The James-Stein Estimator for Eigenvectors

Alec Kercheval (akercheval@fsu.edu)

CDAR Seminar

UC Berkeley

October 4, 2022

Department of Mathematics
Florida State University

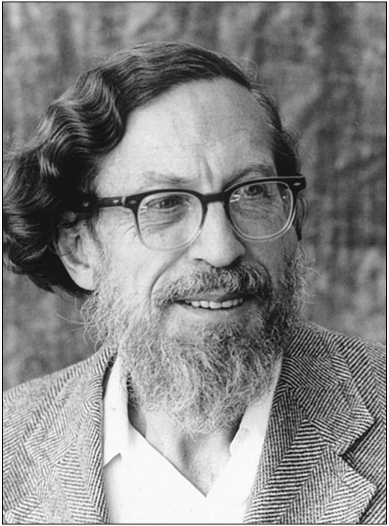
Collaborators

Lisa Goldberg, UC Berkeley CDAR, and Aperio by Blackrock

Hubeyb Gurdogan, UC Berkeley CDAR

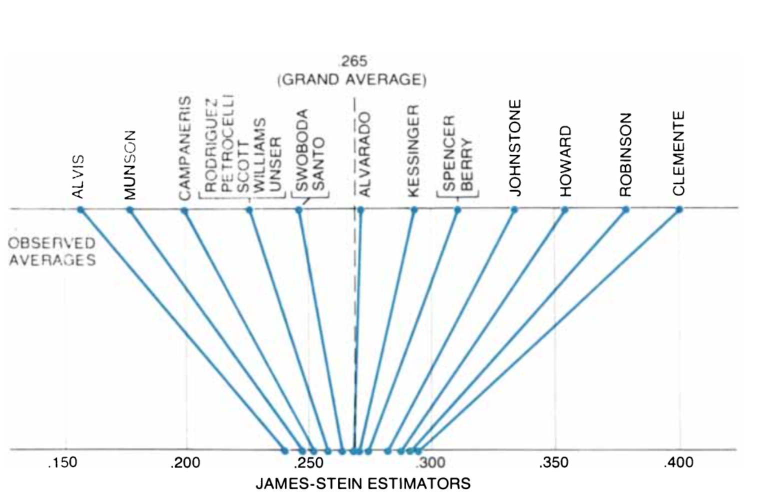
Alex Shkolnik, UC Santa Barbara

Charles Stein, 1920–2016



http://www.a3mreunion.org/remembrance/stein-c.html

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:

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

where

$$c = 1 - \frac{\nu^2}{\tau^2 + \nu^2}. \tag{3}$$

$$E[m(z)] = m \tag{4}$$

and, with some analysis, conditional on ν^2,

$$E\left[\frac{\nu^2}{s^2(z)}\right] = \frac{\nu^2}{\tau^2 + \nu^2}. \tag{5}$$

where

$$s^2(z) = \sum_{i=1}^{p} (z_i - m(z))^2/(p-3). \tag{6}$$

Define the JS (shrinkage) estimator

$$\hat{\mu}^{JS} = m(z)\mathbf{1} + c^{JS}(z - m(z)\mathbf{1}), \tag{7}$$

$$c^{JS} = 1 - \frac{\nu^2}{s^2(z)}. \tag{8}$$

James-Stein estimator

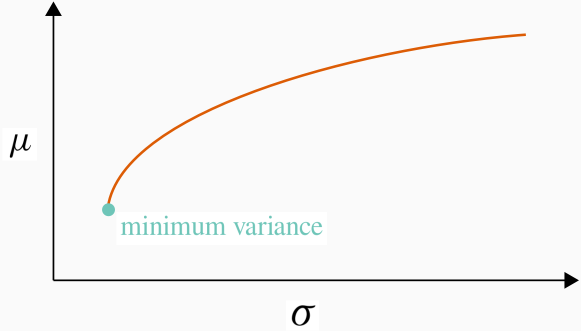
James and Stein (1961): The sample average z is inadmissible as an estimator of μ when $p > 3$.

$$E_{\mu,\nu}\left[|\hat{\mu}^{JS} - \mu|^2\right] < E_{\mu,\nu}\left[|z - \mu|^2\right]. \tag{9}$$

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

Portfolio selection in theory

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



A simple quadratic program expressed in terms of a covariance matrix Σ.

Portfolio selection in practice

The minimum variance portfolio w^* is the solution to:

$$\min_{w \in \mathbb{R}^p} w^\top \Sigma w$$
$$w^\top \mathbf{1}_p = 1$$

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

Theory versus practice

How do w^* and $\widehat{w}$ differ?

MV optimizers are "estimation error maximizers".

A variance minimizing optimizer will overweight securities for which variances or correlations are underforecast, hence under-forecast portfolio variance.

Statistician: errors arise from additive noise in covariance matrix

Math Finance: errors arise from prices used to form the sample covariance matrix

Agenda: Mitigate the effect of sampling error on optimization

Determine the mechanism by which sampling error enters an estimated covariance matrix and distorts optimized portfolios.

Mitigate distortion, to the extent possible.

Improve the accuracy of optimized portfolios.

How to estimate Σ: One-factor model

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

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

where $r, \beta, \epsilon \in R^p$ and $x \in R$. Assume $m(\beta) > 0$.

With common assumptions,

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

with $\sigma^2 = \mathrm{var}(x)$ and $\delta^2 = \mathrm{var}(\epsilon_i)$.

OR:

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

where $\eta^2 = \sigma^2 |\beta|^2$.

A principal component analysis (PCA) estimate of Σ

An estimate of Σ amounts to estimates of η^2, b and δ^2 :

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

Consider a time series of $n << p$ observations of r, with sample covariance matrix S.

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

Findings

Variance error in the optimized portfolio stems mostly from bias in the leading eigenvector of the sample covariance matrix.

Errors in the leading eigenvalue have no impact asymptotically in p.

The data-driven JSE (James-Stein for eigenvectors) estimator improves eigenvector estimation bias, leading to improvement in the optimized portfolio.

James Stein Estimators

$S =$ sample covariance matrix, $h =$ leading eigenvector, $Sh = \lambda^2 h$.

$$
\begin{aligned}
\mu^{JS} &= m(z)\mathbf{1} + c^{JS}(z - m(z)\mathbf{1}), \quad c^{JS} = 1 - \frac{\nu^2}{s^2(z)} \\
s^2(z) &= \frac{1}{p-3}\sum_{i=1}^{p}(z_i - m(z))^2 \\
h^{JSE} &= m(h)\mathbf{1} + c^{JSE}(h - m(h)\mathbf{1}), \quad c^{JSE} = 1 - \frac{\nu^2}{s^2(h)} \\
s^2(h) &= \frac{1}{p}\sum_{i=1}^{p}(\lambda h_i - \lambda m(h))^2 \\
\nu^2 &= \frac{tr(S) - \lambda^2}{p \cdot (n-1)} \to \delta^2/n
\end{aligned}
$$

Errors in portfolio weights

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

$$\mathcal{TE}^2 = (\widehat{w} - w^*)^\top \Sigma (\widehat{w} - w^*).$$

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

Ideally, tracking error is 0.

Errors in variance forecasts

Variance forecast ratio measures the error in the risk forecast as:

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

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

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

Asymptotic Results

Recall λ^2 is the leading eigenvalue of S, $p\nu^2 \to \delta^2/n =$ average of remaining non-zero eigenvalues of S. Define:

$$
\begin{aligned}
\Sigma_{\mathrm{raw}} &= (\lambda^2 - \frac{n-1}{p-1} p\nu^2) hh^\top + \frac{n-1}{p-1} p\nu^2 I \\
\Sigma_{\mathrm{PCA}} &= (\lambda^2 - p\nu^2) hh^\top + n\nu^2 I \\
\Sigma_{\mathrm{JSE}} &= (\lambda^2 - p\nu^2) \bar{h}^{\mathrm{JSE}} (\bar{h}^{\mathrm{JSE}})^\top + n\nu^2 I.
\end{aligned}
$$

Let $w_{\mathrm{raw}}, w_{\mathrm{PCA}}, w_{\mathrm{JSE}}$ be the estimated minimum variance portfolios, resp.

Asymptotic Results

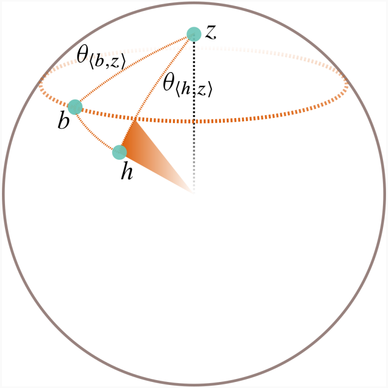
Theorem
Assume $0 < \lim_{p\to\infty} m(\beta) < \infty$.

Asymptotically as $p \to \infty$ *with* n *fixed, almost surely,*

1. $$\left|\bar{h}^{\mathrm{JSE}} - b\right|^2 < |h - b|^2$$
2. $\mathcal{V}(w_{\mathrm{PCA}}) \to 0$ *but* $\mathcal{V}(w_{\mathrm{JSE}}) > 0$
3. $\mathcal{T}\mathcal{E}^2(w_{\mathrm{PCA}}) > \mathcal{T}\mathcal{E}^2(w_{\mathrm{JSE}})$

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

Dispersion bias in the HL regime



source: Goldberg, Papanicalaou & Shkolnik (2022)

Here $z = \mathbf{1}/\sqrt{p}$. With high probability, $\langle z, b \rangle > \langle z, h \rangle$, meaning that the entries of h are more dispersed than the entries of b.

More general version

Let τ be any fixed unit p-vector (e.g. $\tau = z$). Define the generalized variance

$$v_\tau^2(y) = \frac{1}{p}|y - \langle y, \tau\rangle\tau|^2, \tag{10}$$

the generalized shrinkage constant

$$c_\tau^{\mathrm{JSE}} = 1 - \frac{\nu^2}{\lambda^2 v_\tau^2(h)}, \tag{11}$$

and the generalized JSE estimator as

$$h_\tau^{\mathrm{JSE}} = \langle h, \tau\rangle\tau + c_\tau^{\mathrm{JSE}}(h - \langle h, \tau\rangle\tau). \tag{12}$$

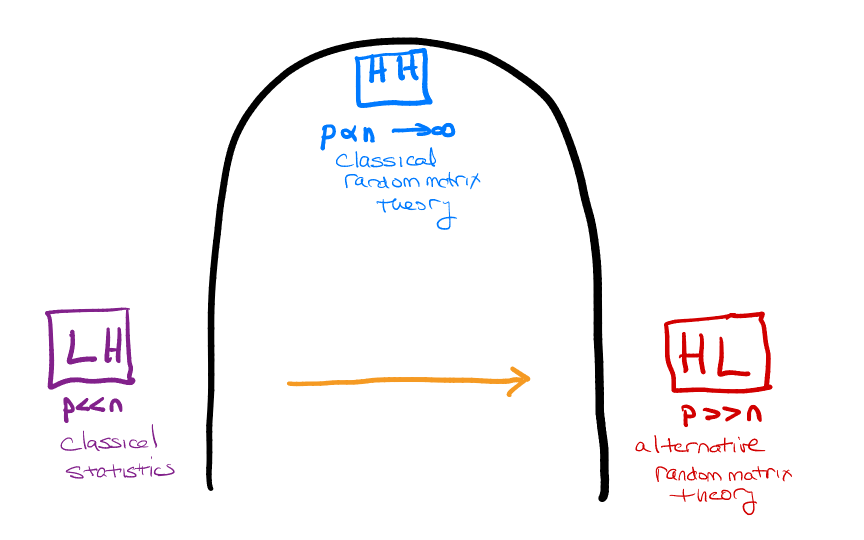
More general version

Theorem (Goldberg and K. 2022)
Asymptotically as $p \to \infty$ with n fixed, almost surely,

$$\angle(h_\tau^{\mathrm{JSE}}, b) \leq \angle(h, b) \tag{13}$$

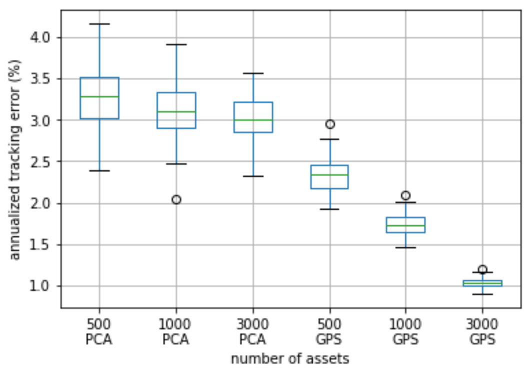
and the inequality is strict if $\liminf \langle b, \tau \rangle > 0$.

Three regimes



drawing: LG

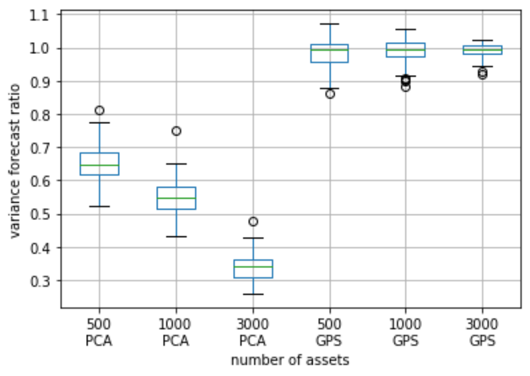
Simulated tracking error



source: Goldberg & Kercheval (2022)

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

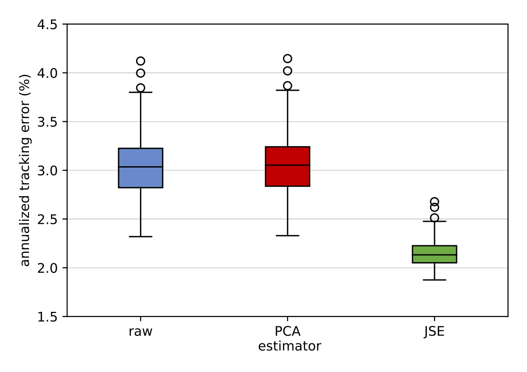
Simulated variance forecast ratio



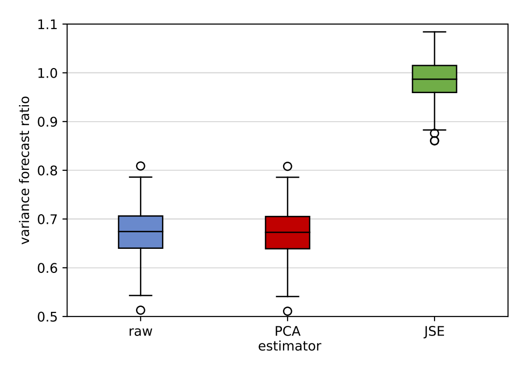
source: Goldberg & Kercheval (2022)

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

Simulated tracking error, $p = 500$



Simulated variance forecast ratio, $p = 500$



Results and problems

- MAPS generalization: arbitrary shrinkage targets; consistency
- In progress: multifactor models, multiple constraints
- In progress: Empirical validation.
- In progress: Admissibility: a finite p version of GPS that, like JS, holds in expectation.

Thank you



Lisa Goldberg, Hubeyb Gurdogan, Alex Shkolnik

References

Goldberg, L. R. & Kercheval, A. (2022), James Stein for eigenvectors. preprint.

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* (forthcoming).

Markowitz, H. (1952), ‘Portfolio selection’, *Journal of Finance* **7**(1), 77–91.

Shkolnik, A. (2021), ‘James-Stein estimation of the first principal component’, *Stat* .