Portfolio optimization via strategy-specific eigenvector shrinkage

Lisa R. Goldberg,* Hubeyb Gurdogan† and Alec Kercheval‡

DRAFT – August 16, 2023
This Version: August 19, 2023

Abstract

We design strategy-specific estimates of covariance matrices for portfolio optimization, tailored to the given constraints. Our factor-based, data-driven construction relies on a generalized version of James-Stein for eigenvectors (JSE), which reduces estimation error in the leading sample eigenvector by shrinking toward a subspace of dimension $\geq 1$. Unchecked, this error gives rise to excess volatility or tracking error 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.

1 Introduction

Since Harry Markowitz launched modern finance in 1952 by constructing portfolios with mean-variance optimization, active research has focused on

*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.
†Department of Mathematics, UCLA, Los Angeles, CA 90024
‡Department of Mathematics, Florida State University, Tallahassee, FL 32306
determining appropriate estimates for expected security returns and covariances. Practical considerations often limit the pool of relevant data, rendering a sample covariance matrix unfit for use in optimization. Prescriptions for empirically sound, well-conditioned covariance matrix estimates required by optimization vary, and the nature of their errors and their impact on optimized portfolios can be obscure.

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

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

We address this problem with the development of high-dimensional covariance matrix estimates that generate relatively low-variance optimized portfolios. Our estimates are based on the assumption that data follow a one-factor model, so covariance matrices are sums of rank-one matrices that drive correlation and diagonal matrices of firm-specific variances. Exploiting recent research that sheds light on how estimation error is transmitted via optimization, the factor underlying the rank-one component of our estimate is obtained by applying James-Stein shrinkage to the leading sample eigenvector, yielding a *James-Stein for eigenvectors* (JSE) estimate of the leading population eigenvector.

We advance the literature in three ways. First, we provide easy-to-code formulas for factor-based covariance matrices that are tailored to specific quadratic optimization problems. When input to optimization, these matrices generate low-variance instances of portfolios that satisfy optimization constraints. The low variance stems from JSE shrinkage of the leading sample eigenvector toward the target subspace generated by optimization constraint gradients, which neutralizes a component of estimation error that is egregiously amplified by optimization. JSE stochastically dominates the leading sample eigenvector as an estimate of ground truth.
The second advancement is a formula for improvement of JSE over the sample leading eigenvector that depends only sample eigenvalues and the angle between the leading population eigenvector and the target subspace. The third advancement concerns the nature of JSE shrinkage. In previous studies, JSE shrinkage is in a specific direction. For the problems considered in this article, the ideal shrinkage target is the projection of the leading population eigenvector, which is unobservable, onto the target subspace. We show that a data-driven shrinkage target obtained by projecting the leading sample eigenvector is sufficient to guarantee reduced variance of the optimized portfolio.

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

2 Context and related literature

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

2.1 Factor models and optimization

The vast literature on the estimation of covariance matrices for use in portfolio construction begins with Markowitz [1952] and Markowitz [1959]. We do not attempt to survey that literature here, but highlight Sharpe [1963], who developed a one-factor or “single index” model. Rosenberg [1974] builds on Sharpe’s idea, establishing Barra’s industry standard fundamental factor models.


2.2 Random matrix theory and covariance matrix estimation

Empirical analysis of public equities in US and Global markets suggests that the leading eigenvalue of a sample stock or bond return covariance matrix tends to grow in proportion to the number of securities in the pool. This means that the widely-studied single-spiked covariance model, introduced in Johnstone [2001] may be an appropriate basis for estimating covariance matrices for portfolio optimization. This paper brings random matrix theory (RMT) to bear on the problem of covariance matrix estimation, with special attention to the limiting spectrum of eigenvalues in the high dimension high sample size (HH) regime, where the number of variables $p$ and the number of observations $n$ tend to infinity in proportion. This type of analysis, launched by Wigner [1955], Wigner [1958] and Marcenko and Pastur [1967] in the 1950s and 1960s, has spawned another vast literature. Applications of this work to financial covariance matrix estimation is in Menchero et al. [2012], Ledoit and Wolf [2012], El Karoui [2010], El Karoui [2013] Ledoit and Wolf [2017] and Wang and Fan [2017], to name a few examples. Donoho et al. [2018] study the effect of different objective functions on the optimal covariance matrix estimate, given that the eigenvectors of the estimate match the sample eigenvectors.

The results in this article rely on a different, less studied branch of RMT concerning high dimension and low sample size (HL), in which the number of variables tends to infinity while the number of observations stays fixed. The relevance of the HL regime to the analysis of scientific data was pioneered in Hall et al. [2005] and Ahn et al. [2007], and Aoshima et al. [2018] surveys results. Data analysis in the HH and HL regimes as well as the low dimension high sample size (LH) regime of classical statistics, is discussed in Jung and Marron [2009],

4
2.3 James-Stein shrinkage for averages and for eigenvectors

Stein [1956] and James and Stein [1961] show that in dimension greater than or equal to 3, the sample average is inadmissible: there is another estimator with lower mean-squared error. That superior estimator is known as James-Stein, and it is obtained by shrinking sample averages toward their collective mean. This work was extended by replacing the grand mean with arbitrary initial guesses in Efron and Morris [1975], and popularized by Efron and Morris [1977]. An overview of James-Stein type shrinkage estimation is in Fourdrinier et al. [2018].

Recent literature, including Shkolnik [2022] and Goldberg and Kercheval [2023], develop James Stein for eigenvectors (JSE). Structurally parallel to James-Stein for averages, JSE improves almost surely on the sample leading eigenvector as an estimate of ground truth when data follow a one-factor spiked model. Its lack of dependence on distributional assumptions distinguishes JSE from classical James Stein as well as most of random matrix theory. The properties of JSE are derived with strong laws of large numbers. Alternative proofs may, however, be constructed with concentration of measures arguments, described in Ball [1997] and applied in Bar and Wells [2023].

2.4 JSE and mean-variance optimization

Estimation error in a covariance matrix leads to optimized portfolios that are sub-optimal. A manifestation is excess variance in an optimized portfolio; see, for example, Klein and Bawa [1976], Jobson and Korkie [1980], Michaud [1989] and Bianchi et al. [2017].

JSE was developed in Goldberg et al. [2022], Goldberg et al. [2020] and Gurdogan and Kercheval [2022] for the purpose of improving optimized minimum variance portfolios. The development rests on a novel analysis of the way estimation error in a spiked covariance model is transmitted via mean-variance analysis. Those articles show that excess dispersion in the leading sample eigenvector contributes material errors in estimated minimum variance and its risk forecasts, and that JSE reduces those errors in the HL regime. In the present article, we show that the original results are a special case of a more general phenomenon. A constrained optimization exacerbates estimation error in the leading sample eigenvector in the direction of the
subspace spanned by constraint gradients. 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 portfolio optimization problem

Here, 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 \( \Sigma \) denote the \( p \times p \)-dimensional covariance matrix of security returns, assumed non-singular. Consider an optimization problem with \( k > 0 \) linear constraints,

\[
\begin{align*}
\min_w & \quad \frac{1}{2} w^\top \Sigma w \\
\text{subject to} & \quad C_1^\top w = a_1 \\
& \quad C_2^\top w = a_2 \\
& \quad \vdots \\
& \quad C_k^\top w = a_k
\end{align*}
\]

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

A simple, explicit formula provides the unique solution to (1) when the inputs to the problem are known. In finance, however, the covariance matrix \( \Sigma \) is never known, so the explicit formula provides a solution that is suboptimal. In what follows, we illuminate the mechanism by which estimation error in a covariance matrix corrupts optimized portfolios, provide estimates
of $\Sigma$ 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, rendering it unsuitable for general use in quadratic programs such as (1). As a synthesis of information from data, however, $S$ can serve as a source of spare parts for estimated empirically reasonable covariance matrices that can be used in optimization.

4 A JSE prescription for a customized, optimization-friendly estimate of $\Sigma$

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

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

4.1 Structure from a factor model

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

$$\Sigma = \eta^2 b b^\top + \delta^2 I,$$

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

We don’t observe $\Sigma$, 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 $\Sigma$, two variances, $\eta^2$ and $\delta^2$, and the unit vector of factor loadings $b$, with functions of eigenvalues and eigenvectors of $S$ in a way that leads to a relatively low variance solution.
to (1). We show in Section 5 that the last of the three estimates is the most consequential.

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

With \(tr(S)\) denoting the trace of the sample covariance matrix \(S\) and \(\lambda^2\) denoting its leading eigenvalue, define

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

the average of the non-zero eigenvalues of \(S\) that are less than \(\lambda^2\), and

\[
\psi^2 = \frac{\lambda^2 - \ell^2}{\lambda^2},
\]

the average leading relative eigengap.

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

\[
c_{JSE} = \frac{\ell^2}{\lambda^2(1 - |h_C|^2)}
\]

and define

\[
H_{JSE} = c_{JSE}h_C + (1 - c_{JSE})h.
\]

The James-Stein for eigenvectors (JSE) estimate\(^2\) of the true eigenvector \(b\) is the unit vector\(^3\)

\[
h_{JSE} = \frac{H_{JSE}}{|H_{JSE}|}.
\]

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

\[
\Sigma_{JSE} = (\lambda^2 - \ell^2)h_{JSE}h_{JSE}^\top + (n/p)\ell^2I.
\]

\(^2\)Formula (6) above is equivalent to formula [6] in Goldberg and Kercheval [2023]. That article and Shkolnik [2022] expose the strong parallel between JSE and classical James-Stein.

\(^3\)Formulas (5), (6) and (7) are identical to formulas (29), (30) and (31) in Section 5.1.4.
Formula (8) is the one-factor covariance matrix designed for use in quadratic optimization (1). Note that the dependence of $\Sigma^{\text{JSE}}$ on $C$ is through the factor loadings $h^{\text{JSE}}$ and not through the estimates of factor and specific variance.

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

### 4.3 The true variance of an optimized portfolio

The benefits of this construction are realized in the portfolio $w^{\text{JSE}}$ generated by (1) when $\Sigma$ is set to $\Sigma^{\text{JSE}}$. Let $\Sigma^{\text{PCA}}$ be the covariance matrix obtained by replacing $h^{\text{JSE}}$ with the sample leading eigenvector $h$ in (8), and let $w^{\text{PCA}}$ denote the portfolio generated by (1) when $\Sigma$ is set to $\Sigma^{\text{PCA}}$.

It will be shown in Theorem 5.7 that the ratio of the true variances of $w^{\text{JSE}}$ and $w^{\text{PCA}}$,

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

(9)
tends to zero as the number of assets grows. When returns to securities in a sufficiently large investment universe are governed by a one-factor model, $w^{\text{JSE}}$ is an improvement on $w^{\text{PCA}}$ by an arbitrarily large factor as measured by true variance.

### 5 JSE stochastically dominates PCA

The formulas in Section 4 prescribe the construction of a strategy-specific covariance matrix based on JSE for use in portfolio construction. Here, we describe in more detail the theory asymptotically guaranteeing that JSE improves eigenvector estimates and lowers variance of optimized portfolios, relative to PCA. We work in the context of a single-factor model. However, experimental results and parallel research efforts indicate that the asymptotic guarantees may apply for certain multi-factor models of practical interest.

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

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

### 5.1 JSE theorem for high-dimensional targets

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

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

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

#### 5.1.1 A one-factor model of returns

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

\[
r = \mu + \beta f + z,
\]

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

\[
r = \beta f + z. \tag{11}
\]

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

**Standing Assumptions.**

A1. The random variable \( f \) is non-zero almost surely, has mean zero, finite fourth moment, and variance \( \sigma^2 > 0 \).

A2. The random variables \( \{z_i^{(p)} : i = 1, 2, \ldots, p; p > 1\} \) are i.i.d. and have mean zero, finite fourth moment, and variance \( \delta^2 > 0 \).

A3. The vector sequence \( \{\beta^{(p)} : p > 1\} \) satisfies the following asymptotic non-degeneracy conditions:

a. The entries \( \beta_i \) of \( \beta \) are uniformly bounded:

\[
\sup_{i,p}\{|\beta_i^{(p)}| : i = 1, 2, \ldots, p; p > 1\} < \infty, \tag{12}
\]

and

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

\[
\lim_{p \to \infty} |\beta^{(p)}|^2/p = B^2 > 0. \tag{13}
\]

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

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

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

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

$$r = \eta bx + z. \tag{14}$$

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

A3’.a $\sup_{i,p} \{\rho |b_i^{(p)}|^2 : i = 1, \ldots, p; p > k\} < \infty$, and

A3’.b $\eta_p^2 / p$ 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, $\delta^2 I$:

$$\Sigma = \eta^2 bb^\top + \delta^2 I. \tag{15}$$

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

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

$$S = \lambda_1^2 hh^\top + \lambda_2^2 v_2 v_2^\top + \lambda_3^2 v_3 v_3^\top \cdots + \lambda_p^2 v_p v_p^\top \tag{16}$$

$\text{Formula (15) is identical to formula (2) in Section 4.1, the starting point of the prescription.}$
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 \( \lambda^2 \) has multiplicity one and \( S \) has rank \( n \).

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

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

Recall
\[ \psi_p^2 = \frac{\lambda^2 - \ell^2}{\lambda^2}. \]  

**Proposition 5.1.** Under assumptions A1 - A3, almost surely, the limits
\[ \theta_{\text{PCA}} = \lim_{p \to \infty} \angle(h, b) \quad \text{and} \quad \psi_\infty^2 = \lim_{p \to \infty} \psi_p^2 \]
exist and
\[ \cos \theta_{\text{PCA}} = \psi_\infty \in (0, 1). \]  

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

The random variable \( \psi_\infty \) can be expressed in terms of the relationship between the relative eigengap and the parameters of the factor model (14). Decomposing, from (14), the \( p \times n \) data matrix of returns \( Y \) into a sum of unobservable components, we have
\[ Y = \eta bX^\top + Z, \]
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 7.4 stated later.
Proposition 5.2. The relative eigengap $\psi_\infty$ is related to the parameters of the factor model by

$$
\psi_\infty^2 = \lim_{p \to \infty} \psi_p^2 = \lim_{p \to \infty} \frac{\chi^2 - \ell^2}{\chi^2} = \frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2} \approx \frac{p \sigma^2 B^2}{p \sigma^2 B^2 + p \delta^2/n}. 
$$

(21)

The term $\psi_\infty^2$, asymptotically equal to the square of the inner product $\langle h, b \rangle$, is a measure of the asymptotic PCA estimation error when using $h$ to estimate $b$. It is random because $|X|^2$ is random, but does not depend on the random matrix $Z$. The approximation symbol $\approx$ in (21) is justified by the fact that $E[|X|^2/n] = 1$ and $|X|^2/n \to 1$ 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 (21). It is the trace of the factor component of the population covariance matrix $\Sigma$, specified in (15), and can be viewed as the variance in the system attributable to the factor. The term $p \delta^2$ is the trace of the specific component of $\Sigma$, 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 5.2 says that the relative eigengap $\psi_\infty^2$ is approximated by a ratio of signal to signal plus $(1/n)$-scaled noise. The ratio on the right hand side of (21) cannot be observed, but it can be estimated in terms of the relative eigengap of $S$.

A consequence of Proposition 5.2 is that the term $\psi_\infty^2$ tends to 1 as $n \to \infty$. Therefore,

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

(22)

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

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

(23)

For a typical value $\delta/(\sigma B) = 4$, this means the angular error $\theta_{\text{PCA}}$ will remain significant even for $n$ as large as 1000 or more, well above the typical values seen in portfolio optimization.
5.1.3 Insight about the relationship between $h$ and $b$ from the perspective of an external reference subspace

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

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

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

**Theorem 5.3.** Suppose the angle $\angle(b, C)$ between $b$ and $C$ tends, as $p \to \infty$, to a limit

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

Then under assumptions A1 – A3, the limit

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

exists almost surely, and

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

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

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

and

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

This theorem is a generalization of Theorem 3.1 of Goldberg et al. [2022]. It implies, asymptotically almost surely, that $h$ is not orthogonal to $C$ if $b$ is not, but the angle $\angle(h, C)$ is greater than both $\angle(b, C)$ and $\angle(h, b)$. Intuitively, this suggests that shrinking $h$ toward $C$ might bring it closer to $b$. This turns out to be correct, as described next.
5.1.4 Shrinkage improves on the leading sample eigenvector $h$ as an estimate of the leading population eigenvector $b$

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

Recall the JSE shrinkage constant $c^{\text{JSE}}$ and estimator $h^{\text{JSE}}$ are defined by\(^5\)

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

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

and

\[
h^{\text{JSE}} = \frac{H^{\text{JSE}}}{|H^{\text{JSE}}|}. \quad (31)
\]

We can show that

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

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

We normalize $h^{\text{JSE}}$ 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

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

**Theorem 5.4.** Suppose the limit

\[
\Theta = \lim_{p \to \infty} \angle(b, C) \quad (34)
\]

exists.

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

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

\(^5\)Formulas (29), (30) and (31) are identical to formulas (5), (6) and (7) in Section 4.2.
exist almost surely, and the asymptotic improvement of $h^{\text{JSE}}$ over $h^{\text{PCA}}$ as an estimate of the leading population eigenvector is

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

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

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

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

$$
\psi^2_\infty = \frac{\phi^2_\infty}{1 + \phi^2_\infty}. \quad (37)
$$

One consequence of Theorem 5.4 is that the angle between $h^{\text{JSE}}$ and $h$ is strictly positive in the limit when $\Theta < \pi/2$. Notice also that this theorem is independent of any optimization problem.

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

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

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

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

almost surely.

The $k$-dimensional target space $C$ may arise in different ways. If chosen at random independently of $b$, we expect $C$ to be asymptotically orthogonal to $b$ as the dimension $p$ tends to infinity (see, for example, Hall et al. [2005] and Ahn et al. [2007]). In this $\Theta = \pi/2$ case, JSE provides no advantage.
The condition $\Theta < \pi/2$ has a Bayesian interpretation in which $C$ represents some mild prior information about the direction of $b$: namely, that $b$ is not orthogonal to $C$.

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

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

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

### 5.2 Estimating Constrained Minimum Variance

We return to the optimization problem (1),

$$\min_w \frac{1}{2} w^\top \Sigma w \quad (41)$$

subject to $C^\top w = a$,

introduced in Section 1, where we have written the constraints in matrix notation. The columns of the $p \times k$ matrix $C$ are the constraint gradients $C_j$ 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, and there are $k \geq 1$ linear constraints $C_1^\top w = a_1, \ldots, C_k^\top w = a_k$.

We apply the results in Section 5.1 to estimate a $p \times p$ covariance matrix $\Sigma = \Sigma^{\text{JSE}}$ for use in (41). The matrix $\Sigma^{\text{JSE}}$ depends on the constraint matrix $C$. Its core is $h^{\text{JSE}}$, 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 $\Sigma^{\text{JSE}}$ and $h^{\text{JSE}}$ on $C$ when possible, but the dependence of $\Sigma^{\text{JSE}}$ on $C$ is a central idea of this section.

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

We assume without loss of generality that the constraint gradient 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 $C^T w = a$ in the asymptotic limit, so we add the following two natural 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. for each $j = 1, \ldots, k$, 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:

$$\liminf_{p \to \infty} \det(C^T C)/p^k > 0. \tag{42}$$

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

The simplest example is the case of the fully invested portfolio, where $k = 1$, there is a single constraint $e^T w = 1$, where $e$ is the column of 1’s, and $C$ is the column matrix $e$. Since $|e|^2 = p$, A4 is satisfied; $C^T C$ is equal to the $1 \times 1$ matrix with determinant $p$, so A5 is satisfied.

5.2.2 Estimating $\Sigma^{\text{JSE}}$

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

$$\Sigma = \eta^2 b b^\top + \delta^2 I.$$  

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

Our estimates of $\eta^2$ and $\delta^2$ are guided, under our standing assumptions, by the relationships between the eigenvalues of $S$ and the factor model structure in the HL regime. These relationships have been identified in numerous sources; as described in Lemma 7.4 below, they are summarized by the almost sure limits

$$\lim_{p \to \infty} (\lambda^2 - \ell^2)/p = \sigma^2 B^2 |X|^2/n \quad \text{(43)}$$

and

$$\lim_{p \to \infty} \ell^2/p = \delta^2/n. \quad \text{(44)}$$

Recall from assumption A3'b 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 $\eta^2$ with $\lambda^2 - \ell^2$. Noting (44), we estimate $\delta^2$ with $n\ell^2/p$. Both $\lambda^2$ and $\ell^2$ are observable from the eigenvalues of the sample covariance matrix $S$. We therefore have an estimated covariance matrix, depending on the choice of unit vector $v$, of the form

$$\Sigma^v = (\lambda^2 - \ell^2) vv^\top + (n/p)\ell^2 I. \quad \text{(45)}$$

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

### 5.2.3 Variance and the optimization bias

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

20
true parameter estimate(s)

\begin{align*}
\eta^2 & \quad \lambda^2 - \ell^2 \\
\delta^2 & \quad n\ell^2/p \\
b & \quad v, h, h^{\text{JSE}}
\end{align*}

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

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

\[
L(w, \Lambda) = (1/2)w^\top \Sigma^v + (a^\top - w^\top C)\Lambda,
\]

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

\[
\begin{align*}
\Lambda^v &= (C^\top (\Sigma^v)^{-1} C)^{-1} a \\
w^v &= (\Sigma^v)^{-1} C \Lambda^v = (\Sigma^v)^{-1} C(C^\top (\Sigma^v)^{-1} C)^{-1} a.
\end{align*}
\]

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

Since \( C \) has rank \( k \), the \( k \times k \) matrix \( C^\top C \) is invertible, so we may define the \( k \times p \) Penrose 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 5.1.** For any nonzero \( a \in \mathbb{R}^k \) and unit vector \( v \in \mathbb{R}^p \) satisfying

\[
|v_C| = \cos(v, C) < 1,
\]

define the unit vector

\[
\alpha = \frac{(C^\dagger)^\top a}{|(C^\dagger)^\top a|}
\]

and define the optimization bias associated to \( v, C, \) and \( a \) by

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

where, as usual, \( b \) denotes the leading population unit eigenvector.
The optimization bias does not depend on the magnitude of $a$, but only on $\alpha$ and the subspace $\text{col}(C)$, and is equal to zero when $v = b$:

$$\mathcal{E}(b, C, a) = 0.$$  \hfill (52)

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

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

which agrees with the optimization bias originally introduced for this case in Goldberg et al. [2022].

**Proposition 5.5.** Let $C, h$ be as above and let $h_C$ denote the orthogonal projection of $h$ onto $C$. If $0 < \Theta < \pi/2$, then

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

and

$$\limsup_{p \to \infty} |(h^\text{JSE})_C| < 1.$$  \hfill (55)

**Theorem 5.6.** Under assumptions A1-A5 above, let $v \in \mathbb{R}^p$ be a unit vector for each $p$ and satisfying

$$\limsup_{p \to \infty} |v_C| < 1.$$  \hfill (56)

Then, for $n, k$ fixed,

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

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

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

asymptotically as $p \to \infty$. 

22
Because of Proposition 5.5, Theorem 5.6 applies to both \( v = h \) and \( v = h^{JSE} \). 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 \( \Sigma^{JSE} \) dominates \( \Sigma^{PCA} \) as measured by the value of the true variance of the estimated portfolios \( w^{JSE} \) and \( w^{PCA} \).

**Theorem 5.7.** Under assumptions A1-A5 above, suppose also that the angle between \( b \) and \( \text{col}(C) \) is asymptotically between 0 and \( \pi/2 \).

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

\[
\lim_{p \to \infty} \cos(\angle(b, (C^\dagger)^\top a)) = \lim_{p \to \infty} \langle b, \alpha \rangle \equiv \langle b, \alpha \rangle_\infty \quad \text{exists.} \tag{59}
\]

Then, almost surely,

\[
\lim_{p \to \infty} \mathcal{E}_p(h^{JSE}, C, a)^2 = 0. \tag{60}
\]

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

\[
\lim_{p \to \infty} \mathcal{E}_p(h, C, a)^2 > 0. \tag{61}
\]

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

\[
\frac{\mathcal{V}(w^{JSE})}{\mathcal{V}(w^{PCA})} \tag{62}
\]

tends to zero asymptotically.

The previous two theorems tell us that \( \mathcal{V}(w^{\text{TRUE}}) \) and \( \mathcal{V}(w^{JSE}) \) are of asymptotic order \( 1/p \), but \( \mathcal{V}(w^{PCA}) \) 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 5.8.** Assume A1-A5 and that the limiting angle \( \Theta \) is less than \( \pi/2 \). Suppose \( a \) does not belong to the orthogonal complement of the unit vector

\[
\lim_{p \to \infty} \frac{C^\dagger b}{|C^\dagger b|} \in \mathbb{R}^k. \tag{63}
\]

Then \( \langle b, \alpha \rangle_\infty \) is not zero.
In this section, we provide numerical examples supporting the results stated above. First, we illustrate (36), which asserts the stochastic dominance of the improvement of \( h_{JSE} \) over \( h_{PCA} \) as an estimate of the leading population eigenvector. Then we illustrate the assertion that the ratio of variances of portfolios \( w_{JSE} \) and \( w_{PCA} \) tends to zero asymptotically almost surely.

### 6.1 Calibration

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

\[
r = \beta f + z,
\]

the \( p \times k \) matrix of constraint gradients \( C \) and \( k \) vector of constraint targets \( a \).

We construct \( \beta \) so that the angle \( \theta \) with \( 1 \) is a prescribed value and \( |\beta|^2/p = 1 \). First draw the components of a vector \( \beta^* \) from the normal
distribution \(N(\cos \theta, \sin^2 \theta)\). Let \(m = m(\beta^*)\) be the realized mean of the entries of \(\beta^*\), and \(s = s(\beta^*)\) the realized standard deviation. Define
\[
c_1 = \frac{\sin \theta}{s} \quad \text{and} \quad c_2 = \cos \theta - \frac{\sin \gamma}{s} m, \tag{64}
\]
and let
\[
\beta = c_1 \beta^* + c_2 \mathbf{1}. \tag{65}
\]
Making use of the identity
\[
|\beta|^2 = p(m(\beta)^2 + s(\beta)^2), \tag{66}
\]
a calculation shows that \(|\beta|^2/p = 1\) and the angle between \(\beta\) and \(\mathbf{1}\) is exactly \(\theta\). Even though the factor loadings \(\beta\) 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 \(\sigma\) to be 16%, and specific return \(z\) to be normally distributed with mean 0 and annualized standard deviation \(\delta\) to be 60%.

Next, we construct an expect return vector \(\mu\) so that
\[
\mu_i = \beta_i + \epsilon_i
\]
where \(\epsilon_i\) 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 \(\mu\) and \(\mathbf{1}\). The angle \(\Theta\) between \(\beta\) and \(C\) is determined by the specification of \(\beta\) and \(\mu\). The 2-vector of constraints targets is \(a = (1, m)^\top\).

Simulation parameters are listed in Table 2.

### 6.2 Stochastic dominance of \(h^{\text{JSE}}\) over \(h^{\text{PCA}}\)

Under Standing Assumptions A1–A3, formula (36) provides an exact expression for the difference between the squared cosines of \(\theta^{\text{PCA}}\) and \(\theta^{\text{JSE}}\):
\[
\cos^2(\theta^{\text{JSE}}) - \cos^2(\theta^{\text{PCA}}) = \left(\frac{1}{\phi^2 + 1}\right) \frac{\cos^2 \Theta}{\phi^2 \sin^2 \Theta + 1}.
\]
This magic formula for the limiting difference between angles \(\angle(\beta, h^{\text{PCA}})\) and \(\angle(\beta, h^{\text{JSE}})\) 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 $\phi^2$.

How well does the asymptotic guidance provided by the magic formula work for finite $p^2$? For $p = 3000$, we report

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

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 $\Theta$ increased. This is consistent with the asymptotic guidance given by the magic formula, which is decreasing in $\Theta$. Even though the asymptotic improvement of JSE over PCA is monotonic in $\Theta$ and distribution independent, the same cannot be said for finite $p$. Further research is required to determine parameter specifications for which monotonicity holds.

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

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

There is much to ponder in Figure 2. While the variance of $w^{JSE}$ is uniformly lower than the variance of $w^{PCA}$, consistent with (62) of Theorem 5.7, the dependence of the degree of improvement of variance ratios on $\Theta$ is not monotonic. Our theory guarantees that $\mathcal{V}(w^{JSE})/\mathcal{V}(w^{PCA})$ and $\mathcal{V}(w^{TRUE})/\mathcal{V}(w^{PCA})$ tend to 0 almost surely as $p$ tends to infinity. The asymptotic behavior of $\mathcal{V}(w^{TRUE})/\mathcal{V}(w^{JSE})$ is not known theoretically, but experiments suggest it may be close to 1 when the angle $\Theta$ between $b$ and $C$ is small.

7 Proofs

We provide mathematical arguments verifying the results stated above. We begin with some preliminary results needed for the subsequent proofs.
Figure 1: Boxplots for $p = 3000$ of 10,000 simulations of the difference between $\cos^2 (\angle(h_{PCA}, b))$ and $\cos^2 (\angle(h_{JSE}, b))$ (finite difference), the asymptotic limit of this difference (magic formula) as well as the path-by-path difference between them (difference). The small, medium and large panels correspond to $\cos \Theta = 0.969, 0.707$ and 0.174, Return data follow (11) with parameters specified in Table 2.
Figure 2: Boxplots for 10,000 simulations of ratios of variances of optimized and optimal portfolios, $w_{\text{PCA}}$, $w_{\text{JSE}}$ and $w_{\text{TRUE}}$, for $p = 3000$. The small, medium and large panels correspond to $\cos \Theta = 0.969, 0.707$ and $0.174$. The expected return target is $m = 0.01$. Return data follow (11) with parameters specified in Table 2.
Lemma 7.1 (Triangular Strong Law of Large Numbers, Tao [2015]). Let 
\((X_{i,p})_{i,p \in \mathbb{N}, i \leq p}\) be a triangular array of scalar random variables such that for each \(p\), the row \(X_{1,p}, \ldots, X_{p,p}\) is a collection of independent random variables. For each \(p\), define the partial sum \(S_p = X_{1,p} + \cdots + X_{p,p}\). Assume all the \(X_{i,p}\) have mean \(\mu\).

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

Lemma 7.2. Let \(\{z_i : i \in \mathbb{N}\}\) be a sequence of independent mean-zero random variables with bounded fourth moments, and let \(Z^{(p)} = (z_1, \ldots, z_p) \in \mathbb{R}^p\) for each \(p\). Let \(u^{(p)} \in \mathbb{R}^p\), \(p \in \mathbb{N}\), be a sequence of unit vectors and let \(u_i^{(p)}\) denote the \(i\)-th coordinate of \(u^{(p)}\).

Assume \(\sup_{i,p} \{p|u_i^{(p)}|^2 : i = 1, \ldots, p; p \in \mathbb{N}\} < \infty\). (67)

Then
\[
\frac{u^{(p)\top} Z^{(p)}}{\sqrt{p}} \to 0
\] (68)

almost surely as \(p \to \infty\).

Proof of Lemma 7.2. Apply Lemma 7.1 with \(X_{i,p} = \sqrt{p}u_i^{(p)}z_i\) and \(S_p = X_{1,p} + \cdots + X_{p,p}\). By the assumptions, the \(X_{i,p}\) have mean zero and uniformly bounded fourth moments. By Lemma 7.1 with \(\mu = 0\),
\[
\frac{u^{(p)\top} Z^{(p)}}{\sqrt{p}} = \frac{1}{p} S_p
\] (69)

converges to zero almost surely.

The following is a version of Proposition 5.2 in Gurdogan and Kercheval [2021], which remains true with a similar proof under our slightly adapted hypotheses:

**Proposition 7.3.** 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. \(\lim_{p \to \infty} \left( \langle h, h_L \rangle - \langle h, b \rangle^2 \langle b, b_L \rangle \right) = 0\),
2. \( \lim_{p \to \infty} \left( \langle b, h_L \rangle - \langle h, b \rangle \langle b, b_L \rangle \right) = 0 \), 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 \( p \to \infty \).

7.1 Proof of Proposition 5.1

Proposition 5.1. Under assumptions A1 - A3, the limits

\[ \theta_{\text{PCA}} = \lim_{p \to \infty} \angle(h, b) \quad \text{and} \quad \psi_\infty^2 = \lim_{p \to \infty} \psi_p^2 \]  \tag{70}

exist almost surely, and

\[ \cos \theta_{\text{PCA}} = \psi_\infty \in (0, 1). \]  \tag{71}

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 \( \lambda_2 \).

Define \( \chi = \chi_p \in \mathbb{R}^n \) such that \( h \) and \( \chi \) 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 \( \chi \) 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 \( h, \chi \), and \( Y \) are related by

\[ \lambda h = Y \chi/\sqrt{n}, \]  \tag{72}

by equation (20) we have

\[ \lambda h = \frac{\eta b^\top \chi + Z \chi}{\sqrt{n}}. \]  \tag{73}

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{74}

\[ \lambda^2/p = \frac{\eta^2 (X^\top \chi)^2}{np} + \frac{X^\top Z^\top Z \chi}{np} + 2(X^\top \chi) \left( \frac{b^\top Z}{\sqrt{p}} \right) \left( \frac{\eta \chi}{n \sqrt{p}} \right). \]  \tag{75}
Applying the independence and distributional assumptions on $Z$ with the strong law of large numbers, we may deduce that $Z^\top Z/p$ tends almost surely to $\delta^2 I$ as $p \to \infty$. This fact, the boundedness of $\eta^2/p$, and Lemma 7.2 applied to $b$ in equation (75) show that $\lambda^2/p$ is eventually bounded between zero and infinity almost surely. Applying Lemma 7.2 to $b$ in the last term of equation (74), we obtain

$$\langle h, b \rangle_\infty = \lim_{p \to \infty} \left( \frac{\eta X^\top X}{\sqrt{\lambda n}} \right) \quad (76)$$

provided the limit in (76) exists almost surely.

Recall $\ell_p^2$ is the average of the non-zero sample eigenvalues less than $\lambda^2$. The proof of the following Lemma is essentially identical to the proof of Lemma A.2 of Goldberg et al. [2022]:

**Lemma 7.4.** Under assumptions A1 - A3 and notation as above, we have the following limits almost surely:

$$\lim_{p \to \infty} \lambda^2/p = \sigma^2 B^2 |X|^2/n + \delta^2/n, \quad (77)$$

$$\lim_{p \to \infty} \chi_p = X/|X|, \text{ and} \quad (78)$$

$$\lim_{p \to \infty} \ell_p^2/p = \delta^2/n. \quad (79)$$

Applying Lemma 7.4 to (76), we obtain

$$\langle h, b \rangle_\infty = \lim_{p \to \infty} \frac{\eta X^\top X}{\lambda \sqrt{n}} = \lim_{p \to \infty} \left( \frac{\eta}{\sqrt{p}} \right) \left( \frac{\sqrt{p}}{\lambda} \right) \left( \frac{X^\top X}{\sqrt{n}} \right) \quad (80)$$

$$= \sigma B \left( \frac{1}{\sqrt{\sigma^2 B^2 |X|^2/n + \delta^2/n}} \right) \left( \frac{|X|}{\sqrt{n}} \right) \quad (81)$$

$$= \sqrt{\frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2}} \in (0, 1). \quad (82)$$

By Lemma 7.4,

$$\psi_p^2 = \frac{\lambda^2 - \ell_p^2}{\lambda^2}, \quad (83)$$

converges to

$$\psi_\infty^2 = \frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2} \quad (84)$$

and hence $\langle h, b \rangle_\infty = \psi_\infty$. This completes the proof of Proposition 5.1.
7.2 Proof of Theorem 5.3

Theorem 5.3. Suppose the angle $\angle(b, C)$ between $b$ and $C$ tends, as $p \to \infty$, to a limit

$$\Theta = \lim_{p \to \infty} \angle(b, C).$$  \hfill (85)

Then under assumptions A1 - A3, the limit

$$\Theta^h = \lim_{p \to \infty} \angle(h, C)$$  \hfill (86)

exists almost surely, and

$$\cos \Theta^h = \cos \theta^{PCA} \cdot \cos \Theta = \psi_\infty \cdot \cos \Theta.$$  \hfill (87)

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

$$0 < \cos \Theta^h < \cos \theta^{PCA}$$  \hfill (88)

and

$$0 < \cos \Theta^h < \cos \Theta.$$  \hfill (89)

Proof. We apply Proposition 7.3(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 5.1 and $\cos \angle(b, C) \to \cos \Theta$ by hypothesis, equation (87) follows immediately.  \hfill $\blacksquare$

7.3 Proof of Theorem 5.4

Theorem 5.4. With notation as above, suppose the limit

$$\Theta = \lim_{p \to \infty} \angle(b, C)$$  \hfill (90)

exists.

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

$$\theta^{JSE} = \lim_{p \to \infty} \angle(h^{JSE}, \beta) \quad \text{and} \quad \theta^{PCA} = \lim_{p \to \infty} \angle(h^{PCA}, \beta)$$  \hfill (91)

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

$$\cos^2(\theta^{JSE}) - \cos^2(\theta^{PCA}) = \left(\frac{1}{\phi_\infty^2 + 1}\right) \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1}.$$  \hfill (92)
If $\Theta = \pi/2$, then $h^{\text{JSE}}$ converges to $h^{\text{PCA}}$, $\theta^{\text{JSE}} = \theta^{\text{PCA}}$ and there is no improvement, while if $\Theta = 0$ then $h^{\text{JSE}}$ converges to $b$. In other cases, $\theta^{\text{JSE}} < \theta^{\text{PCA}}$ almost surely, with the improvement given by (36).

Proof. The existence of the limit $\theta^{\text{PCA}}$ has already been established in Proposition 5.1. The JSE estimator $h^{\text{JSE}}$ relative to the subspace $C$ is an example of the “MAPS” estimator defined and studied in Gurdogan and Kercheval [2022]. 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

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

and define the unit vector

$$h^o = \frac{b_U}{|b_U|}. \tag{93}$$

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 $h^{\text{JSE}}$, 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 [2022], for slightly adjusted assumptions:

**Proposition 7.5.** Under assumptions A1 - A3,

$$\lim_{p \to \infty} |h^o - h^{\text{JSE}}| = 0 \tag{94}$$

almost surely.

Next, let

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

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

$$b_U = b_C + \langle b, u \rangle u,$$
and so
\[ \langle h^o, b \rangle^2 = \left\langle \frac{b_U}{|b_U|}, b \right\rangle^2 = |b_U|^2 \]  
(95)
\[ = |b_C|^2 + \langle u, b \rangle^2 \]  
(96)
\[ = |b_C|^2 + \frac{(\langle h, b \rangle - \langle h_C, b \rangle)^2}{1 - |h_C|^2}. \]  
(97)

All the terms in the right hand side have previously been shown to have limits as \( p \to \infty \):  
\[ |b_C|^2 \to \cos^2 \Theta, \]  
(98)
\[ |h_C|^2 \to \psi_\infty^2 \cos^2 \Theta, \]  
(99)
\[ \langle h, b \rangle \to \psi_\infty = \cos \theta_{\text{PCA}}, \]  
(100)
\[ \langle h_C, b \rangle \to \psi_\infty \cos^2 \Theta. \]  
(101)

Therefore \( \lim_{p \to \infty} \langle h^o, b \rangle^2 \) exists and by Proposition 7.5,
\[ \lim_{p \to \infty} \langle h^o, b \rangle^2 = \lim_{p \to \infty} \langle h_{\text{JSE}}, b \rangle^2 = \cos^2 \theta_{\text{JSE}}. \]  
(102)

Writing \( \psi_\infty^2 = \psi^2 \) and \( \phi_\infty^2 = \phi^2 \) for the remainder of this proof only, and recalling
\[ \psi^2 = \frac{\phi^2}{1 + \phi^2}, \]  
in the limit,
\[ \cos^2 \theta_{\text{JSE}} - \cos^2 \theta_{\text{PCA}} = \cos^2 \Theta + \frac{\psi^2(1 - \cos^2 \Theta)^2}{1 - \psi^2 \cos^2 \Theta} - \psi^2 \]  
(103)
\[ = (1 - \psi^2)^2 \frac{\cos^2 \Theta}{1 - \psi^2 \cos^2 \Theta} \]  
(104)
\[ = \left( \frac{1}{\phi^2 + 1} \right) \frac{\cos^2 \Theta}{\phi^2 \sin^2 \Theta + 1}. \]  
(105)

This is positive when \( \Theta < \pi/2 \). In case \( \Theta = \pi/2 \), Theorem 5.3 implies that \( h_C \) tends to zero and \( h_{\text{JSE}} \) tends to \( h = h_{\text{PCA}} \), so \( \theta_{\text{JSE}} = \theta_{\text{PCA}} \) and JSE provides no improvement over PCA.

If \( \Theta = 0 \), it follows from equation (103) that \( \theta_{\text{JSE}} = 0 \) and so \( h_{\text{JSE}} \) tends to \( b \) itself.
7.4 Proof of Proposition 5.5

Proposition 5.5. Let $C, h$ be as above and $h_C$ denote the orthogonal projection of $h$ onto $C$. If $0 < \Theta < \pi/2$, then

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

and

$$\limsup_{p \to \infty} |(h^{JSE})_C| < 1.$$  \hfill (106)

\hfill (107)

\textbf{Proof.} From part 3 of Proposition 7.3 with $L = C$, we have, in the asymptotic limit,

$$|h_C|^2 = \langle h, b \rangle^2 |b_C|^2 = \psi^2 \infty |b_C|^2. \hfill (108)$$

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

We can write

$$h^{JSE} = \frac{\Gamma_p h + h_C}{|\Gamma_p h + h_C|} \hfill (109)$$

where

$$\Gamma_p = \frac{\psi_p^2 - |h_C|^2}{1 - \psi_p^2}. \hfill (110)$$

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

$$\liminf_{p \to \infty} \Gamma_p > 0. \hfill (111)$$

This follows from equation (108) and the standing assumption that the angle between $b$ and $C$ is asymptotically strictly between 0 and $\pi/2$.

7.5 Proof of Theorem 5.6

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

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

35
A2. The random variables \( \{z_i^{(p)} : i = 1,2,\ldots,p; p > 1\} \) are i.i.d. and have mean zero, finite fourth moment, and variance \( \delta^2 > 0. \)

A3. The vector sequence \( \{\beta^{(p)} : p > 1\} \) satisfies the following asymptotic non-degeneracy conditions:

a. \( \sup_{i,p} (|\beta_i^{(p)}| : i = 1,2,\ldots,p; p > 1) < \infty, \) and

b. the sequence \( |\beta^{(p)}|^2/p \) tends to a finite limit \( B^2 > 0 \) as \( p \to \infty. \)

A4. The constraint matrix \( C \in \mathbb{R}^{p \times k} \) satisfies the following asymptotic non-degeneracy conditions:

a. \( \sup_p \max\{|C_{ij}| : 1 \leq i \leq p, 1 \leq j \leq k\} < \infty, \) and

b. the sequence \( |C_j|^2/p \) tends to a finite limit as \( p \to \infty, \) for \( j = 1,\ldots,k. \)

A5. the constraint matrix \( C \) does not become singular in the high dimensional limit: \( \det(C^\top C)/p^k \) tends to a finite positive constant as \( p \to \infty. \)

Theorem 5.6. Under assumptions A1-A5 above, let \( v \in \mathbb{R}^p \) be a unit vector for each \( p \) and satisfying

\[
\limsup_{p \to \infty} |v_C| < 1. \tag{112}
\]

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

\[
\mathcal{V}(w^v) \equiv (w^v)^\top \Sigma w^v = \eta^2 |(C^\top)^\top a|^2 \mathcal{E}_p(v,C,a)^2 + O(1/p) \tag{113}
\]
asymptotically as \( p \to \infty. \)

Furthermore,

\[
0 < \limsup_{p \to \infty} \eta^2 |(C^\top)^\top a|^2 < \infty. \tag{114}
\]

Proof. Recall that

\[
\Sigma^v = (\lambda^2 - \ell^2)vv^\top + (n\ell^2/p)I,
\]
and define
\[ \kappa^2 = \frac{n\ell^2/p}{\lambda^2 - \ell^2}, \]
noting that \( \kappa^2 = O(1/p) \).

A computation making use of the Woodbury identity shows that
\[ w^v = \left( I + \frac{(v_C - v)v^\top}{1 + \kappa^2 - |v_C|^2} \right) (C^\dagger)^\top a. \tag{115} \]

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
\[ 0 < \limsup_{p \to \infty} \eta^2|\langle C^\dagger\rangle^\top a|^2 < \infty, \]
which establishes the last assertion of the theorem.

To obtain an expression for true variance, first notice that
\[ \mathcal{V}(w^v) = (w^v)^\top \Sigma w^v = \eta^2 \langle w^v, b \rangle^2 + \delta^2 |w^v|^2. \tag{116} \]

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 (115) and recalling
\[ \alpha = \langle C^\dagger \rangle^\top a/|\langle C^\dagger \rangle^\top a|, \quad \limsup_{p \to \infty} |v_C| < 1, \quad \text{and} \quad \kappa^2 = O(1/p), \tag{117} \]
we have
\[ \eta^2 \langle w^v, b \rangle^2 = \eta^2 |\langle C^\dagger \rangle^\top a|^2 \left( \langle b, \alpha \rangle - \frac{\langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 + \kappa^2 - |v_C|^2} \right)^2 \]
\[ = \eta^2 |\langle C^\dagger \rangle^\top a|^2 \left( \frac{\langle b, \alpha \rangle (1 - |v_C|^2) - \langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 - |v_C|^2} \right)^2 + O(1/p) \tag{118} \]
\[ = \eta^2 |\langle C^\dagger \rangle^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p). \tag{119} \]

\[ \mathcal{E}_p(v, C, a)^2 + O(1/p). \tag{120} \]
7.6 Proof of Theorem 5.7

Theorem 5.7. Under assumptions A1-A5 above, suppose also that the angle between \( b \) and \( \text{col}(C) \) is asymptotically between 0 and \( \pi/2 \).

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

\[
\lim_{p \to \infty} \cos(\angle(b, (C^\dagger)^\top a)) = \lim_{p \to \infty} \langle b, \alpha \rangle \equiv \langle b, \alpha \rangle_\infty \text{ exists.} \tag{121}
\]

Then, almost surely,

\[
\lim_{p \to \infty} \mathcal{E}_p(h^{\text{JSE}}, C, a)^2 = 0. \tag{122}
\]

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

\[
\lim_{p \to \infty} \mathcal{E}_p(h, C, a)^2 > 0. \tag{123}
\]

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

\[
\frac{\mathcal{V}(w^{\text{JSE}})}{\mathcal{V}(w^{\text{PCA}})} \tag{124}
\]

tends to zero asymptotically.

Proof. By Proposition 5.5, we know that

\[
\limsup_{p \to \infty} |v_C| < 1 \tag{125}
\]

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

\[
\mathcal{E}_p(v, C, a) = \frac{\langle b, \alpha \rangle (1 - |v_C|^2) - \langle b, v - v_C \rangle \langle v, \alpha \rangle}{1 - |v_C|^2}, \tag{126}
\]

stays away from zero in both cases. For the first statement (122) of the theorem, it then suffices to show that the numerator

\[
\langle b, \alpha \rangle (1 - |(h^{\text{JSE}})_C|^2) - \langle b, h^{\text{JSE}} - (h^{\text{JSE}})_C \rangle \langle h^{\text{JSE}}, \alpha \rangle \tag{127}
\]

vanishes asymptotically. In light of Proposition 7.5, it suffices to show that \( \mathcal{E}_p(h^o, C, a) = 0 \) for the oracle \( h^o = b_U/|b_U| \) defined previously, where \( U = \text{span}(h, C) \). This is a consequence of the fact that \( \langle b_C, \alpha \rangle = \langle b, \alpha \rangle \) and
following straightforward identities:

\[ \langle b, h^\alpha - (h^\alpha)_C \rangle = |b_U| - \frac{|b_C|^2}{|b_U|}, \quad (128) \]

\[ \langle (h^\alpha)_C, \alpha \rangle = \frac{\langle h, \alpha \rangle}{|b_U|}, \quad \text{and} \]

\[ |(h^\alpha)_C|^2 = \frac{|b_C|^2}{|b_U|^2}. \quad (130) \]

Turning to the second statement (123), first note that Proposition 7.3 applied to the subspace \( L = \text{span}(\alpha) \), implies, asymptotically, \( \langle h, \alpha \rangle = \langle h, b \rangle \langle b, \alpha \rangle \), where we omit the subscripts on \( \langle h, \alpha \rangle_\infty \), etc., to unclutter the notation. Also, setting \( L = C \) in the same proposition yields the asymptotic equalities \( |h_C|^2 = \langle h, b \rangle^2 |b_C|^2 \), and \( \langle b, h_C \rangle = \langle h, b \rangle \langle b, b_C \rangle \).

Making use of these facts and simplifying leads to

\[ \lim_{p \to \infty} E_p(h, C, a) = \langle b, \alpha \rangle \left( 1 - \langle h, b \rangle^2 \right) \]

\[ \frac{1 - \langle h, b \rangle^2 |b_C|^2}{1 - \psi^2_\infty |b_C|^2}. \quad (131) \]

\[ = \frac{\langle b, \alpha \rangle (1 - \psi^2_\infty)}{1 - \psi^2_\infty |b_C|^2}. \quad (132) \]

When \( E(h, C, a) \) is positive but \( E(h^{\text{JSE}}, C, a) \) tends to zero, equation (113) implies that \( \mathcal{V}(w^{\text{PCA}}) \) remains bounded above zero while \( \mathcal{V}(w^{\text{JSE}}) \) tends to zero. This establishes the last claim.

\[ \blacksquare \]

### 7.7 Proof of Lemma 5.8

**Lemma 5.8:** Assume A1-A5 and that the limiting angle \( \Theta \) is less than \( \pi/2 \). Suppose \( a \) does not belong to the orthogonal complement of the unit vector

\[ \lim_{p \to \infty} \frac{C^\dagger b}{|C^\dagger b|} \in \mathbb{R}^k. \quad (133) \]

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

We express the singular value decomposition of \( C \) as

\[ C^{(p)} = U^{(p)} Z^{(p)} V^{(p)}, \quad (134) \]

39
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 $s^2_j/p$ 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 $Z_\infty$ and $V_\infty$, respectively, where $V_\infty$ is orthogonal and $Z_\infty$ 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

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

and

$$\langle b, (C^\dagger)^\top a \rangle = \langle C^\dagger b, a \rangle = \langle Z^{-1}U^\top b, V a \rangle.$$  \hfill (136)

Hence

$$\frac{\langle b, (C^\dagger)^\top a \rangle}{\left| (C^\dagger)^\top a \right|} = \frac{\langle Z^{-1}U^\top b, V a \rangle}{\sqrt{\langle Z^{-2}V a, V a \rangle}} \to \frac{\langle Z_\infty^{-1}(U^\top b)_\infty, V_\infty a \rangle}{\sqrt{\langle Z_\infty^{-2}V_\infty a, V_\infty a \rangle}}.$$  \hfill (137)

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

\[\blacksquare\]

References


Bar, H. and Wells, M. T. [2023], ‘On graphical models and convex geometry’, forthcoming in *Computational Statistics and Data Analysis*.


Goldberg, L. R. and Kercheval, A. N. [2023], ‘James-Stein for the leading eigenvector’, *Proceedings of the National Academy of Sciences* 120(2).


Markowitz, H. M. [1959], Portfolio Selection: Efficient Diversification of Investments, Yale University Press.


Stein, C. [1956], Some problems in multivariate analysis part i, Technical report no. 6, Office of Naval Research.

