James Stein for Eigenvectors

Lisa R. Goldberg and Alec N. Kercheval

March 23, 2022

Abstract

Recent research identifies and corrects the dispersion bias in the leading eigenvector of a covariance matrix estimated from a high dimension low sample size (HL) data set. This is the tendency toward greater dispersion in the entries of an estimated eigenvector relative to its population counterpart. The correction is via the data-driven GPS estimator, which is the structural analog of the James-Stein estimator for a collection of averages. A generalization called MAPS can further correct for other biases and even provide a consistent estimator for the leading eigenvector in a factor model setting. In this article, we expose GPS and MAPS to highlight their connection to James-Stein. We review applications of GPS and MAPS to quantitative portfolio construction, emphasizing potential extensions and open questions.

Significance Statement

Eigenvectors are used throughout the physical and social sciences to reduce the dimension of complex problems to manageable levels, and to distinguish signal from noise. Our research identifies and corrects substantial biases in the leading eigenvector of a covariance matrix estimated in the high dimension low sample size (HL) regime. Our analysis sheds light on aspects of how estimation error corrupts an estimated covariance matrix and is transmitted via quadratic optimization. Applications to quantitative portfolio construction are established, while the benefits of our bias correction to genome-wide association studies (GWAS) and machine learning algorithms await exploration.

Introduction

Averaging is the most important tool for distilling information from data. For example, season average is a standard measure of the likelihood that a baseball player will get on base. An average of squared security returns is commonly used to estimate the variance of a portfolio of stocks.

In statistical terms, the average can be the best estimator of the mean in the sense of having the smallest mean squared error. But a strange thing happens when considering a collection of many averages simultaneously. The aggregate mean squared error is no longer minimized by the collection of averages. Instead, the error can be improved by shrinking the averages toward a target, even if, paradoxically, there is no underlying relation among the quantities.

For baseball players, since an individual batting average incorporates both the true mean and estimation error from sampling, the largest observed batting average is likely to be over-estimated and the smallest under-estimated. This line of reasoning shows that the aggregate mean square error is reduced when the collection of observed averages are all moved toward their center.

Charles Stein surprised the community of statisticians with a sequence of papers about this phenomenon beginning in the 1950s. Stein showed that a collection of three or more averages is inadmissible, and that it is possible to lower their aggregate squared error by a formula shrinking the collection of values toward a common center. In 1961, Stein improved and simplified the analysis in collaboration with Willard James. The resulting empirical James-Stein estimator (JS) launched a new era of statistics.

This article describes James-Stein for eigenvectors.

It was originally developed to locate an unobserved position on a sphere, inspiring the name "GPS" for "Global Positioning System", and used to improve accuracy of minimum variance portfolios. GPS is an empirical shrinkage operator, and it turns out to be structurally parallel to JS.

A sample leading eigenvector is a direction in a high dimensional data set that maximizes explained variance. Originally developed to study axes of rotation of rigid bodies, eigenvectors are used today to identify points of centrality on the world wide web, as financial risk factors, and as control variables in genome-wide association studies, to name just a few examples. Like a collection of averages, a sample eigenvector is a collection of values that may be improved by shrinkage.

The GPS estimator corrects excess dispersion in the entries of an eigenvector estimated from a high-dimensional data set, where the number of variables vastly exceeds the number of observations. These noisy regimes fall outside the realm of classical statistics, and they arise in machine learning, genetics, and finance, where a relatively small number of observations is used to explain or predict complex phenomena.

Below, we review the connection between James-Stein for averages and for eigenvectors. The latter sheds light on aspects of how estimation error corrupts an estimated covariance matrix and is transmitted to portfolios via quadratic optimization. Along the way we provide ideas for extensions and applications.

What is the James-Stein estimator?

Suppose there are $p > 3$ unknown means $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ to be estimated. We observe a fixed number of samples, and compute the corresponding sample averages $z = (z_1, z_2, \ldots, z_p)$.

It is common practice to use z_i as an estimate for the unobserved mean value μ_i, and this may be the best one can do if estimating only a single mean. With certain normality assumptions, the discovery of Stein (1956) and James & Stein (1961), elaborated by Efron & Morris (1975), Efron & Morris (1977), Efron (2010), is that a better estimate is obtained by shrinking the sample averages toward their collective average in a specific way.

Let $m(z) = \sum_{i=1}^{p} z_i/p$ denote the collective average, and $\mathbf{1} = (1, 1, \ldots, 1)$. The winning recipe, which defines the JS estimator, is

$$\hat{\mu}^{JS} = m(z)\mathbf{1} + c^{JS}(z - m(z).\mathbf{1}) \qquad (1)$$

The shrinkage factor c^{JS} is given by

$$c^{JS} = 1 - \frac{\nu^2}{s^2(z)}, \qquad (2)$$

where

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

is a measure of the variation of the sample averages z_i around their collective average $m(z)$, and ν^2 is an estimate of the conditional variance of each sample average around its unknown mean. It measures the noise affecting each observation. The value of ν^2 must be estimated independently of $s^2(z)$ or assumed, and it is sometimes taken to be 1 without comment.

The observable quantity $s^2(z)$ incorporates both the unobserved variation of the means and the noise ν^2. The term $\nu^2/s^2(z)$ in equation (2) can be thought of as an estimated ratio of noise to the sum of signal and noise. Equation (1) calls for a lot of shrinkage when the noise dominates the variation of the sample averages around their collective average, and only a little shrinkage when the reverse is true.

The JS estimator is better in the sense of expected mean squared error,

$$E_{\mu,\nu}\left[|\hat{\mu}^{JS} - \mu|^2\right] < E_{\mu,\nu}\left[|z - \mu|^2\right]. \qquad (4)$$

For any fixed μ and ν, the conditional expected mean squared error is improved when using $\hat{\mu}^{JS}$ instead of z. This result comes with an unavoidable caveat: z remains the optimal estimate when $p = 1$ and $p = 2$, and sometimes when $p = 3$.

Suppose we have $p > 3$ baseball players, and, for $i = 1, \ldots, p$, player i has true batting average μ_i, meaning that in any at-bat the player has a probability μ_i of getting a hit. This probability is not observable, but we do observe, say over the first 50 at-bats of the season, the realized proportion z_i of hits. Assuming we know ν^2 or have an independent way to estimate it, equation (1) improves on the z_i as estimates of the true means μ_i.

This example lends intuition to the role of the noise to signal-plus-noise ratio $\nu^2/s^2(z)$ in the JS shrinkage factor. If the true batting averages differ widely, but the sample averages tend to be close to the true values, then equation (1) calls for little shrinkage, as appropriate. Alternatively, if the true averages are close together but the sampling error is large, a lot of shrinkage makes sense. The JS estimator properly quantifies the shrinkage and interpolates between these extremes.

What is the GPS estimator?

The GPS estimator is an empirical approximation of a leading eigenvector of an unobserved covariance matrix in a high-dimension low-sample-size (HL) setting, where the number of variables vastly exceeds the number of observations. It improves on the sample leading eigenvector by having lower squared error with high probability, and leading to better estimates of covariance matrices for use in quadratic optimization.

GPS is, like JS, a shrinkage estimator, and shares many of its characteristics. Suppose we have n independent observations of $p >> n$ variables whose unobserved covariance matrix Σ has leading normalized eigenvector b. We suppose the entries of b have a non-zero average, $m(b) = \sum_{i=1}^{p} b_i/p$, which we are free to assume is positive by change of sign if needed.

Denote by S the $p \times p$ sample covariance matrix constructed from our n observations, with leading eigenvalue λ^2 and corresponding eigenvector h, which we may assume has unit length and positive average entry $m(h) = \sum_{i=1}^{p} h_i/p$ ¿0.

The GPS estimator h^{GPS} is obtained by shrinking the entries of h toward their average,

$$h^{GPS} = m(h)\mathbf{1} + c^{GPS}(h - m(h)\mathbf{1}). \qquad (5)$$

The shrinkage constant c^{GPS} is given by

$$c^{GPS} = 1 - \frac{\nu^2}{s^2(h)}, \qquad (6)$$

where

$$s^2(h) = \frac{1}{p}\sum_{i=1}^{p}\left(\lambda h_i - \lambda m(h)\right)^2 \qquad (7)$$

is a measure of the variation of the entries of λh around their average $\lambda m(h)$, and ν^2 is equal to the average of the non-zero smaller eigenvalues of S, scaled by $1/p$,

$$\nu^2 = \frac{tr(S) - \lambda^2}{p \cdot (n-1)}. \qquad (8)$$

GPS calls for a lot of shrinkage when the average of the non-zero smaller eigenvalues dominates the variation of the entries of λh around their average and only a little shrinkage when the reverse is true.

Using ideas developed Goldberg et al. (2022) and Goldberg et al. (2020), Shkolnik (2021) proves, with high probability, the angle between h^{GPS} and b is smaller than the angle between h and b:

$$\angle\left(h^{GPS}, b\right) < \angle(h, b), \qquad (9)$$

and mathematically justifies the statement that GPS is James Stein for eigenvectors.

We illustrate (9) in Figure 1. The left panel shows GPS shrinkage as defined by equation (5). The right panel shows an equivalent formulation of GPS shrinkage in terms of angles between the corresponding vectors on the unit sphere obtained by normalization.

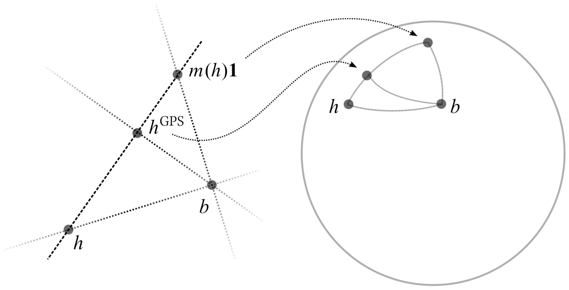


Figure 1: Shrinkage of the sample eigenvector h along the line connecting h and $m(h)\mathbf{1}$ in Euclidean space (left panel) and projected on the unit sphere (right panel, which illustrates the central angles $\angle(h, b)$ and $\angle(h^{GPS}, b)$.

As with JS, an example illustrating how to apply GPS is helpful, and we do that in the next section.

GPS corrects an optimization bias

The GPS estimator originated as an improvement on standard implementations of mean-variance optimization of a portfolio of public equities. In this section we illustrate how GPS mitigates an optimization bias, the impact of estimation error in a covariance matrix on the results of quadratic optimization.

Quantitative portfolio construction

From a universe of p securities, there are countless ways to construct a portfolio. We focus on quantitative portfolio construction, which has relied on mean-variance optimization since Markowitz (1952). In this framework, a portfolio is represented by a vector whose ith entry is the fraction or *weight* of investment in security i. A non-singular estimate of the covariance matrix is required by the program. A portfolio is *efficient* if it has minimum forecast variance subject to constraints and the simplest efficient portfolio is minimum variance.

A minimum variance portfolio

A fully invested but otherwise unconstrained minimum variance portfolio is the solution to the "mean-variance optimization" problem

$$\begin{aligned} &\min\nolimits_{w\in\mathbb{R}^p} w^\top \widehat{\Sigma} w \\ &\text{subject to:} \\ &\; w^\top \mathbf{1} = 1, \end{aligned} \tag{10}$$

where the $p \times p$ matrix $\widehat{\Sigma}$ is a non-singular estimate of the true security covariance matrix Σ. While it is precisely specified, the solution $\hat{w}^*$ to (10) is not optimal: the true optimum w^* is the solution to (10) with $\widehat{\Sigma}$ replaced by Σ.

The impact of estimation error on optimization

The estimation error in the matrix $\widehat{\Sigma}$ is transmitted to the resulting portfolios. From Michaud (1989) and other sources, we know that mean-variance optimizers are "estimation error maximizers". Here, we review two metrics for the effect of the optimization bias, the impact of covariance matrix estimation error on weights and risk forecasts of optimized portfolios.

Since a variance-minimizing optimization tends to place excess weight on securities whose variances and correlations with other securities are under-forecast, variance forecasts for optimized portfolios are biased downward. We measure variance bias with the *variance forecast ratio,* defined for our minimum variance portfolio by

$$\mathrm{VFR}(\hat{w}^*) = \frac{\hat{w}^* \widehat{\Sigma} \hat{w}^*}{\hat{w^*}^\top \Sigma \hat{w}^*}. \tag{11}$$

A variance forecast ratio less than 1 indicates an underforecast while a variance forecast ratio greater than 1 indicates is overforecast.

Another measure of the distance between an optimized and optimal portfolio is *tracking error*, which we define as

$$\mathrm{TE}^2(\hat{w}^*) = (\hat{w}^* - w^*)^\top \Sigma (\hat{w}^* - w^*) \tag{12}$$

for the minimum variance portfolio. Tracking error is used throughout mathematical finance to measure the width of the distribution of the difference in return of two portfolios, and it is commonly applied to measure the distance between a portfolio and its benchmark. All else equal, a smaller tracking error is better.

Since they require knowledge of the true covariance matrix Σ, neither tracking error nor variance forecast ratio can be used in an empirical study. In simulation, they illuminate the transmission of error from $\widehat{\Sigma}$ to $\hat{w}^*$.

Factor models and optimization

In the HL regime where $p >> n$, the sample covariance matrix S is singular, and is therefore not a candidate for $\widehat{\Sigma}$. Factor models of security returns have emerged as a standard tool to generate full-rank estimated covariance matrices. The prototype is a one-factor model of returns:

$$r = \beta f + \epsilon, \tag{13}$$

where r is a p-vector of security returns, β is a p-vector of factor loadings, f is a random variable serving as a common factor through which returns are correlated, and ϵ is a p-vector of specific returns that are uncorrelated with f and each other.

In this situation, the true covariance matrix of r takes the form

$$\Sigma = \sigma^2 \beta\beta^\top + \delta^2 I, \tag{14}$$

where f has variance σ^2 and each entry of ϵ has variance δ^2. Estimating Σ reduces to finding estimates of σ^2, β, and δ^2, so that

$$\widehat{\Sigma} = \hat{\sigma}^2 \hat{\beta}\hat{\beta}^\top + \hat{\delta}^2 I. \tag{15}$$

A standard implementation of (15) relies on principal component analysis (PCA), where the vector of factor loadings $\hat{\beta}$ is taken to be a scalar multiple of the leading eigenvector h of the sample covariance matrix. Goldberg et al. (2022) show that with high probability the entries of h are overly dispersed. Further, in the HL regime, setting β to a multiple of h^{GPS} instead of h can generate optimized portfolios that are closer to optimal, and diminishes the downward bias in forecasts of variance for these portfolios. We provide an example below.

Numerical illustration

Consider a hypothetical market driven by the one-factor model (13). Our calibration is taken approximately from Goldberg et al. (2022) and Goldberg et al. (2020), which explain how to tune simulations to empirical data. We assume the factor model is latent, meaning the components β, f and ϵ are not observed. We draw factor and specific returns f and ϵ independently from mean 0 normal distributions with standard deviations 16% and 60%, respectively. The entries of β, or factor loadings, are loosely inspired by market betas. Even though they are not random quantities, we draw entries of β independently from a normal distribution with mean 1 and variance 0.25. We set the number of securities p to 500, 1000, and 3000. For each p, we simulate $n = 252$ observations 100 times, so each boxplot in Figures 2 and 3 is based on 100 outcomes.

Figure 2 shows errors in the approximation of the leading eigenvector b by the sample leading eigenvector h on the left and its GPS correction h^{GPS} on the right. As the number p of securities increases, we observe median errors diminish along with the widths of their distributions. The GPS correction provides modest but discernible improvement over PCA by lowering the excess dispersion.

While the reduction in total angular error $\angle(h, b)$ is modest, the GPS correction has a profound impact on the optimized portfolio. In Figure 3, we show tracking error (panel a) and variance forecast ratio (panel b) for portfolios optimized with a one-factor PCA model and its GPS correction. Theory predicts that as p increases to infinity, the variance forecast ratio for a PCA model tends to 0 while tracking error is bounded below. With the GPS correction, tracking error tends to 0 and the variance forecast ratio is bounded below. Numerical evidence supports the assertion that variance forecast ratio tends to 1 as p increases to infinity.

The asymptotic theory illustrated in Figures 2

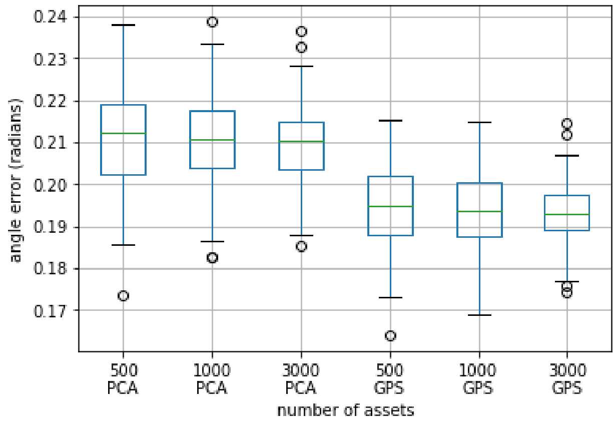
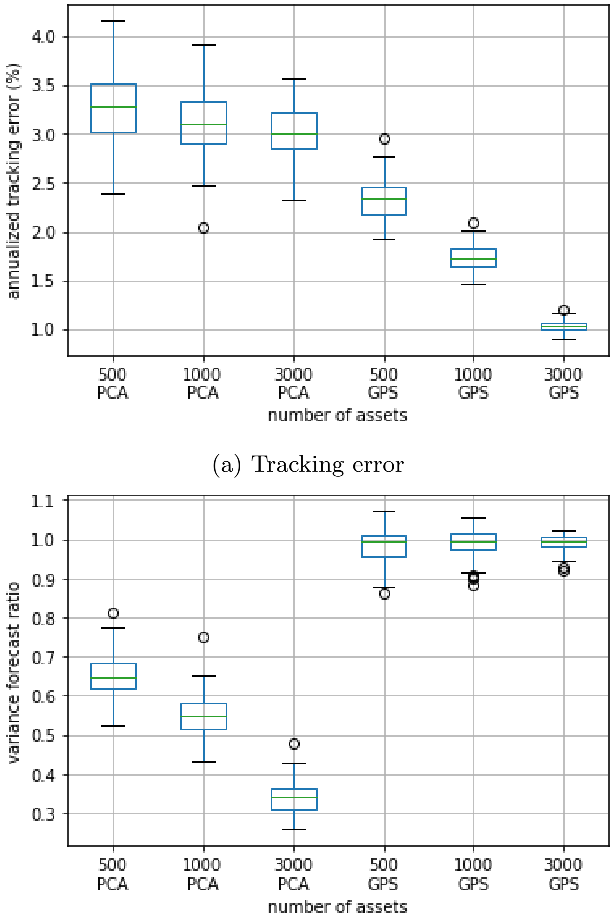


Figure 2: Angle between the leading eigenvector b of the true covariance matrix and its estimators, h and h^{GPS} in simulated markets. The estimated covariance matrix is based on $n = 252$ observations of $p = 500$ securities. Each boxplot is generated by 100 simulations. The GPS correction materially diminishes the angle between estimated and true eigenvectors.

and 3 does not depend on the normal distribution, even though our example is based on normal returns. More empirically realistic simulations in which specific returns are generated by heavy-tailed distributions will still have good asymptotic properties, so long as variance is finite. For fixed numbers of securities p and observations n, however, such simulations generate more outliers than the normal simulations. It would be valuable to frame this issue quantitatively, and to develop improvements to GPS when data are heavy-tailed.

What is a MAPS estimator?

While GPS can correct excess dispersion of the leading sample eigenvector in the HL regime, further improvements are possible using a generalization called a MAPS (Multiple Anchor Point Shrinkage) estimator developed in Gurdogan & Kercheval (2021). In particular, when certain order information about the betas is available, MAPS can provide a consistent es-



(a) Tracking error

(b) Variance forecast ratio

Figure 3: Portfolio-level accuracy metrics for simulated minimum variance portfolios optimized with a PCA model and a GPS correction. The estimated covariance matrix is based on $n = 252$ observations of $p = 500$ securities. Each boxplot is generated by 100 simulations. Tracking error is materially diminished by the GPS correction, indicating greater accuracy of portfolio weights. Variance forecast ratio is increased toward 1 by the GPS correction, indicating greater accuracy of variance forecasts.

timator of the true leading normalized eigenvector b of Σ, that is, an estimate of b with asymptotically zero angular error for fixed n as $p \to \infty$.

MAPS shrinkage

The GPS estimator is created by shrinking the sample covariance leading eigenvector h toward the constant vector $q_1 = m(h)\mathbf{1}$. If we call q_1 an "anchor point" for the estimation, we can ask whether there are additional anchor points $\{q_2, \ldots, q_k\}$ containing useful information. Denote by $L = \langle q_1, \ldots, q_k \rangle$ the linear subspace of R^p spanned by $\{q_1, \ldots, q_k\}$.

With the subspace L in hand, the "MAPS shrinkage target" $P_L(h)$ is the orthogonal projection of h onto L, and the corresponding MAPS estimator shrinks h toward $P_L(h)$:

$$h^{MAPS} = P_L(h) + c^{MAPS}(h - P_L(h)) \qquad (16)$$

where the shrinkage factor is

$$c^{MAPS} = 1 - \frac{\nu^2}{s^2(h)} \qquad (17)$$

with

$$s^2(h) = (\lambda^2/p)(1 - ||P_L(h)||^2) \qquad (18)$$

and

$$\nu^2 = \frac{tr(S) - \lambda^2}{p(n-1)}. \qquad (19)$$

The GPS estimator corresponds to the special case $L = \langle m(h)\mathbf{1} \rangle$. In that case direct calculation verifies that $P_L(h) = m(h)\mathbf{1}$ and formula (18) reduces to the corresponding GPS formula (7).

MAPS as a consistent estimator

For example, suppose we know the rank ordering of the betas $\beta_1, \ldots, \beta_p$, but not their actual values. Let $[x]$ denote the greatest integer less than or equal to x. Order the betas by size and divide them into $k = [\sqrt{p}]$ groups, with the largest $[\sqrt{p}]$ betas in the first group, the next largest in the second group, etc., and any extras added to the last group.

For $i = 1, \ldots, k$, the anchor point q_i is defined as the vector $(a_1, \ldots, a_p)$ where $a_j = 1$ if β_j belongs to group i, and zero otherwise. The subspace $L = \langle q_1, \ldots, q_k \rangle$ and formula (16) define a consistent MAPS estimator in the sense that

$$\lim_{p \to \infty} ||h^{MAPS} - b|| = 0 \qquad (20)$$

almost surely.

It is not necessary that the full rank ordering be known, only that the groups are "ordered" in the sense that no element of any group lies between the minimum and maximum elements of another group.

To illustrate how this could work, we continue in the setting of a public equity market, where analysts sort securities into sectors, such as energy, information technology, financials, and utilities. Empirically, securities in the same sector tend to have similar loadings b_i on the common factor f. For example, utility stocks have had relatively low loadings on the common factor, while energy stock loadings have been relatively high.

To the extent that the sectors organize the betas into ordered groups, the MAPS estimator as defined above will be a consistent estimator of b. In practice, sector groupings of betas are not perfectly ordered, so this will be only approximately true.

From oracles to data-driven estimators

The JS, GPS and MAPS estimators arise as data-driven versions of ideal "oracle" estimators that are not themselves observable. For JS, the oracle is a Bayes estimator, described next. For GPS, the oracle is the point h^O along the line in Euclidean space through h and $m(h)\mathbf{1}$ that is closest to the true eigenvector b. The MAPS oracle is an analogous point also defined in terms of the unknown b. Here we discuss the relationship between the empirical and oracle quantities in more detail.

James-Stein and Bayes

The original formulation of the James-Stein estimator, as well as many modern renditions, relies on the normal distribution, which allows us to see JS as a partially empirical version of a Bayesian estimator.

Suppose that the pairs (μ_i, z_i) and (μ_j, z_j) are independent for $i \neq j$, and satisfy

$$\mu \sim \mathcal{N}(m\mathbf{1}, \tau^2 I) \quad \text{and} \quad z|\mu \sim \mathcal{N}(\mu, \nu^2 I), \tag{21}$$

where $\mathcal{N}$ indicates the normal distribution. In this setting, the unobserved means μ_i are centered around m with variance τ^2, and ν^2 quantifies the noise independently affecting each observed z_i.

Computations using Bayes Rule tell us that the best estimate of the true mean μ, conditional on observing z, is obtained by shrinking z toward $m\mathbf{1}$,

$$\mu^{Bayes} = m\mathbf{1} + c(z - m\mathbf{1}), \tag{22}$$

where

$$c = 1 - \frac{\nu^2}{\tau^2 + \nu^2}. \tag{23}$$

These Bayesian formulas look a lot like the James-Stein shrinkage formulas (1) and (2), but rely on the unobserved parameters m, τ^2 and ν^2. Conditional on ν^2,

$$E[m(z)] = m \tag{24}$$

and, with some analysis,

$$E\left[\frac{\nu^2}{s^2(z)}\right] = \frac{\nu^2}{\tau^2 + \nu^2}. \tag{25}$$

Hence the JS formulas can be viewed as empirical versions of (22) and (23), where m and $\nu^2/((\tau^2 + \nu^2)$ are replaced by empirical unbiased estimators $m(z)$ and $\nu^2/s^2(z)$

The JS framework does not include an estimate of ν^2, which justifies the description of James-Stein as "partially empirical Bayes."

Formula (25) explains why we describe $\nu^2/s^2(z)$ as a noise to signal-plus-noise ratio. The signal in question is the true variance τ^2 of the means μ_i around their collective mean m, and it is obscured by the noise ν^2 contaminating the observations z_i.

Eigenvalues, stability, and factor models

For JS, specification of the noise term ν^2 requires additional assumptions, and along with τ^2 can be considered "oracle parameters" requiring empirical substitutes. Likewise, the unobserved covariance parameters $\sigma^2|\beta|^2$ and δ^2 of (14) can be considered oracle parameters. The observed eigenvalues of S are additional empirical ingredients in estimating these parameters.

It is useful to notice first that in the HL regime, when data are explained by a one-factor model like (13), the eigenvalues of S remain stable after division by the number of variables p as it increases. The stability can be explained with classical statistics. Consider the $p \times n$ matrix Y holding n observations of p variables, assumed to have zero mean. When $p > n$, the sample covariance matrix $S = YY^\top/n$ is singular. Assuming no exceptional dependence among the observations of the variables, exactly $p - n$ of its eigenvalues are zero. Now consider the $n \times n$ dual sample covariance matrix $S^D = Y^\top Y/p$, which measures cross-sectional average co-movement at pairs of times. Every nonzero eigenvalue of S is obtained by scaling an eigenvalue of S^D by p/n. Since the roles of p and n are reversed in S^D compared to S, S^D is a covariance matrix of a small number of variables estimated from a large number of observations. This is the low-high (LH) domain of classical statistics.

The eigenvalues of S help us estimate the oracle parameters in the factor model (14). Under standard factor model assumptions, the leading eigenvalue λ^2 of S is approximated, for large p, by

$$\lambda^2 \approx \frac{|\beta|^2|f|^2}{n} + \frac{p}{n}\delta^2 \tag{26}$$

where $f = (f_1, \ldots, f_n)$ is the vector of realizations of the common factor return corresponding to the n observation times. The trace of S is approximated by

$$Tr(S) \approx \frac{|\beta|^2|f|^2}{n} + p\delta^2. \tag{27}$$

From the definition (8), it follows that

$$\nu^2 \approx \delta^2/n. \tag{28}$$

Noise to corrupted signal

Like JS, the shrinkage constant for GPS and MAPS can be described in terms of a ratio $\nu^2/s^2(h)$ of noise to a "corrupted signal". The numerator ν^2 is a scaled average of sample non-zero smaller eigenvalues. Formula (28) says that ν^2 is, through the lens of a factor model, an estimate of scaled specific variance. Both representations identify ν^2 as noise if the true leading eigenvector and eigenvalue b and λ are viewed as the primary carriers of information. The denominator $s^2(h)$ is expressed in (7) and (18) as the variation of the sample leading eigenvector scaled by the sample leading eigenvalue. This term is driven, in part, by the the variance of the true leading eigenvector. Further analysis is required to fully understand the way in which this signal is obscured by the smaller sample eigenvalues that determine ν^2. A clue to the mystery may be in Denton et al. (2022).

Consistency

Unlike MAPS, neither the data-driven GPS estimator nor the oracle to which it aspires can be consistent estimators of the true eigenvector b. Nevertheless, GPS shrinkage can eliminate the impact of the dispersion bias on quadratic optimization with a one-factor covariance matrix like the one in (15). The elimination of the dispersion bias is complete at finite p for the oracle h^O and asymptotic for the data-driven estimator h^{GPS}.

MAPS may or may not generate a consistent estimator of b, depending on the anchor points used. By correcting systematic errors beyond the dispersion bias, a MAPS-based covariance matrix can generate more accurate minimum variance portfolios than a GPS-based covariance matrix. Eigenvector bias and its relationship to quadratic optimization in the HL regime is an open area of research.

Outlook

The GPS and MAPS eigenvector estimators, conceived originally as corrections to PCA models in the HL regime for use in quadratic optimization, are analogs of the JS estimator for a collection of averages. We've highlighted essential similarities and also important differences between James Stein shrinkage for averages and for eigenvectors. The growing prevalence of HL data sets in finance, genetics and machine learning, where the number of variables vastly exceeds the number of observations, calls for a deeper understanding of biases in estimated eigenvectors and methods to correct them.

Historical notes

Primary sources for the James-Stein estimator are Stein (1956) and James & Stein (1961), and a later overview is Efron & Morris (1977). Sir Francis Galton in the 19th century formulated the concepts of correlation and regression to mediocrity, more commonly known today as regression to the mean; see Galton (1886). Notable contributions on the role of factor models in financial economics include Sharpe (1963), Sharpe (1964) Rosenberg (1974) and Ross (1976). Stein (1986) discussed the currently popular practice of shrinking eigenvalues in 1986, while, to the best of our knowledge, eigenvector shrinkage was developed in 2017. The eigenvector shrinkage formulas presented in this article are linear, as in Goldberg et al. (2020) and Shkolnik (2021). Equivalent, norm-preserving versions of the shrinkage formulas are featured in Goldberg et al. (2022) and Gurdogan & Kercheval (2021).

Acknowledgements

The results presented in this article were developed in collaboration with Hubeyb Gurdogan, Alex Papanicolaou, Alex Shkolnik and Simge Ulucam. The authors are grateful to Ola Mahmoud, Caroline Ribet, Ken Ribet, Stephanie Ribet and Michelle Shkedi for insightful comments on an early draft of our paper.

References

Denton, P. B., Parke, S. J., Tao, T. & Zhang, X. (2022), 'Eigenvectors from eigenvalues: A survey of a basic identity in linear algebra', *Bulletin of the American Mathematical Society* **59**, 31–58.

Efron, B. (2010), *Large Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Cambridge Univ Press.

Efron, B. & Morris, C. (1975), 'Data analysis using stein's estimator and its generalizations', *J. of the American Statistical Assoc.* **70**(350), 311–319.

Efron, B. & Morris, C. (1977), 'Stein's paradox in statistics', *Scientific American* **236**(5), 119–127.

Galton, F. (1886), 'Regression Towards Mediocrity in Hereditary Stature.'.
URL: *https://doi.org/10.2307/2841583*

Goldberg, L. R., Papacinicolau, A., Shkolnik, A. & Ulucam, S. (2020), 'Better betas', *The Journal of Portfolio Management* **47**(1), 119–136.

Goldberg, L. R., Papanicolau, A. & Shkolnik, A. (2022), 'The dispersion bias', *SIAM Journal of Financial Mathematics, forthcoming* .

Gurdogan, H. & Kercheval, A. (2021), Multi anchor point shrinkage for the sample covariance matrix. working paper.

James, W. & Stein, C. (1961), Estimation with quadratic loss, *in* 'Proc. Fourth Berkeley Symp. Math. Stat. Prob.', pp. 361–397.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77–92.

Michaud, R. O. (1989), 'The Markowitz optimization enigma: Is 'optimized' optimal?', *Financial Analysts Journal* **45**(1), 31–43.

Rosenberg, B. (1974), 'Extra-market components of covariance in security returns', *The Journal of Financial and Quantitative Analysis* **9**(2), 263–274.

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

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

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **XIX**(3), 425–442.

Shkolnik, A. (2021), 'James-stein shrinkage for principal components', *Stat* .

Stein, C. (1956), Inadmissibility of the usual estimator for the mean of a multivariate distribution, *in* 'Proc. Third Berkeley Symp. Math. Stat. Prob.', pp. 197–206.

Stein, C. (1986), 'Lectures on the theory of estimation of many parameters', *Journal of Soviet Mathematics* **34**, 1973–1403.