James–Stein for the leading eigenvector

Lisa R. Goldberg[a,b,1] [iD] and Alec N. Kercheval[c,1,2] [iD]

Recent research identifies and corrects bias, such as excess dispersion, in the leading sample eigenvector of a factor-based covariance matrix estimated from a high-dimension low sample size (HL) data set. We show that eigenvector bias can have a substantial impact on variance-minimizing optimization in the HL regime, while bias in estimated eigenvalues may have little effect. We describe a data-driven eigenvector shrinkage estimator in the HL regime called "James–Stein for eigenvectors" (JSE) and its close relationship with the James–Stein (JS) estimator for a collection of averages. We show, both theoretically and with numerical experiments, that, for certain variance-minimizing problems of practical importance, efforts to correct eigenvalues have little value in comparison to the JSE correction of the leading eigenvector. When certain extra information is present, JSE is a consistent estimator of the leading eigenvector.

asymptotic regime | shrinkage | factor model | optimization | covariance matrix

Significance

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 mitigates bias in the leading eigenvector of a sample factor-based covariance matrix estimated in the high-dimension low sample size (HL) regime. The analysis illuminates how estimation error in a covariance matrix can affect quadratic optimization. Eigenvector estimation in the HL regime may be useful for disciplines, such as finance, machine learning, or genomics, in which high-dimensional variables need to be analyzed from a limited number of observations.

Averaging is the most important tool for distilling information from data. To name just two of countless examples, batting average is a standard measure of the likelihood that a baseball player will get on base, and an average of squared security returns is commonly used to estimate the variance of a portfolio of stocks.

The average can be the best estimator of a 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 sum of mean squared errors is no longer minimized by the collection of averages. Instead, the error can be reduced by shrinking the averages toward a common 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 prone to be overestimated and the smallest underestimated. That is why the aggregate mean squared error is reduced when the collection of observed averages are all moved toward their center.

This line of thinking has been available at least since Sir Francis Galton introduced "regression towards mediocrity" in 1886. Still, Charles Stein surprised the community of statisticians with a sequence of papers about this phenomenon beginning in the 1950s. Stein showed that it is always possible to lower the aggregate squared error of a collection of three or more averages by explicitly shrinking them toward their collective average. In 1961, Stein improved and simplified the analysis in collaboration with Willard James. The resulting empirical James–Stein shrinkage estimator (JS) launched a new era of statistics.

This article describes "James–Stein for eigenvectors" (JSE), a recently discovered shrinkage estimator for the leading eigenvector of an unknown covariance matrix. A leading eigenvector is a direction in a multidimensional data set that maximizes explained variance. The variance explained by the leading eigenvector is the leading eigenvalue.

Like a collection of averages, a sample eigenvector is a collection of values that may be overly dispersed. This can happen in the high-dimension low sample size (HL) regime when the number of variables is much greater than the number of observations. In this situation, the JSE estimator reduces excess dispersion in the entries of the leading sample eigenvector. The HL regime arises when a relatively small number of observations are used to explain or predict complex high-dimensional phenomena, and it falls outside the realm of classical statistics. Examples of such settings include genome-wide association studies (GWAS), such as (1) and (2), in which characteristics of a relatively small number of individuals might be explained by millions of single nucleotide polymorphisms (SNPs); machine learning in domains with a limited number of high-dimensional observations, such as in (3); and finance, in which the number of assets in a portfolio can greatly exceed the number of useful observations.

We work in the context of factor models and principal component analysis, which are used throughout the physical and social sciences to reduce dimension and identify

the most important drivers of complex outcomes. Principal component analysis (PCA) is a statistical technique that uses eigenvectors as factors. The results in this article are set in the context of a one-factor model that generates a covariance matrix with a single spike. This means that the leading eigenvalue is substantially larger than the others. We do not provide a recipe for practitioners working in higher-rank contexts; our goal is to describe these ideas in a setting in which we can report the current state of the theory. However, similar results are reported experimentally for multifactor models by Goldberg et al. (4), and continuing theoretical work indicates that the success of this approach is not limited to the one-factor case.

We begin this article by describing the JS and JSE shrinkage estimators side by side, in order to highlight their close relationship. We then describe three asymptotic regimes, low-dimension high sample size (LH), high-dimension high sample size (HH), and high-dimension low sample size (HL), in order to clarify the relationship between our work and the literature. Subsequently, we describe an optimization-based context in which a high-dimensional covariance matrix estimated with the JSE estimator performs substantially better than eigenvalue correction estimators coming from the HH literature. We describe both theoretical and numerical supporting results for performance metrics relevant to minimum variance optimization.

This article focuses on high-dimensional covariance matrix estimation via shrinkage of eigenvectors, rather than eigenvalues or the entire covariance matrix. It relies on results from the HL regime and emphasizes optimization-based performance metrics. The bulk of the existing high-dimensional covariance estimation literature concerns correction of biased eigenvalues or provides results only in the HH regime or focuses on metrics that do not take account of the use of covariance matrices in optimization.

James–Stein for Averages

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 only a single mean is estimated. The discovery of Stein (5) and James and Stein (6) is that a better estimate is obtained by shrinking the sample averages toward their collective average.

Let $m(z) = \sum_{i=1}^{p} z_i/p$ denote the collective average, and $\mathbf{1} = (1, 1, \ldots, 1)$, the p-dimensional vector of 1s. With certain normality assumptions, James and Stein define:

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

The shrinkage constant c^{JS} is given by

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

where

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

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. The value of ν^2, a measure of the noise affecting each observed average,

must be either assumed or estimated independently of $s^2(z)$, and is sometimes tacitly taken to be 1.

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 Eq. 2 can be thought of as an estimated ratio of noise to the sum of signal and noise. Eq. 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. Readers may consult Efron and Morris (7), (8), and Efron (9) for more complete discussion and motivation behind formula [1] as an empirical Bayes estimator.

James and Stein showed that the JS estimator $\hat{\mu}^{JS}$ is superior to z 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 $i = 1, 2, \ldots, p$, 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, Eq. 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 constant. If the true batting averages differ widely, but the sample averages tend to be close to the true values, then Eq. 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.

James–Stein for Eigenvectors

Consider a sequence of n independent observations of a variable of dimension p, drawn from a population with unknown covariance matrix Σ. The $p \times p$ sample covariance matrix S has the spectral decomposition:

$$S = \lambda^2 h h^\top + \lambda_2^2 v_2 v_2^\top + \lambda_3^2 v_3 v_3^\top \cdots + \lambda_p^2 v_p v_p^\top. \qquad [5]$$

in terms of the nonnegative eigenvalues $\lambda^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_p^2 \geq 0$ and orthonormal eigenvectors $\{h, v_2, \ldots, v_p\}$ of S. Our interest is primarily in the leading eigenvalue λ^2 and its corresponding eigenvector h when $p >> n$. In what follows, the sample eigenvector h plays the role of the collection of sample averages z in the previous discussion.

In classical statistics with fixed p, the sample eigenvalues and eigenvectors are consistent estimators of their population counterparts when the population eigenvalues are distinct. This means that the sample estimates converge to the population values as n tends to infinity. However, this may fail when the dimension tends to infinity. The purpose of JSE is to provide an empirical estimator improving on the sample eigenvector h in the HL setting.

JSE is a shrinkage estimator, analogous to JS, that improves on h by having a lower squared error with high probability and leading to better estimates of covariance matrices for use in

quadratic optimization. Goldberg, Papanicolaou, and Shkolnik introduced and analyzed the JSE estimator in (10) as a means to improve the output of quadratic optimization. It is further developed and extended by Goldberg et al. (4) and Gurdogan and Kercheval (11). The connection between JSE and JS first appears in Shkolnik (12) in the context of a single spiked covariance model.

The JSE estimator h^{JSE} is defined by shrinking the entries of h toward their average $m(h)$, just as in Eq. **1**:

$$h^{\mathrm{JSE}} = m(h)\mathbf{1} + c^{\mathrm{JSE}}(h - m(h)\mathbf{1}), \qquad [6]$$

where the shrinkage constant c^{JSE} is

$$c^{\mathrm{JSE}} = 1 - \frac{v^2}{s^2(h)}, \qquad [7]$$

where

$$s^2(h) = \frac{1}{p} \sum_{i=1}^{p} (\lambda h_i - \lambda m(h))^2 \qquad [8]$$

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

$$v^2 = \frac{tr(S) - \lambda^2}{p \cdot (n-1)}. \qquad [9]$$

As with JS, JSE calls for a lot of shrinkage when the average of the nonzero smaller eigenvalues dominates the variation of the entries of λh around their average and only a little shrinkage when the reverse is true. The estimator h^{JSE} improves on the sample leading eigenvector h of S, as we describe below, by reducing its angular distance to the population eigenvector.

To state a precise result, we introduce the factor model framework in which we are applying JSE, as initiated in (10) and elaborated in (11). Factor models are widely used to reduce dimension in settings where there are a relatively small number of drivers of a complex outcome. The prototype is a one-factor model:

$$r = \beta f + \epsilon, \qquad [10]$$

where r is a p-vector that is the sole observable, β is a p-vector of factor loadings, the scalar f is a common factor through which the observable variables are correlated, and ϵ is a p-vector of variable-specific effects that are not necessarily small but are homogeneous and uncorrelated with f and each other. Setting the factor variance to be σ^2 and the specific variance to be δ^2, the population covariance matrix takes the form:

$$\Sigma = \sigma^2 \beta \beta^\top + \delta^2 I, \qquad [11]$$

and β is its leading eigenvector.

Our theoretical results are asymptotic in the number of variables p, so we introduce a fixed sequence of scalars $\{\beta_i\}_{i=1}^{\infty}$, from which we draw factor loadings. Suppressing dependence on dimension in our notation, let β be the p-vector whose entries are the first p elements of the fixed sequence. To prevent asymptotic degeneracy of the p-indexed sequence of models, we impose the normalizing condition that $|\beta|^2/p = (1/p) \sum_{i=1}^{p} \beta_i^2$ tends to a finite positive limit as $p \to \infty$.

Any nonzero multiple of an eigenvector is an eigenvector, so we define the distance between population and estimated eigenvectors as the smallest positive angle, denoted $\angle$, between representatives.

Theorem 1 (10). *Assume that the angle $\angle(\beta, \mathbf{1})$ tends to a limit strictly between zero and $\pi/2$.*

Then, in the limit as $p \to \infty$ with n fixed,

$$\angle(h^{\mathrm{JSE}}, \beta) < \angle(h, \beta) \qquad [12]$$

almost surely.

The proofs in (10) assume the equivalent hypotheses that the mean $m(\beta)$ and dispersion $d(\beta)$ have finite positive limits, where

$$d^2(\beta) = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{\beta_i - m(\beta)}{m(\beta)} \right)^2. \qquad [13]$$

A limiting mean of zero corresponds to a limiting angle between β and $\mathbf{1}$ of $\pi/2$, in which case h^{JSE} reduces to h and the strict inequality of *Theorem 1* becomes a weak inequality.

The unit eigenvector $b = \beta/|\beta|$ is featured in our illustration of [**12**] in Fig. 1. The left panel shows JSE shrinkage as defined by Eq. **6**. The right panel shows an equivalent formulation of JSE shrinkage in terms of vectors on the unit sphere obtained by normalization.

The conclusion of *Theorem 1* is equivalent to the statement that the JSE estimator reduces the Euclidean distance between normalized representatives

$$\left| \frac{h^{\mathrm{JSE}}}{|h^{\mathrm{JSE}}|} - b \right| < |h - b|, \qquad [14]$$

when they are chosen to lie in the same hemisphere. This is due to the elementary relation $(1/2)|u - v|^2 = 1 - \cos\angle(u, v)$ for any unit vectors u, v.

Theorem 1 guarantees that the angle between h^{JSE} and b becomes smaller than the angle between h and b for p sufficiently large as long as $\angle(\beta, \mathbf{1})$ tends to a value in the interval $(0, \pi/2)$ and $|\beta|^2/p$ tends to a positive value as p tends to infinity. We explore the magnitude of improvement offered by JSE on a data set of $n = 40$ observations and $p = 50, 100, 200,$ and 500 variables. Gaussian data are simulated with the factor model Eq. **10**, with $\sigma = 0.16$, $\delta = 0.60$, and β generated by applying an appropriate affine transformation to pseudorandom draws from a normal distribution so that $|\beta|^2/p = 1$ and $\angle(\beta, \mathbf{1})$ is as desired. The choice of these parameters is motivated by equity markets, as described in (4). We consider small, medium, and large angles, $\angle(\beta, \mathbf{1}) = 0.174, 0.785,$ and 1.40 radians, equivalently, 10, 45, and 80°. For each fixed p and β, our experiment relies on 1,000 simulated paths.

Table 1 shows the mean and median difference,

$$D = \angle(h, b) - \angle(h^{\mathrm{JSE}}, b), \qquad [15]$$
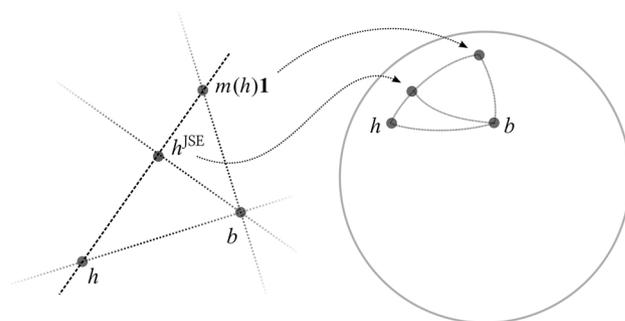
Fig. 1. Shrinkage of the sample eigenvector h along the line connecting h and $m(h)\mathbf{1}$ in Euclidean space (*Left*) and projected on the unit sphere (*right*).

Table 1. Improvement (D) measured in radians, by JSE over the sample eigenvector as an estimator of the population eigenvector.

$\angle(\beta, \mathbf{1})$ (Radians)	p	Mean D	Median D	Interquartile Range D	P(D > 0)
0.174	50	0.276	0.289	0.051	0.996
	100	0.316	0.345	0.055	0.989
	200	0.276	0.367	0.066	0.967
	500	0.328	0.370	0.071	0.983
0.785	50	0.066	0.066	0.025	0.995
	100	0.069	0.069	0.028	0.998
	200	0.070	0.068	0.027	1.000
	500	0.069	0.067	0.025	1.000
1.396	50	0.000	0.002	0.004	0.724
	100	0.001	0.003	0.004	0.762
	200	0.002	0.003	0.004	0.757
	500	0.002	0.002	0.003	0.867

Gaussian data are generated from the factor model [10] with $\sigma = 0.16$, $\delta = 0.60$, and $|\beta|^2/p = 1$. Results are based on 1,000 simulations of $n = 40$ observations for each value of $\angle(\beta, \mathbf{1})$ and p. Average and median improvement are uniformly positive, and they increase as $\angle(\beta, \mathbf{1})$ decreases. Results are consistent across values of p considered.

along with its interquartile range and the probability that D is positive. The mean, median, and interquartile range of improvement D by JSE are small and positive for the largest angle we consider, $\angle(\beta, \mathbf{1}) = 1.40$ radians, close to a right angle, and increase materially as that angle diminishes. The probability that D is positive exceeds 0.72 in all cases and exceeds 0.96 for the two smaller angles. The results are stable across values of p, consistent with the hypothesis that $n = 40$ and $p = 50$ are effectively in the asymptotic regime for the factor model that we specified.

A More General Shrinkage Target. In Eqs. **1** and **6**, JS and JSE reduce excess dispersion in an estimated vector of interest relative to a shrinkage target, $\tau = m(\cdot)\mathbf{1}$, with constant entries. Efron and Morris (7) describe the JS estimator for a more general shrinkage target, where the dispersionless vector $m(\cdot)\mathbf{1}$ is replaced by an initial guess $\tau \in R^p$ for the unknown μ. In that case, the JS estimator becomes

$$\hat{\mu}^{JS} = \tau + c^{JS}(z - \tau), \quad [16]$$

where c^{JS} is defined relative to τ, with

$$s^2(h) = \sum_{i=1}^{p} (z_i - \tau_i)^2/(p-2). \quad [17]$$

We describe a similar generalization of *Theorem 1*. As we did for factor loadings β, we introduce a fixed sequence of scalars $\{\tau_i\}_{i=1}^{\infty}$, from which we draw coordinates of a shrinkage target vector τ. In the previous case, $\tau_i = 1$ for all i. Continuing to suppress dimension in our notation, let τ be the p-vector whose entries are the first p elements of the sequence. To avoid degeneracy, we again impose the normalizing assumption that $|\tau|^2/p$ tends to a finite positive limit as $p \to \infty$.

For any p-vector y, denote the the orthogonal projection of y onto τ by

$$P_\tau(y) = \langle y, \tau \rangle \frac{\tau}{|\tau|^2}. \quad [18]$$

Define the generalized variance relative to τ as

$$v_\tau^2(y) = \frac{1}{p}|y - P_\tau(y)|^2, \quad [19]$$

and define the generalized shrinkage constant

$$c_\tau^{JSE} = 1 - \frac{v^2}{\lambda^2 v_\tau^2(h)}, \quad [20]$$

where v^2 is defined as before and we assume $h \neq P_\tau(h)$. We may now define the generalized JSE estimator as

$$h_\tau^{JSE} = P_\tau(h) + c_\tau^{JSE}(h - P_\tau(h)), \quad [21]$$

which depends only on the line determined by τ.

Theorem 2 (10). *Assume that the angle $\angle(\beta, \tau)$ tends to a limit strictly between zero and $\pi/2$.*

Then, in the limit as $p \to \infty$ with n fixed,

$$\angle(h_\tau^{JSE}, \beta) < \angle(h, \beta) \quad [22]$$

almost surely.

The proof of *Theorem 2* is a formal generalization of the proof of theorem 3.1 in (10), with the original target $\mathbf{1}$ replaced by τ, as long as the nondegeneracy condition on $|\tau|^2/p$ is satisfied. When the entries of τ are all ones, we recover *Theorem 1* as a special case of *Theorem 2*.

The analogy of JSE with JS suggests viewing τ as a guess at the identity of the true eigenvector β. An alternative is to think of τ as an exogenously imposed constraint in a variance-minimizing optimization. In this situation, JSE corrects the sample eigenvector in the direction of τ to reduce optimization error. The effectiveness of this correction is controlled by the angle between β and τ, $\angle(\beta, \tau)$ as well as $|\beta^2|/p$ and $|\tau|^2/p$. This alternative perspective allows us to think of a τ-indexed family of biases in the sample eigenvector h.

A Consistent Estimator. An extension of the generalized JSE estimator developed by Gurdogan and Kercheval in (11) incorporates multiple targets to further reduce estimation error. The result depends on a specific collection of $k = k(p) < p$ linearly independent target vectors $\{\tau^1, \tau^2, \ldots, \tau^k\}$. Letting τ denote the $(p \times k)$-dimensional matrix whose columns are the τ^is, the orthogonal projection of a p-vector y onto the k-dimensional space spanned by the columns of τ is

$$P_\tau(y) = \tau(\tau^\top \tau)^{-1}\tau^\top y. \quad [23]$$

Suppose we know the rank ordering of the betas $\beta_1, \beta_2, \ldots, \beta_p$, but not their actual values. Group the betas into k ordered quantiles, where k is approximately $\sqrt{p}$. For $i = 1, 2 \ldots, k$, define the target vector $\tau^i = (a_1, a_2, \ldots, a_p)$, where $a_j = 1$ if β_j belongs to group i, and zero otherwise.

Theorem 3 (11). *Let the number n of observations be fixed. For τ equal to the $(p \times k)$-dimensional matrix whose columns are the τ^is defined from the rank ordering of betas as above, the JSE estimator defined by Eq. **21** is a consistent estimator of b in the sense that*

$$\lim_{p \to \infty} \angle(h_\tau^{JSE}, b) = 0 \quad [24]$$

almost surely.

In (11), it is shown that the full rank ordering is not needed; only the ordered groupings are used.

Three Regimes

The two James–Stein estimators, for averages and for the leading eigenvector, are structurally parallel, but the current state of theory guarantees their performance in different settings. The dominance of JS over the sample mean expressed in inequality 4 holds in expectation, typically under normality assumptions, for finite $p > 3$. In contrast, the JSE theory of *Theorems 1* and *3* is asymptotic in the HL regime and is nonparametric, courtesy of the strong law of large numbers.

The relevance of the HL regime to the analysis of scientific data was recognized as early as 2005, by Hall et al. (13). The 2018 article by Aoshima et al. (14) surveys results on the HL regime.

The HL regime stands in contrast to the low-dimension high sample size (LH) regime of classical statistics, where the number of variables p is fixed and the number of observations n tends to infinity. In the LH regime, a sample covariance matrix based on identically distributed, independent observations is a consistent estimator of the population covariance matrix, converging in expectation as n tends to infinity. Different effects emerge in the high-dimension high sample size (HH) regime, in which both p and n tend to infinity. The HH regime is part of random matrix theory, dating back to the 1967 work of Marčenko and Pastur (15). This three-regime classification of data analysis is discussed by Jung and Marron in their 2009 article (16).

Placing any particular finite problem into an asymptotic context, whether LH, HL, HH, or something in between, requires specifying how the model is to be extended asymptotically. For LH, this means letting the number of independent observations grow, but the HH and HL regimes require defining a sequence of models of increasing dimension. This extension was natural in early works from random matrix theory that characterized the limiting spectra of standard Gaussian variables in the HH regime. Johnstone (17) looks at the HH spectrum of eigenvalues in a spiked model, where the eigenvalues of a fixed-dimensional set of eigenvectors are substantially larger than the remaining eigenvalues. The covariance matrix corresponding to the factor model, Eq. **10** is spiked. In some settings, it can be beneficial to estimate the spiked covariance model guided by *Theorems 1* and *3* from the HL regime.
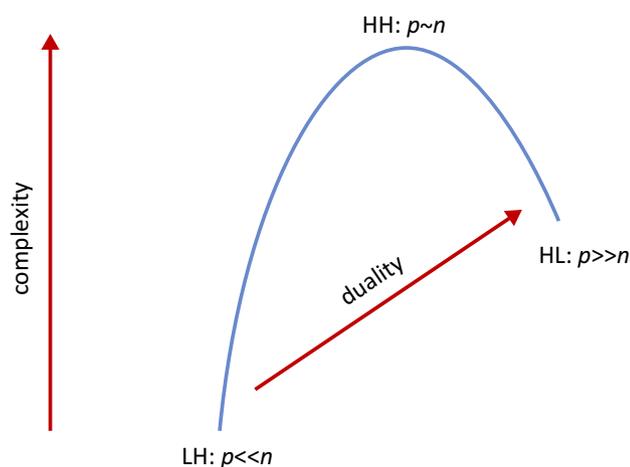
Fig. 2. Three asymptotic regimes for data analysis. LH is the low-dimension high sample size regime of classical statistics. HH is the high-dimension high sample size regime of classical random matrix theory. HL is the high-dimension low sample size regime of alternative random matrix theory. HH tends to be more complex than HL because duality arguments allow some features of classical statistics to emerge in the HL regime.

A schematic diagram of the three regimes is in Fig. 2. Duality enables us to use classical statistics to obtain results in the HL regime. This has been observed by various researchers, including Shen et al. (18) and Wang and Fan (19) and used in (10). For example, if Y is our $p \times n$ data matrix with $p > n$, the $p \times p$ sample covariance matrix $YY^\top/n$ has rank at most n. If we consider the $n \times n$ dual matrix $S^D = Y^\top Y/p$, it has a fixed dimension in the HL regime. The nonzero eigenvalues of S^D and S are related by the multiplicative factor p/n, and the eigenvectors are related by left multiplication by Y or $Y^\top$. Since, for S^D, the roles of p and n are reversed, methods from classical statistics apply.

High-Dimensional Covariance Matrix Estimation

Eigenvalue adjustment to improve covariance performance metrics, or loss functions, goes back at least to Stein's 1956 and 1986 articles (20) and (21). In this section, we discuss aspects of the literature.

In their 2018 article (22), Donoho, Gavish, and Johnstone emphasize the dependence of the optimal estimator on the choice of performance metric. Like Stein (21), they consider estimators obtained by varying the eigenvalues while keeping the sample eigenvectors fixed. In describing an oracle estimator for their spiked covariance model in the HH regime, they write:

> The oracle procedure does not attain zero loss since it is "doomed" to use the eigenbasis of the empirical covariance, which is a random basis corrupted by noise, to estimate the population covariance.

This situation is reasonable in the context they consider in which there is no prior information, other than data, about the eigenvectors. As indicated in (11), prior information can allow for the correction of a wide range of eigenvector biases in the HL regime.

Similar themes emerge from a series of articles (23–28), by Ledoit and Wolf. Beginning in 2003, these papers explore high-dimensional covariance matrix estimation with applications to financial portfolio construction and other disciplines. As in the paper by Donoho et al. (22), Ledoit and Wolf (28), consider "the class of rotation-equivariant estimators".

Ledoit and Wolf write:

> Rotation equivariance is appropriate in the general case where the statistician has no a priori information about the orientation of the eigenvectors of the covariance matrix. . .

> 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.

In earlier papers, Ledoit and Wolf consider estimators that shrink a sample covariance matrix toward a target. Some of these estimators modify the sample eigenvectors. By implementing a spiked shrinkage target in (25), Ledoit and Wolf provide prior structural information to the estimator. For the JSE estimator, that structural information is in the form of a factor model and the positive mean assumption on the leading population eigenvector.

In their 2017 article, Wang and Fan (19) develop the S-POET eigenvalue shrinkage estimator, which can be applied to the spiked covariance model in the HH and certain HL regimes. They evaluate S-POET with performance metrics based on the

relative spectral norm, the relative Frobenius norm, the spectral norm, and the max norm. Their candidate estimators, again, use the sample eigenvectors. In the absence of structural information, they also remark that "correction for the biases of estimating eigenvectors is almost impossible."

Despite the challenges of characterizing or correcting sample eigenvectors in high dimensions, there are streams of literature on the subject in both the HH and HL regimes. Some of the literature concerns consistency of sample eigenvectors under different modeling assumptions. HH references include Paul (29), Nadler (30), Mestre (31), and Johnstone and Lu (32). A 2018 survey by Johnstone and Paul (33) has an extensive reference list. HH results that are partial analogs of our findings include Montanari (34) and Montanari and Venkataramanan (35), who study estimation of singular vectors for low-rank matrices using approximate message passing (AMP) algorithms. In a 2022 article (36), Zhong, Su, and Fan describe a Bayes AMP algorithm to estimate principal components in the HH regime. Techniques from the HH regime have been applied to improve optimized portfolios; see, for example, the 2012 paper by Menchero and Orr (37), and the 2013 publication by El Karoui (38).

For the HL regime, asymptotics and estimation of eigenvectors have been studied in work previously cited and, among others, Ahn et al. (39), Jung et al. (40), Lee et al. (41), and Jung (42).

In the next section, we introduce a focus on optimization error and relevant performance metrics. We show that JSE eigenvector shrinkage, perhaps surprisingly, can substantially dominate the gains due to eigenvalue correction in optimization-based performance metrics.

JSE Corrects an Optimization Bias

Estimated covariance matrices are used in quadratic optimization, which chooses coefficients to minimize the variance of a linear combination of random variables subject to constraints. In what follows, we evaluate estimators of high-dimensional spiked covariance matrices with performance metrics that measure the accuracy of optimized quantities.

We present simulations of practical situations where JSE materially improves optimization-based performance metrics while eigenvalue corrections can have little effect. Our simulations illustrate results from (10) and (11) showing the dependence of optimization-based performance metrics on the optimization bias as the number of variables p tends to infinity and the lack of dependence of these metrics on errors in eigenvalues. Our context and examples are taken from financial economics but our results apply in any discipline where spiked covariance models are used as inputs to quadratic optimization.

Quantitative Portfolio Construction. From a universe of p financial securities, there are countless ways to construct a portfolio. We focus on quantitative portfolio construction, which has relied on mean–variance optimization since Markowitz (43). In this framework, a portfolio is represented by a vector whose ith entry is the fraction or weight of the portfolio invested in security i. A portfolio is efficient if it has minimum forecast variance subject to constraints, and the search for efficient portfolios is central to quantitative finance. The simplest efficient portfolio is minimum variance.

A fully invested but otherwise unconstrained minimum variance portfolio is the solution $\hat{w}^*$ to the mean-variance optimization problem

$$\min_{w \in \mathbb{R}^p} w^\top \widehat{\Sigma} w$$
$$\text{subject to:} \quad\quad\quad\quad\quad\quad [25]$$
$$w^\top \mathbf{1} = 1,$$

where the $p \times p$ matrix $\widehat{\Sigma}$ is a nonsingular estimate of the unknown true security covariance matrix Σ. If the estimate $\widehat{\Sigma}$ is derived from observed data, then $\hat{w}^*$ is a data-driven approximation of the true optimum w^*, defined as the solution to [25] with $\widehat{\Sigma}$ replaced by Σ.

Performance Metrics and Optimization. We review three performance metrics that are sensitive to different aspects of the impact of covariance matrix estimation error on optimization.

The variance forecast ratio (VFR) is the quotient of estimated by true variance of a linear combination of random variables. Considered in 1956 by Stein (20) for arbitrary combinations, the VFR can be substantially less than the maximum value 1 when it is applied to an optimized quantity like a minimum variance portfolio:

$$\text{VFR}(\hat{w}^*) = \frac{\hat{w}^{*\top} \widehat{\Sigma} \hat{w}^*}{\hat{w}^{*\top} \Sigma \hat{w}^*}. \quad\quad [26]$$

This is because a variance-minimizing optimization tends to place excess weight on securities whose variances and correlations with other securities are underforecast. In the words of Richard Michaud (44), mean–variance optimizers are "estimation error maximizers." Bianchi et al. (45) use the VFR to assess risk underforecasting in optimized portfolios. By considering the additional metrics described next, we are able to gauge the accuracy of optimized portfolios themselves, not merely the accuracy of their risk forecasts.

Unlike the VFR, the true variance ratio (TVR) makes sense only for optimized combinations of random variables. TVR is the quotient of the true variance of the true optimum by the true variance of the estimated optimum, and it measures excess variance in the latter:

$$\text{TVR}(\hat{w}^*) = \frac{w^* \Sigma w^*}{\hat{w}^{*\top} \Sigma \hat{w}^*}. \qu\quad [27]$$

A more direct measure of the accuracy of an optimized quantity is tracking error, which we define as:

$$\text{TE}^2(\hat{w}^*) = (\hat{w}^* - w^*)^\top \Sigma (\hat{w}^* - w^*), \qu\quad [28]$$

for the minimum variance portfolio. Tracking error is widely used by portfolio managers 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.

Since these performance metrics require knowledge of the true covariance matrix Σ, they cannot be used directly in an empirical study. However, the denominator of VFR, the true variance of the optimized quantity, can be approximated in out-of-sample empirical tests.

Factor Models, Eigenvalues, and Eigenvectors. When $p > n$, the sample covariance matrix S is singular and so is not a candidate for $\widehat{\Sigma}$. Factor models are used throughout the financial services industry and the academic literature to generate full-rank estimates of security return covariance matrices. In the

discussion below, we rely on the one-factor model specified in Eq. **10**. Similar results are obtained numerically in the case of multiple factors and nonhomogeneous specific risk in (4) and are supported by theoretical work currently in development.

Writing the factor loadings β as a product $|\beta|b$ of a scale factor and a unit vector, the population covariance matrix Eq. **11** takes the form

$$\Sigma = (\sigma^2 |\beta|^2) bb^\top + \delta^2 I. \qquad [29]$$

The quantities σ^2 and $|\beta|^2$ are not identifiable from data, but their product $\eta^2 = \sigma^2 |\beta|^2$ is. Thus, we specify an estimator $\widehat{\Sigma}$ in terms of an estimator $\hat{b} \in R^p$ of unit length and positive estimators $\hat{\eta}^2, \hat{\delta}^2 \in R$ so that

$$\widehat{\Sigma} = \hat{\eta}^2 \hat{b}\hat{b}^\top + \hat{\delta}^2 I. \qquad [30]$$

In what follows, we use guidance from the HL regime to estimate the identifiable but unobservable quantities η^2 and δ^2 from a data set.

We assume, without loss of generality, that the sample covariance matrix S has rank n. The leading eigenvalue is denoted λ^2 as before, and we set ℓ^2 to be the average of the remaining nonzero eigenvalues,

$$\ell^2 = \frac{tr(S) - \lambda^2}{n - 1}, \qquad [31]$$

where tr denotes trace. Under the assumptions of *Theorem 1*, Lemma A.2 of (10) provides the asymptotic relationships between eigenvalues of S and factor model parameters. If p is sufficiently large,

$$\lambda^2 \approx \frac{|\beta|^2 |f|^2}{n} + \frac{p}{n}\delta^2, \qquad [32]$$

where $f = (f_1, f_2 \ldots, f_n)$ is the vector of realizations of the common factor return, and

$$\ell^2 \approx \frac{p}{n}\delta^2, \qquad [33]$$

where $\approx$ means equality after division by p, in the limit as $p \to \infty$. An immediate consequence is an approximate expression for the trace of S in terms of the elements of the factor model:

$$tr(S) \approx \frac{|\beta|^2 |f|^2}{n} + p\delta^2. \qquad [34]$$

Although we do not have access to $|f|^2/n$, it is an unbiased estimator of the true factor variance σ^2. Relabelling $|f|^2/n$ by $\hat{\sigma}^2$ and applying formulas **32** and **33** gives us estimators:

$$\hat{\eta}^2 = \hat{\sigma}^2 |\beta|^2 \approx \lambda^2 - \ell^2. \qquad [35]$$

$$\hat{\delta}^2 = (n/p)\ell^2, \qquad [36]$$

that determine, for any choice of eigenvector estimator $\hat{b}$, the covariance estimator

$$\hat{\Sigma}(\hat{b}) = (\lambda^2 - \ell^2)\hat{b}\hat{b}^\top + (n/p)\ell^2 I, \qquad [37]$$

with leading eigenvalue $\lambda^2 - \ell^2 + (n/p)\ell^2$ and trace $\lambda^2 + (n-1)\ell^2$. The leading sample eigenvalue is approximately equal to the leading population eigenvalue $\sigma^2 |\beta|^2 + \delta^2$. It also agrees, for $p >> n$, with the S-POET leading eigenvalue estimate of

Wang and Fan (19), developed in a regime that includes our spiked model in the HL setting.

The leading population eigenvector b remains to be estimated. To help quantify the effect of estimation error on our performance metrics, we use the following two quantities defined for any nonzero eigenvector estimate $\hat{b}$ of b. The "optimization bias" $\mathcal{E}(\hat{b})$, introduced in (10), is

$$\mathcal{E}^2(\hat{b}) = \frac{(b, q) - (b, \hat{b})(\hat{b}, q)}{1 - (\hat{b}, q)^2}. \qquad [38]$$

and the "eigenvector bias" $\mathcal{D}(\hat{b})$, introduced in (11), is

$$\mathcal{D}(\hat{b}) = \frac{(\hat{b}, q)^2 (1 - (\hat{b}, b)^2)}{(1 - (\hat{b}, q)^2)(1 - (b, q)^2)}, \qquad [39]$$

where q is the unit vector $\mathbf{1}/\sqrt{p}$ and $(\cdot, \cdot)$ denotes the Euclidean inner product. Note $\mathcal{E}^2(b) = 0$, meaning the population eigenvector has zero bias, as desired.

As shown in (10) and (11), and discussed below, these bias measures are substantial contributors to the optimization-based performance metrics VFR, TVR, and TE. A lesson from (10) is that eigenvalue estimates can be less important, for the purpose of optimization in the HL regime, than estimating the leading eigenvector. This is especially true when considering the true variance $(\hat{w}^*)^\top \Sigma \hat{w}^*$ of an estimated minimum risk portfolio $\hat{w}^*$ defined by Eq. **25** using the estimated covariance matrix.

Correcting the Optimization Bias. In a factor model in the HL regime, JSE can correct the optimization bias, Eq. **38**, leading to greater accuracy in optimized quantities.

Theoretical guarantees of this assertion are expressed in terms of $\eta^2 = \sigma^2 |\beta|^2$, δ^2, b, and their estimates $\hat{\eta}^2$, $\hat{\delta}^2$, and $\hat{b}$ from Eq. **30**.

As a consequence of our assumptions on β, η^2 is of order p asymptotically, so the covariance matrix of data generated by our factor model is spiked. As in the setting of *Theorem 1*, we assume the nondegeneracy condition that $|\beta|^2/p$ tends to a finite positive limit as $p \to \infty$.

Theorem 4 (10 and 11). *Assume that the angle $\angle(\beta, \mathbf{1})$ tends to a limit strictly between zero and $\pi/2$. Assume that the population covariance matrix is given by Eq. **29**.*

1. *Asymptotically, the true variance of the estimated minimum variance portfolio is*

$$(\hat{w}^*)^\top \Sigma \hat{w}^* = (\eta^2/p)\mathcal{E}^2(\hat{b}) + o(p). \qquad [40]$$

   In particular, the true variance of the estimated minimum variance portfolio is asymptotically independent of eigenvalue estimates but depends only on the eigenvector estimate $\hat{b}$ and the true covariance matrix Σ.
2. $\lim_{p\to\infty} \mathcal{E}(h^{\mathrm{JSE}}) = 0$ *and* $\lim_{p\to\infty} \mathcal{E}(h) > 0$ *almost surely, where h is the leading eigenvector of S.*
3. *Asymptotically, the tracking error of the estimated minimum variance portfolio $\hat{w}^*$ is*

$$\mathrm{TE}^2(\hat{w}) = \frac{\eta^2}{p}\mathcal{E}^2(\hat{b}) + \frac{\delta^2}{p}\mathcal{D}(\hat{b}) + \frac{C}{p}\mathcal{E}(\hat{b}) + o(p), \qquad [41]$$

   where C is a constant depending on the population covariance matrix, the data, $\hat{\eta}^2$, and $\hat{\delta}^2$, but not on $\hat{b}$ (see (11)).

If we denote by w_{PCA} the minimum variance portfolio constructed using the sample eigenvector h in Eq. 37, and w_{JSE} using h^{JSE}, parts 1 and 2 of *Theorem 4* imply that $TVR(w_{PCA})$ tends to zero as the dimension p tends to infinity, but $TVR(w_{JSE})$ does not. From parts 2 and 3, it follows that $TE^2(w_{PCA})$ is bounded below, and $TE^2(w_{JSE})$ tends to zero.

Simulations calibrated to financial markets in refs. (4), (10), and (11) illustrate that these asymptotic properties are already present for values of p and n that are typical in financial markets. In addition, we observe that the variance forecast ratio is drastically improved by the JSE estimator.

Numerical Illustration. Consider the problem of estimating a covariance matrix with a year's worth of daily observations for stocks in an index like the S&P 500. The observation frequency and size of the data window are limited by empirical considerations: stocks enter and exit the index, markets undergo changes in volatility, and intraday sampling magnifies serial correlation.

In the case at hand, we have approximately $n = 252$ daily observations to estimate a covariance matrix for approximately $p = 500$ variables. Since $p > n$, this problem falls outside the realm of classical statistics. Whether it falls under the HH or HL regime and which performance metrics should be used depend on application details. The example described here illustrates a realistic context in which substantial performance improvements can be achieved using results from the HL regime to correct the leading eigenvector, while corrections of the leading eigenvalue have little value.

We examine a hypothetical market driven by the one-factor model, Eq. **10** with covariance matrix, Eq. **29**. Because the diagonal elements of S are unbiased estimators of the population variances, the trace $tr(S)$ is an unbiased estimator of the sum $tr(\Sigma)$ of the population variances. As a consequence, we preserve $tr(S)$ in our covariance matrix estimators.

We consider the following three data-driven, trace-preserving estimators:

$$\Sigma_{raw} = \left(\lambda^2 - \frac{n-1}{p-1}\ell^2\right)hh^\top + \frac{n-1}{p-1}\ell^2 I, \quad [42]$$

$$\Sigma_{PCA} = (\lambda^2 - \ell^2)hh^\top + (n/p)\ell^2 I, \quad [43]$$

$$\Sigma_{JSE} = (\lambda^2 - \ell^2)\frac{h^{JSE}(h^{JSE})^\top}{|h^{JSE}|^2} + (n/p)\ell^2 I. \quad [44]$$

Here, Σ_{raw} matches the leading eigenvalue and eigenvector of S without correction. Σ_{PCA} has the corrected leading eigenvalue but still uses the leading eigenvector h to estimate b; Σ_{JSE} improves further by substituting h^{JSE} of Eq. 6 for h.

Our factor model parameters are taken approximately from (4) and (10), which contain detailed information about calibration to financial markets. We draw factor and specific returns f and ϵ independently with mean 0 and standard deviations 16% and 60%, respectively. In the simulation, factor returns are normal, and specific returns are drawn from a t-distribution with 5 degrees of freedom. We use this fat-tailed t-distribution to illustrate that the results do not require Gaussian assumptions; repeating the experiment with several different distributions including the normal gives similar results.

The factor loadings β are inspired by market betas. We draw entries of β independently from a normal distribution with mean 1 and variance 0.25 and hold them fixed across time and simulations.

We compare the effect of eigenvalue vs. eigenvector correction on our portfolio performance metrics. In the experiment summarized in Fig. 3, we fix $p = 500$, $n = 252$, and examine the tracking error, variance forecast ratio, and true variance ratio for each of the three estimators Σ_{raw}, Σ_{PCA}, and Σ_{JSE}, with box plots summarizing the values for 400 simulations.

Correcting the leading eigenvalue, from λ^2 to the asymptotically correct $\lambda^2 - (1 - n/p)\ell^2$, has little effect compared to the JSE eigenvector correction. Related experiments described in (4) and (10) confirm that improving the accuracy of optimized
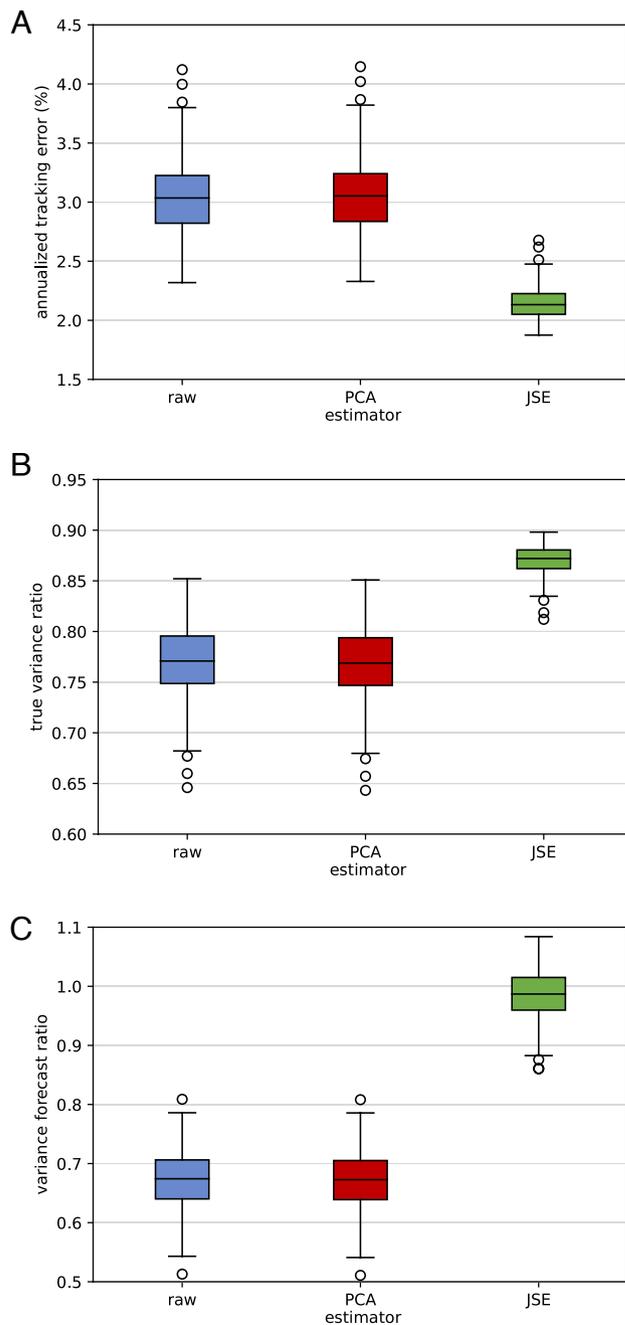
Fig. 3. Portfolio-level accuracy metrics for simulated minimum variance portfolios optimized with Σ_{raw}, Σ_{PCA}, and Σ_{JSE}: (*A*) annualized tracking error, (*B*) variance forecast ratio, and (*C*) true variance ratio. A perfect tracking error is equal to zero, and perfect variance forecast ratios and true variance ratios are equal to one. The estimated covariance matrix is based on $n = 252$ observations of $p = 500$ securities. Each boxplot summarizes 400 simulations. The experiments show that eigenvalue correction (PCA) makes no improvement, but the eigenvector correction (JSE) is substantial.

quantities has negligible dependence on the eigenvalue estimator and relies almost entirely on the choice of eigenvector. All else equal, the magnitude of the improvement in accuracy increases as the dispersion of beta decreases.

Comparing our experiment to the numerical study in (19) illustrates a conclusion from (22): The choice of performance metric materially affects the optimal covariance matrix estimator.

Summary and Outlook

This article concerns James–Stein for eigenvectors, a shrinkage method that is structurally identical to classical James–Stein. JSE has asymptotic guarantees to improve optimization-based performance metrics in the high-dimension low sample size HL regime. In the context of an empirically motivated one-factor model with a spiked covariance matrix, we show theoretically and illustrate numerically that optimization error is materially reduced by the JSE estimator, while relatively unaffected by eigenvalue correction.

Next steps are to extend the theoretical results to multifactor models and further develop the link between constrained optimization and eigenvector bias. Open problems include an empirical Bayes formulation of JSE for finite p and n and a more comprehensive understanding of the relationship between performance metrics and errors in eigenvectors and eigenvalues. The notion of "three regimes" is a simplified framework that allows us to organize results, but, in reality, the three regimes belong to a family of largely uninvestigated possibilities. Applications of JSE to GWAS studies, machine learning, and other high-dimension low sample size empirical problems await exploration.

1. G. Li, S. Jung, Incorporating covariates into integrated factor analysis of multi-view data. *Biometrics* **73**, 1433–1442 (2017).
2. The 1000 genomes project consortium: A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
3. A. Vabalas, E. Gowen, E. Poliakoff, A. Casson, Machine learning algorithm validation with a limited sample size. *PLoS One* **14** (2019).
4. L. R. Goldberg, A. Papanicolaou, A. Shkolnik, S. Ulucam, Better betas. *J. Portfolio Manage.* **47**, 119–136 (2020).
5. C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate distribution" in *Proceedings of Third Berkeley Symposium Mathematical Statistical Probabilities* (1956), pp. 197–206.
6. W. James, C. Stein, "Estimation with quadratic loss" in *Proceedings of Fourth Berkeley Symposium Mathematical Statistical Probabilities* (1961), pp. 361–397.
7. B. Efron, C. Morris, Data analysis using Stein's estimator and its generalizations. *J. Am. Stat. Assoc.* **70**, 311–319 (1975).
8. B. Efron, C. Morris, Stein's paradox in statistics. *Sci. Am.* **236**, 119–127 (1977).
9. B. Efron, *Large Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction* (Cambridge University Press, 2010).
10. L. R. Goldberg, A. Papanicolaou, A. Shkolnik, The dispersion bias. *SIAM J. Financ. Math.* **13**, 521–550 (2022).
11. H. Gurdogan, A. Kercheval, Multi anchor point shrinkage for the sample covariance matrix. *SIAM J. Financ. Math.* **13**, 1112–1143 (2022).
12. A. Shkolnik, James-Stein shrinkage for principal components. *Statistics* (2021).
13. P. Hall, J. Marron, A. Neeman, Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **67**, 427–444 (2005).
14. M. Aoshima *et al.*, A survey of high dimension low sample size asymptotics. *Aust. N. Z. J. Stat.* **60**, 4–19 (2018).
15. V. Marčenko, L. Pastur, Distribution of eigenvalues for some sets of random matrices. *Matemat. Sbornik* **114**, 507–536 (1867).
16. S. Jung, J. Marron, PCA consistency in high dimension, low sample size context. *Ann. Stat.* **37**, 4104–4130 (2009).
17. I. Johnstone, On the distribution of the largest eigenvalue in principal components analysis. *Ann. Stat.* **29**, 295–327 (2001).
18. D. Shen, H. Shen, H. Zhu, J. S. Marron, The statistics and mathematics of high dimension low sample size asymptotics. *Stat. Sin.* **26**, 1747–1770 (2016).
19. W. Wang, J. Fan, Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann. Stat.* **45**, 1342–1374 (2017).
20. C. Stein, Some problems in multivariate analysis part i (Tech. rep. no. 6, Office of Naval Research, 1956).
21. C. Stein, Lectures on the theory of estimation of many parameters. *J. Soviet Math.* **34**, 1403–1973 (1986).
22. D. Donoho, M. Gavish, I. Johnstone, Optimal shrinkage of eigenvalues in the spiked covariance model. *Ann. Stat.* **46**, 1742–1778 (2018).
23. O. Ledoit, M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance* **10**, 603–621 (2003).
24. O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **88**, 365–411 (2004).
25. O. Ledoit, M. Wolf, Honey, I shrunk the sample covariance matrix. *J. Portfolio Manage.* **30**, 110–119 (2004).
26. O. Ledoit, M. Wolf, Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Stat.* **40**, 1024–1060 (2012).
27. O. Ledoit, M. Wolf, Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *J. Multivar. Anal.* **139**, 360–384 (2015).
28. O. Ledoit, M. Wolf, Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* **30**, 4349–4388 (2017).
29. D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Stat. Sin.* **17**, 1617–1642 (2007).
30. B. Nadler, Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Stat.* **36**, 2791–2817 (2008).
31. X. Mestre, Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Trans. Inf. Theory* **54**, 5113–5129 (2008).
32. I. M. Johnstone, A. Y. Lu, On consistency and sparsity for principal components analysis in high dimensions. *J. Am. Stat. Assoc.* **104**, 682–693 (2009).
33. I. Johnstone, D. Paul, PCA in high dimensions: An orientation. *Proc. IEEE Inst. Electr. Electron Eng.* **106**, 1277–1292 (2018).
34. A. Montanari, Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62**, 1458–1484 (2016).
35. A. Montanari, R. Venkataramanan, Estimation of low-rank matrices via approximate message passing. *Ann. Stat.* **49**, 321–345 (2021).
36. X. Zhong, C. Su, Z. Fan, Empirical Bayes PCA in high dimensions. *J. R. Stat. Soc. Ser. B* (2022).
37. J. Menchero, J. Wang, D. Orr, Improving risk forecasts for optimized portfolios. *Financ. Anal. J.* **68**, 40–50 (2012).
38. N. E. Karoui, On the realized risk of high-dimensional Markowitz portfolios. *SIAM J. Financ. Math.* **4**, 737–783 (2013).
39. J. Ahn, J. S. Marron, K. M. Muller, Y. Y. Chi, The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrics* **94**, 76–766 (2007).
40. S. Jung, A. Sen, J. S. Marron, Boundary behavior in high dimension, low sample size asymptotics of PCA. *J. Multivar. Anal.* **109**, 190–203 (2012).
41. S. Lee, F. Zou, F. A. Wright, Convergence of sample eigenvalues, eigenvectors, and principal component scores for ultra-high dimensional data. *Biometrika* **101**, 484–490 (2014).
42. S. Jung, Adjusting systematic bias in high dimensional principal component scores. *Stat. Sin.* **32**, 939–959 (2022).
43. H. Markowitz, Portfolio selection. *J. Finance* **7**, 77–92 (1952).
44. R. O. Michaud, The Markowitz optimization enigma: Is 'optimized' optimal? *Financ. Anal. J.* **45**, 31–43 (1989).
45. S. W. Bianchi, L. R. Goldberg, A. Rosenberg, The impact of estimation error on latent factor model forecasts of portfolio risk. *J. Portfolio Manage.* **43**, 147–156 (2017).