Custom portfolio construction with James-Stein for eigenvectors

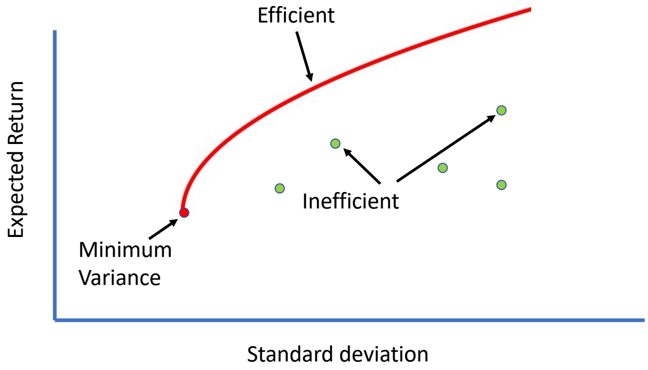
Lisa Goldberg(lrg@berkeley.edu)

EcoSta

August 2, 2023

Aperio Group, LLC, a wholly-owned, indirect subsidiary of BlackRock, Inc.
University of California, Berkeley

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



Portfolio optimization in practice

A Markowitz portfolio is a solution to a quadratic optimization with linear constraints. More generally, consider a $p \times k$ matrix of constraint gradients C and k-vector of constraint targets a:

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

Typical Markowitz portfolio:

$$C = \begin{pmatrix} 1 & \mu_1 \\ 1 & \mu_2 \\ \vdots & \vdots \\ 1 & \mu_p \end{pmatrix} \qquad a = \begin{pmatrix} 1 \\ \mu \end{pmatrix}$$

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

Estimating Σ when returns follow one-factor model

Structure reduces the dimension of the estimation problem

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

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

Then:

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

We need to estimate two positive scalars, η^2 and δ^2, and a unit p-vector b.

Only r is observed. $r \in \mathbb{R}^p$, $f \in \mathbb{R}$ and $\epsilon \in \mathbb{R}^p$ are random (but not necessarily Gaussian) and $\beta \in \mathbb{R}^p$ is an unknown parameter. $\sigma^2 = \mathrm{var}(f)$, $\delta^2 = \mathrm{var}(\epsilon)$, $\eta^2 = \sigma^2 |\beta|^2$ and $\mathrm{corr}(f, \epsilon) = 0$.

One-factor estimates of Σ

A sample covariance matrix synthesizes information from data but cannot be used in high-dimensional optimization

$$S = (\ \sigma_{ij}\ )$$

Volatility

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Correlation

Spectral decomposition, or principal component analysis (PCA), of the sample covariance matrix provides spare parts for constructing estimates of Σ

Leading eigenvalue

Leading eigenvector

$$S = (\ \sigma_{ij}\ ) = \lambda^2 h h^\top + \ \ldots$$

Volatility

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Correlation

The leading sample eigenvector and the difference between the leading sample eigenvalue and the average of its lesser, nonzero counterparts play crucial roles

Leading eigenvalue

Leading eigenvector

$$S = (\,\sigma_{ij}\,) = \lambda^2 hh^\top + \ \dots$$

Volatility

$$\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Correlation

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

Average non-zero lesser eigenvalue

A family of one-factor estimates of Σ

For a unit p vector , set:

$$\Sigma^{v} = (\lambda^2 - \ell^2) v v^\top + \frac{n}{p} \ell^2 I,$$

The estimate Σ^{v} varies with the unit vector v.

Since $\lim_{p\to\infty}(\lambda^2 - \ell^2)/p = \lim_{p\to\infty} \eta^2/p$ is finite and $\lim_{p\to\infty} n\ell^2/p = \delta^2$, we keep the estimates of η^2 and δ^2 fixed. For any v, $\mathrm{Trace}(\Sigma^v) = \mathrm{Trace}(S)$ is an unbiased estimate of $\mathrm{Trace}(\Sigma)$.

PCA and JSE estimates of Σ

PCA estimate of Σ

Setting v to the leading sample eigenvector $h = h^{\mathrm{PCA}}$ yields

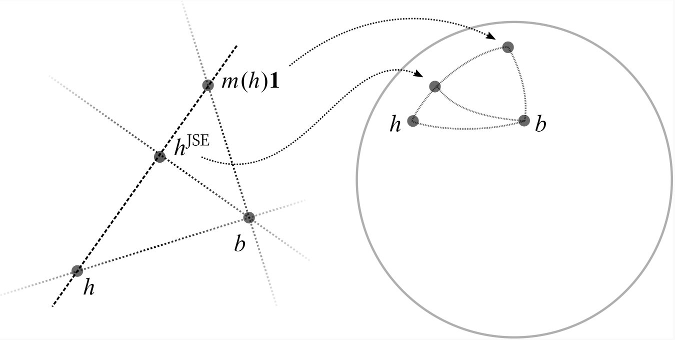
$$\Sigma^{\mathrm{PCA}} = (\lambda^2 - \ell^2) h^{\mathrm{PCA}} h^{\mathrm{PCA}\top} + \frac{n}{p} \ell^2 I$$

This is the S-POET estimate of Weichen Wang and Jianqing Fan.

Knowing they were poor, researchers and analysts used PCA estimates of eigenvectors in high dimensions because they didn't have a better idea.

> *The fact that we keep the sample eigenvectors does not mean that we assume they are close to the population eigenvectors. It only means that we do not know how to improve upon them.*
>
> *–Olivier Ledoit and Michael Wolf*

James Stein shrinkage transforms h^{PCA} to h^{JSE}, a better estimate of b



$$h_C = \underset{C}{proj}(h), c^{JSE} = \frac{\ell^2/\lambda^2}{1 - |h_C|^2}, h^{\mathrm{JSE}} \propto (1 - c^{JSE})h + c^{JSE} h_C$$

$$C = k - \text{dimensional subpace of} R^p, \angle(b, C) < \pi/2$$

source: Goldberg & Kercheval (2023), drawing by Alex Shkolnik

JSE estimate of Σ

Setting v to the leading sample eigenvector h^{JSE} yields

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

As we discuss next, JSE shrinkage leads to two types of stochastic dominance.

Asymptotic stochastic dominance of h^{JSE} over h^{PCA}

The magic formula

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,

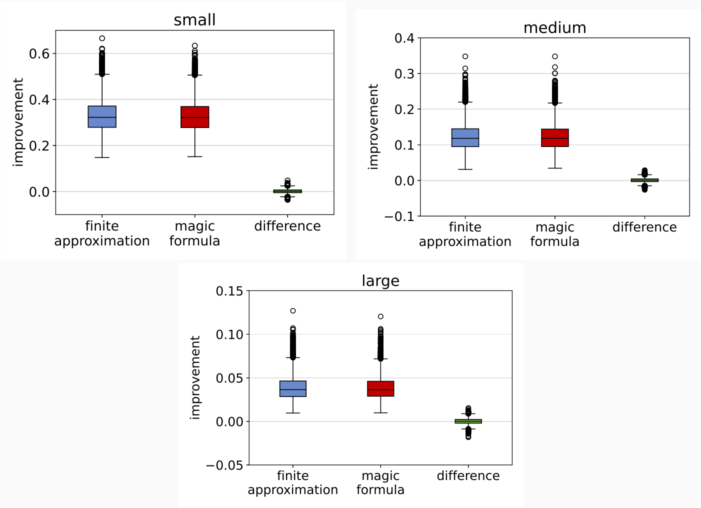
$$\lim_{p\to\infty} |h^{\text{JSE}} - b| < \lim_{p\to\infty} |h^{\text{PCA}} - b| \quad \text{almost surely}$$

Asymptotically, the improvement of JSE over PCA is:

$$\cos^2(\angle(h^{\text{JSE}}, b)) - \cos^2(\angle(h^{\text{PCA}}, b)) = \frac{1}{1+\phi_\infty^2}\left(\frac{\cos^2\Theta}{\phi_\infty^2 \sin^2\Theta + 1}\right) > 0.$$

ϕ_∞^2 is a relative eigengap equal to $\lim_{p\to\infty} \frac{\lambda^2 - \ell^2}{\ell^2}$.

How well does the magic formula work for finite p?



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

Asymptotic stochastic dominance of w^{JSE} over w^{PCA}

Construct estimates w^{PCA} and w^{JSE} of Markowitz portfolios by solving a quadratic program

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

$$w^\top C = a$$

with $\Sigma = \Sigma^{\mathrm{PCA}}$ and $\Sigma = \Sigma^{\mathrm{JSE}}$

$$C = \begin{pmatrix} 1 & \mu_1 \\ 1 & \mu_2 \\ \vdots & \vdots \\ 1 & \mu_p \end{pmatrix} \qquad a = \begin{pmatrix} 1 \\ \mu \end{pmatrix}$$

For large p, the true variance of a portfolio optimized with Σ^v is approximately equal to the optimization bias

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

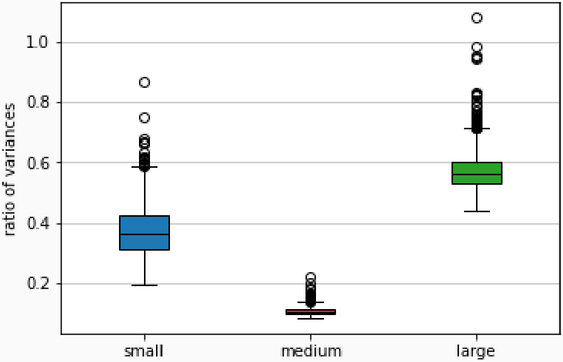
$$\mathcal{V}(w^v) = K\mathcal{E}_p^2(v, C, a) + O(1/p)$$

Set $v = h^{\mathrm{PCA}}$ and $v = h^{\mathrm{JSE}}$ for applications.

$\mathcal{E}_p(v, C, a) = \dfrac{\langle b, w_C^v \rangle (1 - ||v_C||^2) - \langle b, v - v_C \rangle \langle v, w_C^v \rangle}{||w_C^v|| (1 - ||v_C||^2)}, \mathcal{V}(w^v) = \eta^2 ||(C^\dagger)^\top a||^2 \mathcal{E}_p^2(v, C, a) + O(1/p)$

$C^\dagger$ is the Penrose inverse of C.

Variances of optimized and optimal portfolios



Box plots are generated from 10000 simulations of $n = 24$ monthly observations of returns to $p = 3000$ securities. See Appendix for calibration details. Graphics by Rahul Vinoth.

For any angle Θ, w^{JSE} stochastically dominates w^{PCA}

Almost surely,

$$\lim_{p \to \infty} \frac{\mathcal{V}(w^{\mathrm{JSE}})}{\mathcal{V}(w^{\mathrm{PCA}})} = 0$$

The result requires that $\lim_{p\to\infty} \angle \left(a, C^{\dagger} b\right)$ exists and is non-zero.

Collaborators

Jeongyoun Ahn

Haim Bar

Stjepan Begušić

Hubeyb Gurdogan

Sungkyu Jung

Alec Kercheval

Youhong Lee

Alex Papanicolaou

Alex Shkolnik

Simge Ulucam

Appendix: calibration for the numerical experiment

- factor and specific volatilities are 16% and 60%
- β is deterministic in our model, but it is generated with a normal, mean 1 distribution, The variance is set so that β has a prescribed angle with the unique, positive dispersionless vector on the sphere. We normalize so that $|\beta|^2/3000 = 1$
- μ is deterministic in our model, but is generated by adding noise from a normal distribution with mean 0.5 and variance 2 to β
- C is the span of the unique, positive dispersionless vector and μ
- for small, medium and large angles, $\cos\Theta = 0.99, 0.75, 0.49.$

References

Ahn, J., J. S. Marron, K. M. M. & Chi, Y.-Y. (2007), 'The high-dimension, low-sample-size geometric representation holds under mild conditions', *Biometrika* **94**(3), 760–766.

Bar, H. & Wells, M. T. (2023), 'On graphical models and convex geometry', forthcoming in *Computational Statistics and Data Analysis* .

Goldberg, L. R. & Kercheval, A. N. (2023), 'James Stein for the leading eigenvector', *Proceedings of the National Academy of Sciences* **120**.

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.

Jung, S. & Marron, J. S. (2009), ‘PCA consistency in high dimension, low sample size context’, *The Annals of Statistics* **37**(6B), 4104–4130.

Ledoit, O. & Wolf, M. (2017), ‘Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks’, *Review of Financial Studies* **30**(12), 4349–4388.

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

Michaud, R. (45), ‘The Markowitz optimization enigma: Is 'optimized' optimal?’, *Financial Analysts Journal* **1**(31–42).

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

Wang, W. & Fan, J. (2017), ‘Asymptotics of empirical eigenstructure for high dimensional spiked covariance’, *Annals of statistics* **45**, 1342–1374.