= (\lambda^2 - \ell^2) h^{\text{PCA}} (h^{\text{PCA}})^\top + (n/p)\ell^2 I \\ \Sigma^{\text{JSE}} &= (\lambda^2 - \ell^2) h^{\text{JSE}} (h^{\text{JSE}})^\top + (n/p)\ell^2 I. \end{aligned}$$

Both of these preserve the trace of S, which is a consistent estimator of $trace(\Sigma)$. It can also be shown (see [6]) that $(\lambda^2 - \ell^2)/p - \eta^2/p \to 0$ and $(n/p)\ell^2 \to \delta^2$ as $p \to 0$. Hence these are natural estimators of the form (7) that only differ in the choice of $\hat{b}$.

For comparison, we also consider

$$\Sigma^{\text{raw}} = (\lambda^2 - \frac{n-1}{p-1}\ell^2) h^{\text{PCA}} h^{\text{PCA}\top} + \frac{n-1}{p-1}\ell^2 I, \tag{15}$$

which matches the leading eigenvalue and eigenvector of S without correction. We will see that eigenvalue correction (passing from Σ^{raw} to Σ^{PCA}) makes little improvement compared to the eigenvector correction implemented in Σ^{JSE}.

As a measure of performance, we examine the tracking error and variance forecast ratio for optimized portfolios $\widehat{w}$, using one of the two estimated covariance matrices.

The *(squared) tracking error* of an optimized portfolio $\widehat{w}$ relative to the true optimal portfolio w^* is

$$\mathscr{T}\mathscr{E}^2(\widehat{w}) = (\widehat{w} - w^*)^\top \Sigma (\widehat{w} - w^*).$$

The *variance forecast ratio* measures error in the risk forecast:

$$\mathscr{V}(\widehat{w}) = \frac{\widehat{w}^\top \widehat{\Sigma} \widehat{w}}{\widehat{w}^\top \Sigma \widehat{w}}.$$

Theorem 2 ([3]) *With assumptions as before, then asymptotically as $p \to \infty$ with n fixed, almost surely,*

1. $\mathcal{TE}^2(w^{\text{PCA}}) > \mathcal{TE}^2(w^{\text{JSE}})$
2. $\mathcal{V}(w^{\text{PCA}}) \to 0$ *but* $\mathcal{V}(w^{\text{JSE}}) > 0$

Simulations show the asymptotic regime is reached for fairly small p like 500.

A third measure of estimation error as it affects the optimization problem is the *true variance ratio*, defined as

$$\frac{w^{*\top}\Sigma w^*}{\hat{w}^\top \Sigma \hat{w}}.$$

The true variance ratio is less than one, larger is better.

In Figure 1, we show boxplots of simulation experiments examining the performance of optimized portfolios using $\Sigma^{\text{raw}}, \Sigma^{\text{PCA}}, \Sigma^{\text{JSE}}$. In this experiment, we draw $n = 252$ consecutive samples of the factor return f and specific return ε independently with mean zero and standard deviation 16% and 60%, respectively, as typical market values. The factor return is drawn from a normal distribution, while the specific returns are drawn independently from a t-distribution with five degrees of freedom. The factor loadings β are drawn once independently from a normal distribution with mean 1 and variance 0.25 to imitate typical market betas, and are kept fixed across time and experiments. We are in dimension $p = 500$, and simulate the 252-period sample a total of 400 times to produce the box plots shown.

These results are typical of many simulation experiments with varying parameters. The tracking error and true variace ratio are significantly improved when employing the JSE estimate as compared to PCA. The variance forecast ratio shows even more dramatic improvement, indicating that the estimated variance is very close to the (unobserved) true variace of the estimated portfolio.

References

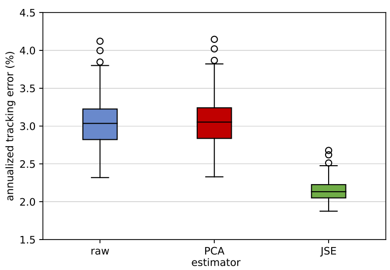
[1] David L Donoho et al. High-dimensional data analysis: The curses and blessings of dimensionality. *AMS math challenges lecture*, 1(2000):32, 2000.

[2] Bradley Efron and Carl Morris. Stein's paradox in statistics. *Scientific American*, 236(5):119–127, 1977.

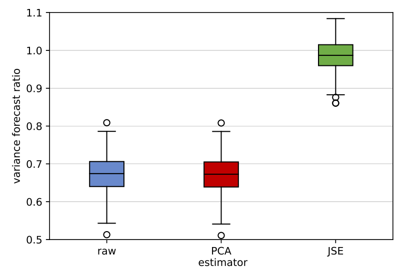
[3] Lisa R. Goldberg, Hubeyb Gurdogan, and Alec Kercheval. Portfolio optimization via strategy-specific eigenvector shrinkage. *Finance and Stochastics*, 25:665–706, 2025.

[4] Lisa R. Goldberg and Alec N. Kercheval. James-Stein for the leading eigenvector. *Proceedings of the National Academy of Sciences*, 120(2), 2023.

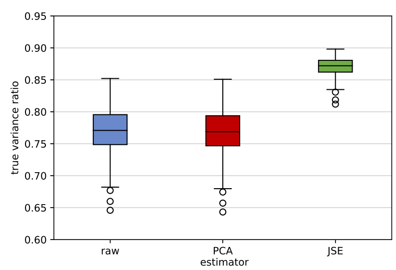
[5] Lisa R. Goldberg, Alex Papacinicolau, Alex Shkolnik, and Simge Ulucam. Better betas. *The Journal of Portfolio Management*, 47(1):119–136, 2020.



(a) Tracking error



(b) Variance forecast ratio



(c) True variance ratio

Figure 1: Portfolio-level accuracy metrics for simulated minimum variance portfolios optimized with Σ^{raw}, Σ^{PCA}, and Σ^{JSE}: (a) annualized tracking error, (b) variance forecast ratio, and (c) true variance ratio. A perfect tracking error is equal to zero, and perfect variance forecast ratios and true variance ratios are equal to one. The estimated covariance matrix is based on $n = 252$ observations of $p = 500$ securities. Each boxplot summarizes 400 simulations. The experiments show that eigenvalue correction (PCA) makes no improvement, but the eigenvector correction (JSE) is substantial. Figures are taken from [4].

[6] Lisa R. Goldberg, Alex Papanicolau, and Alex Shkolnik. The dispersion bias. *SIAM Journal of Financial Mathematics*, 13(2):521–550, 2022.

[7] Hubeyb Gurdogan and Alec Kercheval. Multi anchor point shrinkage for the sample covariance matrix. *SIAM Journal of Financial Mathematics*, 13(3):1112–1143, 2022.

[8] W. James and Charles Stein. Estimation with quadratic loss. In *Proc. Fourth Berkeley Symp. Math. Stat. Prob.*, pages 361–397, 1961.

[9] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–92, 1952.

[10] Alex Shkolnik. James-Stein estimation of the first principal component. *Stat*, 11(1), 2022.

[11] Alex Shkolnik, Alec Kercheval, Hubeyb Gurdogan, Lisa R. Goldberg, and Haim Bar. Portfolio selection revisited. *The Annals of Operations Research*, 346:137–155, 2024.

[12] Charles Stein. Inadmissibility of the usual estimator for the mean of a multivariate distribution. In *Proc. Third Berkeley Symp. Math. Stat. Prob.*, pages 197–206, 1956.

[13] Michel Talagrand. A new look at independence. *The Annals of probability*, pages 1–34, 1996.