Portfolio optimization via strategy-specific eigenvector shrinkage

Alec Kercheval (akercheval@fsu.edu)

October 18, 2023

UC Berkeley CDAR

Department of Mathematics
Florida State University

Collaborators

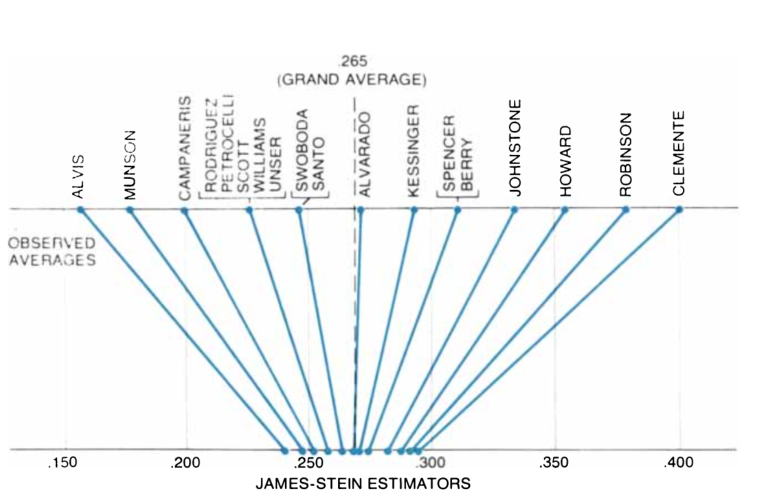
Lisa Goldberg, UC Berkeley CDAR, and Aperio by Blackrock

Hubeyb Gurdogan, UCLA Math

Alex Shkolnik, UCSB

Chapter 1: James-Stein Shrinkage

The James-Stein estimator



source: Efron and Morris, 1977

Given p players, let μ_i, $i = 1, \ldots, p$, be the probability that player i gets a hit when at bat.

Let z_i be the batting average over the first 45 games of the season.

Suppose

$$\mu \sim \mathcal{N}(m\mathbf{1}, \tau^2 I) \quad \text{and} \quad z|\mu \sim \mathcal{N}(\mu, \nu^2 I). \tag{1}$$

Bayes Rule calculation:

$$\mu^{Bayes} \equiv E[\mu|z] = (1-c)z + c(m\mathbf{1}), \tag{2}$$

where $c = \nu^2/(\tau^2 + \nu^2)$.

Observables:

Let $\bar{z} = (1/p) \sum z_i$ and $s^2(z) = \sum_{i=1}^{p} (z_i - \bar{z})^2/(p-3)$.

Then

$$E[\bar{z}] = m \text{ and } E_\nu \left[\frac{\nu^2}{s^2(z)} \right] = \frac{\nu^2}{\tau^2 + \nu^2} = c \tag{3}$$

Define the JS (shrinkage) estimator

$$\mu^{JS} = (1 - c^{JS}) z + c^{JS} (\bar{z}\mathbf{1}) \tag{4}$$

where

$$c^{JS} = \frac{\nu^2}{s^2(z)} \tag{5}$$

– "Empirical Bayes"

James-Stein estimator

James and Stein (1961): The sample average z is inadmissible as an estimator of μ when $p > 3$.

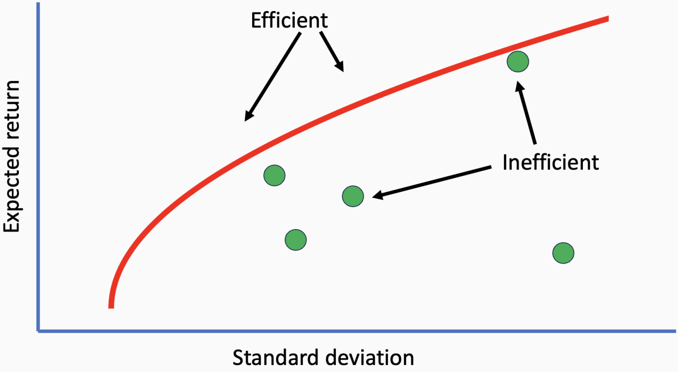
$$E_{\mu,\nu}\left[|\mu^{JS} - \mu|^2\right] < E_{\mu,\nu}\left[|z - \mu|^2\right]. \tag{6}$$

for all $\mu \in R^p$, $\nu \in R$.

Chapter 2: Estimating Covariance in the HL Regime

In 1952, Harry Markowitz framed investment as a tradeoff between expected return and variance

Efficient Frontier



Portfolio optimization in practice

An efficient portfolio is a solution to a quadratic optimization with k linear constraints.

C: $p \times k$ matrix of constraint gradients
a: k-vector of constraint targets

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

Efficient frontier $(k = 2)$:

$$C = \begin{pmatrix} 1 & \mu_1 \\ \vdots & \vdots \\ 1 & \mu_p \end{pmatrix} \qquad a = \begin{pmatrix} 1 \\ \mu \end{pmatrix}$$

Problem: Σ is unknown so we use an estimate, yielding an optimized portfolio that is not optimal.

How to estimate Σ: One-factor model

$p =$ number of securities $>> n =$ number of observations.

Suppose returns $r \in \mathbb{R}^p$ follow a one-factor model:

$$r = \beta f + \epsilon$$

Only r is observed. $f \in \mathbb{R}$ and $\epsilon \in \mathbb{R}^p$ are random, uncorrelated. $\beta \in \mathbb{R}^p$ is an unknown parameter, $|\beta|^2/p$ has a finite limit.

If $\mathrm{var}(f) = \sigma^2$, $x = f/\sigma$, $\mathrm{var}(\epsilon) = \delta^2 I$, $\eta^2 = \sigma^2|\beta|^2$, then

$$r = \eta b x + \epsilon \text{ and } \Sigma = \eta^2 b b^\top + \delta^2 I$$

To estimate: two scalars, η^2 and δ^2, and a unit p-vector b.
Note: b is the leading eigenvector of Σ, with eigenvalue $\eta^2 + \delta^2$.

Inputs: Sample covariance matrix

$Y = p \times n$ data matrix, columns $= n$ observations of $r \in \mathbb{R}^p$

$r^i = \eta b x^i + \epsilon^i$, $i = 1, 2, \ldots, n$

Sample covariance matrix $S = YY^\top/n$.

rank$(S) = n < p$, singular

Define

$$\lambda^2 = \text{leading eigenvalue of } S$$

$$h = h^{\text{PCA}} \text{ leading unit eigenvector of } S$$

$$\ell^2 = (\text{Trace}(S) - \lambda^2)/(n-1)$$

A family of estimates of Σ

HL = high dimension low sample size regime.

Asymptotic limits:

$$\lim_{p\to\infty}(\lambda^2-\ell^2)/p = \left(\frac{|X|^2}{n}\right)\lim_{p\to\infty}\eta^2/p < \infty$$

where $X = (x^1, \ldots, x^n)$.

$$\lim_{p\to\infty} n\ell^2/p = \delta^2, \text{ but } \lim_{p\to\infty}|h-b| \neq 0$$

Family of estimators of Σ:

$$\Sigma^v = (\lambda^2-\ell^2)vv^\top + (n/p)\ell^2 I.$$

Note: $\text{Trace}(\Sigma^v) = \lambda^2 + (n-1)\ell^2 = \text{Trace}(S)$

PCA covariance estimator

Set $v = h \equiv h^{\mathrm{PCA}}$

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

Σ_{PCA} is an S-POET estimator of W. Wang and J. Fan. 2017

Now set $v = h^{\mathrm{JSE}}$

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

JSE eigenvector estimator h^{JSE}

C = k-dim linear subspace, $\angle(\beta, C) < \pi/2$.

Define h^{JSE}, an eigenvector shrinkage estimator:

$$h_C = \mathop{proj}_{C}(h), \qquad c^{JSE} = \frac{\ell^2/\lambda^2}{1 - |h_C|^2}$$

$$H^{JSE} = (1 - c^{JSE})h + c^{JSE} h_C$$

and

$$h^{\mathrm{JSE}} = \frac{H^{JSE}}{|H^{JSE}|}.$$

Why do we call it JSE?

Baseline special case: $C = span(\mathbf{1})$

Then $h_C = \bar{h}\mathbf{1}$ and $1 - |h_C|^2 = |h - \bar{h}\mathbf{1}|^2$

$$c^{JS} = \frac{\nu^2}{s^2(z)} \qquad c^{JSE} = \frac{\ell^2/\lambda^2}{|h - \bar{h}\mathbf{1}|^2}$$

$$\begin{aligned} \mu^{JS} &= (1 - c^{JS})z + c^{JS}(\bar{z}\mathbf{1}) \\ H^{JSE} &= (1 - c^{JSE})h + c^{JSE}(\bar{h}\mathbf{1}) \end{aligned}$$

HL Regime eigenvector estimation

Theorem (Goldberg, Gurdogan, K. 2023)
Assume, as $p \to \infty$, $|\beta_i|$ is bounded, $|\beta|^2/p$ has a positive limit, and $\angle(\beta, C)$ has a positive limit $\Theta < \pi/2$. Then,

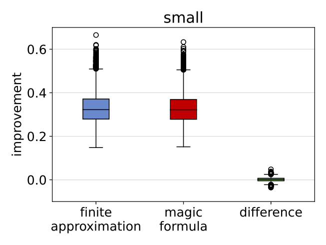
$$\lim_{p\to\infty} |h^{\mathrm{JSE}} - b| < \lim_{p\to\infty} |h - b| \quad \textit{a.s.}$$

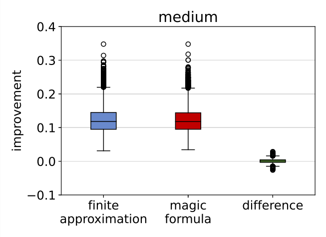
Asymptotically,

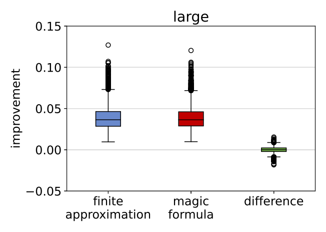
$$\cos^2(\angle(h^{\mathrm{JSE}}, b)) - \cos^2(\angle(h, b)) = \frac{(1-\psi_\infty^2)^2 \cos^2\Theta}{\sin^2\Theta + (1-\psi_\infty^2)} > 0$$

where $\psi_\infty^2 = \lim_{p\to\infty} \frac{\lambda^2 - \ell^2}{\lambda^2} \in (0,1)$.

How well does the improvement formula work for finite p and different angles Θ?







Box plots generated from 10000 simulations of $n = 24$ monthly observations of returns to $p = 3000$ securities, $k = 2$. Note the difference in scale for small, medium and large angles. Graphics by Stephanie Ribet.

Idea: Oracle estimator

An oracle estimator:

Let $U = span(h, C)$, $h^o = b_U / |b_U|$.

Note: h^o is the unit vector in $span(h, C)$ closest to b (hence closer than h).

Lemma (Gurdogan - K, 2022): $|h^o - h^{\mathrm{JSE}}| \to 0$ as $p \to \infty$.

Back to the Optimization Problem

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

- Portfolio variance error is mostly due to bias in the leading sample eigenvector not the leading eigenvalue.
- JSE using the constraint C improves Markowitz portfolio risk estimates in high dimensions.
- $\angle(\beta, C) < \pi/2$, if one column of C is $\mathbf{1}$ since empirically $\angle(\beta, \mathbf{1}) < \pi/2$.

Two covariance matrix estimators

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

w_{PCA}: $(w^\top \Sigma_{\text{PCA}} w)$-minimizing portfolio

w_{JSE}: $(w^\top \Sigma_{\text{JSE}} w)$-minimizing portfolio

w^*: $(w^\top \Sigma w)$-minimizing portfolio

We compare minimized variance of w_{PCA}, w_{JSE}, and w^* theoretically and in simulation experiments.

Errors in optimal portfolios

True variance ratio

$$\frac{{w^*}^\top \Sigma w^*}{w_{\mathrm{PCA}}^\top \Sigma w_{\mathrm{PCA}}} \text{ or } \frac{{w^*}^\top \Sigma w^*}{w_{\mathrm{JSE}}^\top \Sigma w_{\mathrm{JSE}}}$$

True variance ratio is less than one, larger is better.

Other measures: tracking error, variance forecast ratio

Asymptotic Theory

For a unit p-vector v the optimization bias $\mathcal{E}_p(v, C, a)$ controls the variance $\mathcal{V}(w^v)$ of a portfolio optimized with Σ^v asymptotically:

$$\mathcal{V}(w^v) = K\mathcal{E}_p^2(v, C, a) + O(1/p)$$

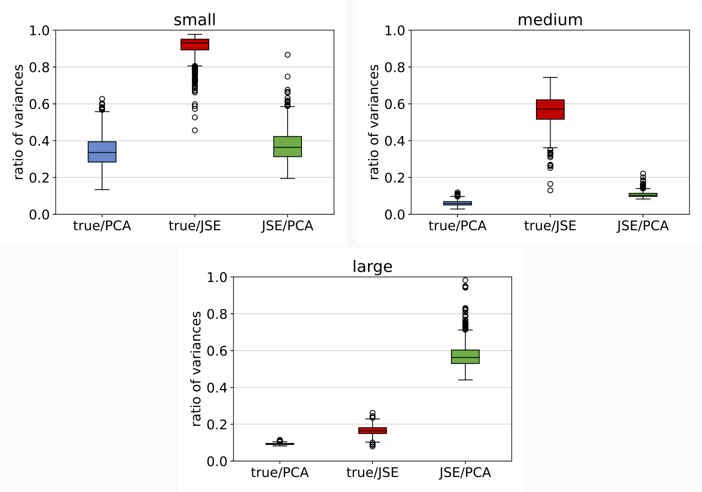
and

$$\mathcal{E}_p(b, C, a) = 0, \ \lim_{p\to\infty} \mathcal{E}_p(h^{\text{JSE}}, C, a) = 0, \ \lim_{p\to\infty} \mathcal{E}_p^2(h^{\text{PCA}}, C, a) > 0.$$

$\mathcal{E}_p(v, C, a)$ is given by a formula involving v, C, a, and b.

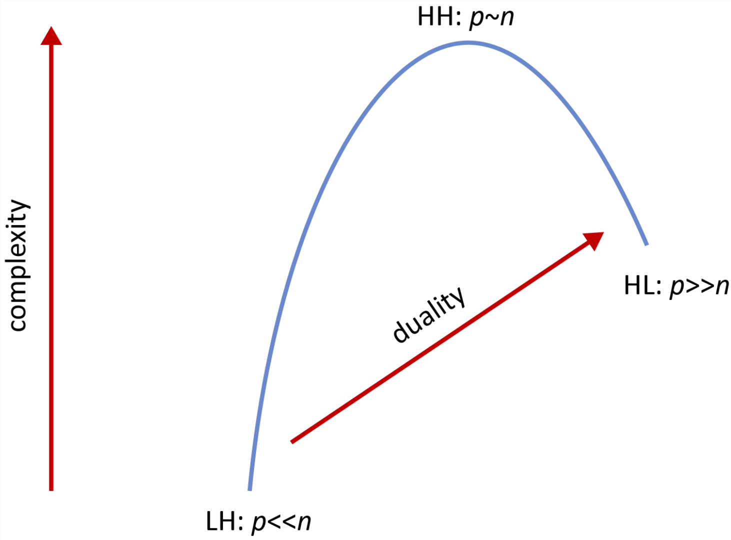
$$\begin{aligned} \mathcal{E}_p(v, C, a) &= \frac{(b^\top \alpha)(1 - ||v_C||^2) - b^\top (v - v_C)(v^\top \alpha)}{(1 - ||v_C||^2)} \\ \alpha &= \frac{(C^\dagger)^\top a}{||(C^\dagger)^\top a||} \end{aligned}$$

True Variance Ratios, $p = 3000$ simulation

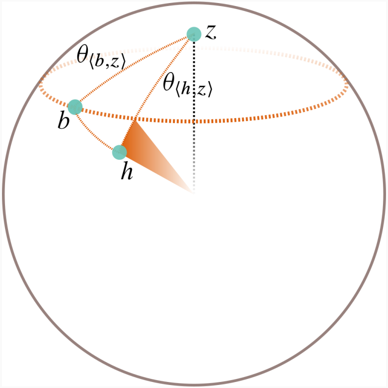


Box plots are generated from 10000 simulations of $n = 24$ monthly observations of returns to $p = 3000$ securities. Graphics by Stephanie Ribet.

The Blessing of Dimensionality



Dispersion bias geometry; Concentration of measure in high dimension



source: Goldberg, Papanicalaou & Shkolnik (2022)

Here $z = \mathbf{1}/\sqrt{p}$. With high probability, $\angle(h, z) > \angle(b, z)$

References

Goldberg, L. R., Gurdogan, H. & Kercheval, A. (2023), 'Portfolio optimization via stategy-specific eigenvector shrinkage', *in submission* .

Goldberg, L. R. & Kercheval, A. (2023), 'James Stein for the leading eigenvector', *Proc. National Academy of Sciences* **120**(2), 1–9.

Goldberg, L. R., Papanicalaou, A. & Shkolnik, A. (2022), 'The dispersion bias', *SIAM Journal on 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.

Gurdogan, H. & Kercheval, A. (2022), 'Multi anchor point shrinkage for the sample covariance matrix', *SIAM Journal on Financial Mathematics* **13**(3), 1112–1143.

Shkolnik, A. (2022), 'James-Stein estimation of the first principal component', *Stat* **11**(1).

Chapter 3: More simulations!

Tracking error

Squared tracking error of an estimated portfolio w relative to the true optimal portfolio w^*:

$$(w - w^*)^\top \Sigma (w - w^*).$$

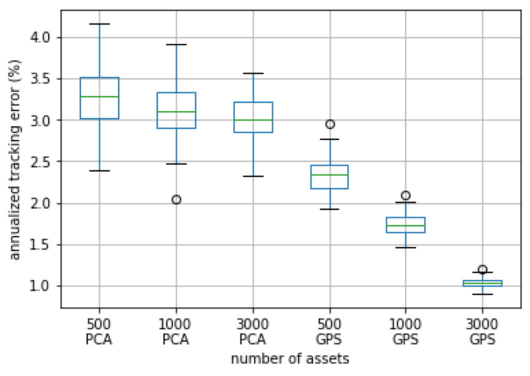
Variance forecast ratio compares estimated variance over true variance of w:

$$\frac{w^\top \widehat{\Sigma} w}{w^\top \Sigma w}$$

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

Ideally, tracking error is 0.

Simulated tracking error



source: Goldberg & Kercheval (2023)

Based on 100 simulations of a year's worth of daily returns following a one-factor model. Ideally, tracking error is 0.

Variance forecast ratio

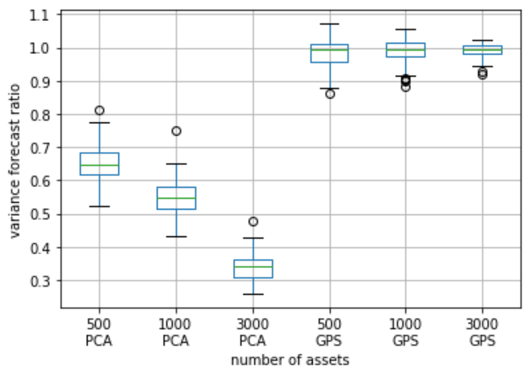
Variance forecast ratio measures error in the risk forecast:

$$\mathcal{V} = \frac{\widehat{w}^\top \widehat{\Sigma} \widehat{w}}{\widehat{w}^\top \Sigma \widehat{w}}.$$

This is the ratio of the estimated risk (variance) over the actual risk of the estimated minimum variance portfolio.

Ideally, the variance forecast ratio is 1. Likely, it is less than 1.

Simulated variance forecast ratio



source: Goldberg & Kercheval (2023)

Based on 100 simulations of a year's worth of daily returns following a one-factor model. Ideally, the variance forecast ratio is 1.