

Portfolio optimization via strategy-specific eigenvector shrinkage

Lisa R. Goldberg · Hubeyb Gurdogan ·
Alec Kercheval

First Draft: August 16, 2023
This Version: August 29, 2023

Abstract We estimate covariance matrices for portfolio optimization that are tailored to given constraints. Our factor-based, data-driven construction relies on a generalized version of James-Stein for eigenvectors (JSE), which reduces estimation error in the leading sample eigenvector by shrinking toward a subspace of dimension ≥ 1 . Unchecked, this error gives rise to excess volatility for optimized portfolios. Our results include a formula for the asymptotic improvement of JSE over the sample leading eigenvector as an estimate of ground truth, and provide improved optimal portfolio estimates when variance is to be minimized subject to finitely many linear constraints.

Keywords eigenvector estimation · high dimension · portfolio optimization · James-Stein

Mathematics Subject Classification (2020) 91G10 · 62H12 · 62P05

JEL classification C38

1 Introduction

Since Harry Markowitz launched modern finance in 1952 by constructing portfolios with mean-variance optimization, active research has focused on determining appropriate estimates for expected security returns and covariances.

Lisa R. Goldberg
Department of Economics and Consortium for Data Analytics in Risk, UC Berkeley, Berkeley, CA 94720 and Aperio Group LLC, a wholly-owned indirect subsidiary of BlackRock, Inc.

Hubeyb Gurdogan
Department of Mathematics, UCLA, Los Angeles, CA 90024

Alec Kercheval
Department of Mathematics, Florida State University, Tallahassee, FL 32306
E-mail: akercheval@fsu.edu

Practical considerations often limit the pool of relevant data, rendering a sample covariance matrix unfit for use in optimization. Prescriptions for empirically sound, well-conditioned covariance matrix estimates required by optimization vary, and the nature of their errors and their impact on optimized portfolios can be obscure.

A standard approach to return covariance matrix estimation begins with a factor model. This is consistent with empirically observed correlations in financial returns and it reduces dimension, leading to conditioning sufficient to stabilize optimization. Principal component analysis (PCA) can be used to identify factors that explain correlation, for example in the arbitrage pricing theory developed by Stephen Ross in 1976. The factor loadings are sample eigenvectors, linear combinations of security returns that maximize in-sample variance.

When securities are numerous and observations are scant, however, sample eigenvectors are poor estimates of their population counterparts. As building blocks of covariance matrices intended for optimization, sample eigenvectors lead to optimized portfolios with variance that is substantially higher than it needs to be.

We address this problem in the setting of a single-factor model that incorporates the most salient features of equity markets in simplest form. We develop high-dimensional covariance matrix estimates that generate low-variance optimized portfolios. Extending recent research that sheds light on how estimation error is transmitted via optimization, the factor underlying the rank-one component of our estimate is obtained by applying James-Stein shrinkage to the leading sample eigenvector, yielding a *James-Stein for eigenvectors* (JSE) estimate of the leading population eigenvector.

We advance the literature in three ways. First, we provide easy-to-code formulas for factor-based covariance matrices that are tailored to specific quadratic optimization problems. When input to optimization, these matrices generate low-variance instances of portfolios that satisfy optimization constraints. The low variance stems from JSE shrinkage of the leading sample eigenvector toward the target subspace generated by optimization constraint vectors. JSE stochastically dominates the leading sample eigenvector as an estimate of ground truth, and neutralizes a significant component of estimation error that arises in optimization.

The second advancement is a formula for improvement of JSE over the sample leading eigenvector that depends only sample eigenvalues and the angle between the leading population eigenvector and the target subspace. The third advancement concerns the nature of JSE shrinkage. In previous studies, JSE shrinkage is in a specific direction. For the problems considered in this article, the ideal shrinkage target is the projection of the leading population eigenvector, which is unobservable, onto the target subspace. We show that a data-driven shrinkage target obtained by projecting the leading sample eigenvector onto the constraint subspace is sufficient to guarantee reduced variance of the optimized portfolio.

In Section 2, we review some background and literature relevant to our results. In Section 3, we set up the problem of finding a low-variance solution to mean-variance optimization with linear constraints when the covariance matrix is estimated. Readers interested in the bottom-line formulas for implementation will find them summarized in Section 4, while Section 5 provides a detailed mathematical discussion of the construction and describes its asymptotic properties. Numerical experiments illustrating our results are in Section 6, and Section 7 contains concluding thoughts. Mathematical proofs are in the Appendix, Section 8.

2 Context and related literature

Our results about JSE draw on factor models and optimization, random matrix theory and James-Stein shrinkage. We extend the connection between eigenvector shrinkage and estimation of minimum variance to a large collection of optimization problems: minimum variance with linear constraints.

2.1 Factor models and optimization

The vast literature on the estimation of covariance matrices for use in portfolio construction begins with Markowitz [32] and [33]. We do not attempt to survey that literature here, but highlight Sharpe [38], who developed a one-factor or “single index” model. Rosenberg [36] builds on Sharpe’s idea, establishing Barra’s industry standard fundamental factor models.

A statistical approach to factor models, based on the Arbitrage Pricing Theory in Ross [37] and developed in Chamberlain and Rothschild [6], Connor [7], Connor and Korajczyk [11] and [9], is an antecedent of the material in this article. Connor [8] and Connor and Korajczyk [10] review roles of different types of factor models in finance.

2.2 Random matrix theory and covariance matrix estimation

Empirical analysis of public equities in US and Global markets suggests that the leading eigenvalue of a sample stock or bond return covariance matrix tends to grow in proportion to the number of securities in the pool. This means that the widely-studied single-spiked covariance model, introduced in Johnstone [26] may be an appropriate basis for estimating covariance matrices for portfolio optimization. That paper brings random matrix theory (RMT) to bear on the problem of covariance matrix estimation, with special attention to the limiting spectrum of eigenvalues in the high dimension high sample size (HH) regime, where the number of variables p and the number of observations n tend to infinity in proportion. This type of analysis, launched by Wigner [43], [44] and Marcenko and Pastur [31] in the 1950s and 1960s, has spawned another vast literature. Applications of this work to financial covariance matrix

estimation is in Menchero et. al.[34], Ledoit and Wolf [29], [30], El Karoui [15], [16], and Wang and Fan [42], to name a few examples. Donoho et. al. [12] study the effect of different objective functions on the optimal covariance matrix estimate, given that the eigenvectors of the estimate match the sample eigenvectors.

The results in this article rely on a different, less studied branch of RMT concerning high dimension and low sample size (HL), in which the number of variables tends to infinity while the number of observations stays fixed. The relevance of the HL regime to the analysis of scientific data was pioneered in Hall et. al. [23] and Ahn et. al. [1], and results are surveyed in Aoshima et. al. [2]. Data analysis in the HH and HL regimes, as well as the low dimension high sample size (LH) regime of classical statistics, is discussed in Jung and Marron [27].

2.3 James-Stein shrinkage for averages and for eigenvectors

Stein [40] and James and Stein [24] show that in dimension greater 3 or greater, the sample average is inadmissible: there is another estimator with lower mean-squared error. That superior estimator is known as James-Stein, and it is obtained by shrinking sample averages toward their collective mean. This work was extended by replacing the grand mean with arbitrary initial guesses in Efron and Morris[13], and popularized by Efron [14]. An overview of James-Stein type shrinkage estimation is in Foudrinier et al. [17].

Recent literature, including Sholnik [39] and Goldberg et al. [18], develop James Stein for eigenvectors (JSE). Structurally parallel to James-Stein for averages, JSE improves almost surely on the sample leading eigenvector as an estimate of ground truth when data follow a one-factor spiked model. The theory rests on strong laws of large numbers, and therefore is free of any special distributional assumptions other than boundedness of the fourth moments. Alternative approaches may be constructed with concentration of measures arguments, described in Ball [3] and applied in Bar and Wells [4].

2.4 JSE and mean-variance optimization

Estimation error in a covariance matrix leads to optimized portfolios that are sub-optimal. A manifestation is excess variance in an optimized portfolio; see, for example, Klein andn Bawa [28], Jobson and Korkie [25], Michaud [35] and Bianchi et al. [5].

JSE was developed in Goldberg et al. [20], Goldberg et al. [19] and Gurdogan and Kercheval [22] for the purpose of improving optimized minimum variance portfolios. The development rests on a novel analysis of the way estimation error in a spiked covariance model is transmitted via mean-variance analysis. Those articles show that excess dispersion in the leading sample eigenvector contributes material errors in estimated minimum variance and its risk

forecasts, and that JSE reduces those errors in the HL regime. In the present article, we show that the original results are a special case of a more general phenomenon. A constrained optimization exacerbates estimation error in the leading sample eigenvector in the direction of the subspace spanned by constraint vectors. By shrinking the sample leading eigenvector toward that subspace, we correct the leading eigenvector in a way that is tailored to the constrained optimization problem, leading to improved results.

2.5 Multi-factor models

While a one-factor model has played an important role in the historical development of portfolio covariance estimation, practitioners usually require the factor models with several or many factors. Our analysis focuses on the case of a one-factor model driving returns in order to draw out the theory most explicitly in its simplest version.

However, similar methods are also known to work, so far with less complete theory, in at least some multi-factor contexts. Goldberg et al. [19] describe the results of simulation experiments showing that JSE, applied to the leading sample eigenvector only, improves risk forecasts in a four-factor model framework.

3 The portfolio optimization problem

We specify the central problem addressed in this article: finding low-variance solutions to variance-minimizing optimization when inputs are corrupted by estimation error.

In a universe of p securities, we specify a portfolio by a p -vector of weights w . The entries of w are the fractions of portfolio value invested in different securities. Alternatively, we can think of w in an active framework, as the difference between portfolio weight and benchmark weight. The second perspective reduces to the first when the benchmark is cash. Here, we explore a widely used framework for quantitative portfolio construction.

Let Σ denote the $p \times p$ -dimensional covariance matrix of security returns, assumed non-singular. Consider an optimization problem with $k > 0$ linear constraints,

$$\begin{aligned} \min_w \quad & \frac{1}{2} w^\top \Sigma w & (1) \\ \text{subject to} \quad & C_1^\top w = a_1 \\ & C_2^\top w = a_2 \\ & \vdots \\ & C_k^\top w = a_k \end{aligned}$$

where the j th constraint coefficient vector C_j is a p -vector, and the j th constraint target value a_j is a scalar. Typical constraints demand full investment, total and active return targets, and factor tilts, and in general are chosen to reflect an investor's specific investment strategy.

A simple, explicit formula provides the unique solution to (1) when the inputs to the problem are known. In finance, however, the covariance matrix Σ is never known, so the explicit formula provides a solution that is sub-optimal. In what follows, we illuminate the mechanism by which estimation error in a covariance matrix corrupts optimized portfolios, provide estimates of Σ tailored to instances of (1) leading to optimized portfolios that have relatively low variance.

We work in a setting where the number of securities p is larger than the number of observations n , which is commonplace for investors. In this situation, the sample covariance matrix S is singular, leading to non-unique solutions to quadratic programs such as (1). As a synthesis of information from data, however, S can serve as a source of spare parts for estimated empirically reasonable covariance matrices that can be used in optimization.

4 A JSE prescription for a customized, optimization-friendly estimate of Σ

This section contains a brief summary of our prescribed estimate of the return covariance matrix Σ that is tailored to mitigate estimation error in the optimization problem (1). The centerpiece of the prescription is an estimate of Σ 's leading eigenvector, which is obtained by applying James-Stein shrinkage to the sample leading eigenvector. Shrinkage improves on the sample leading eigenvector as an estimate of ground truth by an amount that we calculate.

In Section 5, we discuss in more detail the good asymptotic properties of both the leading eigenvector estimate and the optimal portfolio estimate when returns follow a one-factor model.

4.1 Structure from a factor model

The persistent, substantial correlations observed across financial returns have led researchers to use factor models to estimate return covariance matrices. In the simplest example of a one-factor model, the true (population) covariance matrix has the structure

$$\Sigma = \eta^2 bb^\top + \delta^2 I, \quad (2)$$

where b is a leading unit eigenvector of Σ with eigenvalue $\eta^2 + \delta^2$. A derivation of (2), repeated as (15), is in Section 5.1.1.

We don't observe Σ , but see instead a time series of n realized values of the returns p -vector r , which determine a sample $p \times p$ covariance matrix S of rank at most $n < p$. We estimate the parameters of Σ , two variances, η^2 and δ^2 , and the unit vector of factor loadings b , with functions of eigenvalues

and eigenvectors of S in a way that leads to a relatively low variance solution to (1). We show in Section 5 that the last of these three estimates is the most consequential.

4.2 A strategy-specific estimator of the vector of factor loadings

For our minimum variance problem, a *strategy* refers to the choice of constraint vectors C_1, C_2, \dots, C_k and constraint values a_1, a_2, \dots, a_k . With $\text{tr}(S)$ denoting the trace of the sample covariance matrix S and λ^2 denoting its leading eigenvalue, define

$$\ell^2 = \frac{\text{tr}(S) - \lambda^2}{n - 1}, \quad (3)$$

the average of the non-zero eigenvalues of S that are less than λ^2 , and

$$\psi^2 = \frac{\lambda^2 - \ell^2}{\lambda^2}, \quad (4)$$

the average leading relative eigengap.

Let C denote the span of the constraint vectors C_1, C_2, \dots, C_k from (1) and let h_C denote the orthogonal projection of the leading sample eigenvector h onto the subspace C . Now define the JSE shrinkage constant

$$c^{\text{JSE}} = \frac{\ell^2}{\lambda^2(1 - |h_C|^2)} \quad (5)$$

and define

$$H^{\text{JSE}} = c^{\text{JSE}}h_C + (1 - c^{\text{JSE}})h. \quad (6)$$

The James-Stein for eigenvectors (JSE) estimate of the true eigenvector b is the unit vector

$$h^{\text{JSE}} = H^{\text{JSE}}/|H^{\text{JSE}}|. \quad (7)$$

Formula (6) above is equivalent to formula [6] of [18]. That article and [39] expose the strong parallel between JSE and classical James-Stein. Formulas (5), (6) and (7) are identical to formulas (29), (30) and (31) in Section 5.1.4.

Setting $\lambda^2 - \ell^2$ and $(n/p)\ell^2$ as estimates of factor variance η^2 and specific variance δ^2 , and h^{JSE} and an estimate of b , an estimate of (2) is given by

$$\Sigma^{\text{JSE}} = (\lambda^2 - \ell^2)h^{\text{JSE}}h^{\text{JSE}\top} + (n/p)\ell^2I. \quad (8)$$

Formula (8) is the one-factor covariance matrix designed for use in quadratic optimization (1). Note that the dependence of Σ^{JSE} on C is through the factor loadings h^{JSE} and not through the estimates of factor and specific variance.

We will see, under the assumptions described in Section 5, that $|h_C|^2$ is strictly less than 1 for large p , so that c^{JSE} is well-defined, and c^{JSE} is strictly between 0 and 1 for large p , so that H^{JSE} is a proper convex combination of h and h_C .

4.3 The true variance of an optimized portfolio

The benefits of this construction are realized in the portfolio w^{JSE} generated by (1) when Σ is set to Σ^{JSE} .

Let Σ^{PCA} be the covariance matrix obtained by replacing h^{JSE} with the sample leading eigenvector h in (8), and let w^{PCA} denote the portfolio generated by (1) when Σ is set to Σ^{PCA} .

We will see in Theorem 4 that the ratio of the true variances w^{JSE} and w^{PCA} ,

$$\frac{\mathcal{V}(w^{\text{JSE}})}{\mathcal{V}(w^{\text{PCA}})}, \quad (9)$$

tends to zero as the number of assets grows. When returns to securities in a sufficiently large investment universe are governed by a one-factor model, w^{JSE} is an improvement on w^{PCA} by an arbitrarily large factor as measured by true variance.

5 JSE stochastically dominates PCA

The formulas in Section 4 prescribe the construction of a strategy-specific covariance matrix based on JSE for use in portfolio construction. Here, we describe in more detail the theory asymptotically guaranteeing that JSE improves eigenvector estimates and lowers variance of optimized portfolios, relative to PCA.

In our asymptotic analysis, we will consider n fixed and p tending to infinity. Therefore we will need to consider a sequence of models of increasing dimension. The variables in question may have a superscript (p) to emphasize the presence of the asymptotic parameter p .

In section 5.1 we show that the JSE estimator asymptotically dominates the PCA estimator in our one-factor setting, in the sense that it is strictly closer, almost surely, to the true unknown leading eigenvector. We provide a formula for the angular improvement. In section 5.2, we apply these results to estimating the variance of a portfolio obtained by minimizing variance under finitely many linear constraints. We obtain an asymptotic formula for the true variance of the portfolio obtained using an estimated covariance matrix, and show that the JSE estimator strongly dominates the PCA estimator for almost all choices of the constraint values.

5.1 JSE theorem for high-dimensional targets

We develop the JSE family of corrections of a leading sample eigenvector and provide a formula for their improvement as estimates of ground truth b when data follow a one-factor model. An estimate h^{JSE} is obtained by shrinking the sample leading eigenvector toward an observable linear subspace, the *shrinkage target* C , by a specified optimal amount. The estimate depends on the choice of

shrinkage target, and the asymptotic degree of improvement provided increases as the angle b and C decreases.

In our one-factor context, the improvement due to a JSE correction depends only on two quantities:

- The angle between the leading population eigenvector b and the shrinkage target C , and
- The relative gap between the leading sample eigenvalue and the average of the lesser, nonzero sample eigenvalues.

5.1.1 A one-factor model of returns

For $p > 1$ we will develop an estimated p -dimensional covariance matrix assuming returns follow a latent one-factor model:

$$r = \mu + \beta f + z, \quad (10)$$

where $r = r^{(p)}$ is a random p -vector that is the sole observable, $\mu = \mu^{(p)}$ is a mean returns vector, $\beta = \beta^{(p)}$ is a p -vector of factor loadings, the random scalar f is a mean-zero common factor through which the observable variables are correlated, and $z = z^{(p)}$ is a mean-zero random p -vector of variable-specific effects that are not necessarily small but are uncorrelated with f .

For the problems we consider in this article, returns are used only to estimate a sample covariance matrix. In practice, this involves subtracting expected return estimates from the observations, and it introduces expected return estimation noise into the sample covariance matrix. To focus on correlation estimation error that is not related to expected return, we assume mean zero, $\mu = 0$, and study the model

$$r = \beta f + z. \quad (11)$$

Replacing r with $r - \mu$ does not affect the covariance matrix, and amounts to the strong assumption that expected returns μ are known, and only the variances and correlations need to be estimated.

Standing Assumptions.

- A1. The random variable f is non-zero almost surely, has mean zero, finite fourth moment, and variance $\sigma^2 > 0$.
- A2. The random variables $\{z_i^{(p)} : i = 1, 2, \dots, p; p > 1\}$ are i.i.d. and have mean zero, finite fourth moment, and variance $\delta^2 > 0$.
- A3. The vector sequence $\{\beta^{(p)} : p > 1\}$ satisfies the following asymptotic non-degeneracy conditions:
 - a. The entries β_i of β are uniformly bounded:

$$\sup_{i,p} \{|\beta_i^{(p)}| : i = 1, 2, \dots, p; p > 1\} < \infty, \quad (12)$$

and

b. the sequence $|\beta^{(p)}|^2/p$ tends, as $p \rightarrow \infty$, to a positive finite limit,

$$\lim_{p \rightarrow \infty} |\beta^{(p)}|^2/p = B^2 > 0. \quad (13)$$

We make no parametric assumptions, Gaussian or otherwise, on the distributions of f or z . This level of generality holds for all the asymptotic results in the prior literature on JSE. Here, we relax JSE assumptions in previous articles by removing the requirement that the β s are drawn from a fixed, infinite list with a limiting non-zero mean and dispersion. Instead, we make weaker assumptions A3.a and A3.b. Together, these two properties guarantee that the factor loadings β are *diversifying*, in other words, they do not concentrate in a few dimensions for large values of p .

Because β and f appear in the model (11) only as a product βf , their respective scales $|\beta|$ and σ cannot be separately identified from observations of r . Therefore we introduce a single combined scale parameter

$$\eta = \eta_p = \sigma|\beta^{(p)}|,$$

and rescaled model parameters $b = \beta/|\beta|$, a unit vector, and $x = f/\sigma$, a random variable with mean zero and unit variance, and rewrite the factor model as

$$r = \eta bx + z. \quad (14)$$

With this formulation, the Standing Assumptions on $\beta^{(p)}$ are equivalent to the conditions

- A3'.a $\sup_{i,p} \{p|b_i^{(p)}|^2 : i = 1, \dots, p; p > 1\} < \infty$, and
 A3'.b η_p^2/p tends to a positive limit $\sigma^2 B^2$ as $p \rightarrow \infty$.

The population covariance matrix is a sum of a factor component, $\eta^2 bb^\top$, and a specific component, $\delta^2 I$:

$$\Sigma = \eta^2 bb^\top + \delta^2 I. \quad (15)$$

Formula (15) is identical to formula (2) in Section 4.1, the starting point of the prescription.

5.1.2 The leading sample eigenvector as an estimate of the leading population eigenvector

Fix $n \geq 2$, assume $p > n$, and consider a sequence of n independent observations r_1, r_2, \dots, r_n of the p -vector r of security returns with factor structure (14) and hence, covariance matrix Σ given by (15). Denote by Y the resulting $p \times n$ matrix whose columns are the observations r_i . The $p \times p$ sample covariance matrix $S = YY^\top/n$ has a spectral decomposition given by:

$$S = \lambda^2 h h^\top + \lambda_2^2 v_2 v_2^\top + \lambda_3^2 v_3 v_3^\top \cdots + \lambda_p^2 v_p v_p^\top \quad (16)$$

in terms of non-negative eigenvalues

$$\lambda^2 > \lambda_2^2 \geq \cdots \geq \lambda_n^2 > \lambda_{n+1}^2 = \cdots = \lambda_p^2 = 0$$

and orthonormal eigenvectors $\{h, v_2, \dots, v_p\}$ of S . We assume the generic conditions that the leading eigenvalue λ^2 has multiplicity one and S has rank n .

Our interest is in the leading sample eigenvalue λ^2 and its corresponding leading unit eigenvector h , with sign chosen, when needed, so that the inner product $\langle h, b \rangle$ is positive.

The eigenvector h of S is a commonly used estimate of the leading population eigenvector b : it is a consistent estimate of b , up to sign, in the LH regime, in the sense that it converges to b for fixed p as $n \rightarrow \infty$. However, for fixed n , the following proposition states that h stays away from b with high probability when $p \gg n$.

Recall

$$\psi_p^2 = \frac{\lambda^2 - \ell^2}{\lambda^2}. \quad (17)$$

Proposition 1 *Under assumptions A1 - A3, almost surely, the limits*

$$\theta^{\text{PCA}} = \lim_{p \rightarrow \infty} \angle(h, b) \quad \text{and} \quad \psi_\infty^2 = \lim_{p \rightarrow \infty} \psi_p^2 \quad (18)$$

exist and

$$\cos \theta^{\text{PCA}} = \psi_\infty \in (0, 1). \quad (19)$$

This means there is a positive limiting angle between h and b almost surely.

The random variable ψ_∞ can be expressed in terms of the relationship between the relative eigengap and the parameters of the factor model (14). Decomposing, from (14), the $p \times n$ data matrix of returns Y into a sum of unobservable components, we have

$$Y = \eta b X^\top + Z, \quad (20)$$

where $X = (X_1, X_2, \dots, X_n)^\top$ is the n -vector of independent realizations of x and Z is the $p \times n$ matrix whose columns are the n independent realizations of the random vector z . Since x is a mean-zero random variable with unit variance and finite fourth moment, $|X|^2$ is a noisy estimate of n . The following proposition is a simple consequence of Lemma 4 stated later.

Proposition 2 *The relative eigengap ψ_∞ is related to the parameters of the factor model by*

$$\psi_\infty^2 = \lim_{p \rightarrow \infty} \psi_p^2 = \lim_{p \rightarrow \infty} \frac{\lambda^2 - \ell^2}{\lambda^2} = \frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2} \approx \frac{p \sigma^2 B^2}{p \sigma^2 B^2 + p \delta^2 / n}. \quad (21)$$

The term ψ_∞^2 , asymptotically equal to the square of the inner product $\langle h, b \rangle$, is a measure of the asymptotic PCA estimation error when using h to estimate b . It is random because $|X|^2$ is random, but does not depend on the random matrix Z . The approximation symbol \approx in (21) is justified by the fact that $E[|X|^2/n] = 1$ and $|X|^2/n \rightarrow 1$ almost surely as $n \rightarrow \infty$. (Although we do not assume the model factor x is normal, if it were, the quantity $|X|^2$ would be chi-squared distributed with n degrees of freedom.)

The term $p\sigma^2 B^2$ appears in the numerator and denominator on the right hand side of (21). It is the trace of the factor component of the population covariance matrix Σ , specified in (15), and can be viewed as the variance in the system attributable to the factor. The term $p\delta^2$ is the trace of the specific component of Σ , and can be viewed as the variance in the system attributable to specific effects.

If we think of factor variance as signal and specific variance as noise, then Proposition 2 says that the relative eigengap ψ_∞^2 is approximated by a ratio of signal to signal plus $(1/n)$ -scaled noise. The ratio on the right hand side of (21) cannot be observed, but it can be estimated in terms of the relative eigengap of S .

A consequence of Proposition 2 is that the term ψ_∞^2 tends to 1 as $n \rightarrow \infty$. Therefore,

$$\lim_{n \rightarrow \infty} \lim_{p \rightarrow \infty} |h - b| = 0. \quad (22)$$

As a result, the defect in the PCA estimate h in applications where $p \gg n$ can be viewed as arising from limitations on the size of n . As n grows, the need for correction diminishes. Measured in radians, the asymptotic angle θ^{PCA} between h and b is, for large n , approximately

$$\theta_{PCA} \approx \frac{1}{\sqrt{n}} \frac{\delta}{\sigma B}. \quad (23)$$

For a typical value $\delta/(\sigma B) = 4$, this means the angular error θ_{PCA} will remain significant even for n as large as 1000 or more, well above the typical values seen in portfolio optimization.

5.1.3 Insight about the relationship between h and b from the perspective of an external reference subspace

Fix $k \geq 1$. For each $p > k$, let $C = C^{(p)}$ be a $p \times k$ matrix of rank k . When there is no risk of confusion, we use C to denote either the matrix or its k -dimensional column space in \mathbb{R}^p .

Notation: We use subscripts to denote orthogonal projection of a vector onto a linear subspace: h_C is the orthogonal projection of h onto C .

For any nonzero vectors $x, y \in \mathbb{R}^p$, we denote the smallest angle between the sub-spaces $\text{span}(x)$ and $\text{span}(y)$ by $\angle(x, y)$, with $0 \leq \angle(x, y) \leq \pi/2$. The angle $\angle(x, C)$ between a vector x and a subspace C is equal to $\angle(x, x_C)$.

Theorem 1 *Suppose the angle $\angle(b, C)$ between b and C tends, as $p \rightarrow \infty$, to a limit*

$$\Theta = \lim_{p \rightarrow \infty} \angle(b, C). \quad (24)$$

Then under assumptions A1 – A3, the limit

$$\Theta^h = \lim_{p \rightarrow \infty} \angle(h, C) \quad (25)$$

exists almost surely, and

$$\cos \Theta^h = \cos \theta^{\text{PCA}} \cdot \cos \Theta = \psi_\infty \cdot \cos \Theta. \quad (26)$$

In particular, if $0 < \Theta < \pi/2$, then

$$0 < \cos \Theta^h < \cos \theta^{\text{PCA}} \quad (27)$$

and

$$0 < \cos \Theta^h < \cos \Theta. \quad (28)$$

This theorem is a generalization of Theorem 3.1 of [20]. It implies, asymptotically almost surely, that h is not orthogonal to C if b is not, but the angle $\angle(h, C)$ is greater than both $\angle(b, C)$ and $\angle(h, b)$. Intuitively, this suggests that shrinking h toward C might bring it closer to b . This turns out to be correct, as described next.

5.1.4 Shrinkage improves on the leading sample eigenvector h as an estimate of the leading population eigenvector b

We will use the notation $h = h^{\text{PCA}}$ when emphasizing the contrast between PCA and JSE estimates. Next, we explore the properties of h^{JSE} , which stochastically dominates h^{PCA} as an estimate of ground truth in the limit as $p \rightarrow \infty$ under Standing Assumptions A1–A3.

Recall the JSE shrinkage constant c^{JSE} and estimator h^{JSE} are defined by

$$c^{\text{JSE}} = \frac{\ell^2}{\lambda^2(1 - |h_C|^2)}, \quad (29)$$

$$H^{\text{JSE}} = c^{\text{JSE}} h_C + (1 - c^{\text{JSE}})h, \quad (30)$$

and

$$h^{\text{JSE}} = H^{\text{JSE}} / |H^{\text{JSE}}|. \quad (31)$$

Formulas (29), (30) and (31) are identical to formulas (5), (6) and (7) in Section 4.2.

We can show that

$$\lim_{p \rightarrow \infty} c^{\text{JSE}} = \frac{1 - \psi_\infty^2}{1 - \psi_\infty^2 \cos^2 \Theta} = \frac{\delta^2}{\sigma^2 B^2 |X|^2 \sin^2 \Theta + \delta^2}. \quad (32)$$

(If now n is taken to infinity, c^{JSE} tends to zero and both h and h^{JSE} converge to b .)

We normalize h^{JSE} for convenience; all that matters is the 1-dimensional subspace it spans, as an estimate of the eigenspace $\text{span}(b)$. The angle between these subspaces is our measure of error.

Define

$$\phi_\infty^2 \equiv \frac{\psi_\infty^2}{1 - \psi_\infty^2} = \frac{\sigma^2 B^2 |X|^2}{\delta^2} = \lim_{p \rightarrow \infty} \frac{\lambda^2 - \ell^2}{\ell^2}. \quad (33)$$

Theorem 2 *Suppose the limit*

$$\Theta = \lim_{p \rightarrow \infty} \angle(b, C) \quad (34)$$

exists.

Then, under the standing assumptions A1 - A3, the limits

$$\theta^{\text{JSE}} = \lim_{p \rightarrow \infty} \angle(h^{\text{JSE}}, \beta) \quad \text{and} \quad \theta^{\text{PCA}} = \lim_{p \rightarrow \infty} \angle(h^{\text{PCA}}, \beta) \quad (35)$$

exist almost surely, and the asymptotic improvement of h^{JSE} over h^{PCA} as an estimate of the leading population eigenvector is

$$\cos^2(\theta^{\text{JSE}}) - \cos^2(\theta^{\text{PCA}}) = \left(\frac{1}{\phi_\infty^2 + 1} \right) \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1} \geq 0. \quad (36)$$

In particular, JSE is never worse asymptotically than PCA, and:

- *if $\Theta < \pi/2$, then $\theta^{\text{JSE}} < \theta^{\text{PCA}}$ almost surely,*
- *if $\Theta = 0$, then h^{JSE} converges to b and JSE is a consistent estimator, and*
- *if $\Theta = \pi/2$ then h^{JSE} converges to h^{PCA} and $\theta^{\text{JSE}} = \theta^{\text{PCA}}$.*

The right hand side of (33) is the ratio of the factor variance to the specific variance in (14). The formula highlights the relationship between the relative eigengap and the factor model parameters. Taken together, (21) and (33) imply

$$\psi_\infty^2 = \frac{\phi_\infty^2}{1 + \phi_\infty^2}. \quad (37)$$

One consequence of Theorem 2 is that the angle between h^{JSE} and h is strictly positive in the limit when $\Theta < \pi/2$. Notice also that this theorem is independent of any optimization problem.

The true asymptotic improvement $\cos^2(\theta^{\text{JSE}}) - \cos^2(\theta^{\text{PCA}})$ cannot be computed from finite data because it depends on the unobservable vector b . An observable indicator I is:

$$I(\angle(h, C), \phi_p^2) = \frac{\cos^2 \angle(h, C)}{(\phi_p^4 + \phi_p^2) \sin^2 \angle(h, C)}. \quad (38)$$

It follows from equations (26) and (36) that

$$\lim_{p \rightarrow \infty} I(\angle(h, C), \phi_p^2) = \cos^2(\theta^{\text{JSE}}) - \cos^2(\theta^{\text{PCA}}) \quad (39)$$

almost surely.

The k -dimensional target space C may arise in different ways. If chosen at random independently of b , we expect C to be asymptotically orthogonal to b as the dimension p tends to infinity (see, for example, Hall et al. [23] and Ahn et al. [1]). In this $\Theta = \pi/2$ case, JSE provides no advantage.

The condition $\Theta < \pi/2$ has a Bayesian interpretation in which C represents some mild prior information about the direction of b : namely, that b is not orthogonal to C .

This condition arises naturally in financial applications when C enters as the span of k constraint vectors. An often used constraint is the full investment condition, $w^\top e = 1$, where $e = (1, 1, 1, \dots, 1)^\top$. Since β will typically have positive mean in equity applications, we obtain

$$\cos \angle(b, C) \geq \langle b, e/|e| \rangle = \frac{1}{|\beta|\sqrt{p}} \sum \beta_i = \frac{\sqrt{p}}{|\beta|} \left(\frac{1}{p} \sum \beta_i \right) > 0 \quad (40)$$

asymptotically, and so $\Theta < \pi/2$ and JSE will improve on PCA by (36).

5.2 Estimating Constrained Minimum Variance

We return to the optimization problem (1),

$$\begin{aligned} \min_w \quad & \frac{1}{2} w^\top \Sigma w \\ \text{subject to} \quad & C^\top w = a, \end{aligned} \quad (41)$$

introduced in Section 1, where we have written the constraints in matrix notation. The columns of the $p \times k$ matrix C are the k constraint vectors C_1, \dots, C_k and $a = (a_1, \dots, a_k) \in \mathbb{R}^k$ is the non-zero vector of constraint values, fixed for all p . As before, the symbol $w = w^{(p)} \in \mathbb{R}^p$ is a vector of weights defining the portfolio holdings.

We apply the results in Section 5.1 to estimate a $p \times p$ covariance matrix $\Sigma = \Sigma^{\text{JSE}}$ for use in (41). The matrix Σ^{JSE} depends on the constraint matrix C ; its core is h^{JSE} , the leading eigenvector of the sample covariance matrix, shrunk by a prescribed amount in the direction of C . To avoid visual clutter, we suppress the dependence of Σ^{JSE} and h^{JSE} on C when possible, but the dependence of Σ^{JSE} on C is a central idea of this section.

When data follow the factor model (14), the solutions to (41) with the estimates Σ^{JSE} form a sequence (in p) of optimized portfolios w^{JSE} whose true variance tends to 0 almost surely as p tends to infinity, a property shared by the optimal solution to (41).

5.2.1 Constraints

We assume without loss of generality that the constraint matrix C has full rank, and the entries of a are non-negative, with at least one positive entry.

We are interested in asymptotic estimation of the constrained minimum variance as p tends to infinity with the number k of constraints fixed. When it is required for clarity, dependence on p is indicated with a superscript. To engage the theory of the previous sections, we accept the standing assumptions A1 - A3 on the underlying factor model described there. In addition, we wish to avoid degeneracy of the constraints $C^\top w = a$ in the asymptotic limit, so we add the following two natural assumptions:

- A4. For each $j = 1, \dots, k$, the columns $C_j^{(p)}$ of $C^{(p)} \in \mathbb{R}^{p \times k}$ satisfy:
- $\sup_{p \geq 1} |C_j^{(p)}|_\infty < \infty$, where $|\cdot|_\infty$ denotes the maximum norm, and
 - the sequence $|C_j^{(p)}|^2/p$ tends to a positive finite limit as $p \rightarrow \infty$.
- A5. The constraint matrix C does not become singular in the high dimensional limit:

$$\liminf_{p \rightarrow \infty} \det(C^\top C)/p^k > 0. \quad (42)$$

Assumption A4 is similar to A3, and says that the average squared entry of the columns doesn't tend to zero or infinity with p . Assumptions A4 and A5 mean that the angle between any two columns of C is bounded above zero, and the singular values of C are bounded above and below by positive constants times p .

The simplest example is the case of the fully invested portfolio, where $k = 1$, there is a single constraint $e^\top w = 1$, where e is the column of 1's, and C is the column matrix e . Since $|e|^2 = p$, A4 is satisfied; $C^\top C$ is equal to the 1×1 matrix with determinant p , so A5 is satisfied.

5.2.2 Estimating Σ^{JSE}

The constraint matrix C and vector of constraint values a in the optimization (41) are known to the user, but the covariance matrix Σ must be estimated. When data follow the one factor model (11), the population covariance matrix Σ takes the form specified in (15):

$$\Sigma = \eta^2 bb^\top + \delta^2 I. \quad (43)$$

As a consequence of this structure, an estimate of Σ amounts to estimates of two positive scalars, η^2 and δ^2 , and a unit-length p -vector b . The estimates we develop are in terms of the sample covariance matrix S of n observed returns to p securities. We build our estimates from the trace of S , $\text{tr}(S)$, the leading eigenvalue λ^2 of S , and its corresponding leading eigenvector h .

Our estimates of η^2 and δ^2 are guided, under our standing assumptions, by the relationships between the eigenvalues of S and the factor model structure in the HL regime. These relationships have been identified in numerous sources; as described in Lemma 4 below, they are summarized by the almost sure limits

$$\lim_{p \rightarrow \infty} (\lambda^2 - \ell^2)/p = \sigma^2 B^2 |X|^2/n \quad (44)$$

and

$$\lim_{p \rightarrow \infty} \ell^2/p = \delta^2/n. \quad (45)$$

Recall from assumption A3'.b that $\eta^2/p \rightarrow \sigma^2 B^2$ as $p \rightarrow \infty$, and, while X itself is not observed, we know $E[|X|^2/n] = 1$. Therefore we estimate η^2 with $\lambda^2 - \ell^2$. Noting (45), we estimate δ^2 with $n\ell^2/p$. Both λ^2 and ℓ^2 are observable from the eigenvalues of the sample covariance matrix S . We therefore have an

true parameter	estimate(s)
η^2	$\lambda^2 - \ell^2$
δ^2	$n\ell^2/p$
b	v, h, h^{JSE}

Table 1: Parameters of a covariance matrix in a one-factor model.

estimated covariance matrix, depending on the choice of unit vector v , of the form

$$\Sigma^v = (\lambda^2 - \ell^2)vv^\top + (n/p)\ell^2 I. \quad (46)$$

It remains to specify an estimator v of b . We examine two competing estimates of Σ^v : Σ^{PCA} and Σ^{JSE} obtained by setting v to h and h^{JSE} , respectively. These estimates differ only in the leading eigenvector. A summary of our parameter estimates is in Table 1.

5.2.3 Variance and the optimization bias

For any choice of principal unit eigenvector v , let w^v denote the unique minimizer of $w^\top \Sigma^v w$ subject to the known constraint $C^\top w = a$. We are interested in the true variance $\mathcal{V}^v = (w^v)^\top \Sigma w^v$ of the optimized portfolio w^v .

The unique solution w^v is obtained via the first order conditions for the Lagrangian

$$L(w, \Lambda) = (1/2)w^\top \Sigma^v w + (a^\top - w^\top C)\Lambda, \quad (47)$$

where $\Lambda \in \mathbb{R}^k$ is the vector of Lagrange multipliers (“shadow prices”). We have

$$\Lambda^v = (C^\top (\Sigma^v)^{-1} C)^{-1} a \quad (48)$$

$$w^v = (\Sigma^v)^{-1} C \Lambda^v = (\Sigma^v)^{-1} C (C^\top (\Sigma^v)^{-1} C)^{-1} a. \quad (49)$$

We use the notation $\angle(v, C)$ to denote the angle between v and $\text{col}(C)$, $\cos(v, C)$ to denote the cosine of that angle, and similarly for other trigonometric functions of the angle.

Since C has rank k , the $k \times k$ matrix $C^\top C$ is invertible, so we may define the $k \times p$ pseudo-inverse C^\dagger by $(C^\dagger)^\top = C(C^\top C)^{-1}$, also of full rank. Therefore $(C^\dagger)^\top a$ is nonzero whenever $a \in \mathbb{R}^k$ is nonzero.

Definition 1 For any nonzero $a \in \mathbb{R}^k$ and unit vector $v \in \mathbb{R}^p$ satisfying

$$|v_C| = \cos(v, C) < 1, \quad (50)$$

define the unit vector

$$\alpha = \frac{(C^\dagger)^\top a}{|(C^\dagger)^\top a|} \quad (51)$$

and define the *optimization bias* associated to v , C , and a by

$$\mathcal{E}_p(v, C, a) = \frac{\langle b, \alpha \rangle (1 - |v_C|^2) - \langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 - |v_C|^2}, \quad (52)$$

where, as usual, b denotes the leading population unit eigenvector.

The optimization bias does not depend on the magnitude of a , but only on α and the subspace $\text{col}(C)$, and is equal to zero when $v = b$:

$$\mathcal{E}(b, C, a) = 0. \quad (53)$$

As described below, the optimization bias represents a measure of the variance error when v is used in place of the true principal eigenvector b .

In the simplest example of the fully invested portfolio, $k = 1$, $a = 1$ and C is the column vector e of ones, so that $e^\top w = 1$. If we choose $v = h$, the leading sample eigenvector, a computation shows

$$\mathcal{E}_p(h, e, 1) = \frac{\langle b, e/|e| \rangle - \langle b, h \rangle \langle h, e/|e| \rangle}{1 - \langle h, e/|e| \rangle^2}, \quad (54)$$

which agrees with the optimization bias originally introduced for this case in Goldberg et al. [20].

Proposition 3 *Let C, h be as above and let h_C denote the orthogonal projection of h onto C . If $0 < \Theta < \pi/2$, then*

$$\limsup_{p \rightarrow \infty} |h_C| < 1 \quad (55)$$

and

$$\limsup_{p \rightarrow \infty} |(h^{\text{JSE}})_C| < 1. \quad (56)$$

Theorem 3 *Under assumptions A1-A5 above, let $v \in \mathbb{R}^p$ be a unit vector for each p and satisfying*

$$\limsup_{p \rightarrow \infty} |v_C| < 1. \quad (57)$$

Then, for n, k fixed,

$$0 < \limsup_{p \rightarrow \infty} \eta^2 |(C^\dagger)^\top a|^2 < \infty, \quad (58)$$

and the true variance $\mathcal{V}(w^v)$ of the estimated portfolio w^v is

$$\mathcal{V}(w^v) \equiv (w^v)^\top \Sigma w^v = \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p) \quad (59)$$

asymptotically as $p \rightarrow \infty$.

Because of Proposition 3, Theorem 3 applies to both $v = h$ and $v = h^{\text{JSE}}$. When $v = b$, the optimization bias is zero and the true minimum variance is asymptotically $O(1/p)$. Otherwise, the limiting value of the optimization bias \mathcal{E}_p^2 controls the large- p variance of the estimated portfolio.

The next theorem states that Σ^{JSE} dominates Σ^{PCA} as measured by the value of the true variance of the estimated portfolios w^{JSE} and w^{PCA} .

Theorem 4 *Under assumptions A1-A5 above, suppose also that the angle between b and $\text{col}(C)$ is asymptotically between 0 and $\pi/2$.*

In addition, assume (by passing to a subsequence if needed) that

$$\lim_{p \rightarrow \infty} \cos(\angle(b, (C^\dagger)^\top a)) = \lim_{p \rightarrow \infty} \langle b, \alpha \rangle \equiv \langle b, \alpha \rangle_\infty \quad \text{exists.} \quad (60)$$

Then, almost surely,

$$\lim_{p \rightarrow \infty} \mathcal{E}_p(h^{\text{JSE}}, C, a)^2 = 0. \quad (61)$$

Moreover, if $\langle b, \alpha \rangle_\infty^2 > 0$, then

$$\lim_{p \rightarrow \infty} \mathcal{E}_p(h, C, a)^2 > 0. \quad (62)$$

Consequently, if $\langle b, \alpha \rangle_\infty^2 > 0$, the true variance ratio

$$\frac{\mathcal{V}(w^{\text{JSE}})}{\mathcal{V}(w^{\text{PCA}})} \quad (63)$$

tends to zero asymptotically.

The previous two theorems tell us that $\mathcal{V}(w^{\text{TRUE}})$ and $\mathcal{V}(w^{\text{JSE}})$ tend to zero as $p \rightarrow \infty$, but $\mathcal{V}(w^{\text{PCA}})$ usually has a positive limit. This means the variance of w^{PCA} is an arbitrarily large factor greater than the optimal variance as p grows. The following lemma shows that the condition $\langle b, \alpha \rangle_\infty \neq 0$ will typically be satisfied when the vector a is unrelated to the other problem parameters.

Lemma 1 *Assume A1-A5 and that the limiting angle Θ is less than $\pi/2$. Suppose a does not belong to the orthogonal complement of the unit vector*

$$\lim_{p \rightarrow \infty} \frac{C^\dagger b}{|C^\dagger b|} \in \mathbb{R}^k. \quad (64)$$

Then $\langle b, \alpha \rangle_\infty$ is not zero.

6 Numerical examples

In this section, we describe the results of simulation experiments supporting the results stated above. First, we illustrate (36), which asserts the stochastic dominance of the improvement of h^{JSE} over h^{PCA} as an estimate of the leading population eigenvector. Then we illustrate the assertion that the ratio of variances of portfolios w^{JSE} and w^{PCA} tends to zero asymptotically almost surely.

These experiments serve two purposes. The first is to show that the asymptotic properties described in the theorems, such as equation (36), are approximately realized when the dimension p has realistic values much less than infinity. The results reported here are for $p = 3000$, but we have observed similar outcomes for p as low as 40.

parameter	value(s)	description
$\cos \theta$	0.969, 0.707, 0.174	cosine of the angle between β and $\mathbf{1}$
β^*	$N(\cos \theta, \sin^2 \theta)$	factor loadings
σ	0.16	annualized factor volatility
δ	0.60	annualized specific volatility
f	$N(0, \sigma^2)$	factor return
z	mean 0, st dev δ	specific return
$\cos \Theta$	0.97, 0.75, 0.49	cosine of the angle between β and C
p	3000	number of securities
n	24	number of observations
k	2	number of constraints
μ	$0.01(\beta + N(0.5, 2))$	3000×1 vector of expected returns
C	$(\mathbf{1}, \mu)$	3000×2 matrix of constraint vectors
m	0.01	monthly expected target return
a	$(1, m)^\top$	constraint target vector

Table 2: Simulation parameters

Second, the variance experiments described in Section 6.3 illustrate the observed strength of the effect of JSE on the variance ratio for this particular choice of parameters. Since we do not have theoretical results about the asymptotic rate of convergence of the true variance ratio, these experiments confirm that JSE can be of material use in at least some reasonable circumstances for a realistic choice of dimension.

6.1 Calibration

We specify the parameters of the return generating process (11), repeated here for convenience,

$$r = \beta f + z, \quad (65)$$

the $p \times k$ matrix of constraint vectors C and k vector of constraint targets a .

We construct β so that the angle θ with $\mathbf{1}$ is a prescribed value and $|\beta|^2/p = 1$. First draw the components of a vector β^* from the normal distribution $N(\cos \theta, \sin^2 \theta)$. Let $m = m(\beta^*)$ be the realized mean of the entries of β^* , and $s = s(\beta^*)$ the realized standard deviation. Define

$$c_1 = \frac{\sin \theta}{s} \quad \text{and} \quad c_2 = \cos \theta - \frac{\sin \theta}{s} m, \quad (66)$$

and let

$$\beta = c_1 \beta^* + c_2 \mathbf{1}. \quad (67)$$

Making use of the identity

$$|\beta|^2 = p(m(\beta)^2 + s(\beta)^2), \quad (68)$$

a calculation shows that $|\beta|^2/p = 1$ and the angle between β and $\mathbf{1}$ is exactly θ . Even though the factor loadings β are deterministic in our model, we specify

them by drawing from a normal distribution as described next. The calibration of the factor model generating returns is completed by setting the factor return f to be normally distributed with mean 0 and annualized standard deviation σ to be 16%, and specific return z to be normally distributed with mean 0 and annualized standard deviation δ to be 60%. The observed qualitative results do not depend on the choice of normal distribution for specific returns; we observe similar outcomes for heavier-tailed specific returns, including double exponential and student- t distributions.

Next, we construct an expected return vector μ so that

$$\mu_i = \beta_i + \epsilon_i$$

where ϵ_i is drawn from a normal distribution with mean 0.5 and variance 2.0, $N(0.5, 2.0)$. Thus, securities with higher betas tend to have higher expected returns. The target expected return is $m = 0.01$.

The two-dimensional shrinkage target C is the span of p -vectors μ and $\mathbf{1}$. The angle Θ between β and C is determined by the specification of β and μ . The 2-vector of constraints targets is $a = (1, m)^\top$.

Simulation parameters are listed in Table 2.

6.2 Stochastic dominance of h^{JSE} over h^{PCA}

Under Standing Assumptions A1–A3, formula (36) provides an exact expression for the difference between the squared cosines of θ^{PCA} and θ^{JSE} :

$$\cos^2(\theta^{\text{JSE}}) - \cos^2(\theta^{\text{PCA}}) = \left(\frac{1}{\phi_\infty^2 + 1} \right) \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1}. \quad (69)$$

This *magic formula* for the limiting difference between angles $\angle(\beta, h^{\text{PCA}})$ and $\angle(\beta, h^{\text{JSE}})$ as $p \rightarrow \infty$ is positive almost surely when $\Theta < \pi/2$. It is expressed in terms of two quantities: the angle between the leading eigenvector and the shrinkage target, $\Theta = \angle(\beta, C)$, and the relative eigengap ϕ^2 .

How well does the asymptotic guidance provided by the magic formula work for finite p ? For $p = 3000$, we report

$$\cos^2(\angle(h^{\text{JSE}}, b)) - \cos^2(\angle(h^{\text{PCA}}, b))$$

as well as the asymptotic limit of that difference as p tends to infinity, given by the magic formula. The results of 10,000 simulations are shown in Figure 1 for small, medium and large angles, $\cos(\Theta) = 0.969, 0.707$ and 0.174 .

In all 10,000 simulations, the improvement was positive, and it declined as the angle Θ increased. This is consistent with the asymptotic guidance given by the magic formula, which is decreasing in Θ .

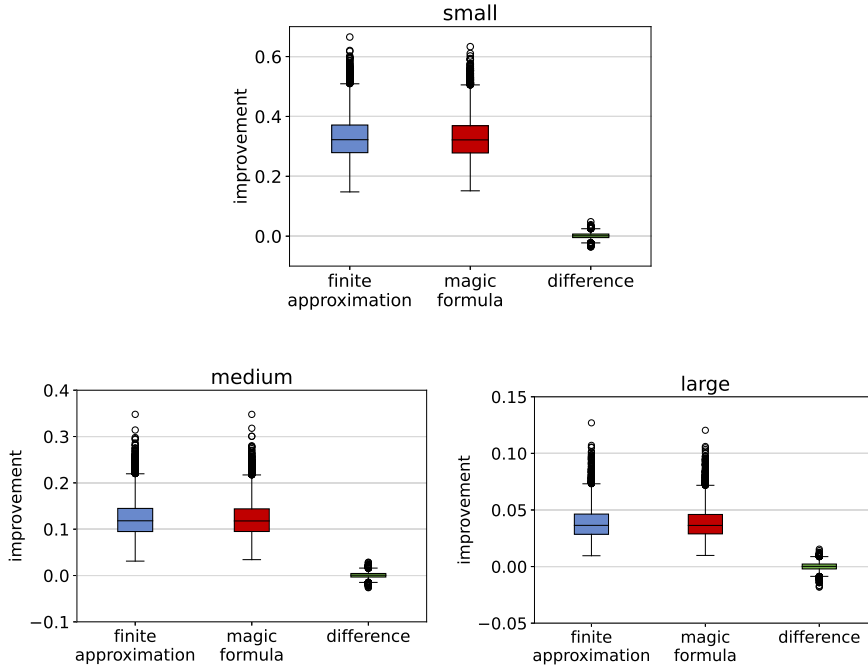


Fig. 1: Boxplots for $p = 3000$ of 10,000 simulations of the difference between $\cos^2(\angle(h^{\text{PCA}}, b))$ and $\cos^2(\angle(h^{\text{JSE}}, b))$ (finite difference), the asymptotic limit of this difference (magic formula) as well as the path-by-path difference between them (difference). The small, medium and large panels correspond to $\cos \Theta = 0.969, 0.707$ and 0.174 , Return data follow (11) with parameters specified in Table 2.

6.3 Stochastic dominance of w^{JSE} over w^{PCA}

We report ratios of variances of portfolios w^{PCA} , w^{JSE} and w^{TRUE} optimized with (1) where Σ is set to Σ^{PCA} , Σ^{JSE} and $\Sigma^{\text{TRUE}} = \Sigma$, the true (population) covariance matrix. The portfolio w^{TRUE} and covariance matrix Σ^{TRUE} are independent of state.

The blue and red boxplots in Figure 2 illustrate the variance comparison of PCA and JSE portfolios: those estimated using Σ^{JSE} have substantially lower true variance for small and medium angles between b and C . As expected, improvement is best when the angle between b and C is small, and declines as this angle increases toward $\pi/2$. (In the limit where b is orthogonal to C , we expect no improvement.)

These results are displayed for $p = 3000$; they are consistent with the asymptotic guarantees that $\mathcal{V}(w^{\text{JSE}})/\mathcal{V}(w^{\text{PCA}})$ and $\mathcal{V}(w^{\text{TRUE}})/\mathcal{V}(w^{\text{PCA}})$ tend to 0 almost surely as p tends to infinity.

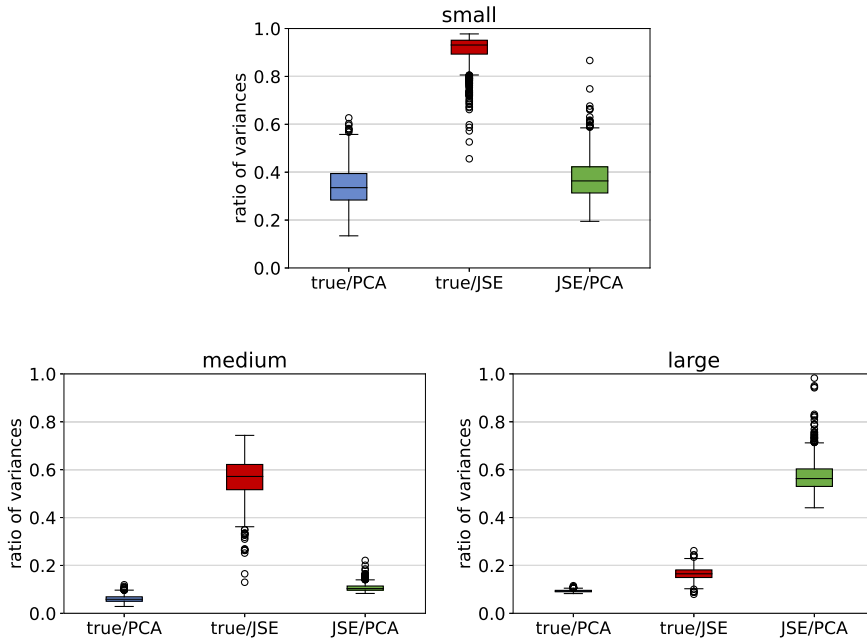


Fig. 2: Boxplots for 10,000 simulations of ratios of variances of optimized and optimal portfolios, w^{PCA} , w^{JSE} and w^{TRUE} , for $p = 3000$. The small, medium and large panels correspond to $\cos \Theta = 0.969, 0.707$ and 0.174 . The expected return target is $m = 0.01$. Return data follow (11) with parameters specified in Table 2.

The asymptotic behavior of $\mathcal{V}(w^{\text{TRUE}})/\mathcal{V}(w^{\text{JSE}})$ is not known theoretically, but related experiments suggest it may be close to 1 when the angle Θ between b and C is small.

7 Conclusion

In this paper, we extend the literature on James-Stein for eigenvectors (JSE), a data driven method for improving the accuracy of a high-dimensional, noisy leading sample eigenvector. For a spiked factor model, prior work guarantees that JSE shrinkage toward a one-dimensional target improves on the leading sample eigenvector as an estimate of ground truth. We show those guarantees persist when we shrink toward a target of dimension greater than one. This generalization greatly enlarges the range of applications of JSE, which can now be used to be build strategy-specific covariance matrices suitable for quadratic optimization with any number of linear constraints. We provide easy-to-code formulas for these covariance matrices as well as a theoretical

guarantee that they lead to relatively low-variance solutions to the optimization. The connection between JSE and the variance of optimized portfolios is via the optimization bias, which was formulated for minimum variance in earlier work and extended to take account of arbitrary linear constraints in this article. The optimization bias asymptotically controls the variance of optimized portfolios, and it tends to zero as the number of securities tends to infinity under JSE optimization.

Also new in this article is a formula for the degree of improvement of JSE over the sample leading eigenvector. The formula depends only on sample eigenvalues and the angle between the leading population eigenvector and the target subspace. Simulations suggest that the asymptotic guarantees apply in situations of practical relevance.

Our research opens a range of intriguing possibilities and questions. These include the use of JSE to generate low-variance solutions to quadratic optimization in a multi-factor setting, which has been shown effective in numerical experiments. Another direction forward is to pursue the theoretical connections between JSE and concentration of measure in high dimensional spheres, understanding of which may provide new, deeper perspectives on these powerful and often counter-intuitive results.

Acknowledgements We are grateful to Jeongyoun Ahn, Stjepan Begušić, Haim Bar, Sungku Jung, Youhong Lee and Alex Shkolnik for support. We thank Richard Michaud and Alex Ulitsky for helpful comments on an early draft of this article.

8 Appendix: Proofs

We provide mathematical arguments verifying the results stated above. We begin with some preliminary results needed for the subsequent proofs.

Lemma 2 (Triangular Strong Law of Large Numbers, Tao [41])

Let $(X_{i,p})_{i,p \in \mathbb{N}, i \leq p}$ be a triangular array of scalar random variables such that for each p , the row $X_{1,p}, \dots, X_{p,p}$ is a collection of independent random variables. For each p , define the partial sum $S_p = X_{1,p} + \dots + X_{p,p}$. Assume all the $X_{i,p}$ have mean μ .

- If $\sup_{i,p} E|X_{i,p}|^2 < \infty$, then S_p/p converges in probability to μ .
- If $\sup_{i,p} E|X_{i,p}|^4 < \infty$, then S_p/p converges almost surely to μ .

Lemma 3 Let $\{z_i : i \in \mathbb{N}\}$ be a sequence of independent mean-zero random variables with bounded fourth moments, and let $Z^{(p)} = (z_1, \dots, z_p) \in \mathbb{R}^p$ for each p . Let $u^{(p)} \in \mathbb{R}^p$, $p \in \mathbb{N}$, be a sequence of unit vectors and let $u_i^{(p)}$ denote the i -th coordinate of $u^{(p)}$.

Assume

$$\sup_{i,p} \{p|u_i^{(p)}|^2 : i = 1, \dots, p; p \in \mathbb{N}\} < \infty. \quad (70)$$

Then

$$\frac{u^{(p)\top} Z^{(p)}}{\sqrt{p}} \rightarrow 0 \quad (71)$$

almost surely as $p \rightarrow \infty$.

Proof of Lemma 3. Apply Lemma 2 with $X_{i,p} = \sqrt{p}u_i^{(p)}z_i$ and $S_p = X_{1,p} + \dots + X_{p,p}$. By the assumptions, the $X_{i,p}$ have mean zero and uniformly bounded fourth moments. By Lemma 2 with $\mu = 0$,

$$\frac{u^{(p)\top} Z^{(p)}}{\sqrt{p}} = \frac{1}{p} S_p \quad (72)$$

converges to zero almost surely. \square

The following is a version of Proposition 5.2 in Gurdogan and Kercheval [21], which remains true with a similar proof under our slightly adapted hypotheses:

Proposition 4 *Under assumptions A1 - A3, let $L = L_p \subset \mathbb{R}^p$ be a sequence of linear subspaces with constant dimension and independent of the random variables z . Then*

1. $\lim_{p \rightarrow \infty} (\langle h, h_L \rangle - \langle h, b \rangle^2 \langle b, b_L \rangle) = 0$,
2. $\lim_{p \rightarrow \infty} (\langle b, h_L \rangle - \langle h, b \rangle \langle b, b_L \rangle) = 0$, and
3. $\lim_{p \rightarrow \infty} |h_L - \langle h, b \rangle b_L| = 0$.

In particular, part 3 implies that $\angle(h_L, b_L) \rightarrow 0$ as $p \rightarrow \infty$.

8.1 Proof of Proposition 1

Proposition 1. *Under assumptions A1 - A3, the limits*

$$\theta^{\text{PCA}} = \lim_{p \rightarrow \infty} \angle(h, b) \quad \text{and} \quad \psi_\infty^2 = \lim_{p \rightarrow \infty} \psi_p^2 \quad (73)$$

exist almost surely, and

$$\cos \theta^{\text{PCA}} = \psi_\infty \in (0, 1). \quad (74)$$

Recall that we have the sample covariance matrix $S = YY^\top/n$ with unit leading eigenvector h , choosing the sign so that $\langle h, b \rangle > 0$, and leading eigenvalue λ^2 .

Define $\chi = \chi_p \in \mathbb{R}^n$ such that h and χ are the left and right singular vectors of Y/\sqrt{n} , respectively, with singular value $\lambda > 0$. We take $|\chi| = 1$ and specify the sign of χ so that $(\chi, X) > 0$. The vector $X \in \mathbb{R}^n$ does not depend on p , and for simplicity in the notation we suppress the dependence of $h, b, \lambda, \chi, Z, Y$ on p .

Since h , χ , and Y are related by

$$\lambda h = Y\chi/\sqrt{n}, \quad (75)$$

by equation (20) we have

$$\lambda h = \frac{\eta b X^\top \chi + Z\chi}{\sqrt{n}}. \quad (76)$$

Taking the dot product of both sides with b and $\lambda h/p$ yields the following identities:

$$\langle h, b \rangle = \left(\frac{\eta X^\top \chi}{\lambda \sqrt{n}} \right) + \left(\frac{b^\top Z}{\sqrt{p}} \right) \left(\frac{\chi \sqrt{p}}{\lambda \sqrt{n}} \right), \quad (77)$$

$$\lambda^2/p = \frac{\eta^2 (X^\top \chi)^2}{np} + \frac{\chi^\top Z^\top Z \chi}{np} + 2(X^\top \chi) \left(\frac{b^\top Z}{\sqrt{p}} \right) \left(\frac{\eta \chi}{n \sqrt{p}} \right). \quad (78)$$

Applying the independence and distributional assumptions on Z with the strong law of large numbers, we may deduce that $Z^\top Z/p$ tends almost surely to $\delta^2 I$ as $p \rightarrow \infty$. This fact, the boundedness of η^2/p , and Lemma 3 applied to b in equation (78) show that λ^2/p is eventually bounded between zero and infinity almost surely. Applying Lemma 3 to b in the last term of equation (77), we obtain

$$\langle h, b \rangle_\infty = \lim_{p \rightarrow \infty} \left(\frac{\eta X^\top \chi}{\lambda \sqrt{n}} \right) \quad (79)$$

provided the limit in (79) exists almost surely.

Recall ℓ_p^2 is the average of the non-zero sample eigenvalues less than λ^2 . The proof of the following Lemma is essentially identical to the proof of Lemma A.2 of Goldberg, et. al.[20]:

Lemma 4 *Under assumptions A1 - A3 and notation as above, we have the following limits almost surely:*

$$\lim_{p \rightarrow \infty} \lambda^2/p = \sigma^2 B^2 |X|^2/n + \delta^2/n, \quad (80)$$

$$\lim_{p \rightarrow \infty} \chi_p = X/|X|, \quad \text{and} \quad (81)$$

$$\lim_{p \rightarrow \infty} \ell_p^2/p = \delta^2/n. \quad (82)$$

Applying Lemma 4 to (79), we obtain

$$\langle h, b \rangle_\infty = \lim_{p \rightarrow \infty} \frac{\eta X^\top \chi}{\lambda \sqrt{n}} = \lim_{p \rightarrow \infty} \left(\frac{\eta}{\lambda} \right) \left(\frac{\sqrt{p}}{\lambda} \right) \left(\frac{X^\top \chi}{\sqrt{n}} \right) \quad (83)$$

$$= \sigma B \left(\frac{1}{\sqrt{\sigma^2 B^2 |X|^2/n + \delta^2/n}} \right) \left(\frac{|X|}{\sqrt{n}} \right) \quad (84)$$

$$= \sqrt{\frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2}} \in (0, 1). \quad (85)$$

By Lemma 4,

$$\psi_p^2 = \frac{\lambda^2 - \ell_p^2}{\lambda^2}, \quad (86)$$

converges to

$$\psi_\infty^2 = \frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2} \quad (87)$$

and hence $\langle h, b \rangle_\infty = \psi_\infty$. This completes the proof of Proposition 1.

8.2 Proof of Theorem 1

Theorem 1. *Suppose the angle $\angle(b, C)$ between b and C tends, as $p \rightarrow \infty$, to a limit*

$$\Theta = \lim_{p \rightarrow \infty} \angle(b, C). \quad (88)$$

Then under assumptions A1 – A3, the limit

$$\Theta^h = \lim_{p \rightarrow \infty} \angle(h, C) \quad (89)$$

exists almost surely, and

$$\cos \Theta^h = \cos \theta^{\text{PCA}} \cdot \cos \Theta = \psi_\infty \cdot \cos \Theta. \quad (90)$$

In particular, if $0 < \Theta < \pi/2$, then

$$0 < \cos \Theta^h < \cos \theta^{\text{PCA}} \quad (91)$$

and

$$0 < \cos \Theta^h < \cos \Theta. \quad (92)$$

Proof. We apply Proposition 4(1) with $L = C$, noting that $\langle h, h_C \rangle = \cos \angle(h, C)$ and $\langle b, b_C \rangle = \cos \angle(b, C)$. Since $\langle h, b \rangle \rightarrow \psi_\infty$ from Proposition 1 and $\cos \angle(b, C) \rightarrow \cos \Theta$ by hypothesis, equation (90) follows immediately. \square

8.3 Proof of Theorem 2

Theorem 2. *With notation as above, suppose the limit*

$$\Theta = \lim_{p \rightarrow \infty} \angle(b, C) \quad (93)$$

exists.

Then, under the standing assumptions A1 – A3, the limits

$$\theta^{\text{JSE}} = \lim_{p \rightarrow \infty} \angle(h^{\text{JSE}}, \beta) \quad \text{and} \quad \theta^{\text{PCA}} = \lim_{p \rightarrow \infty} \angle(h^{\text{PCA}}, \beta) \quad (94)$$

exist almost surely, and the asymptotic improvement of h^{JSE} over h^{PCA} as an estimate of the leading population eigenvector is

$$\cos^2(\theta^{\text{JSE}}) - \cos^2(\theta^{\text{PCA}}) = \left(\frac{1}{\phi_\infty^2 + 1} \right) \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1}. \quad (95)$$

If $\Theta = \pi/2$, then h^{JSE} converges to h^{PCA} , $\theta^{\text{JSE}} = \theta^{\text{PCA}}$ and there is no improvement, while if $\Theta = 0$ then h^{JSE} converges to b . In other cases, $\theta^{\text{JSE}} < \theta^{\text{PCA}}$ almost surely, with the improvement given by (36).

Proof. The existence of the limit θ^{PCA} has already been established in Proposition 1. The JSE estimator h^{JSE} relative to the subspace C is an example of the ‘‘MAPS’’ estimator defined and studied in Gurdogan and Kercheval [22]. We make further use of some results in that paper, first defining for each p , the oracle estimator $h^\circ = h^\circ(C)$ as follows. Let

$$U = \text{span}(h, C),$$

and define the unit vector

$$h^\circ = \frac{b_U}{|b_U|}. \quad (96)$$

The oracle h° is the normalized orthogonal projection of b onto the linear subspace spanned by h and C . We use the name ‘‘oracle’’ because, unlike h^{JSE} , it is not observable from the data, but requires knowledge of b , precisely the quantity we are trying to estimate.

The proof of the following proposition is a simpler version of the proof of Theorem 5.1 of Gurdogan and Kercheval [22], for slightly adjusted assumptions:

Proposition 5 *Under assumptions A1 - A3,*

$$\lim_{p \rightarrow \infty} |h^\circ - h^{\text{JSE}}| = 0 \quad (97)$$

almost surely.

Next, let

$$u = \frac{h - h_C}{\|h - h_C\|}$$

Then $U \equiv \text{span}(h, C) = \text{span}(u, C)$ and u is a unit vector orthogonal to C (assuming, with probability one, that h does not belong to C). Hence

$$b_U = b_C + \langle b, u \rangle u,$$

and so

$$\langle h^\circ, b \rangle^2 = \left\langle \frac{b_U}{|b_U|}, b \right\rangle^2 = |b_U|^2 \quad (98)$$

$$= |b_C|^2 + \langle u, b \rangle^2 \quad (99)$$

$$= |b_C|^2 + \frac{(\langle h, b \rangle - \langle h_C, b \rangle)^2}{1 - |h_C|^2}. \quad (100)$$

All the terms in the right hand side have previously been show to have limits as $p \rightarrow \infty$:

$$|b_C|^2 \rightarrow \cos^2 \Theta, \quad (101)$$

$$|h_C|^2 \rightarrow \psi_\infty^2 \cos^2 \Theta, \quad (102)$$

$$\langle h, b \rangle \rightarrow \psi_\infty = \cos \theta^{\text{PCA}}, \quad (103)$$

$$\langle h_C, b \rangle \rightarrow \psi_\infty \cos^2 \Theta. \quad (104)$$

Therefore $\lim_{p \rightarrow \infty} \langle h^o, b \rangle^2$ exists and by Proposition 5,

$$\lim_{p \rightarrow \infty} \langle h^o, b \rangle^2 = \lim_{p \rightarrow \infty} \langle h^{\text{JSE}}, b \rangle^2 = \cos^2 \theta^{\text{JSE}}. \quad (105)$$

Writing $\psi_\infty^2 = \psi^2$ and $\phi_\infty^2 = \phi^2$ for the remainder of this proof only, and recalling

$$\psi^2 = \frac{\phi^2}{1 + \phi^2}, \quad (106)$$

in the limit,

$$\cos^2 \theta^{\text{JSE}} - \cos^2 \theta^{\text{PCA}} = \cos^2 \Theta + \frac{\psi^2(1 - \cos^2 \Theta)^2}{1 - \psi^2 \cos^2 \Theta} - \psi^2 \quad (107)$$

$$= (1 - \psi^2)^2 \frac{\cos^2 \Theta}{1 - \psi^2 \cos^2 \Theta} \quad (108)$$

$$= \left(\frac{1}{\phi^2 + 1} \right) \frac{\cos^2 \Theta}{\phi^2 \sin^2 \Theta + 1}. \quad (109)$$

This is positive when $\Theta < \pi/2$. In case $\Theta = \pi/2$, Theorem 1 implies that h_C tends to zero and h^{JSE} tends to $h = h^{\text{PCA}}$, so $\theta^{\text{JSE}} = \theta^{\text{PCA}}$ and JSE provides no improvement over PCA.

If $\Theta = 0$, it follows from equation (107) that $\theta^{\text{JSE}} = 0$ and so h^{JSE} tends to b itself.

□

8.4 Proof of Proposition 3

Proposition 3. *Let C, h be as above and h_C denote the orthogonal projection of h onto C . If $0 < \Theta < \pi/2$, then*

$$\limsup_{p \rightarrow \infty} |h_C| < 1 \quad (110)$$

and

$$\limsup_{p \rightarrow \infty} |(h^{\text{JSE}})_C| < 1. \quad (111)$$

Proof From part 3 of Proposition 4 with $L = C$, we have, in the asymptotic limit,

$$|h_C|^2 = \langle h, b \rangle_\infty^2 |b_C|^2 = \psi_\infty^2 |b_C|^2. \quad (112)$$

This establishes the first statement. For the second, it suffices to show that the angle $\angle(h^{\text{JSE}}, C)$ is positive in the limit.

We can write

$$h^{\text{JSE}} = \frac{\Gamma_p h + h_C}{|\Gamma_p h + h_C|} \quad (113)$$

where

$$\Gamma_p = \frac{\psi_p^2 - |h_C|^2}{1 - \psi_p^2}. \quad (114)$$

Since $\angle(h^{\text{JSE}}, C) = \angle(h^{\text{JSE}}, h_C)$, it suffices to show that

$$\liminf_{p \rightarrow \infty} \Gamma_p > 0. \quad (115)$$

This follows from equation (112) and the standing assumption that the angle between b and C is asymptotically strictly between 0 and $\pi/2$.

8.5 Proof of Theorem 3

As a reminder, our assumptions regarding our factor model $r = \beta f + z$ and our linear constraints $C^\top w = a$ are:

- A1. The random variable f is non-zero almost surely, and has mean zero, finite fourth moment, and variance $\sigma^2 > 0$.
- A2. The random variables $\{z_i^{(p)} : i = 1, 2, \dots, p; p > 1\}$ are i.i.d. and have mean zero, finite fourth moment, and variance $\delta^2 > 0$.
- A3. The vector sequence $\{\beta^{(p)} : p > 1\}$ satisfies the following asymptotic non-degeneracy conditions:
 - a. $\sup_{i,p} \{|\beta_i^{(p)}| : i = 1, 2, \dots, p; p > 1\} < \infty$, and
 - b. the sequence $|\beta^{(p)}|^2/p$ tends to a finite limit $B^2 > 0$ as $p \rightarrow \infty$.
- A4. The constraint matrix $C \in \mathbb{R}^{p \times k}$ satisfies the following asymptotic non-degeneracy conditions:
 - a. $\sup_p \max\{|C_{ij}| : 1 \leq i \leq p, 1 \leq j \leq k\} < \infty$, and
 - b. the sequence $|C_j|^2/p$ tends to a finite limit as $p \rightarrow \infty$, for $j = 1, \dots, k$.
- A5. the constraint matrix C does not become singular in the high dimensional limit: $\det(C^\top C)/p^k$ tends to a finite positive constant as $p \rightarrow \infty$.

Theorem 3. *Under assumptions A1-A5 above, let $v \in \mathbb{R}^p$ be a unit vector for each p and satisfying*

$$\limsup_{p \rightarrow \infty} |v_C| < 1. \quad (116)$$

Recall that w^v denotes the unique vector in \mathbb{R}^p minimizing $w^\top \Sigma^v w$ subject to the constraint $C^\top w = a$.

Then, for n, k fixed, the true variance of the estimated portfolio w^v is

$$\mathcal{V}(w^v) \equiv (w^v)^\top \Sigma w^v = \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p) \quad (117)$$

asymptotically as $p \rightarrow \infty$.

Furthermore,

$$0 < \limsup_{p \rightarrow \infty} \eta^2 |(C^\dagger)^\top a|^2 < \infty. \quad (118)$$

Proof. Recall that

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

and define

$$\kappa^2 = \frac{n\ell^2/p}{\lambda^2 - \ell^2},$$

noting that $\kappa^2 = O(1/p)$.

A computation making use of the Woodbury identity shows that

$$w^v = \left(I + \frac{(v_C - v)v^\top}{1 + \kappa^2 - |v_C|^2} \right) (C^\dagger)^\top a. \quad (119)$$

Let $C = UZV$ be the singular value decomposition of C , where V is $k \times k$ orthogonal, Z is a $k \times k$ diagonal matrix with entries equal to the singular values of C , and U is a $p \times k$ matrix with orthonormal columns. This means $(C^\dagger)^\top = UZ^{-1}V$.

Assumptions A4 and A5 imply that the squared singular values of C are bounded above and below by a constant times p . Therefore the singular values of C^\dagger are bounded above and below by a constant times $1/\sqrt{p}$. Since $\eta^2 = O(p)$, this implies

$$0 < \limsup_{p \rightarrow \infty} \eta^2 |(C^\dagger)^\top a|^2 < \infty,$$

which establishes the last assertion of the theorem.

To obtain an expression for true variance, first notice that

$$\mathcal{V}(w^v) = (w^v)^\top \Sigma w^v = \eta^2 \langle w^v, b \rangle^2 + \delta^2 |w^v|^2. \quad (120)$$

For the second term, it follows from A4 and $C^\top w^v = a$ that $|w^v|^2 \leq O(1/p)$. It remains to analyze the first term.

Making use of equation (119) and recalling

$$\alpha = (C^\dagger)^\top a / |(C^\dagger)^\top a|, \quad \limsup_{p \rightarrow \infty} |v_C| < 1, \quad \text{and} \quad \kappa^2 = O(1/p), \quad (121)$$

we have

$$\eta^2 \langle w^v, b \rangle^2 = \eta^2 |(C^\dagger)^\top a|^2 \left(\langle b, \alpha \rangle - \frac{\langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 + \kappa^2 - |v_C|^2} \right)^2 \quad (122)$$

$$= \eta^2 |(C^\dagger)^\top a|^2 \left(\frac{\langle b, \alpha \rangle (1 - |v_C|^2) - \langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 - |v_C|^2} \right)^2 + O(1/p) \quad (123)$$

$$= \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p). \quad (124)$$

□

8.6 Proof of Theorem 4

Theorem 4. *Under assumptions A1-A5 above, suppose also that the angle between b and $\text{col}(C)$ is asymptotically between 0 and $\pi/2$.*

In addition, assume (by passing to a subsequence if needed) that

$$\lim_{p \rightarrow \infty} \cos(\angle(b, (C^\dagger)^\top a)) = \lim_{p \rightarrow \infty} \langle b, \alpha \rangle \equiv \langle b, \alpha \rangle_\infty \quad \text{exists.} \quad (125)$$

Then, almost surely,

$$\lim_{p \rightarrow \infty} \mathcal{E}_p(h^{\text{JSE}}, C, a)^2 = 0. \quad (126)$$

Moreover, if $\langle b, \alpha \rangle_\infty^2 > 0$, then

$$\lim_{p \rightarrow \infty} \mathcal{E}_p(h, C, a)^2 > 0. \quad (127)$$

Consequently, if $\langle b, \alpha \rangle_\infty^2 > 0$, the true variance ratio

$$\frac{\mathcal{V}(w^{\text{JSE}})}{\mathcal{V}(w^{\text{PCA}})} \quad (128)$$

tends to zero asymptotically.

Proof. By Proposition 3, we know that

$$\limsup |v_C| < 1 \quad (129)$$

for both $v = h$ and $v = h^{\text{JSE}}$. Hence the denominator of

$$\mathcal{E}_p(v, C, a) = \frac{\langle b, \alpha \rangle (1 - |v_C|^2) - \langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 - |v_C|^2}, \quad (130)$$

stays away from zero in both cases. For the first statement (126) of the theorem, it then suffices to show that the numerator

$$\langle b, \alpha \rangle (1 - |(h^{\text{JSE}})_C|^2) - \langle b, h^{\text{JSE}} - (h^{\text{JSE}})_C \rangle \langle h^{\text{JSE}}, \alpha \rangle \quad (131)$$

vanishes asymptotically. In light of Proposition 5, it suffices to show that $\mathcal{E}_p(h^o, C, a) = 0$ for the oracle $h^o = b_U/|b_U|$ defined previously, where $U = \text{span}(h, C)$. This is a consequence of the fact that $\langle b_C, \alpha \rangle = \langle b, \alpha \rangle$ and following straightforward identities:

$$\langle b, h^o - (h^o)_C \rangle = |b_U| - \frac{|b_C|^2}{|b_U|}, \quad (132)$$

$$\langle (h^o)_C, \alpha \rangle = \frac{\langle b, \alpha \rangle}{|b_U|}, \quad \text{and} \quad (133)$$

$$|(h^o)_C|^2 = \frac{|b_C|^2}{|b_U|^2}. \quad (134)$$

Turning to the second statement (127), first note that Proposition 4 applied to the subspace $L = \text{span}(\alpha)$, implies, asymptotically, $\langle h, \alpha \rangle = \langle h, b \rangle \langle b, \alpha \rangle$, where we omit the subscripts on $\langle h, \alpha \rangle_\infty$, etc., to unclutter the notation. Also,

setting $L = C$ in the same proposition yields the asymptotic equalities $|h_C|^2 = \langle h, b \rangle^2 |b_C|^2$, and $\langle b, h_C \rangle = \langle h, b \rangle \langle b, b_C \rangle$.

Making use of these facts and simplifying leads to

$$\lim_{p \rightarrow \infty} \mathcal{E}_p(h, C, a) = \frac{\langle b, \alpha \rangle (1 - \langle h, b \rangle^2)}{1 - \langle h, b \rangle^2 |b_C|^2} \quad (135)$$

$$= \frac{\langle b, \alpha \rangle (1 - \psi_\infty^2)}{1 - \psi_\infty^2 |b_C|^2}. \quad (136)$$

When $\mathcal{E}(h, C, a)$ is positive but $\mathcal{E}(h^{\text{JSE}}, C, a)$ tends to zero, equation (117) implies that $\mathcal{V}(w^{\text{PCA}})$ remains bounded above zero while $\mathcal{V}(w^{\text{JSE}})$ tends to zero. This establishes the last claim.

□

8.7 Proof of Lemma 1

Lemma 1: *Assume A1-A5 and that the limiting angle Θ is less than $\pi/2$. Suppose a does not belong to the orthogonal complement of the unit vector*

$$\lim_{p \rightarrow \infty} \frac{C^\dagger b}{|C^\dagger b|} \in \mathbb{R}^k. \quad (137)$$

Then $\langle b, \alpha \rangle_\infty$ is not zero.

We express the singular value decomposition of C as

$$C^{(p)} = U^{(p)} Z^{(p)} V^{(p)}, \quad (138)$$

where $Z = Z^{(p)}$ is a $k \times k$ diagonal matrix with diagonal entries equal to the positive singular values s_1, s_2, \dots, s_k of C ; $V = V^{(p)}$ is $k \times k$ orthogonal, and $U = U^{(p)}$ is $p \times k$ orthonormal. Note $(C^\dagger)^\top = U Z^{-1} V$.

Assumptions A4 and A5 imply, for each j , that s_j^2/p is bounded away from zero and infinity. By taking subsequences if necessary, we may assume that $(1/\sqrt{p})Z^{(p)}$ and $V^{(p)}$ tend to $k \times k$ limits Z_∞ and V_∞ , respectively, where V_∞ is orthogonal and Z_∞ is diagonal with positive diagonal entries.

By taking a further subsequence if needed, we assume that the inner product $U^\top b$ tends to a non-zero limit $(U^\top b)_\infty \in \mathbb{R}^k$ as $p \rightarrow \infty$.

A short calculation shows

$$|(C^\dagger)^\top a|^2 = \langle Z^{-2} V a, V a \rangle \quad (139)$$

and

$$\langle b, (C^\dagger)^\top a \rangle = \langle C^\dagger b, a \rangle = \langle Z^{-1} U^\top b, V a \rangle. \quad (140)$$

Hence

$$\frac{\langle b, (C^\dagger)^\top a \rangle}{|(C^\dagger)^\top a|} = \frac{\langle Z^{-1} U^\top b, V a \rangle}{\sqrt{\langle Z^{-2} V a, V a \rangle}} \rightarrow \frac{\langle Z_\infty^{-1} (U^\top b)_\infty, V_\infty a \rangle}{\sqrt{\langle Z_\infty^{-2} V_\infty a, V_\infty a \rangle}}. \quad (141)$$

This limit is nonzero whenever a does not belong to the orthogonal complement of the non-zero vector $V_\infty^\top Z_\infty^{-1} (U^\top b)_\infty$.

□

References

1. Ahn, J., Marron, J.S., Muller, K.M., Chi, Y.Y.: The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika* **94**(3), 76–766 (2007)
2. Aoshima, M., Shen, D., Shen, H., Yata, K., Zhou, Y.H., , Marron, J.S.: A survey of high dimension low sample size asymptotics. *Australia & New Zealand Journal of Statistics* **60**(1), 4–19 (2018)
3. Ball, K.: An elementary introduction to modern convex geometry. In: *Flavors of Geometry*, no. 31 in Mathematics Sciences Research Institute Publications, pp. 1–58. Cambridge University Press (1997)
4. Bar, H., Wells, M.T.: On graphical models and convex geometry. forthcoming in *Computational Statistics and Data Analysis* (2023)
5. Bianchi, S.W., Goldberg, L.R., Rosenberg, A.: The impact of estimation error on latent factor model forecasts of portfolio risk. *The Journal of Portfolio Management* **43**(5), 147–156 (2017)
6. Chamberlain, G., Rothschild, M.: Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**(5), 1281–1304 (1983)
7. Connor, G.: A unified beta pricing theory. *Journal of Economic Theory* **34**, 13–31 (1984)
8. Connor, G.: The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal* **51**(3), 42–46 (1995)
9. Connor, G., Korajczyk, R.: Risk and return in equilibrium APT. *Journal of Financial Economics* **21**, 255–289 (1988)
10. Connor, G., Korajczyk, R.: Factor models. In: *Encyclopedia of Quantitative Finance*. Wiley (2010)
11. Connor, G., Korajczyk, R.A.: Performance measurement with the arbitrage pricing theory. *Journal of Financial Economics* **15**, 373–394 (1986)
12. Donoho, D., Gavish, M., Johnstone, I.: Optimal shrinkage of eigenvalues in the spiked covariance model. *The Annals of Statistics* **46**(4), 1742–1778 (2018)
13. Efron, B., Morris, C.: Data analysis using stein’s estimator and its generalizations. *J. of the American Statistical Assoc.* **70**(350), 311–319 (1975)
14. Efron, B., Morris, C.: Stein’s paradox in statistics. *Scientific American* **236**(5), 119–127 (1977)
15. El Karoui, N.: High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: Risk underestimation1 high-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation1. *The Annals of Statistics* **38**(6), 3487–3566 (2010)
16. El Karoui, N.: On the realized risk of high-dimensional markowitz portfolios. *SIAM Journal of Financial Mathematics* **4**, 737–783 (2013)
17. Fourdrinier, D., Strawderman, W.E., Wells, M.T.: *Shrinkage Estimation*. Springer Series in Statistics. Springer Nature, Cham, Switzerland (2018)
18. Goldberg, L.R., Kercheval, A.N.: James-Stein for the leading eigenvector. *Proceedings of the National Academy of Sciences* **120**(2) (2023)
19. Goldberg, L.R., Papacicolau, A., Shkolnik, A., Ulucam, S.: Better betas. *The Journal of Portfolio Management* **47**(1), 119–136 (2020)
20. Goldberg, L.R., Papanicolau, A., Shkolnik, A.: The dispersion bias. *SIAM Journal of Financial Mathematics* **13**(2), 521–550 (2022)
21. Gurdogan, H., Kercheval, A.: Multi anchor point shrinkage for the sample covariance matrix. Tech. rep., working paper (2021)
22. Gurdogan, H., Kercheval, A.: Multi anchor point shrinkage for the sample covariance matrix. *SIAM Journal on Financial Mathematics* **13**(3), 1112–1143 (2022)
23. Hall, P., Marron, J., Neeman, A.: Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(3), 427–444 (2005)
24. James, W., Stein, C.: Estimation with quadratic loss. In: *Proc. Fourth Berkeley Symp. Math. Stat. Prob.*, pp. 361–397 (1961)
25. Jobson, J., Korkie, B.: Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association* **75**(371), 544–554 (1980)

26. Johnstone, I.: On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics* **29**(2), 295–327 (2001)
27. Jung, S., Marron, J.: PCA consistency in high dimension, low sample size context. *The Annals of Statistics* **37**(6B), 4104–4130 (2009)
28. Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* **3**, 215–231 (1976)
29. Ledoit, O., Wolf, M.: Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* **40**(2), 1024–1060 (2012)
30. Ledoit, O., Wolf, M.: Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies* **30**(12), 4349–4388 (2017)
31. Marcenko, V., Pastur, L.: Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik* **114**(4), 507–536 (1967)
32. Markowitz, H.: Portfolio selection. *The Journal of Finance* **7**(1), 77–92 (1952)
33. Markowitz, H.M.: *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press (1959)
34. Menchero, J., Wang, J., Orr, D.: Improving risk forecasts for optimized portfolios. *Financial Analysts Journal* **68**(3), 40–50 (2012)
35. Michaud, R.O.: The Markowitz optimization enigma: Is ‘optimized’ optimal? *Financial Analysts Journal* **45**(1), 31–43 (1989)
36. Rosenberg, B.: Extra-market components of covariance in security returns. *The Journal of Financial and Quantitative Analysis* **9**(2), 263–274 (1974)
37. Ross, S.: The arbitrage theory of capital asset pricing. *Journal of Economic Theory* **13**, 341–360 (1976)
38. Sharpe, W.F.: A simplified model for portfolio analysis. *Management Science* **9**(2), 277–293 (1963)
39. Shkolnik, A.: James-Stein estimation of the first principal component. *Stat* **11**(1) (2022)
40. Stein, C.: Some problems in multivariate analysis part i. Technical report no. 6, Office of Naval Research (1956)
41. Tao, T.: 275a, notes 3: The weak and strong law of large numbers (2015). URL <https://terrytao.wordpress.com/2015/10/23/275a-notes-3-the-weak-and-strong-law-of-large-numbers/>
42. Wang, W., Fan, J.: Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics* **45**(3), 1342–1374 (2017)
43. Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics* **62**, 548–564 (1955)
44. Wigner, E.: On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics* **67**, 325–328 (1958)