Noname manuscript No.
(will be inserted by the editor)

Portfolio optimization via strategy-specific eigenvector shrinkage

In memory of Harry Markowitz, 1927–2023

Lisa R. Goldberg · Hubeyb Gurdogan · Alec Kercheval

First Draft: August 16, 2023
This Version: April 21, 2024

Abstract We estimate covariance matrices that are tailored to portfolio optimization constraints. We rely on a generalized version of James-Stein for eigenvectors (JSE), a data-driven operator that reduces estimation error in the leading sample eigenvector by shrinking toward a target subspace determined by constraint gradients. 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

In 1952, Harry Markowitz launched modern finance by framing portfolio construction as a tradeoff between risk, which he characterized as variance, and

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

expected or *mean* return. A standard tool for asset allocation, for constructing quantitative exchange traded funds, mutual funds and active strategies, and for customizing separately managed accounts, Markowitz's optimization remains the workhorse of financial services today. So-called Markowitz portfolios are efficient in the sense they minimize variance subject to a return target and constraints, and they are industry standard for asset allocation and the construction of exchange traded funds, mutual funds and some indexes.

In his early work, Markowitz considered practical challenges to implementing mean-variance optimization, including the lack of reliable algorithms, the complexity of inequality constraints required to preclude short positions, and the impact of data limitations on estimated inputs. Evidently concerned that classical statistical methods alone would not yield estimates suitable for mean-variance optimization, Markowitz [48] wrote in 1952:

> Perhaps there are ways, by combining statistical techniques and the judgment of experts, to form reasonable probability beliefs (μ_i, σ_{ij}).

This query preceded works by Eugene Wigner, Charles Stein, Volodymyr Marchenko and Leonid Pastur that launched statistical estimation in high dimensions and random matrix theory.

Since 1952, the problem of estimating suitable inputs to mean-variance optimization has been an active area of research. Prescriptions for estimates of means and covariances vary, and the nature of their errors and their impact on optimized portfolios can be obscure.

Almost universally, scholars and practitioners use factor models to reduce the number of parameters required to estimate large covariance matrices. This is consistent with empirically observed correlations in financial returns and generates estimated covariance matrices that are well-conditioned enough for use in 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 [57] 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 estimated optimized portfolios with variance that tend to be far larger than the true optimum.

We address this problem in the context of a single-factor model, which incorporates the most salient features of equity markets in simplest form. In this setting, 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, we apply a form of James-Stein shrinkage to the leading sample eigenvector, yielding a *James-Stein for eigenvectors* (JSE) estimate for the leading population eigenvector.

We advance the literature in four ways. First, we provide, novel, explicit and easy-to-code formulas for factor-based covariance matrices that are tailored to specific quadratic optimization problems with multiple linear constraints. By neutralizing the component of estimation error that is amplifed in optimization, our methods produce relatively low-variance instances of portfolios satisfying optimization constraints. This distinguishes our work from much of the literature, which focuses almost exclusively on the fully-invested (single-constraint) minimum variance portfolio. While that simple case is instructive, it fails to cover subtle and important issues that arise when multiple constraints are specified, as they are in all practical settings. Consider, for example, the minimum variance exchange traded fund, USMV. That so-called minimum variance portfolio includes benchmark-relative sector constraints, position limits and long-only constraints, in addition to the simple fully invested constraint on which the literature is largely based. As of March 15, 2024, multi-constraint USMV accounted for more $24 billion in assets. The value of the assets in the single-constraint minimum variance portfolio featured in the literature is $0.

The second advancement is a new asymptotic formula for improvement of JSE over the sample leading eigenvector that depends only on limiting ratios of sample eigenvalues and the angle between the leading population eigenvector and the constraint subspace. This novel formula characterizes JSE's asymptotic stochastic dominance over PCA, and opens the way to rate of convergence analysis that determines the utility of JSE in practical applications.

The third advancement concerns the target of JSE shrinkage. In previous studies, JSE shrinkage is toward a known fixed direction. To account for multiple constraints, we generalize the theory to accommodate a stochastic, data-dependent shrinkage target vector lying in the constraint subspace.

The fourth advancement is to extend the analysis to the more realistic case of a factor model with heterogeneous specific variances, and further to the "approximate factor model" setting in which specific returns are allowed to be correlated.

For the problems considered in this article, we show that the ideal shrinkage target vector is the orthogonal 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. Beyond a finite fourth moment, none of our theoretical results rely on parametric distributional assumptions on the underlying data.

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 3.2, while Section 4 provides a detailed mathematical discussion of the construction and describes its asymptotic properties. Numerical experiments illustrating our results are in Section 5, and Section 6 contains concluding thoughts. Mathematical proofs are in the Appendix, Section 7.

2 Financial and statistical context

The use of covariance matrices in portfolio construction dates back to work by Markowitz [48, 49] in the 1950s. Effective estimation of the high dimensional covariance matrices required by Markowitz's mean-variance optimization rests on an expansive mathematical literature and is informed by empirical and practical guidance from finance professionals. Here, we review aspects of the literature that are relevant to our results. Topics include factor models, random matrix theory, statistical consistency, and James-Stein shrinkage.

2.1 Factor models

Introduced in 1904 by Charles Spearman [61], factor models provide a framework for analyzing high dimensional data that is parsimonious and, in some cases, interpretable. When calibrated to equity markets, factor-based covariance matrices are generally well conditioned and, paradoxically, are both sufficiently stable over time and sufficiently responsive to changing market conditions for practical purposes.

In 1963, William Sharpe [58] developed the one-factor or "single index" market model whose covariance matrix is expressed as a sum of rank one and diagonal matrices. Empirical evidence of the importance of non-market factors along with issues of market non-stationarity led to Rosenberg and McKibben [56] and [55], which develops multi-factor models based on cross-sectional regressions and forms the basis of Barra's industry standard fundamental factor models. A statistical approach to factor models with roots in the Arbitrage Pricing Theory pioneered by Ross [57] and developed in Chamberlain and Rothschild [9], Connor [11], and Connor and Korajczyk [13, 15] is an antecedent of the material in this article. The strengths and weaknesses of statistical and fundamental factor models are complementary. The former respond dynamically to changing markets but can mistake noise for signal and can rely on factors that are hard to interpret. The latter are based on interpretable factors but require explicit re-architecting to incorporate new factors. Connor [12] and Connor and Korajczyk [14] review roles of different types of factor models in finance.

The results in this paper are framed in terms of latent, single-factor model, which allows for heterogenous specific variance and even mild correlations across specific returns. The focus on the single factor allows us to showcase novel estimation methods in a simple setting, while the allowance for heterogeneous variances and correlations for specific returns expand the scope of applicability of the model.

2.2 Regimes of random matrix theory

A set of methods used to contend with the scarcity of security return data comes from random matrix theory, which originated in the 1950s with the work of Wigner [68,69] and Stein [62]. In the 1960s, Marcenko and Pastur [47] characterized distributions of the eigenvalues of covariance matrices of standard Gaussian variables as the observations n and the number of parameters p tend to in infinity in proportion. This work spawned a large literature identifying and correcting biases in high dimensional eigenvalues when the number of parameters p and the number of observations n tend to infinity. We denote this asymptotic setting by HH for "high dimension high sample size" and refer to [3], Edelman and Rao [18], Bai and Silverstein [4], Tao [65], and Paul [54] for more information.

In the 2000s, Hall, Marron, and Neeman [34] and Ahn et. al [1] explored a different asymptotic framework in which the number of parameters p tends to infinity while the number of observations n stays fixed. This asymptotic regime, which we denote HL for "high dimension low sample size", is surveyed in the 2018 article by Aoshima et al. [2] and it is the setting for the present article. It is relevant to practical problems where data are limited by experimental constraints or non-stationarity of time series.

Random matrix theory overlaps with classical statistics, where asymptotic guidance is obtained by letting the number of observations n tend to infinity as the number of parameters p stays fixed, the LH regime. Results on random matrices can be organized around LH, HH and HL as discussed, for example, in Jung and Marron [40], and Goldberg and Kercheval [28]. Since any particular problem involves some specific n and p, it can be a matter of judgment or experimentation to decide which asymptotic regime provides the best guidance. The choice can be consequential since HL offers novel methods for correction of eigenvector biases, which demonstrably affect optimized quantities in simulations calibrated to financial markets.

2.3 Consistency

Sample eigenvalues and eigenvectors are used throughout the sciences to reduce the dimension of complex problems and distinguish signal from noise. The basis for this is the classical fact that sample estimates are consistent in the sense that they converge to their population counterparts as the number of independent observations tends to infinity, so long as the total dimension is fixed.

In high dimensional asymptotic regimes, the situation is more nuanced. For the HH regime, where both p and n tend to infinity, consistency of sample eigenvalues or eigenvectors can depend on the limit of $\lambda^2 n/p$, where λ^2 is a sample eigenvalue. Wang and Fan [66] show that if data are assumed sub-Gaussian, then a sample eigenvalue-eigenvector pair (λ^2, v) is a consistent estimator of its population counterpart if and only if $\lambda^2 n/p$ tends to infinity as $p \to \infty$. For example, in the case that p/n tends to a positive constant and the leading eigenvalue λ^2 is bounded in p, sample eigenvectors are inconsistent. This occurs in the spiked models developed in 2001 by Johnstone [38] and further studied by Johnstone and Lu [39] in the 2000s. For more analyses of consistency of sample eigenvalues and eigenvectors in high dimension, see Paul's 2007 article [53], the 2013 article by paper Fan, Liao, and Mincheva [24], and the 2016 article by Shen, Shen, Zhu and Marron [59].

In our HL regime, a bounded sample size n prevents consistency because the sampling error cannot be averaged out. So long as n remains bounded, there is a need for asymptotic correction of the sample eigenvector (see Theorem 1 below). This is the JSE correction, which makes use of laws of large numbers and concentration of measure (see Ball [5] and Tao [63]).

2.4 James-Stein shrinkage for averages and for eigenvectors

Shrinkage operators dampen the effects of extreme observations in data sets, which occur routinely in finance. The concept of shrinkage dates back at least to Stein [62] and James and Stein [36] in the 1950s and 1960s. They show that in dimension 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 average. This work was extended by replacing the collective average with arbitrary initial guesses in Efron and Morris [19], and popularized by Efron [20]. An overview of James-Stein type shrinkage estimation is in Foudrinier et al. [26].

Recent literature, including Shkolnik [60] and Goldberg et al. [28], develop James Stein for eigenvectors (JSE). Structurally identical 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 laws of large numbers, and therefore is free of special distributional assumptions other than boundedness of fourth moments.

2.5 Covariance matrices, extreme factors, estimation error, shrinkage and portfolio optimization

Our work has roots in two streams of literature that explain how attributes of a covariance matrix are propagated by optimization. The first considers how 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 and Bawa [41], Jobson and Korkie [37], Michaud [52] and Bianchi et al. [6]. In 2010 and 2013 articles [21] and [22] El Karoui documents how risk of optimized portfolios is underforecast by covariance matrices estimated using methods from the HH regime.

The second stream begins with Green and Hollifield's 1992 article [31], which explains how dispersion in exposures of a dominant factor can generate concentration in an optimized portfolio. In 2003 Jagannathan and Ma [35] show that this type of concentration is mitigated by imposing no-short-sale constraints, which effectively act as a shrinkage operator on a covariance matrix. In 2011, Clarke, de Silva and Thorley [10] give insightful, useful formulas for weights of long-short and long-only minimum variance portfolios when returns follow a one factor model. While estimation error is not the focus of these papers—Green and Hollifield [31] argue that estimation error is not the cause of the concentration in optimized portfolios—they are, nevertheless, foundational to a a large literature that attempts to mitigate estimation error with shrinkage.

Prompted by these developments, Ledoit and Wolf develop shrinkage-based schemes for constructing well-conditioned security return covariance matrices suitable for use in optimization. In 2003 and 2004, Ledoit and Wolf published three articles that impose structure and conditioning on an estimated covariance matrix by expressing it as a weighted sum of a sample covariance matrix and a single index matrix [42], a constant correlation matrix [44] and a scalar matrix [43]. In 2012, relying on guidance from the HH regime, Ledoit and Wolf [45] show that shrinkage of a sample covariance matrix toward a scalar amounts to linear shrinkage of sample eigenvalues toward their grand mean while preserving sample eigenvectors. They apply non-linear shrinkage to sample eigenvalues and combine the result with sample eigenvectors to generate estimated covariance matrices, which they evaluate with matrix norms. Also in 2012, Menchero, Wang and Orr [51] use guidance from the HH regime to adjust sample eigenvalues of a covariance matrix with simulation. In their 2017 article, Ledoit and Wolf [46], like many other researchers, compare realized variance and information ratios of single-constraint minimum variance portfolios constructed with different covariance matrices, some based on the non-linear shrinkage of eigenvalues from their 2012 article. In a lucid 2024 discussion of out-of-sample tests of covariance matrices developed for optimization, Menchero and Lazanas [50] argue that volatility is an appropriate out-of-sample metric, but not information ratio.

Much of the literature on high dimensional covariance matrices of financial returns relies on an empirically observed spiked structure: data suggest that one or several leading eigenvalues grow roughly in proportion to the number of securities in the pool, while the other eigenvalues stay bounded. Covariance matrix estimation for spiked models is further developed in 2011 and 2013 by Fan, Liao and Mincheva [23], [24], in 2017 by Wang and Fan [66] and in 2021 by Ding Li and Zheng [16]. In their 2018 article [8], Bodnar, Parolya and Schmid apply shrinkage to the weights of a minimum variance portfolio optimized with a sample covariance matrix. In 2021, the results are extended to include estimates of security means by Bodnar, Okhrin and Parolya [7].

The recent works by Ledoit, Wolf Fan, Liao, Mincheva, Wang, Ding, Li and Zheng and many others share several common themes. First, they apply shrinkage to estimated eigenvalues but still use the sample eigenvectors. In the language of the 2018 article by Donoho, Gavish, and Johnstone [17], these covariance matrix estimates are "orthogonally-equivariant." Ledoit and Wolf call them "rotationally equivariant." With the exception of Wang and Fan [66], these articles rely on the HH regime. In all cases these models are tested on single-constraint, fully invested minimum variance portfolios.

By contrast, with their use of James-Stein for eigenvectors, the covariance matrix estimates discussed in this article rely on distribution-free eigenvector shrinkage in the HL regime, and can be customized to any quadratic minimization with linear constraints. James-Stein for eigenvectors was developed in Goldberg et al. [30], Goldberg et al. [29] and Gurdogan and Kercheval [33] 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 estimation errors 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.

3 The optimization problem and a JSE prescription

3.1 Constrained optimization

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,

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

where the jth constraint coefficient vector C_j is a p-vector, and the jth 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. 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. 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.

3.2 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 make explicit.

In this section, we consider first the simplified situation in which returns have identical specific risk. In Section 4, we discuss the more general one-factor case, and provide more complete mathematical details.

3.2.1 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 with homogeneous specific risk, the true (population) covariance matrix has the structure

$$\Sigma = \eta^2 bb^\top + \delta^2 I, \tag{2}$$

where b is a leading unit eigenvector of Σ with eigenvalue $\eta^2 + \delta^2$.

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 4 that the last of these three estimates is the most consequential.

3.2.2 A strategy-specific estimate of the vector of factor loadings

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

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

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

$$\psi^2 = \frac{\lambda^2 - \ell^2}{\lambda^2}, \tag{4}$$

the average relative leading eigengap.

Let C denote the span of the constraint vectors $C_1, C_2, \ldots, C_k$ 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

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

and define

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

The James-Stein for eigenvectors (JSE) estimate of the true eigenvector b is the unit vector[1]

$$h^{\mathrm{JSE}} = H^{\mathrm{JSE}}/|H^{\mathrm{JSE}}|. \tag{7}$$

The James-Stein estimate h^{JSE} is a better approximation to the true leading eigenvector b than than the principal component estimate $h = h^{\mathrm{PCA}}$. Let θ^{JSE} and θ^{PCA} denote the angles from b to h^{JSE} and h^{PCA}, and letting Θ denote the angle between b and the C, then

$$\cos^2(\theta^{\mathrm{JSE}}) - \cos^2(\theta^{\mathrm{PCA}}) = \left(\frac{1}{\phi_\infty^2 + 1}\right) \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1} > 0 \tag{8}$$

in the context of the portfolio construction problems studied in this article, where $\Theta < \pi/2$.[2]

[1] Formula (6) is equivalent to formula [6] of [28]. That article and [60] expose the parallel between JSE and classical James-Stein. Formulas (5), (6) and (7) are identical to formulas (36), (37) and (38) in Section 4.1.4.

[2] The asymptotic context in which formula (8) holds is described precisely in Theorem 2. Here, we have overloaded the notation for θ^{JSE}, θ^{PCA} and Θ.

3.2.3 A strategy-specific estimate of the covariance matrix

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

$$\Sigma^{\text{JSE}} = (\lambda^2 - \ell^2) h^{\text{JSE}} h^{\text{JSE}^\top} + (n/p)\ell^2 I. \tag{9}$$

Formula (9) 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 4, 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.

3.3 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 (9), and let w^{PCA} denote the portfolio generated by (1) when Σ is set to Σ^{PCA}.

Theorem 4 below shows that the ratio of the true variances w^{JSE} and w^{PCA},

$$\frac{\mathcal{V}(w^{\text{JSE}})}{\mathcal{V}(w^{\text{PCA}})}, \tag{10}$$

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.

4 JSE stochastically dominates PCA

The formulas in Section 3.2 prescribe the construction of a strategy-specific covariance matrix based on JSE for use in portfolio construction. Here, we describe in more precise detail the theory asymptotically guaranteeing that JSE improves eigenvector estimates and lowers variance of optimized portfolios, relative to PCA.

In our asymptotic analysis, we 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 4.1 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 4.2, 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.

4.1 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. In the 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.

A smaller angle and a larger relative gap translate to greater effectiveness of the JSE correction.

4.1.1 A one-factor model of returns and standing assumptions

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

$$r = \mu + \beta f + z, \tag{11}$$

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

$$r = \beta f + z. \tag{12}$$

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.

For the asymptotic theory we need to define a sequence of models of increasing dimension. If we imagine that increasing the dimension corresponds to adding new assets to the model, this will be described by a nested sequence

$$r^{(p)} = \beta^{(p)} f + z^{(p)}, \quad p = 1, 2, 3, \dots . \tag{13}$$

The nested property means that the models are defined by an infinite sequence of scalars $\{\beta_i\}$ and an infinite sequence random variables $\{z_i\}$ such that truncation at p forms the p-vectors $\beta^{(p)}$ and $z^{(p)}$, respectively.[3]

We list below our standing assumptions on the factor model (12).

Standing Assumptions.

A1. The random variable f representing factor returns is non-zero almost surely, and has mean zero and variance $\sigma^2 > 0$.

A2. (a) The random variables $\{z_i : i = 1, 2, \dots\}$ representing security specific returns have mean zero, are uncorrelated with f, and have uniformly bounded second moments, with variances $Var(z_i) = \delta_i^2$ tending on average to a limit $\delta^2 > 0$:

$$\lim_{p\to\infty} (1/p) \sum_{i=1}^{p} \delta_i^2 = \delta^2 > 0. \tag{14}$$

(b) In addition, we assume either

i. the variables $\{z_i\}$ are mutually independent, or

ii. the variables $\{z_i\}$ have uniformly bounded fourth moments and satisfy the following correlation decay conditions

$$\frac{1}{p^2} \sum_{i,j=1}^{p} Cov(z_i, z_j)^2 \to 0 \quad \text{and} \quad \frac{1}{p^2} \sum_{i,j=1}^{p} Cov(z_i^2, z_j^2)^2 \to 0 \tag{15}$$

as $p \to \infty$.

A3. The sequence $\{\beta_i : i = 1, 2, 3, \dots\}$ of security exposures to the factor is bounded and the average of the squared entries tends, as $p \to \infty$, to a positive limit,

$$\lim_{p\to\infty} (1/p) \sum_{i=1}^{p} \beta_i^2 = B^2 > 0, \tag{16}$$

or, equivalently, $|\beta^{(p)}|^2/p \to B^2$ as $p \to \infty$.

In particular, if $b^{(p)} = \beta^{(p)}/|\beta^{(p)}|$, these assumptions imply that $\{p(b_i^{(p)})^2 : p > 1, i = 1, 2, 3, \dots, p\}$ is bounded.

Importantly, we make no parametric assumptions, Gaussian, sub-Gaussian, or otherwise, on the distributions of f or z. The finite moment assumptions on f and z_i allow for heavy-tailed distributions.

[3] The nested property is not required for our results if we accept a bound on higher moments, but it simplifies the discussion.

The assumption that the random variable f representing factor returns and the random variables z_i representing security specific returns have finite variances is standard in the financial literature, and estimating those variances is central to financial practice. The optional assumption of finite fourth moments is common in the literature, but its empirical justification for returns to public equities is weak. Security returns in public equity markets exhibit heavy tails, with power law coefficients estimated, in some studies, to be below 4; see, for example, Gabaix [27] and Warusawitharana [67].

The assumption that factor returns f and specific returns are uncorrelated embodies the essence of a "factor model", and implies that the covariance matrix decomposes as a sum $\Sigma = \eta^2 bb^\top + \Omega$ of factor and specific covariance components. Assumptions on the joint distribution of specific returns have deeper implications, as they are needed for our application of laws of large numbers to prove asymptotic results.

The condition in A2 and A3 that the sequences $\{\delta_i^2\}$ and $\{\beta_i^2\}$ have positive limiting averages (called *pervasiveness* in Fan et. al. [25]) means that a non-negligible fraction of the entries are non-vanishing. This is a basic and mild non-degeneracy condition on our asymptotic sequence of models. It means that a non-negligible fraction of extra assets added to increase the model dimension have non-negligible exposure to the factor, and a non-negligible fraction have non-negligible specific risk. (The existence of the limit is a matter of convenience, since otherwise we could pass to subsequences.)

For the factor model (12), under our assumptions the population covariance matrix of returns takes the form

$$\Sigma = \sigma^2 \beta\beta^\top + \Omega, \tag{17}$$

where Ω is the covariance matrix of the specific returns z_i, which by A2 has bounded eigenvalues.

Assumption A2(b)i implies that Ω is diagonal and we have a strict factor model. The alternative A2(b)ii allows the specific returns to be correlated, so that we are in the setting of an *approximate factor model* in the sense of Chamberlain and Rothschild [9]. This allows for the presence of additional weak factors provided their corresponding eigenvalues are bounded. The correlation decay conditions are satisfied, if, for example, the variables depend only on a bounded number of other z's.

If we strengthen assumption A2 to

A2* The random variables z_i satisfy assumption A2(a) and in addition are mutually independent and have uniformly bounded fourth moments,

then the limiting theorems in this paper hold almost surely instead of in probability.

Note: In this article, what follows will be a series of limit theorems as $p \to \infty$. All results assume our standing assumptions A1, A2, and A3 hold. All limits of random variables will be in the sense of convergence in probability. In addition, when assumption A2 also holds, then the limits will hold almost surely.*

Because β and f appear in the model (12) 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

$$r = \eta b x + z. \tag{18}$$

With this formulation, A3 tells us that η_p^2/p tends to a positive limit $\sigma^2 B^2$ as $p \to \infty$.
The population covariance matrix is a sum of a factor component, $\eta^2 bb^\top$, and a specific component, Ω:

$$\Sigma = \eta^2 bb^\top + \Omega. \tag{19}$$

4.1.2 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 $r_1, r_2 \ldots, r_n$ of the p-vector r of security returns with factor structure (18) and hence, covariance matrix Σ given by (19). 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:

$$S = \lambda^2 hh^\top + \lambda_2^2 v_2 v_2^\top + \lambda_3^2 v_3 v_3^\top \cdots + \lambda_p^2 v_p v_p^\top \tag{20}$$

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 $\{h, v_2, \ldots, v_p\}$ 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.

In our context, natural for portfolio theory, we have fixed n and λ^2/p bounded as $p \to \infty$. The following proposition states that h stays away from b with high probability when $p >> n$.

Recall

$$\ell^2 = \frac{tr(S) - \lambda^2}{n-1} \tag{21}$$

and

$$\psi_p^2 = \frac{\lambda^2 - \ell^2}{\lambda^2}. \tag{22}$$

Proposition 1 *The limits*

$$\theta^{\mathrm{PCA}} = \lim_{p\to\infty} \angle(h,b) \quad \textit{and} \quad \psi^2_\infty = \lim_{p\to\infty} \psi^2_p \tag{23}$$

exist, and

$$\cos\theta^{\mathrm{PCA}} = \psi_\infty \in (0,1). \tag{24}$$

These limits hold in probability under assumptions A1, A2, and A3, and hold almost surely if A2 is replaced by A2.*

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

The random variable ψ_∞ can be expressed in terms of the relationship between the relative eigengap and the parameters of the factor model (18). Decomposing, from (18), the $p \times n$ data matrix of returns Y into a sum of unobservable components, we have

$$Y = \eta b X^\top + Z, \tag{25}$$

where $X = (X_1, X_2, \ldots, X_n)^\top$ 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 7 stated later.

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

$$\psi^2_\infty = \lim_{p\to\infty} \psi^2_p = \lim_{p\to\infty} \frac{\lambda^2 - \ell^2}{\lambda^2} = \frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2} \approx \frac{p\sigma^2 B^2}{p\sigma^2 B^2 + p\delta^2/n}. \tag{26}$$

These limits hold in probability under assumptions A1, A2, and A3, and hold almost surely if A2 is replaced by A2.*

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 (26) is justified by the fact that $E[|X|^2/n] = 1$ and $|X|^2/n \to 1$ almost surely as $n \to \infty$. (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 (26). It is the asymptotic trace of the factor component of the population covariance matrix Σ, specified in (19), and can be viewed as the variance in the system attributable to the factor. The term $p\delta^2$ is the asymptotic 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 (26) cannot be observed, but it can be estimated in terms of the relative eigengap of S.

A consequence of Proposition 2 is that, after first taking the limit $p \to \infty$ and then allowing $n \to \infty$ (a weak version of the HH regime), the term ψ_∞^2 tends to 1. Therefore,

$$\lim_{n\to\infty} \lim_{p\to\infty} |h - b| = 0. \tag{27}$$

As a result, the defect in the PCA estimate h in applications where $p >> n$ 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

$$\theta_{PCA} \approx \frac{1}{\sqrt{n}} \frac{\delta}{\sigma B}. \tag{28}$$

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.

We note that Wang and Fan [66] provide a stronger HH version of equation (27), under the additional assumption that the population variables are all sub-Gaussian: in our factor model context, if n and p both tend to infinity in any manner, then

$$\lim_{n,p\to\infty} |h - b| = 0. \tag{29}$$

4.1.3 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 span(x) and 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 $p \to \infty$, to a limit*

$$\Theta = \lim_{p\to\infty} \angle(b, C). \tag{30}$$

Then the limit

$$\Theta^h = \lim_{p\to\infty} \angle(h, C) \tag{31}$$

exists, and

$$\cos\Theta^h = \cos\theta^{\text{PCA}} \cdot \cos\Theta = \psi_\infty \cdot \cos\Theta. \tag{32}$$

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

$$0 < \cos\Theta^h < \cos\theta^{\text{PCA}} \tag{33}$$

and

$$0 < \cos \Theta^h < \cos \Theta. \tag{34}$$

These limits hold in probability under assumptions A1, A2, and A3, and hold almost surely if A2 is replaced by A2.*

This theorem is a generalization of Theorem 3.1 of [30]. 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.

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. [34] and Ahn et al. [1]). The condition $\Theta < \pi/2$ thus has a Bayesian interpretation in which C represents some mild prior information about the direction of b.

In our context, the condition $\Theta < \pi/2$ 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 $e = (1, 1, 1, \ldots, 1)^\top$. Since stock betas tend to be positive, β will typically have positive mean in equity applications, we obtain

$$\cos \angle(b, C) \geq \langle b, e/|e| \rangle = \frac{1}{|\beta|\sqrt{p}} \sum \beta_i = \frac{\sqrt{p}}{|\beta|} \left(\frac{1}{p} \sum \beta_i \right) > 0 \tag{35}$$

asymptotically, and so we can expect that $\Theta < \pi/2$ in typical financial settings.

The assumption that $\lim_{p\to\infty} \angle(b, C)$ exists is a matter of convenience. It could be replaced by assuming that $\limsup \angle(b, C) < \pi/2$, and then the subsequent discussion would apply to any convergent subsequence.

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

We will use the notation $h = h^{\mathrm{PCA}}$ 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 $p \to \infty$ under Standing Assumptions A1–A3.

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

$$c^{\mathrm{JSE}} = \frac{\ell^2}{\lambda^2(1 - |h_C|^2)}, \tag{36}$$

$$H^{\mathrm{JSE}} = c^{\mathrm{JSE}} h_C + (1 - c^{\mathrm{JSE}}) h, \tag{37}$$

and

$$h^{\mathrm{JSE}} = H^{JSE}/|H^{JSE}|. \tag{38}$$

Formulas (36), (37) and (38) are identical to formulas (5), (6) and (7) in Section 3.2.2.

We can show that

$$\lim_{p\to\infty} c^{\mathrm{JSE}} = \frac{1-\psi_\infty^2}{1-\psi_\infty^2 \cos^2\Theta} = \frac{\delta^2}{\sigma^2 B^2 |X|^2 \sin^2\Theta + \delta^2}. \tag{39}$$

(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} solely for convenience; all that matters is the 1-dimensional subspace it spans, as an estimate of the eigenspace span(b). The angle between these subspaces is our measure of error.

Define

$$\phi_\infty^2 \equiv \frac{\psi_\infty^2}{1-\psi_\infty^2} = \frac{\sigma^2 B^2 |X|^2}{\delta^2} = \lim_{p\to\infty} \frac{\lambda^2-\ell^2}{\ell^2}, \tag{40}$$

and recall that the angle between two vectors is, by definition, always nonnegative.

Theorem 2 *Suppose the limit*

$$\Theta = \lim_{p\to\infty} \angle(b, C) \tag{41}$$

exists.

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

$$\theta^{\mathrm{JSE}} = \lim_{p\to\infty} \angle(h^{\mathrm{JSE}}, \beta) \quad \textit{and} \quad \theta^{\mathrm{PCA}} = \lim_{p\to\infty} \angle(h^{\mathrm{PCA}}, \beta) \tag{42}$$

exist in probabilty, and hold almost surely under additional assumption A2.*

The asymptotic improvement of h^{JSE} over h^{PCA} as an estimate of the leading population eigenvector is

$$\cos^2(\theta^{\mathrm{JSE}}) - \cos^2(\theta^{\mathrm{PCA}}) = \left(\frac{1}{\phi_\infty^2+1}\right) \frac{\cos^2\Theta}{\phi_\infty^2 \sin^2\Theta + 1}. \tag{43}$$

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

- *if $\Theta < \pi/2$, then $\theta^{\mathrm{JSE}} < \theta^{\mathrm{PCA}}$*
- *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 $\theta^{\mathrm{JSE}} = \theta^{\mathrm{PCA}}$.*

The right hand side of (40) is the ratio of the factor variance to the specific variance in (18). The formula highlights the relationship between the relative eigengap and the factor model parameters. Taken together, (26) and (40) imply

$$\psi_\infty^2 = \frac{\phi_\infty^2}{1+\phi_\infty^2}. \tag{44}$$

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

$$I(\angle(h,C),\phi_p^2) = \frac{\cos^2\angle(h,C)}{(\phi_p^4+\phi_p^2)\sin^2\angle(h,C)}. \tag{45}$$

It follows from equations (32) and (43) that

$$\lim_{p\to\infty} I(\angle(h,C),\phi_p^2) = \cos^2(\theta^{\mathrm{JSE}}) - \cos^2(\theta^{\mathrm{PCA}}) \tag{46}$$

almost surely.

4.2 Estimating Constrained Minimum Variance

We return to the optimization problem (1),

$$\begin{aligned} &\min_w \frac{1}{2} w^\top \Sigma w \\ &\text{subject to} \quad C^\top w = a, \end{aligned} \tag{47}$$

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 $C_1, \ldots, C_k$ and $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$ 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 4.1 to estimate a $p \times p$ covariance matrix $\Sigma = \Sigma^{\mathrm{JSE}}$ for use in (47). The matrix Σ^{JSE} depends on the constraint matrix C; its core is h^{JSE}, the leading eigenvector of the sample covariance matrix, shrunken 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.

4.2.1 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 from now on we add the following two natural standing assumptions:

A4. For each $j = 1, \ldots, k$, the columns $C_j^{(p)}$ of $C^{(p)} \in \mathbb{R}^{p\times k}$ satisfy:

a. $\sup_{p\geq 1} |C_j^{(p)}|_\infty < \infty$, where $|.|_\infty$ denotes the maximum norm, and
b. the sequence $|C_j^{(p)}|^2/p$ tends to a positive finite limit as $p \to \infty$.

A5. The constraint matrix C does not become singular in the high dimensional limit:

$$\liminf_{p\to\infty} \det(C^\top C)/p^k > 0. \tag{48}$$

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.

4.2.2 Estimating Σ^{JSE}

The constraint matrix C and vector of constraint values a in the optimization (47) are known to the user, but the covariance matrix Σ must be estimated. When data follow the one factor model (12), the population covariance matrix Σ takes the form specified in (19):

$$\Sigma = \eta^2 bb^\top + \Omega. \tag{49}$$

As a consequence of this structure, an estimate of Σ amounts to estimates of a positive scalar η^2, a unit-length p-vector b, and the diagonal entries of Ω. 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, $tr(S)$, the leading eigenvalue λ^2 of S, and its corresponding leading eigenvector h.

Under our spiked model assumptions, it will turn out that for minimum variance estimation it suffices to estimate Ω with a multiple of the identity converging to $\delta^2 I$. 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. As described in Lemma 7 below, they are summarized by the limits

$$\lim_{p\to\infty} (\lambda^2 - \ell^2)/p = \sigma^2 B^2 |X|^2/n \tag{50}$$

and

$$\lim_{p\to\infty} \ell^2/p = \delta^2/n. \tag{51}$$

Recall from assumption A3 that $\eta^2/p \to \sigma^2 B^2$ as $p \to \infty$, 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 (51), 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

$$\Sigma^v = (\lambda^2 - \ell^2)vv^\top + (n/p)\ell^2 I. \tag{52}$$

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.

4.2.3 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

$$L(w, \Lambda) = (1/2)w^\top \Sigma^v w + (a^\top - w^\top C)\Lambda, \tag{53}$$

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

$$\Lambda^v = (C^\top (\Sigma^v)^{-1} C)^{-1} a \tag{54}$$
$$w^v = (\Sigma^v)^{-1} C \Lambda^v = (\Sigma^v)^{-1} C (C^\top (\Sigma^v)^{-1} C)^{-1} a. \tag{55}$$

We use the notation $\angle(v, C)$ to denote the angle between v and $col(C)$, $\cos(v, C)$ 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

$$|v_C| = \cos(v, C) < 1, \tag{56}$$

define the unit vector

$$\alpha = \frac{(C^\dagger)^\top a}{|(C^\dagger)^\top a|} \tag{57}$$

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

$$\mathcal{E}_p(v, C, a) = \frac{\langle b, \alpha \rangle (1 - |v_C|^2) - \langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 - |v_C|^2}, \tag{58}$$

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 $col(C)$, and is equal to zero when $v = b$:

$$\mathcal{E}(b, C, a) = 0. \tag{59}$$

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

$$\mathcal{E}_p(h, e, 1) = \frac{\langle b, e/|e|\rangle - \langle b, h\rangle \langle h, e/|e|\rangle}{1 - \langle h, e/|e|\rangle^2}, \tag{60}$$

which agrees with the optimization bias originally introduced for this case in Goldberg et al. [30].

The limits in the following two statements hold in probability under assumptions A1-A3, and almost surely if A2* is added.

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*

$$\limsup_{p\to\infty} |h_C| < 1 \tag{61}$$

and

$$\limsup_{p\to\infty} |(h^{\mathrm{JSE}})_C| < 1. \tag{62}$$

Theorem 3 *Let $v \in \mathbb{R}^p$ be a unit vector for each p and satisfying*

$$\limsup_{p\to\infty} |v_C| < 1. \tag{63}$$

Then, for n, k fixed,

$$0 < \limsup_{p\to\infty} \eta^2 |(C^\dagger)^\top a|^2 < \infty, \tag{64}$$

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

$$\mathcal{V}(w^v) \equiv (w^v)^\top \Sigma w^v = \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p) \tag{65}$$

asymptotically as $p \to \infty$.

Because of Proposition 3, Theorem 3 applies to both $v = h$ and $v = h^{\mathrm{JSE}}$. 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 *Suppose that the angle between b and $col(C)$ is asymptotically between 0 and $\pi/2$.*

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

$$\lim_{p\to\infty} \cos(\angle(b, (C^\dagger)^\top a)) = \lim_{p\to\infty} \langle b, \alpha\rangle \equiv \langle b, \alpha\rangle_\infty \quad \textit{exists.} \tag{66}$$

Then

$$\lim_{p\to\infty} \mathcal{E}_p(h^{\mathrm{JSE}}, C, a)^2 = 0. \tag{67}$$

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

$$\lim_{p\to\infty} \mathcal{E}_p(h, C, a)^2 > 0. \tag{68}$$

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

$$\frac{\mathcal{V}(w^{\mathrm{JSE}})}{\mathcal{V}(w^{\mathrm{PCA}})} \tag{69}$$

tends to zero asymptotically.

The limits are in probability under assumptions A1 – A3, and hold almost surely if A2 is added.*

The previous two theorems tell us that $\mathcal{V}(w^{\mathrm{TRUE}})$ and $\mathcal{V}(w^{\mathrm{JSE}})$ tend to zero as $p \to \infty$, but $\mathcal{V}(w^{\mathrm{PCA}})$ 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 (passing to a subsequence if needed) that a does not belong to the orthogonal complement of the unit vector*

$$\lim_{p\to\infty} \frac{C^\dagger b}{|C^\dagger b|} \in \mathbb{R}^k. \tag{70}$$

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

5 Numerical examples

In this section, we describe the results of simulation experiments supporting the results stated above. First, we illustrate (43), 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 (43), are approximately realized when the dimension p has realistic values much less than

parameter	value(s)	description
$\cos\theta$	0.969, 0.707, 0.174	cosine of the angle between β and $e = (1, \ldots, 1)^\top$
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	(e, μ)	3000×2 matrix of constraint vectors
m	0.01	monthly expected target return
a	$(1, m)^\top$	constraint target vector

Table 2: Simulation parameters

infinity. The results reported here are for $p = 3000$, but we have observed similar outcomes for p as low as 40.

Second, the variance experiments described in Section 5.3 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.

5.1 Calibration

We specify the parameters of the return generating process (12), repeated here for convenience,

$$r = \beta f + z, \tag{71}$$

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 $e = (1, \ldots, 1)^\top$ 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

$$c_1 = \frac{\sin\theta}{s} \quad \textit{and} \quad c_2 = \cos\theta - \frac{\sin\theta}{s} m, \tag{72}$$

and let

$$\beta = c_1 \beta^* + c_2 e. \tag{73}$$

Making use of the identity

$$|\beta|^2 = p(m(\beta)^2 + s(\beta)^2), \tag{74}$$

a calculation shows that $|\beta|^2/p = 1$ and the angle between β and e 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 expect 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.

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

Under Standing Assumptions A1–A3, formula (43) provides an exact expression for the difference between the squared cosines of θ^{PCA} and θ^{JSE}:

$$\cos^2(\theta^{\mathrm{JSE}}) - \cos^2(\theta^{\mathrm{PCA}}) = \left(\frac{1}{\phi_\infty^2 + 1}\right) \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1}. \tag{75}$$

This *magic formula* for the limiting difference between angles $\angle(\beta, h^{\mathrm{PCA}})$ and $\angle(\beta, h^{\mathrm{JSE}})$ as $p \to \infty$ 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

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

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.

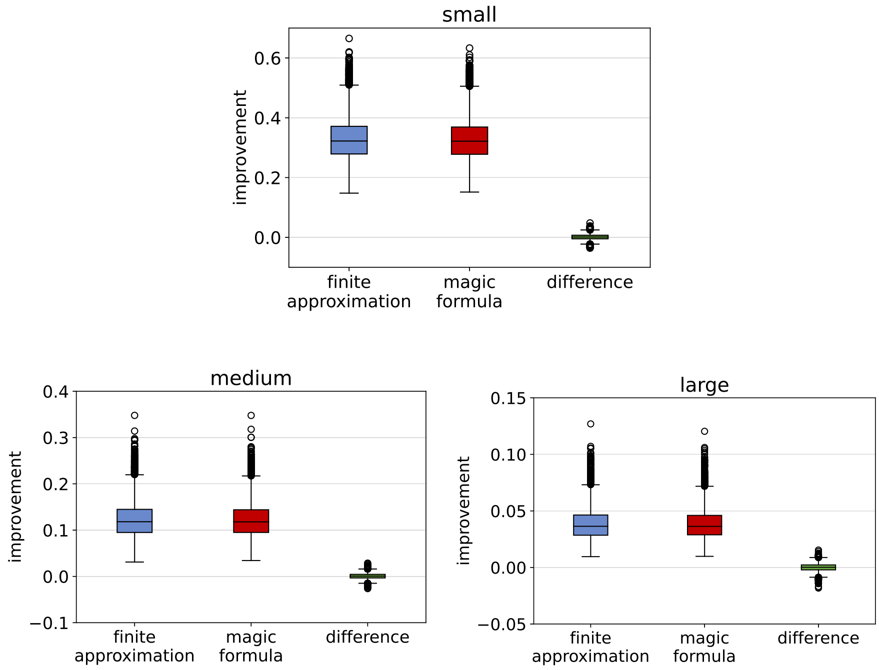

Fig. 1: Boxplots for $p = 3000$ of 10,000 simulations of the difference between $\cos^2\left(\angle(h^{\mathrm{PCA}}, b)\right)$ and $\cos^2\left(\angle(h^{\mathrm{JSE}}, b)\right)$ (finite difference), the asymptotic limit of this difference (magic formula) as well as the path-by-path difference between them (difference). The small, medium and large panels correspond to $\cos\Theta = 0.969, 0.707$ and 0.174, Return data follow (12) with parameters specified in Table 2.

5.3 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 $\Sigma^{\mathrm{TRUE}} = \Sigma$, 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 $\mathcal{V}(w^{\mathrm{JSE}})/\mathcal{V}(w^{\mathrm{PCA}})$ and $\mathcal{V}(w^{\mathrm{TRUE}})/\mathcal{V}(w^{\mathrm{PCA}})$ tend to 0 almost surely as p tends to infinity.

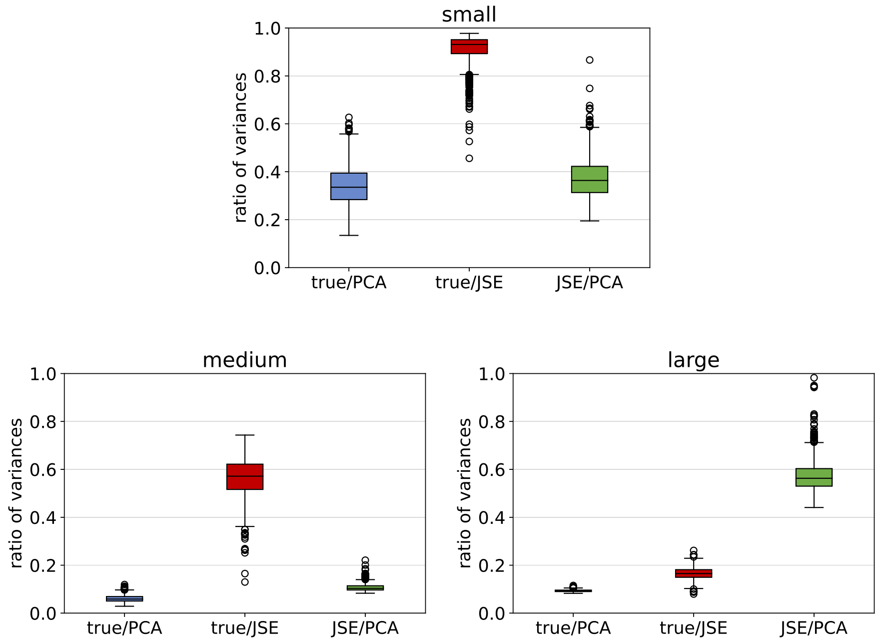

Fig. 2: Boxplots for 10,000 simulations of ratios of variances of optimized and optimal portfolios, w^{PCA}, w^{JSE} and w^{TRUE}, for $p = 3000$. The small, medium and large panels correspond to $\cos \Theta = 0.969, 0.707$ and 0.174. The expected return target is $m = 0.01$. Return data follow (12) with parameters specified in Table 2.

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

6 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.

7 Appendix: Proofs

7.1 Lemmas

We begin with some preliminary results needed for the subsequent proofs.

Lemma 2 (Triangular Strong Law of Large Numbers, Tao [64]) *Let $(X_{i,p})_{i,p\in N, i\leq p}$ be a triangular array of scalar random variables such that for each p, the row $X_{1,p},\ldots,X_{p,p}$ is a collection of independent random variables. For each p, define the partial sum $S_p = X_{1,p} + \ldots + X_{p,p}$. Assume all the $X_{i,p}$ have mean μ.*

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

Lemma 3 (Kolmogorov Strong Law of Large Numbers)

Suppose $X_1, X_2, \ldots$ is a sequence of independent mean-zero random variables with finite variance and such that

$$\sum_{i=1}^{\infty} \frac{Var(X_i)}{i^2} < \infty, \tag{76}$$

and for each p define the partial sum $S_p = X_1 + \ldots + X_p$.

Then S_p/p converges almost surely to zero.

Lemma 4 *Let $\{z_i : i \in \mathbb{N}\}$ be a sequence of independent mean-zero random variables with uniformly bounded fourth moments, and let $\{b_i : i \in \mathbb{N}\}$ be a sequence of scalars satisfying*

$$\sup_i \{p|b_i|^2 : i \in \mathbb{N}\} < \infty. \tag{77}$$

Then

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} b_i z_i \to 0 \tag{78}$$

almost surely as $p \to \infty$.

Proof of Lemma 4. Let $X_{i,p} = \sqrt{p} b_i z_i$ and $S_p = X_{1,p} + \ldots + X_{p,p}$. By the assumptions, the $X_{i,p}$ have mean zero and uniformly bounded fourth moments. By Lemma 2 with $\mu = 0$,

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} b_i z_i = \frac{1}{p} S_p \tag{79}$$

converges to zero almost surely. □

Lemma 5 *Let $\{z_i : i \in \mathbb{N}\}$ be a sequence of independent mean-zero random variables with uniformly bounded fourth moments. Suppose*

$$\lim_{p\to\infty} (1/p) \sum_{i=1}^{p} E(z_i^2) = \delta^2. \tag{80}$$

Then, almost surely,

$$\lim_{p\to\infty} (1/p) \sum_{i=1}^{p} z_i^2 = \delta^2. \tag{81}$$

Proof of Lemma 5.

Let $X_i = z_i^2 - E(z_i^2)$; it suffices to prove that $(1/p) \sum X_i \to 0$ as $p \to \infty$. The X_i have uniformly bounded variance because the z_i have uniformly bounded fourth moment. Hence

$$\sum_{i=1}^{\infty} \frac{Var(X_i)}{i^2} < \infty \tag{82}$$

and the result follows by Lemma 3. □

Lemma 6 *Recall our $p \times n$ data matrix of returns*

$$Y = \beta X^\top + Z. \tag{83}$$

Let $Z^k \in \mathbb{R}^p, k = 1, \ldots, n$, denote the kth column (observation) of Z. Then we have the following limits in probability:

$$\lim_{p\to\infty} \frac{1}{p} b^T Z^k = 0,$$

and

$$\lim_{p\to\infty} \frac{1}{p} Z^T Z = \delta^2 I_n.$$

These limits hold in probability under assumptions A1, A2, and A3, and hold almost surely if A2 is replaced by A2.*

Proof of Lemma 6.

The limits in probability follow from straightforward calculation using A2 and Markov's inequality. The almost sure limits follow from A2*, A3, and Lemmas 4 and 5. □

The following is a version of Proposition 5.2 in Gurdogan and Kercheval [32], 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. $\lim_{p\to\infty} \big(\langle h, h_L \rangle - \langle h, b \rangle^2 \langle b, b_L \rangle) \big) = 0,$
2. $\lim_{p\to\infty} \big(\langle b, h_L \rangle - \langle h, b \rangle \langle b, b_L \rangle) \big) = 0,$ *and*
3. $\lim_{p\to\infty} |h_L - \langle h, b \rangle\, b_L| = 0.$

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

7.2 Proof of Proposition 1

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

$$\theta^{\mathrm{PCA}} = \lim_{p\to\infty} \angle(h, b) \quad \textit{and} \quad \psi_\infty^2 = \lim_{p\to\infty} \psi_p^2 \tag{84}$$

exist, and

$$\cos \theta^{\mathrm{PCA}} = \psi_\infty \in (0, 1). \tag{85}$$

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

$$\lambda h = Y\chi/\sqrt{n}, \tag{86}$$

by equation (25) we have

$$\lambda h = \frac{\eta b X^\top \chi + Z\chi}{\sqrt{n}}. \tag{87}$$

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

$$\langle h, b\rangle = \left(\frac{\eta X^\top \chi}{\lambda\sqrt{n}}\right) + \left(\frac{b^\top Z}{\sqrt{p}}\right)\left(\frac{\chi\sqrt{p}}{\lambda\sqrt{n}}\right), \tag{88}$$

$$\lambda^2/p = \frac{\eta^2(X^\top\chi)^2}{np} + \frac{\chi^\top Z^\top Z\chi}{np} + 2(X^\top\chi)\left(\frac{b^\top Z}{\sqrt{p}}\right)\left(\frac{\eta\chi}{n\sqrt{p}}\right). \tag{89}$$

Applying Lemma 6, we may deduce that $Z^\top Z/p$ tends to $\delta^2 I$ and $b^\top Z/\sqrt{p}$ tends to zero as $p \to \infty$. This means that λ^2/p is eventually bounded between zero and infinity, and

$$\langle h, b\rangle_\infty = \lim_{p\to\infty}\left(\frac{\eta X^\top\chi}{\lambda\sqrt{n}}\right) \tag{90}$$

provided the limit in (90) exists.

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. [30]:

Lemma 7 *Under assumptions A1 - A3 and notation as above, we have the following limits:*

$$\lim_{p\to\infty}\lambda^2/p = \sigma^2 B^2|X|^2/n + \delta^2/n, \tag{91}$$

$$\lim_{p\to\infty}\chi_p = X/|X|, \quad \textit{and} \tag{92}$$

$$\lim_{p\to\infty}\ell_p^2/p = \delta^2/n. \tag{93}$$

Applying Lemma 7 to (90), we obtain

$$\langle h, b\rangle_\infty = \lim_{p\to\infty}\frac{\eta X^T\chi}{\lambda\sqrt{n}} = \lim_{p\to\infty}\left(\frac{\eta}{\sqrt{p}}\right)\left(\frac{\sqrt{p}}{\lambda}\right)\left(\frac{X^\top\chi}{\sqrt{n}}\right) \tag{94}$$

$$= \sigma B\left(\frac{1}{\sqrt{\sigma^2 B^2|X|^2/n + \delta^2/n}}\right)\left(\frac{|X|}{\sqrt{n}}\right) \tag{95}$$

$$= \sqrt{\frac{\sigma^2 B^2|X|^2}{\sigma^2 B^2|X|^2 + \delta^2}} \in (0,1). \tag{96}$$

By Lemma 7,

$$\psi_p^2 = \frac{\lambda^2 - \ell_p^2}{\lambda^2} \tag{97}$$

converges to

$$\psi_\infty^2 = \frac{\sigma^2 B^2|X|^2}{\sigma^2 B^2|X|^2 + \delta^2} \tag{98}$$

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

7.3 Proof of Theorem 1

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

$$\Theta = \lim_{p\to\infty} \angle(b, C). \tag{99}$$

Then the limit

$$\Theta^h = \lim_{p\to\infty} \angle(h, C) \tag{100}$$

exists, and

$$\cos\Theta^h = \cos\theta^{\mathrm{PCA}} \cdot \cos\Theta = \psi_\infty \cdot \cos\Theta. \tag{101}$$

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

$$0 < \cos\Theta^h < \cos\theta^{\mathrm{PCA}} \tag{102}$$

and

$$0 < \cos\Theta^h < \cos\Theta. \tag{103}$$

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 $\langle h, b\rangle \to \psi_\infty$ from Proposition 1 and $\cos\angle(b, C) \to \cos\Theta$ by hypothesis, equation (101) follows immediately. □

7.4 Proof of Theorem 2

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

$$\Theta = \lim_{p\to\infty} \angle(b, C) \tag{104}$$

exists.

Then the limits

$$\theta^{\mathrm{JSE}} = \lim_{p\to\infty} \angle(h^{\mathrm{JSE}}, \beta) \quad \text{and} \quad \theta^{\mathrm{PCA}} = \lim_{p\to\infty} \angle(h^{\mathrm{PCA}}, \beta) \tag{105}$$

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

$$\cos^2(\theta^{\mathrm{JSE}}) - \cos^2(\theta^{\mathrm{PCA}}) = \left(\frac{1}{\phi_\infty^2 + 1}\right) \frac{\cos^2\Theta}{\phi_\infty^2 \sin^2\Theta + 1}. \tag{106}$$

If $\Theta = \pi/2$, then h^{JSE} converges to h^{PCA}, $\theta^{\mathrm{JSE}} = \theta^{\mathrm{PCA}}$ and there is no improvement, while if $\Theta = 0$ then h^{JSE} converges to b. In other cases, $\theta^{JSE} < \theta^{PCA}$ almost surely, with the improvement given by (43).

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 [33].

We make further use of some results in that paper, first defining for each p, the oracle estimator $h^o = h^o(C)$ as follows. Let

$$U = span(h, C),$$

and define the unit vector

$$h^o = \frac{b_U}{|b_U|}. \tag{107}$$

The oracle h^o 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 [33], for slightly adjusted assumptions:

Proposition 5

$$\lim_{p\to\infty} |h^o - h^{\mathrm{JSE}}| = 0. \tag{108}$$

Next, let

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

Then $U \equiv span(h, C) = span(u, C)$ 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

$$\langle h^o, b\rangle^2 = \left\langle \tfrac{b_U}{|b_U|}, b\right\rangle^2 = |b_U|^2 \tag{109}$$

$$= |b_C|^2 + \langle u, b\rangle^2 \tag{110}$$

$$= |b_C|^2 + \tfrac{(\langle h,b\rangle - \langle h_C, b\rangle)^2}{1-|h_C|^2}. \tag{111}$$

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

$$|b_C|^2 \to \cos^2 \Theta, \tag{112}$$

$$|h_C|^2 \to \psi_\infty^2 \cos^2 \Theta, \tag{113}$$

$$\langle h, b\rangle \to \psi_\infty = \cos\theta^{\mathrm{PCA}}, \tag{114}$$

$$\langle h_C, b\rangle \to \psi_\infty \cos^2 \Theta. \tag{115}$$

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

$$\lim_{p\to\infty} \langle h^o, b\rangle^2 = \lim_{p\to\infty} \left\langle h^{\mathrm{JSE}}, b\right\rangle^2 = \cos^2\theta^{\mathrm{JSE}}. \tag{116}$$

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

$$\psi^2 = \frac{\phi^2}{1+\phi^2}, \tag{117}$$

in the limit,

$$\begin{aligned}\cos^2\theta^{\mathrm{JSE}} - \cos^2\theta^{\mathrm{PCA}} &= \cos^2\Theta + \tfrac{\psi^2(1-\cos^2\Theta)^2}{1-\psi^2\cos^2\Theta} - \psi^2 &(118)\\ &= (1-\psi^2)^2\tfrac{\cos^2\Theta}{1-\psi^2\cos^2\Theta} &(119)\\ &= \left(\tfrac{1}{\phi^2+1}\right)\tfrac{\cos^2\Theta}{\phi^2\sin^2\Theta+1}. &(120)\end{aligned}$$

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 $h = h^{\mathrm{PCA}}$, so $\theta^{\mathrm{JSE}} = \theta^{\mathrm{PCA}}$ and JSE provides no improvement over PCA.

If $\Theta = 0$, it follows from equation (118) that $\theta^{\mathrm{JSE}} = 0$ and so h^{JSE} tends to b itself.

□

7.5 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*

$$\limsup_{p\to\infty} |h_C| < 1 \tag{121}$$

and

$$\limsup_{p\to\infty} |(h^{\mathrm{JSE}})_C| < 1. \tag{122}$$

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

$$|h_C|^2 = \langle h, b\rangle_\infty^2 \, |b_C|^2 = \psi_\infty^2 |b_C|^2. \tag{123}$$

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

We can write

$$h^{\mathrm{JSE}} = \frac{\Gamma_p h + h_C}{|\Gamma_p h + h_C|} \tag{124}$$

where

$$\Gamma_p = \frac{\psi_p^2 - |h_C|^2}{1-\psi_p^2}. \tag{125}$$

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

$$\liminf_{p\to\infty} \Gamma_p > 0. \tag{126}$$

This follows from equation (123) and the standing assumption that the angle between b and C is asymptotically strictly between 0 and $\pi/2$.

7.6 Proof of Theorem 3

Theorem 3. *Let $v \in \mathbb{R}^p$ be a unit vector for each p and satisfying*

$$\limsup_{p\to\infty} |v_C| < 1. \tag{127}$$

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

$$\mathcal{V}(w^v) \equiv (w^v)^\top \Sigma w^v = \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p) \tag{128}$$

asymptotically as $p \to \infty$.

Furthermore,

$$0 < \limsup_{p\to\infty} \eta^2 |(C^\dagger)^\top a|^2 < \infty. \tag{129}$$

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

$$w^v = \big(I + \frac{(v_C - v)v^\top}{1 + \kappa^2 - |v_C|^2}\big)(C^\dagger)^\top a. \tag{130}$$

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

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

which establishes the last assertion of the theorem.

To obtain an expression for true variance, first notice that

$$\mathcal{V}(w^v) = (w^v)^\top \Sigma w^v = \eta^2 \langle w^v, b\rangle^2 + \delta^2 |w^v|^2. \tag{131}$$

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 (130) and recalling

$$\alpha = (C^\dagger)^\top a / |(C^\dagger)^\top a|, \quad \limsup_{p\to\infty} |v_C| < 1, \quad \textit{and} \quad \kappa^2 = O(1/p), \tag{132}$$

we have

$$\begin{aligned} \eta^2 \langle w^v, b\rangle^2 &= \eta^2 |(C^\dagger)^\top a|^2 \big(\langle b,\alpha\rangle - \tfrac{\langle b, v-v_C\rangle \langle v,\alpha\rangle}{1+\kappa^2-|v_C|^2} \big)^2 && (133)\\ &= \eta^2 |(C^\dagger)^\top a|^2 \big(\tfrac{\langle b,\alpha\rangle (1-|v_C|^2) - \langle b, v-v_C\rangle\langle v,\alpha\rangle}{1-|v_C|^2}\big)^2 + O(1/p) && (134)\\ &= \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v,C,a)^2 + O(1/p). && (135)\end{aligned}$$

□

7.7 Proof of Theorem 4

Theorem 4. *Suppose that the angle between b and $col(C)$ is asymptotically between 0 and $\pi/2$.*

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

$$\lim_{p\to\infty} \cos(\angle(b, (C^\dagger)^\top a)) = \lim_{p\to\infty} \langle b,\alpha\rangle \equiv \langle b,\alpha\rangle_\infty \quad \textit{exists}. \tag{136}$$

Then

$$\lim_{p\to\infty} \mathcal{E}_p(h^{\mathrm{JSE}}, C, a)^2 = 0. \tag{137}$$

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

$$\lim_{p\to\infty} \mathcal{E}_p(h, C, a)^2 > 0. \tag{138}$$

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

$$\frac{\mathcal{V}(w^{\mathrm{JSE}})}{\mathcal{V}(w^{\mathrm{PCA}})} \tag{139}$$

tends to zero asymptotically.

Proof. By Proposition 3, we know that

$$\limsup |v_C| < 1 \tag{140}$$

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

$$\mathcal{E}_p(v,C,a) = \frac{\langle b,\alpha\rangle (1-|v_C|^2) - \langle b, v-v_C\rangle \langle v,\alpha\rangle}{1-|v_C|^2}, \tag{141}$$

stays away from zero in both cases. For the first statement (137) of the theorem, it then suffices to show that the numerator

$$\langle b,\alpha\rangle (1 - |(h^{\mathrm{JSE}})_C|^2) - \left\langle b, h^{\mathrm{JSE}} - (h^{\mathrm{JSE}})_C \right\rangle \left\langle h^{\mathrm{JSE}}, \alpha\right\rangle \tag{142}$$

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 $U = span(h, C)$. This is a consequence of the fact that $\langle b_C, \alpha\rangle = \langle b, \alpha\rangle$ and following straightforward identities:

$$\langle b, h^o - (h^o)_C\rangle = |b_U| - \tfrac{|b_C|^2}{|b_U|}, \tag{143}$$

$$\langle (h^o)_C, \alpha\rangle = \tfrac{\langle b,\alpha\rangle}{|b_U|}, \quad \textit{and} \tag{144}$$

$$|(h^o)_C|^2 = \tfrac{|b_C|^2}{|b_U|^2}. \tag{145}$$

Turning to the second statement (138), first note that Proposition 4 applied to the subspace $L = span(\alpha)$, 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

$$\lim_{p\to\infty} \mathcal{E}_p(h, C, a) = \tfrac{\langle b,\alpha\rangle(1-\langle h,b\rangle^2)}{1-\langle h,b\rangle^2|b_C|^2} \tag{146}$$

$$= \tfrac{\langle b,\alpha\rangle(1-\psi_\infty^2)}{1-\psi_\infty^2|b_C|^2}. \tag{147}$$

When $\mathcal{E}(h, C, a)$ is positive but $\mathcal{E}(h^{\mathrm{JSE}}, C, a)$ tends to zero, equation (128) implies that $\mathcal{V}(w^{\mathrm{PCA}})$ remains bounded above zero while $\mathcal{V}(w^{\mathrm{JSE}})$ tends to zero. This establishes the last claim.

□

7.8 Proof of Lemma 1

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

$$\lim_{p\to\infty} \frac{C^\dagger b}{|C^\dagger b|} \in \mathbb{R}^k. \tag{148}$$

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

We express the singular value decomposition of C as

$$C^{(p)} = U^{(p)} Z^{(p)} V^{(p)}, \tag{149}$$

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

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 $p \to \infty$.

A short calculation shows

$$|(C^\dagger)^\top a|^2 = \left\langle Z^{-2} V a, V a \right\rangle \tag{150}$$

and

$$\left\langle b, (C^\dagger)^\top a \right\rangle = \left\langle C^\dagger b, a \right\rangle = \left\langle Z^{-1} U^\top b, V a \right\rangle . \tag{151}$$

Hence

$$\frac{\left\langle b, (C^\dagger)^\top a \right\rangle}{|(C^\dagger)^\top a|} = \frac{\left\langle Z^{-1} U^\top b, V a \right\rangle}{\sqrt{\langle Z^{-2} V a, V a \rangle}} \to \frac{\left\langle Z_\infty^{-1} (U^\top b)_\infty, V_\infty a \right\rangle}{\sqrt{\left\langle Z_\infty^{-2} V_\infty a, V_\infty a \right\rangle}} . \tag{152}$$

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 Jorurnal of Statistics **60**(1), 4–19 (2018)
3. Bai, Z.: Methodologies in spectral analysis of large dimensional random matrices: A review. Statistica Sinica **9**, 611–677 (1999)
4. Bai, Z., Silverstein, J.W.: Spectral Analayis of Large Dimensional Random Matrices, second edn. Spriger Series in Statistics. Springer (2010)
5. 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)
6. 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)
7. Bodnar, T., Okhrin, Y., Parolya, N.: Optimal shrinkage-based portfolio selection in high dimensions. Journal of Business & Economic Statistics **41**(1), 14–00156 (2021)
8. Bodnar, T., Parolya, N., Schmid, W.: Estimation of the global minimum variance portfolio in high dimensions. European Journal of Operational Research **266**, 371–390 (2018)
9. Chamberlain, G., Rothschild, M.: Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica **51**(5), 1281–1304 (1983)
10. Clarke, R., De Silva, R., Thorley, S.: Minimum-variance portfolio composition. Journal of Portfolio Management **2**(37), 31–45 (2011)
11. Connor, G.: A unified beta pricing theory. Journal of Economic Theory **34**, 13–31 (1984)
12. Connor, G.: The three types of factor models: A comparison of their explanatory power. Financial Analysts Journal **51**(3), 42–46 (1995)
13. Connor, G., Korajczyk, R.: Risk and return in equilibrium APT. Journal of Financial Economics **21**, 255–289 (1988)
14. Connor, G., Korajczyk, R.: Factor models. In: Encyclopedia of Quantitative Finance. Wiley (2010)
15. Connor, G., Korajczyk, R.A.: Performance measurement with the arbitrage pricing theory. Journal of Financial Economics **15**, 373–394 (1986)

16. Ding, Y., Li, Y., Zheng, X.: High dimensional minimum variance portfolio estimation under statistical factor models. Journal of Econometrics **222** (2021)
17. Donoho, D., Gavish, M., Johnstone, I.: Optimal shrinkage of eigenvalues in the spiked covariance model. The Annals of Statistics **46**(4), 1742–1778 (2018)
18. Edelman, A., Rao, N.R.: Random matrix theory. Acta Numerica **14**(233–297) (2005)
19. Efron, B., Morris, C.: Data analysis using stein's estimator and its generalizations. J. of the American Statistical Assoc. **70**(350), 311–319 (1975)
20. Efron, B., Morris, C.: Stein's paradox in statistics. Scientific American **236**(5), 119–127 (1977)
21. 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)
22. El Karoui, N.: On the realized risk of high-dimensional markowitz portfolios. SIAM Journal of Financial Mathematics **4**, 737–783 (2013)
23. Fan, J., Liao, Y., Mincheva, M.: High-dimensional covariance matrix estimation in approximate factor models. Annals of Statistics **39**(6), 3320–3356 (2011)
24. Fan, J., Liao, Y., Mincheva, M.: Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society, Series B **75**(4), 603–680 (2013)
25. Fan, J., Liao, Y., Mincheva, M.: Large covariance estimation by thresholding principal orthogonal complements. J R Stat Soc Series B **75**(4), 603–680 (2013)
26. Fourdrinier, D., Strawderman, W.E., Wells, M.T.: Shrinkage Estimation. Springer Series in Statistics. Springer Nature, Cham, Switzerland (2018)
27. Gabaix, X.: Power laws in economics and finance. Annual Review of Economics **1**, 255–294 (2009)
28. Goldberg, L.R., Kercheval, A.N.: James-Stein for the leading eigenvector. Proceedings of the National Academy of Sciences **120**(2) (2023)
29. Goldberg, L.R., Papacinicolau, A., Shkolnik, A., Ulucam, S.: Better betas. The Journal of Portfolio Management **47**(1), 119–136 (2020)
30. Goldberg, L.R., Papanicolau, A., Shkolnik, A.: The dispersion bias. SIAM Journal of Financial Mathematics **13**(2), 521–550 (2022)
31. Green, R.C., Hollifield, B.: When will mean-variance efficient portfolios be well diversified? The Journal of Finance **47**(5), 1785–1809 (1992)
32. Gurdogan, H., Kercheval, A.: Multi anchor point shrinkage for the sample covariance matrix. Tech. rep., working paper (2021)
33. Gurdogan, H., Kercheval, A.: Multi anchor point shrinkage for the sample covariance matrix. SIAM Journal on Financial Mathematics **13**(3), 1112–1143 (2022)
34. 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)
35. Jagananthan, R., Ma, T.: Risk reduction in large portfolios: Why imposing the wrong constraints helps. Journal of Finance **58**(4), 1651–1683 (2003)
36. James, W., Stein, C.: Estimation with quadratic loss. In: Proc. Fourth Berkeley Symp. Math. Stat. Prob., pp. 361–397 (1961)
37. Jobson, J., Korkie, B.: Estimation for markowitz efficient portfolios. Journal of the American Statistical Association **75**(371), 544–554 (1980)
38. Johnstone, I.: On the distribution of the largest eigenvalue in principal components analysis. The Annals of Statistics **29**(2), 295–327 (2001)
39. Johnstone, I.M., Lu, A.Y.: On consistency and sparsity for principal components analysis in high dimensions. Journal of the American Statistical Association **104**(486), 682–693 (2009). DOI 10.1198/jasa.2009.0121. URL `https://doi.org/10.1198/jasa.2009.0121`
40. Jung, S., Marron, J.: PCA consistency in high dimension, low sample size context. The Annals of Statistics **37**(6B), 4104–4130 (2009)
41. Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. Journal of Financial Economics **3**, 215–231 (1976)
42. Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock retursn with an application to portfolio selection. J. of Empirical Finance **10**, 603–621 (2003)

43. Ledoit, O., Wolf, M.: Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management **30**(4), 110–119 (2004)
44. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis **88**, 365–411 (2004)
45. Ledoit, O., Wolf, M.: Nonlinear shrinkage estimation of large-dimensional covariance matrices. The Annals of Statistics **40**(2), 1024–1060 (2012)
46. 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)
47. Marcenko, V., Pastur, L.: Distribution of eigenvalues for some sets of random matrices. Matematicheskii Sbornik **114**(4), 507–536 (1967)
48. Markowitz, H.: Portfolio selection. The Journal of Finance **7**(1), 77–92 (1952)
49. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments. Yale University Press (1959)
50. Menchero, J., Lazaas, A.: Evaluating and comparing risk model performance. The Journal of Portfolio Management **50**(3), 90–110 (2024)
51. Menchero, J., Wang, J., Orr, D.: Improving risk forecasts for optimized portfolios. Financial Analysts Journal **68**(3), 40–50 (2012)
52. Michaud, R.O.: The Markowitz optimization enigma: Is 'optimized' optimal? Financial Analysts Journal **45**(1), 31–43 (1989)
53. Paul, D.: Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Statistica Sinica **17**, 1617–1642 (2007)
54. Paul, D., Aue, A.: Random matrix theory in statistics: a review. Journal of Statistical Planning and Inference **150**, 1–29 (2014)
55. Rosenberg, B.: Extra-market components of covariance in security returns. The Journal of Financial and Quantitative Analysis **9**(2), 263–274 (1974)
56. Rosenberg, B., McKibben, W.: The prediction of systematic and specific risk in common stocks. Journal of Financial and Quantitative Analysis **8**(2), 317–333 (1973)
57. Ross, S.: The arbitrage theory of capital asset pricing. Journal of Economic Theory **13**, 341–360 (1976)
58. Sharpe, W.F.: A simplified model for portfolio analysis. Management Science **9**(2), 277–293 (1963)
59. Shen, D., Shen, H., Zhu, H., Marron, J.S.: The statistics and mathematics of high dimension low sample size asymptotics. Statistica Sinica **26**(4), 1747–1770 (2016)
60. Shkolnik, A.: James-Stein estimation of the first principal component. Stat **11**(1) (2022)
61. Spearman, C.: General intelligence, objectively determined and measured. American Journal of Psychology **15**(201–293) (1904)
62. Stein, C.: Some problems in multivariate analysis part I. Technical report no. 6, Office of Naval Research (1956)
63. Tao, T.: 254a, notes 1: Concentration of measure. URL `https://terrytao.wordpress.com/2010/01/03/254a-notes-1-concentration-of-measure/`
64. Tao, T.: 275a, notes 3: The weak and strong law of large numbers. URL `https://terrytao.wordpress.com/2015/10/23/275a-notes-3-the-weak-and-strong-law-of-large-numbers/`
65. Tao, T.: Topics in Random Matrix Theory, *Graduate Studies in Mathematics*, vol. 132. American Mathematical Society (2012)
66. Wang, W., Fan, J.: Asymptotics of empirical eigenstructure for high dimensional spiked covariance. The Annals of Statistics **45**(3), 1342–1374 (2017)
67. Warusawitharana, M.: Time-varying volatility and the power law distribution of stock returns. Journal of Empirical Finance **49**, 123–141 (2018)
68. Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. Annals of Mathematics **62**, 548–564 (1955)
69. Wigner, E.: On the distribution of the roots of certain symmetric matrices. Annals of Mathematics **67**, 325–328 (1958)