The JSE Backstory

Lisa Goldberg(lrg@berkeley.edu)

Seminar 217

January 31, 2025

Main points that may be useful to projects you are working on: part I

Portfolio construction with quadratic optimization, conceived and developed by Harry Markowitz, has been central to financial services for decades and shows no sign of going away. On the contrary, its applications seem to be growing.

The large universe of securities used to construct portfolios and the **non-stationarity of markets** mean that statistics 101 techniques cannot be used to create usable inputs to quadratic optimization.

Factor models provide the dimension reduction required to create the inputs. Fortunately for us, they give empirically sound descriptions of returns to public equities.

Main points that may be useful to projects you are working on: part II

Random matrix theory improves statistics 101 estimates of factor models.

It is crucial to choose improvement metrics that align with project goals.

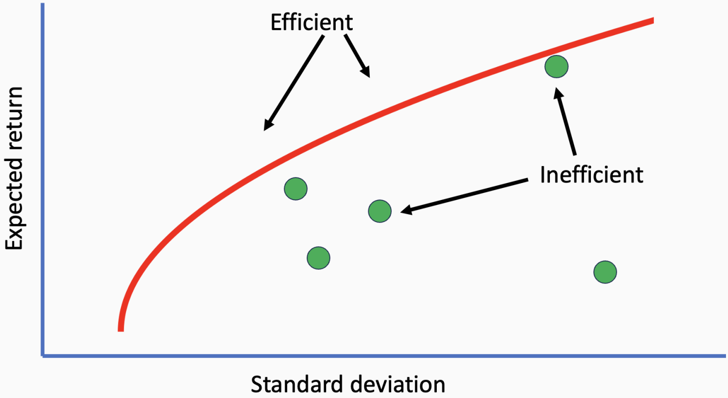
The interaction between factors and constraints is an essential consideration. It is far from enough to consider global minimum variance portfolios.

JSE, a newcomer to random matrix theory, shows promise at meaningful improvement to quadratic optimization on the basis of theory and simulation. JSE also may have applications to training neural networks.

We need empirical studies.

The efficient frontier

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



Markowitz did not originally draw the efficient frontier the way we do today

Historical note: Markowitz's initial focus was long-only portfolios, as long/short investing may impossible or at least very hard in the 1950s. Even today, long-only constraints are widely used by many investors. Seminar 217 work long-only constraints we are doing may be of practical use.

Portfolio optimization in practice typically includes large universes of securities and vast numbers of constraints, many of which are not convex

An efficient 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 efficient portfolio:

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

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

Estimating Σ for use in optimization: part I

Evidently realizing that classical statistics would not provide what he needed, Markowitz considered alternative ways to estimate optimization inputs

> *The calculation of efficient surfaces might possibly be of practical use. Perhaps there are ways, by combining statistical techniques and the judgment of experts, to form reasonable probability beliefs* (μ, Σ).

Factor models are standard tools for creating estimates of Σ when the number of returns p is large relative to the number of observations n.

An estimated covariance matrix is used to construct an optimized portfolio, which can be far from optimal

In 1989 Richard Michaud famously wrote that MV optimizers are in a fundamental sense, "estimation error maximizers."

Optimization selectively amplifies errors in an estimated covariance matrix.

The problem can be severe when the number of securities p exceeds the number of observations n, which is frequently the case.

Factor models reduce estimation error by reducing dimension

A large number of variables are, in many situations, driven by a relatively small number of factors.

1904: Charles Spearman's two-factor theory of intelligence
1963: Bill Sharpe's one-factor market model
1974: Barr Rosenberg's Barra models
1976: Stephen Ross's arbitrage pricing theory

An overview of three factor model architectures commonly used in finance is in Connor (1995). He calls them fundamental (favored by practitioners), macroeconomic (used by finance academics) and statistical (typically chosen by science and statistics academics). There are pros and cons for all three. We'll focus on the third.

Today, we'll assume returns to p securities follow a one-factor model

The structural model is given by:

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

Then:

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

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 $\mathrm{corr}(f, \epsilon) = 0$.

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

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

Volatility

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

Correlation

The Bianchi Experiment

The Bianchi experiment

In simulation:

Replace the estimated eigenvalue with the truth, leaving the corrupted eigenvector in place

Switch the roles of eigenvalue and eigenvector in the previous step

Data are generated from a one factor model, $r = \beta f + \epsilon$. Factor returns f are drawn independently from $N(0, 0.16^2)$ and specific returns are drawn independently from $N(0, 0.50)$. While values of β are parameters to be estimated, they are drawn from $N(1, 0.5)$. Boxplot generated with 200 simulated paths. Analysis and graphics by Rahul Vinoth.

We borrow metrics from financial practitioners to quantify errors in optimized portfolios and their risk forecasts

Tracking error of an optimized portfolio $\mathbf{w}^*$ measures its distance from the optimal portfolio $\mathbf{w}^*$:

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

Variance forecast ratio is estimated variance as a fraction of the truth:

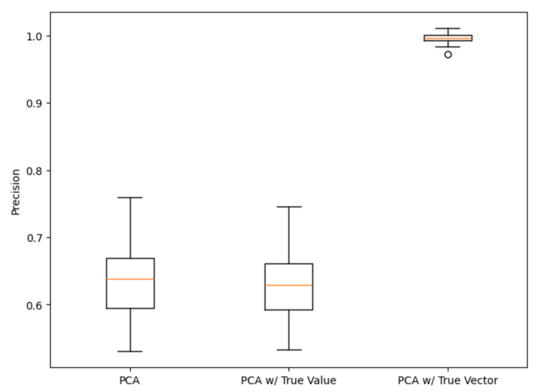
$$\mathcal{V} = \frac{\mathbf{w}^\top \Sigma^{est} \mathbf{w}}{\mathbf{w}^\top \Sigma \mathbf{w}}$$

Ideally, tracking error is 0 and the variance forecast ratio is 1.

These metrics are great for simulation where we know ground truth. What metrics should we use in an empirical study, where we don't have access to the covariance matrix or the optimal portfolio? There is disagreement among experts.

It seems that errors in eigenvectors rather than errors in eigenvalues that lower variance forecast ratio

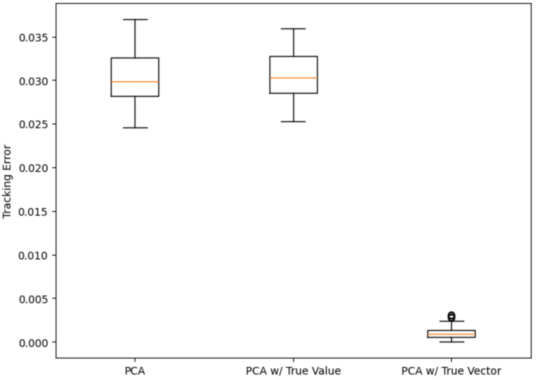
$$\mathcal{V} = \frac{\mathbf{w}^\top \Sigma^{est} \mathbf{w}}{\mathbf{w}^\top \Sigma \mathbf{w}}$$



As we improve variance forecast ratio by correcting eigenvectors, we also move the optimized portfolio closer to the optimum

Tracking error, the workhorse of financial services, is the distance between two portfolios

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



Data are generated from a one factor model, $r = \beta f + \epsilon$. Factor returns f are drawn independently from $N(0, 0.16^2)$ and specific returns are drawn independently from $N(0, 0.50^2)$. While values of β are parameters to be estimated, they are drawn from $N(1, 0.5)$. Analysis and graphics by Rahul Vinoth.

Estimating Σ for use in optimization: part II

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

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

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

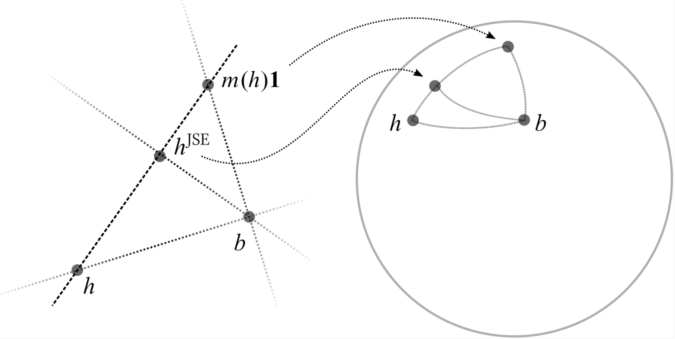
$$\Sigma^{\mathrm{PCA}} = (\lambda^2 - \ell^2) h^{\mathrm{PCA}} h^{\mathrm{PCA}\top} + \frac{n}{p} \ell^2 I$$

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



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

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

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,

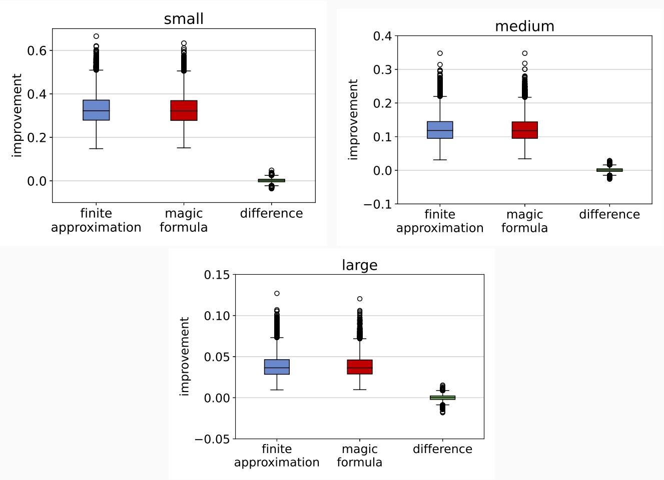
$$\lim_{p\to\infty} |h^{\mathrm{JSE}} - b| < \lim_{p\to\infty} |h^{\mathrm{PCA}} - b| \quad \text{almost surely}$$

Asymptotically, the improvement of JSE over PCA is:

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

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 efficient 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.

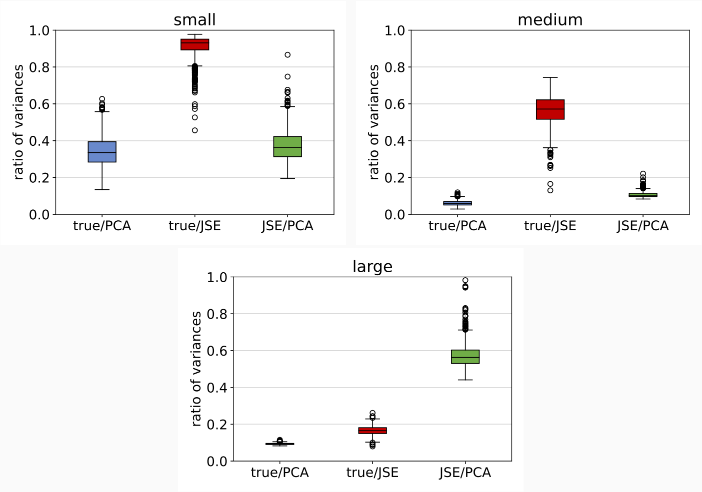
$$\mathcal{V}(w^v) = K\mathcal{E}_p^2(v, C, a) + O(1/p)$$

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

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

$C^\dagger$ 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 Stephanie Ribet.

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

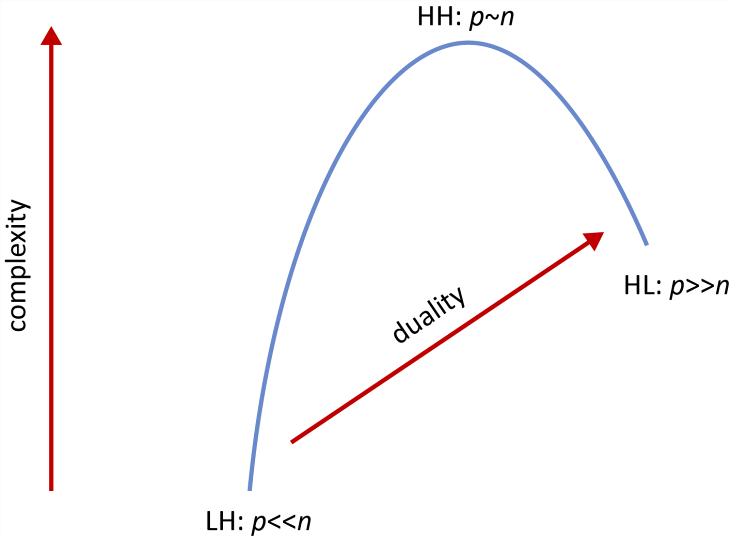
Almost surely,

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

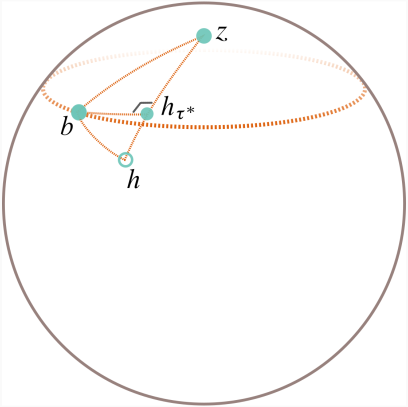
The result requires that $\lim_{p\to\infty} \angle\left(a, C^{\dagger} b\right)$ exists and is non-zero.

Why this works

Our results benefit from the blessing of dimensionality, present in the HL regime of new random matrix theory



In the HL regime, concentration of measure arguments imply that the sample leading eigenvector h is overwhelming likely to be further away from an anchor point z than the population leading eigenvector b



source: Goldberg, Papanicalaou & Shkolnik (2022)

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

Connor, G. (1995), 'The three types of factor models: A comparison of their explanatory power', *Financial Analysts Journal* **51**(3), 42–46.

Goldberg, L. R., Gurdogan, H. & Kercheval, A. (2024), 'Portfolio optimization via strategy-specific eigenvector shrinkage', forthcoming in *Finance and Stochastics* .

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.

Gurdogan, H. & Shkolnik, A. (2024), The quadratic optimization bias of large covariance matrices.

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.

Rosenberg, B. & McKibben, W. (1973), ‘The prediction of systematic and specific risk in common stocks’, *Journal of Financial and Quantitative Analysis* **8**(2), 317–333.

Ross, S. (1976), ‘The arbitrage theory of capital asset pricing’, *Journal of Economic Thoery* **13**, 341–360.

Sharpe, W. F. (1963), 'A simplified model for portfolio analysis', *Management Science* **9**(2), 277–293.

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

Shkolnik, A., Kercheval, A., Gurdogan, H., Goldberg, L. R. & Bar, H. (2024), 'Portfolio selection revisited', *Annals of Operations Research* . **URL:** *https://doi.org/10.1007/s10479-024-06340-7*

Spearman, C. (1904), 'General intelligence, objectively determined and measured', *American Journal of Psychology* **15**(201–293).

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