Finance and Stochastics
https://doi.org/10.1007/s00780-025-00566-4

Portfolio optimisation via strategy-specific eigenvector shrinkage

Lisa R. Goldberg[1,2] · **Hubeyb Gurdogan**[3] · **Alec Kercheval**[4]

Received: 29 August 2023 / Accepted: 15 October 2024

Abstract
We estimate covariance matrices that are tailored to portfolio optimisation constraints. We rely on a generalised version of James–Stein for eigenvectors (JSE), a data-driven operator that reduces estimation error in the leading sample eigenvector by shrinking towards a target subspace determined by constraint gradients. Unchecked, this error gives rise to excess volatility for optimised portfolios. Our results include a formula for the asymptotic improvement of JSE over the leading sample eigenvector as an estimate of ground truth, and provide improved optimal portfolio estimates when variance is to be minimised subject to finitely many linear constraints.

Keywords Eigenvector estimation · High dimension · Portfolio optimisation · James–Stein

Mathematics Subject Classification 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 characterised as variance, and expected or

✉ L.R. Goldberg
lrg@berkeley.edu

A. Kercheval
akercheval@fsu.edu

[1] Department of Economics and Consortium for Data Analytics in Risk, UC Berkeley, Berkeley, CA 94720, USA

[2] Aperio Group LLC, a wholly-owned indirect subsidiary of BlackRock, Inc., New York, USA

[3] Department of Mathematics, UCLA, Los Angeles, CA 90024, USA

[4] Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA

Published online: 21 May 2025

mean return. A standard tool for asset allocation, for constructing quantitative exchange traded funds, mutual funds and active strategies, and for customising separately managed accounts, Markowitz's optimisation remains the workhorse of financial services today. Mean–variance optimised portfolios are efficient in the sense that they minimise 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 optimisation, 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 optimisation, Markowitz [51] 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 optimisation 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 optimised 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 conditioned well enough for use in optimisation. Principal component analysis (PCA) can be used to identify factors that explain correlation, for example in the arbitrage pricing theory developed by Ross [61] in 1976. The factor loadings are sample eigenvectors, linear combinations of security returns that maximise 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 optimisation, sample eigenvectors lead to estimated optimised portfolios with variances that tend to be 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 optimised portfolios. Extending recent research that sheds light on how estimation error is transmitted via optimisation, 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 optimisation problems with multiple linear constraints. By neutralising the component of estimation error that is amplifed in optimisation, our methods produce relatively low-variance instances of portfolios satisfying optimisation 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, which in 2024 accounted for more than \$24 billion in assets. That robust 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. The single-constraint minimum-variance portfolio featured in the literature is rarely used because it is extremely sensitive to small changes in the covariance matrix and in particular exhibits highly variable leverage.

In practice, virtually all quantitatively constructed *investable* portfolios include numerous constraints, which stabilise the behaviour over time. These constraints may impose strategy considerations such as a return target or factor tilt, but they also take account of realistic considerations such as leverage, turnover and transaction costs. This fact underscores the importance of one of the central premises of our paper: it is important to go beyond the global minimum-variance portfolio.

The second advancement is a new asymptotic formula for improvement of JSE over the leading sample 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 characterises JSE's asymptotic stochastic dominance over PCA and opens the way to a 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 towards a known fixed direction. To account for multiple constraints, we generalise 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.

A distinguishing feature of this article and the works on which we build is an analysis of how estimation error is transmitted by optimisation. In cases of practical importance, errors in eigenvectors substantially distort optimised portfolios, while errors in eigenvalues may be less important.

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 optimised portfolio. Beyond a finite fourth moment, none of our theoretical results rely on parametric distributional assumptions on the underlying data.

In Sect. 2, we review some background and literature relevant to our results. In Sect. 3, we set up the problem of finding a low-variance solution to mean–variance optimisation with linear constraints when the covariance matrix is estimated. Readers interested in the bottom-line formulas for implementation will find them summarised in Sect. 3.2, while Sect. 4 provides a detailed mathematical discussion of the construction and describes its asymptotic properties. Numerical experiments illustrating our results are in Sect. 5, and Sect. 6 contains concluding thoughts. Mathematical proofs are in the Appendix.

2 Financial and statistical context

The use of covariance matrices in portfolio construction dates back to Markowitz [51], [52, Chap. 5] in the 1950s. Effective estimation of the high-dimensional covariance matrices required by Markowitz's mean–variance optimisation rests on an extensive 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 Spearman [67], factor models provide a framework for analysing 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, Sharpe [63] 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 [60] and [59], which develop multi-factor models based on cross-sectional regressions and form 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 [61] and developed in Chamberlain and Rothschild [10], Connor [12] and Connor and Korajczyk [14, 16] 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 [13] and Connor and Korajczyk [15] review roles of different types of factor models in finance.

The results in this paper are framed in terms of a latent, single-factor model, which allows heterogeneous 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 allowing heterogeneous variances and correlations for specific returns expands 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 [73, 74] and Stein [68]. In the 1960s, Marchenko and Pastur [50] characterised distributions of the eigenvalues of covariance matrices of standard Gaussian variables as the number n of observations and the number p of parameters tend to infinity in proportion. This work spawned a large literature identifying and correcting biases in high-dimensional eigenvalues when p and n tend to infinity. We denote this asymptotic setting by HH for "high dimension high sample size" and refer to Bai [3], Edelman and Rao [19], Bai and Silverstein [4, Chap. 3], Tao [70, Chap. 2] and Paul [58] for more information.

In the 2000s, Hall et al. [36] and Ahn et al. [1] explored a different asymptotic framework in which p tends to infinity while n stays fixed. This asymptotic regime, which we denote by HL for "high dimension low sample size", is surveyed in the 2018 article by Aoshima et al. [2] and 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 n tend to infinity as p stays fixed, the LH regime. Results on random matrices can be organised around LH, HH and HL, as discussed for example in Jung and Marron [42] and Goldberg and Kercheval [30]. 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 optimised 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, as 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 [71] 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 discussed in 2001 by Johnstone [40] and further studied by Johnstone and Lu [41] and Donoho et al. [18]. For more analyses of consistency of sample eigenvalues and eigenvectors in high dimension, see Paul's 2007 article [57], the 2013 article by Fan et al. [25] and the 2016 article by Shen et al. [64].

In our HL factor model setting, we have $\lambda^2 n/p$ tending to a finite limit due to the prevalence condition on beta discussed below after Assumption 4.1, 3), so that $\lambda^2 n/p$ is bounded. In this setting, 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 4.4 below). This is the JSE correction, which makes use of laws of large numbers and concentration of measure (see Ball [5]).

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 [68] and James and Stein [38] 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 towards their collective average. This work was extended by replacing the collective average with arbitrary initial guesses in Efron and Morris [20], and popularised by Efron [21]. An overview of James–Stein type shrinkage estimation is in Foudrinier et al. [26, Chap. 2].

Recent literature, including Shkolnik [65] and Goldberg et al. [30], develops James–Stein for eigenvectors (JSE). Structurally identical to James–Stein for averages, JSE improves almost surely on the leading sample 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 optimisation

Our work has roots in two streams of literature that explain how attributes of a covariance matrix are propagated by optimisation. The first considers how estimation error in a covariance matrix leads to optimised portfolios that are suboptimal. A manifestation is excess variance in an optimised portfolio; see for example Klein and Bawa [43], Jobson and Korkie [39], Michaud [56] and Bianchi et al. [6]. In her 2010 and 2013 articles [22] and [23], El Karoui documents how risk of optimised portfolios is underforecast by covariance matrices estimated using methods from the HH regime.

The second stream begins with Green and Hollifield's 1992 article [34], which explains how dispersion in exposures of a dominant factor can generate concentration in an optimised portfolio. In 2003, Jagannathan and Ma [37] 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 et al. [11] 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 [34] argue that estimation error is not the cause of the concentration in optimised portfolios—, they are nevertheless foundational to a large literature that attempts to mitigate estimation error with shrinkage.

Ledoit and Wolf develop schemes for constructing well-conditioned security return covariance matrices suitable for use in optimisation. In 2003 and 2004, they 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 either (a) a single index matrix [44], (b) a constant correlation matrix [46] or (c) a scalar matrix [45]. In 2012, relying on guidance from the HH regime, Ledoit and Wolf [47] show that shrinkage of a sample covariance matrix towards a scalar amounts to linear shrinkage of sample eigenvalues towards their grand mean while preserving sample eigenvectors. They apply nonlinear 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 et al. [55] use guidance from the HH-regime to adjust sample eigenvalues of a covariance matrix with simulation. In their 2017 article, Ledoit and Wolf [48], like many other researchers, compare realised variance and information ratios of single-constraint minimum-variance portfolios constructed with different covariance matrices, some based on the nonlinear shrinkage of eigenvalues from [47]. In a lucid 2024 discussion of out-of-sample tests of covariance matrices developed for optimisation, Menchero and Lazanas [54] 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 et al. [24, 25], in 2017 by Wang and Fan [71], and in 2021 by Ding et al. [17]. In their 2018 article, Bodnar et al. [8] apply shrinkage to the weights of a minimum-variance portfolio optimised with a sample covariance matrix. In 2021, the results are extended by Bodnar et al. [7] to include estimates of security means.

Recent works, including those above, share several common themes. First, they attempt to correct estimated eigenvalues, but still use the sample eigenvectors. In the language of [18], these covariance matrix estimates are “orthogonally-equivariant.” Ledoit and Wolf call them “rotationally equivariant.” With the exception of [71], 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 the present article rely on distribution-free eigenvector shrinkage in the HL regime, and can be customised to any quadratic minimisation with linear constraints. James–Stein for eigenvectors was developed in Goldberg et al. [32], Goldberg et al. [31] and Gurdogan and Kercheval [35] for the purpose of improving optimised 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 contribute 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 optimisation exacerbates estimation error in the leading sample eigenvector in the direction of the subspace spanned by constraint vectors. By shrinking the leading sample eigenvector towards that subspace, we correct the leading eigenvector in a way that is tailored to the constrained optimisation problem, leading to improved results.

2.6 Constraints, risk factors and estimation error

There is an extensive literature that looks at the interaction between constraints and risk factors in an optimised portfolio without considering estimation error. In their 2003 article, Jagannathan and Ma [37] show an equivalence between a fully-invested, long-only, position-limited quadratic optimisation and an optimisation with a shrunken covariance matrix subject only to the full-investment constraint. This result foreshadows the "robustification" of the simplest Markowitz optimisation problems, a topic that is explored in generality in the 2024 paper [9] by Boyd, Johansson, Kahn, Schiele and Schmelzer. The 2008 article by Lee and Stefek [49] and the 2012 article by Saxena and Stubbs [62], with insightful commentary by Markowitz [53], look at problems associated with the misalignment of alpha constraints and risk factors. In contrast, Garvey et al. [29] argue in their 2017 article for the benefits of complete misalignment: alpha constraints that are orthogonal to factors. Consider this against the backdrop of Ross's 1976 paper [61] showing that in an idealised setting, alpha orthogonal to factors must asymptotically imply arbitrage opportunities.

Robust optimisation takes account of uncertainty around inputs. In a 2007 article that is widely cited by academics and also used in industry, Garlappi et al. [28] take account of uncertainty around expected returns in an optimised Markowitz portfolio. See the 2020 article by Xidonas et al. [75] for a survey of some of the applications of robust optimisation to portfolio construction. In a 2024 article, Shkolnik et al. [66] begin to analyse the interaction between estimation errors in constraints and risk factors using James–Stein type shrinkage methods, as in the present article.

3 The optimisation problem and a JSE prescription

3.1 Constrained optimisation

We specify the central problem addressed in this article: finding low-variance solutions to variance-minimising optimisation 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 the 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 $p \times p$ covariance matrix of security returns, assumed nonsingular. Consider an optimisation problem with $k > 0$ linear constraints, namely

$$\begin{aligned} \min_{w} \frac{1}{2} w^\top \Sigma w \\ \text{subject to } C_1^\top w &= a_1, \\ C_2^\top w &= a_2, \\ &\vdots \\ C_k^\top w &= a_k, \end{aligned} \tag{3.1}$$

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 (3.1) 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 optimised portfolios and provide estimates of Σ tailored to instances of (3.1) leading to optimised portfolios that have relatively low variance.

We work in a setting where the number p of securities is larger than the number n of observations, 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 optimisation.

3.2 A JSE prescription for a customised, optimisation-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 optimisation problem (3.1). The centerpiece of the prescription is an estimate of Σ's leading eigenvector, which is obtained by applying James–Stein shrinkage to the leading sample eigenvector. Shrinkage improves on the leading sample 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 Sect. 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

$$\Sigma = \eta^2 b b^\top + \delta^2 I, \tag{3.2}$$

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

We do not observe Σ, but see instead a time series of n realised 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, the two variances η^2 and δ^2 and the unit vector b of factor loadings, with functions of eigenvalues and eigenvectors of S in a way that leads to a relatively low-variance solution to (3.1). We show in Sect. 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 $C_1, C_2, \dots, C_k$ and constraint values $a_1, a_2, \dots, a_k$. With $\mathrm{tr}(S)$ denoting the trace of the sample covariance matrix S and λ^2 denoting its leading eigenvalue, define

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

the average of the nonzero eigenvalues of S that are less than λ^2, and

$$\phi^2 = \frac{\lambda^2 - \ell^2}{\ell^2},$$

a measure of the average relative leading eigengap.

Let C denote the span of the constraint vectors $C_1, C_2, \dots, C_k$ from (3.1) and h_C the orthogonal projection of the leading sample eigenvector h onto the subspace C. Now define the JSE shrinkage constant

$$c^{\mathrm{JSE}} = \frac{\ell^2}{\lambda^2(1 - |h_C|^2)} \tag{3.3}$$

and define

$$H^{\mathrm{JSE}} = c^{\mathrm{JSE}} h_C + (1 - c^{\mathrm{JSE}}) h. \tag{3.4}$$

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

$$h^{\mathrm{JSE}} = H^{\mathrm{JSE}} / |H^{\mathrm{JSE}}|. \tag{3.5}$$

The James–Stein estimate h^{JSE} is a better approximation to the true leading eigenvector b than the principal component estimate $h = h^{\mathrm{PCA}}$. Let θ^{JSE} and θ^{PCA} denote the angles from b to h^{JSE} and to h^{PCA}, and Θ the angle between b and the subspace C. Then asymptotically as p tends to infinity, we have

$$\cos^2 \theta^{\mathrm{JSE}} - \cos^2 \theta^{\mathrm{PCA}} = \frac{1}{\phi^2 + 1} \frac{\cos^2 \Theta}{\phi^2 \sin^2 \Theta + 1} > 0 \tag{3.6}$$

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

Formula (3.4) is equivalent to Goldberg and Kercheval [30, formula (6)]. That article and Shkolnik [65] expose the parallel between JSE and classical James–Stein. Formulas (3.3)–(3.5) are identical to formulas (4.8)–(4.10) in Sect. 4.1.4.

The asymptotic context in which formula (3.6) holds is described precisely in Theorem 4.5. Here, the quantities θ^{JSE}, θ^{PCA}, Θ, ϕ refer to their asymptotic limits.

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 with h^{JSE} and an estimate of b, an estimate of (3.2) is given by

$$\Sigma^{\mathrm{JSE}} = (\lambda^2 - \ell^2) h^{\mathrm{JSE}} (h^{\mathrm{JSE}})^\top + (n/p)\ell^2 I. \tag{3.7}$$

Formula (3.7) is the one-factor covariance matrix designed for use in the quadratic optimisation (3.1). 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 shall see that under the assumptions described in Sect. 4, $|h_C|^2$ 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 optimised portfolio

The benefits of this construction are realised in the portfolio w^{JSE} generated by (3.1) when Σ is set to Σ^{JSE}. Let Σ^{PCA} be the covariance matrix obtained by replacing h^{JSE} with the leading sample eigenvector h in (3.7), and w^{PCA} the portfolio generated by (3.1) when Σ is set to Σ^{PCA}. Theorem 4.9 below shows that the ratio of the true variances of w^{JSE} and w^{PCA},

$$\frac{\mathrm{Var}[w^{\mathrm{JSE}}]}{\mathrm{Var}[w^{\mathrm{PCA}}]},$$

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 the true variance.

4 JSE stochastically dominates PCA

The formulas in Sect. 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 the variance of optimised portfolios, relative to PCA.

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

In Sect. 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 Sect. 4.2, we apply these results to estimating the variance of a portfolio obtained by minimising the 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 the ground truth b when the data follow a one-factor model. An estimate h^{JSE} is obtained by shrinking the leading sample eigenvector towards 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 develop an estimated $p \times p$ 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

$$r = \beta f + z. \tag{4.1}$$

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 can be described by a nested sequence

$$r^{(p)} = \beta^{(p)} f + z^{(p)}, \qquad p = 1, 2, 3, \ldots$$

The nestedness property means that the models are defined by an infinite sequence (β_i) of scalars and an infinite sequence (z_i) of random variables such that truncation at p forms the p-vectors $\beta^{(p)}$ and $z^{(p)}$, respectively. (The nestedness property is not required for our results if we accept a bound on higher moments, but it simplifies the discussion.)

We list below our **standing assumptions** on the factor model (4.1).

Assumption 4.1 1) The random variable f representing factor returns is nonzero almost surely, and has mean zero and variance $\sigma^2 > 0$.

2) (a) The random variables $(z_i)_{i \in \mathbb{N}}$ representing security-specific returns have mean zero, are uncorrelated with f and have uniformly bounded second moments with variances $\operatorname{Var}[z_i] = \delta_i^2$ tending on average to a limit $\delta^2 > 0$, i.e.,

$$\lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^{p} \delta_i^2 = \delta^2 > 0.$$

(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 correlation decay conditions

$$\frac{1}{p^2} \sum_{i,j=1}^{p} \operatorname{Cov}(z_i, z_j)^2 \longrightarrow 0 \qquad \text{and} \qquad \frac{1}{p^2} \sum_{i,j=1}^{p} \operatorname{Cov}(z_i^2, z_j^2)^2 \longrightarrow 0$$

as $p \to \infty$.

3) The sequence $(\beta_i)_{i \in \mathbb{N}}$ of security exposures to the factor is bounded and the average of the squared entries tends to a positive limit as $p \to \infty$, i.e.,

$$\lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^{p} \beta_i^2 = B^2 > 0,$$

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

In particular, if $b^{(p)} = \beta^{(p)} / |\beta^{(p)}|$, Assumption 4.1, 3) implies that the set $\{p(b_i^{(p)})^2 : p > 1, i = 1, 2, 3, \ldots, p\}$ 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 the z_i allow heavy-tailed distributions.

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 [72].

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 Assumptions 4.1, 2) and 3) that the sequences (δ_i^2) and (β_i^2) have positive limiting averages (called *pervasiveness* in Fan et al. [25]) means that a nonnegligible fraction of the entries are nonvanishing. This is a basic and mild asymptotic nondegeneracy condition on our sequence of models. It means that a nonnegligible fraction of extra assets added to increase the model dimension have nonnegligible exposure to the factor, and a nonnegligible fraction have nonnegligible specific risk. (The existence of the limit is a matter of convenience, since otherwise we could pass to subsequences.)

For the factor model (4.1), under our assumptions, the population covariance matrix of returns takes the form

$$\Sigma = \sigma^2 \beta\beta^\top + \Omega,$$

where Ω is the covariance matrix of the specific returns z_i, which by Assumption 4.1, 2) has bounded eigenvalues.

Assumption 4.1, 2)(b)i) implies that Ω is diagonal and we have a strict factor model. The alternative Assumption 4.1, 2)(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 [10]. This allows the presence of additional weak factors, provided their corresponding eigenvalues are bounded. The correlation decay conditions are satisfied if for example each specific return is correlated with only a uniformly bounded number of other specific returns.

If we strengthen Assumption 4.1, 2) to

2*) The random variables z_i satisfy Assumption 4.1, 2)(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 is a series of limit theorems as $p \to \infty$. All results assume that our standing Assumptions 4.1, 1)–3) hold. All limits of random variables are in the sense of convergence in probability. In addition, when Assumption 4.1, 2*) also holds, the limits hold almost surely.

Because β and f appear in the model (4.1) 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

$$r = \eta b x + z. \tag{4.2}$$

With this formulation, Assumption 4.1, 3) 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 then a sum of a factor component $\eta^2 bb^\top$ and a specific component Ω, i.e.,

$$\Sigma = \eta^2 bb^\top + \Omega. \tag{4.3}$$

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 (4.2) and hence covariance matrix Σ given by (4.3). 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

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

in terms of nonnegative 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. Let $\angle(h, b)$ denote the angle between the vectors h and b.

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

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

and

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

Proposition 4.2 *The limits*

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

exist, and

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

These limits hold in probability under Assumptions 4.1, 1)–3), and hold almost surely if Assumption 4.1, 2) is replaced by Assumption 4.1, 2).*

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 (4.2). Decomposing from (4.2) the $p \times n$ data matrix of returns Y into a sum of unobservable components, we have

$$Y = \eta b X^\top + Z, \tag{4.4}$$

where $X = (X_1, X_2, \ldots, X_n)^\top$ is the n-vector of independent realisations of x, and Z is the $p \times n$ matrix whose columns are the n independent realisations 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 result is a simple consequence of Lemma A.7 stated later.

Proposition 4.3 *The relative eigengap ψ_∞ is related to the parameters of the factor model by*

$$\psi_\infty^2 = \lim_{p\to\infty} \psi_p^2 = \lim_{p\to\infty} \frac{\lambda^2 - \ell^2}{\lambda^2} = \frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2 |X|^2 + \delta^2} \approx \frac{p\sigma^2 B^2}{p\sigma^2 B^2 + p\delta^2/n}. \tag{4.5}$$

These limits hold in probability under Assumptions 4.1, 1)–3), and hold almost surely if Assumption 4.1, 2) is replaced by Assumption 4.1, 2).*

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 (4.5) 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-square 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 (4.5). It is the asymptotic trace of the factor component of the population covariance matrix Σ specified in (4.3) 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 4.3 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 (4.5) cannot be observed, but it can be estimated in terms of the relative eigengap of S.

A consequence of Proposition 4.3 is that after first taking the limit $p \to \infty$ and then allowing $n \to \infty$, the term ψ_∞^2 tends to 1. Therefore,

$$\lim_{n\to\infty} \lim_{p\to\infty} |h - b| = 0. \tag{4.6}$$

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

$$\theta^{\mathrm{PCA}} \approx \frac{1}{\sqrt{n}} \frac{\delta}{\sigma B}.$$

For a typical value $\delta/(\sigma B) = 4$, this means that the angular error θ^{PCA} will remain significant even for n as large as 1'000 or more, well above the typical values seen in portfolio optimisation.

We note that Wang and Fan [71] provide an HH version of (4.6) 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

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

Central to the ideas underlying Propositions 4.2 and 4.3 is a duality between the $p\times n$ problem and an $n \times p$ problem in which the roles of p and n are interchanged. If we consider the $p\times p$ sample covariance matrix $S = YY^\top/n$, there is a corresponding dual (or "Gram") $n \times n$ matrix $S^* = Y^\top Y/p$. Making use of (4.4),

$$S^* = \frac{\eta^2}{p} XX^\top + \frac{1}{p} Z^\top Z + \frac{\eta}{\sqrt{p}} \left(X \frac{b^\top Z}{\sqrt{p}} + \frac{Z^\top b}{\sqrt{p}} X^\top \right),$$

and with arguments that appear in the proofs of the propositions, we can show that

$$\lim_{p\to\infty} S^* = \sigma^2 B^2 XX^\top + \delta^2 I_n.$$

This limit takes place for fixed dimension n and helps us evaluate the limiting behaviour of our p-dimensional problem as the dimension p tends to infinity. We note that in our setting, the $n \times p$ problem does not correspond to a simple LH regime limit because the p dual "observations" are not independent due to the common factor connecting the returns of different assets. Considerations of independence aside, the leading eigenvector of S^* is a consistent estimator of the unobserved factor returns vector X, while the leading eigenvector of S is, as shown in Proposition 4.2, not a consistent estimator of the population factor exposure vector β. This last fact is a central theme of this work.

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 $\mathbb{R}^p$. We use subscripts to denote the orthogonal projection of a vector onto a linear subspace; so 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 subspaces span(x) and span(y) by $\angle(x, y)$, with $0 \le \angle(x, y) \le \pi/2$. The angle $\angle(x, C)$ between a vector x and a subspace C is equal to $\angle(x, x_C)$.

Theorem 4.4 *Suppose the angle $\angle(b, C)$ between b and C tends to a limit*

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

Then the limit

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

exists, and

$$\cos\Theta^h = \cos\theta^{\mathrm{PCA}} \cos\Theta = \psi_\infty \cos\Theta. \tag{4.7}$$

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

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

and

$$0 < \cos\Theta^h < \cos\Theta.$$

These limits hold in probability under Assumptions 4.1, 1)–3), *and hold almost surely if Assumption* 4.1, 2) *is replaced by Assumption* 4.1, 2*).

This theorem is a generalisation of Goldberg et al. [32, Theorem 3.1]. 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 towards 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. [36] 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

$$\cos\angle(b, C) \ge \langle b, e/|e|\rangle = \frac{1}{|\beta|\sqrt{p}} \sum_{i=1}^{p} \beta_i = \frac{\sqrt{p}}{|\beta|} \frac{1}{p} \sum_{i=1}^{p} \beta_i > 0$$

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 $\limsup_{p\to\infty} \angle(b, C) < \pi/2$, 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 use the notation $h = h^{\text{PCA}}$ when emphasising 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 Assumptions 4.1, 1)–3).

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

$$c^{\text{JSE}} = \frac{\ell^2}{\lambda^2(1 - |h_C|^2)}, \tag{4.8}$$

$$H^{\text{JSE}} = c^{\text{JSE}} h_C + (1 - c^{\text{JSE}})h, \tag{4.9}$$

$$h^{\text{JSE}} = H^{\text{JSE}}/|H^{\text{JSE}}|. \tag{4.10}$$

Formulas (4.8)–(4.10) are identical to formulas (3.3)–(3.5) in Sect. 3.2.2. We can show that

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

(If now n is taken to infinity, then by (4.6) and since $|X|^2$ tends to infinity, we have that c^{JSE} tends to zero and both h and h^{JSE} converge to b.)

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

Define

$$\phi_\infty^2 := \frac{\psi_\infty^2}{1 - \psi_\infty^2} = \frac{\sigma^2 B^2 |X|^2}{\delta^2} = \lim_{p\to\infty} \frac{\lambda^2 - \ell^2}{\ell^2}, \tag{4.11}$$

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

Theorem 4.5 *Suppose the limit*

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

exists. Then under Assumptions 4.1, 1)–3), *the limits*

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

exist in probability, and hold almost surely under the additional Assumption 4.1, 2*). *The asymptotic improvement of* h^{JSE} *over* h^{PCA} *as an estimate of the leading population eigenvector is*

$$\cos^2 \theta^{\text{JSE}} - \cos^2 \theta^{\text{PCA}} = \frac{1}{\phi_\infty^2 + 1} \frac{\cos^2 \Theta}{\phi_\infty^2 \sin^2 \Theta + 1}. \tag{4.12}$$

In particular, JSE is never worse asymptotically than PCA, and
- *if* $\Theta < \pi/2$, *then* $\theta^{\text{JSE}} < \theta^{\text{PCA}}$;
- *if* $\Theta = 0$, *then* h^{JSE} *converges to* b *and JSE is a consistent estimator*;
- *if* $\Theta = \pi/2$, *then* h^{JSE} *converges to* h^{PCA} *and* $\theta^{\text{JSE}} = \theta^{\text{PCA}}$.

The right-hand side of (4.11) is the ratio of the factor variance and the specific variance in (4.2). The formula highlights the relationship between the relative eigengap and the factor model parameters. Taken together, (4.5) and (4.11) imply that

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

One consequence of Theorem 4.5 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 optimisation problem.

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

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

It follows from (4.7) and (4.12) that

$$\lim_{p \to \infty} I\big(\angle(h, C), \phi_p^2\big) = \cos^2 \theta^{\text{JSE}} - \cos^2 \theta^{\text{PCA}} \qquad \text{almost surely.}$$

4.2 Estimating the constrained minimum variance

We return to the optimisation problem (3.1) introduced in Sect. 3.1, namely

$$\begin{aligned} &\min_w \frac{1}{2} w^\top \Sigma w \\ &\text{subject to } C^\top w = a, \end{aligned} \tag{4.13}$$

where we have now 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 nonzero 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 Sect. 4.1 to estimate a $p \times p$ covariance matrix $\Sigma = \Sigma^{\text{JSE}}$ for use in (4.13). The matrix Σ^{JSE} depends on the constraint matrix C; its core is h^{JSE}, the leading eigenvector of the sample covariance matrix, shrunk by a prescribed amount in the direction of C. To avoid visual clutter, we suppress the dependence of Σ^{JSE} and h^{JSE} on C when possible, but the dependence of Σ^{JSE} on C is a central idea of this section.

4.2.1 Constraints

We assume without loss of generality that the constraint matrix C has full rank and the entries of a are nonnegative, 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 impose Assumptions 4.1, 1)–3) on the underlying factor model described there. In addition, we wish to avoid degeneracy of the constraints $C^\top w = a$ in the limit; so **from now on**, we add to Assumption 4.1 the following two natural **standing assumptions**:

4) For each $j = 1, \dots, k$, 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 $|\cdot|_\infty$ denotes the maximum norm;

b) the sequence $(|C_j^{(p)}|^2/p)_{p\in\mathbb{N}}$ tends to a positive finite limit as $p \to \infty$.

5) The constraint matrix C does not become singular in the high-dimensional limit, i.e.,

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

Assumption 4.1, 4) is similar to Assumption 4.1, 3); it says that the average squared entry of the columns does not tend to zero or infinity with p. Assumptions 4.1, 4) and 5) imply that the angle between any two columns of C is bounded away from zero as p tends to infinity 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 $e = (1, \dots, 1)^\top$ and C is the $p \times 1$ matrix whose only column is e. Since $|e|^2 = p$, Assumption 4.1, 4) is satisfied, and $C^\top C$ is equal to the 1×1 matrix with determinant p so that Assumption 4.1, 5) is satisfied.

4.2.2 Estimating Σ^{JSE}

The constraint matrix C and the vector a of constraint values in the optimisation problem (4.13) are known to the user, but the covariance matrix Σ must be estimated. When data follow the one-factor model (4.1), the population covariance matrix Σ takes the form specified in (4.3), namely

$$\Sigma = \eta^2 bb^\top + \Omega.$$

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 for p securities. We build our estimates from the trace $\mathrm{tr}(S)$ of S, the leading eigenvalue λ^2 of S and its corresponding leading eigenvector h.

Under our spiked model assumptions, it turns out that for minimum-variance estimation, it suffices to estimate Ω with a multiple of the identity converging to $\delta^2 I$. Our

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

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

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 A.7 below, they are summarised by the limits

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

and

$$\lim_{p\to\infty} \ell^2/p = \delta^2/n. \tag{4.14}$$

Recall from Assumption 4.1, 3) 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 (4.14), 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 estimated covariance matrix, depending on the choice of unit vector v, of the form

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

It remains to specify an estimator v of the unit vector b. We examine two competing estimates of Σ^v, namely 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 optimisation bias

For any choice of principal unit eigenvector v, let w^v denote the unique minimiser 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 optimised portfolio w^v.

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

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

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

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

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

Because 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 4.6 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 *optimisation 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 optimisation bias does not depend on the magnitude of a, but only on α and the subspace $\mathrm{col}(C)$, and is equal to zero when $v = b$, i.e.,

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

As described below, the optimisation 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 that

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

which agrees with the optimisation bias originally introduced for this case in Goldberg et al. [32].

The limits in the following two results hold in probability under Assumptions 4.1, 1)–3), and almost surely if Assumption 4.1, 2*) is added.

Proposition 4.7 *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

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

Theorem 4.8 *Let $v \in \mathbb{R}^p$ be a unit vector for each p and satisfying*

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

Then for n, k fixed,

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

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

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

as $p \to \infty$.

Because of Proposition 4.7, Theorem 4.8 applies to both $v = h$ and $v = h^{\mathrm{JSE}}$. When $v = b$, the optimisation bias is zero and the true minimum variance is asymptotically $O(1/p)$. Otherwise, the limiting value of the optimisation bias $\mathcal{E}_p^2$ controls the large-p variance of the estimated portfolio.

The next result 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 4.9 *Suppose the angle between b and $\mathrm{col}(C)$ tends, as $p \to \infty$, to a limit between 0 and $\pi/2$. In addition, assume (by passing to a subsequence if needed) that*

$$\lim_{p\to\infty} \cos\angle\big(b, (C^\dagger)^\top a\big) = \lim_{p\to\infty} \langle b, \alpha\rangle =: \langle b, \alpha\rangle_\infty \qquad \textit{exists.}$$

Then

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

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

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

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

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

tends to zero as $p \to \infty$. The limits are in probability under Assumptions 4.1, 1)–3), and hold almost surely if Assumption 4.1, 2) is added.*

The previous two theorems tell us that $\mathcal{V}(w^b)$ 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 result shows that the condition $\langle b, \alpha\rangle_\infty \neq 0$ typically is satisfied when the vector a is unrelated to the other problem parameters.

Lemma 4.10 *Impose Assumptions 4.1, 1)–5) and suppose 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*

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

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 above. First, we illustrate (4.12), 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 the portfolios w^{JSE} and w^{PCA} tends to zero almost surely.

These experiments serve two purposes. The first is to show that the asymptotic properties described in the theorems, such as (4.12), are approximately realised when the dimension p has realistic values much less than infinity. The results reported here are for $p = 3'000$, but we have observed similar outcomes for p as low as 40.

Second, the variance experiments described in Sect. 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 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 (4.1), repeated here for convenience,

$$r = \beta f + z,$$

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

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

$$c_1 = \frac{\sin\theta}{s} \qquad \text{and} \qquad c_2 = \cos\theta - \frac{\sin\theta}{s} m,$$

and let

$$\beta = c_1\beta^* + c_2 e.$$

Making use of the identity

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

Table 2 Simulation parameters

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^*	$\mathcal{N}(\cos\theta, \sin^2\theta)$	factor loadings
σ	0.16	annualised factor volatility
δ	0.60	annualised specific volatility
f	$\mathcal{N}(0, \sigma^2)$	factor return
z	mean 0, std dev δ	specific return
$\cos\Theta$	0.97, 0.75, 0.49	cosine of the angle between β and C
p	3'000	number of securities
n	24	number of observations
k	2	number of constraints
μ	$0.01\big(\beta + \mathcal{N}(0.5, 2)\big)$	3'000-vector of expected returns
C	(e, μ)	3'000 × 2 matrix of constraint vectors
m	0.01	monthly expected target return
a	$(1, m)^\top$	constraint target vector

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 annualised standard deviation σ to be 16%, and specific return z to be normally distributed with mean 0 and annualised 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 $\mathcal{N}(0.5, 2.0)$ with mean 0.5 and variance 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 the p-vectors μ and e. The angle Θ between β and C is determined by the specification of β and μ. The 2-vector of constraints targets is $a = (1, m)^\top$.

The simulation parameters are listed in Table 2.

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

Under Assumptions 4.1, 1)–3), (4.12) provides an exact expression for the difference between the squared cosines of θ^{PCA} and θ^{JSE}, namely

$$\cos^2\theta^{\mathrm{JSE}} - \cos^2\theta^{\mathrm{PCA}} = \frac{1}{\phi_\infty^2 + 1}\,\frac{\cos^2\Theta}{\phi_\infty^2 \sin^2\Theta + 1}.$$

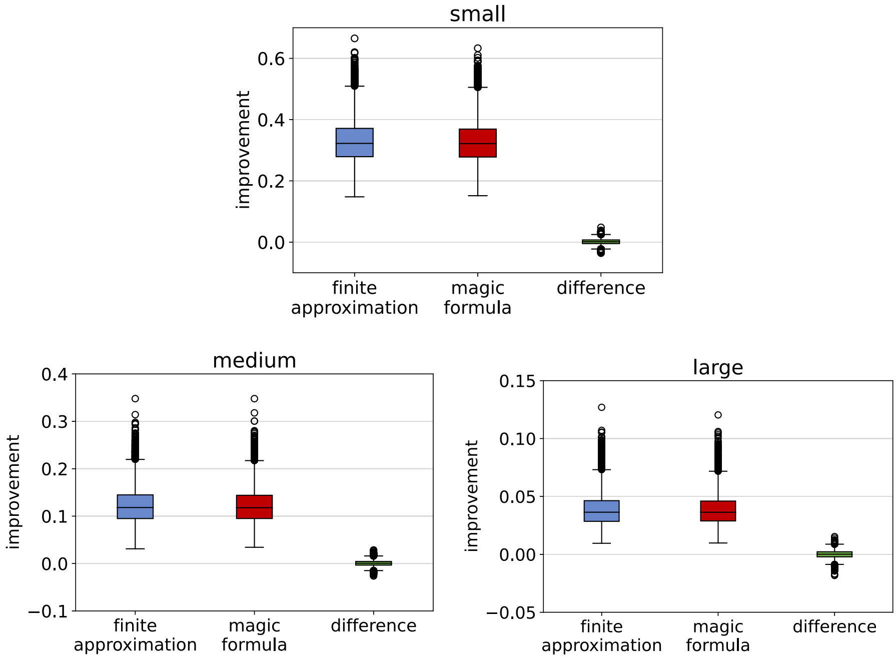

Fig. 1 Box plots for $p = 3'000$ of 10'000 simulations of the difference between $\cos^2 \angle(h^{\text{PCA}}, b)$ and $\cos^2 \angle(h^{\text{JSE}}, b)$ (finite approximation), the limit of this difference (magic formula) as well as the path-by-path difference between the finite approximation and the magic formula (difference). The small, medium and large panels correspond to $\cos\Theta = 0.969, 0.707$ and 0.174. Return data follow (4.1) with parameters specified in Table 2

This *magic formula* for the limiting difference between the two 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 $\Theta = \angle(\beta, C)$ between the leading eigenvector and the shrinkage target, and the relative eigengap ϕ^2.

How well does the asymptotic guidance provided by the magic formula work for finite p? For $p = 3'000$, we report

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

as well as the limit of that difference as p tends to infinity, given by the magic formula. The results of 10'000 simulations are shown in Fig. 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.

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

We report ratios of variances of portfolios w^{PCA}, w^{JSE} and w^{TRUE}, optimised with (3.1), where Σ is set to Σ^{PCA}, Σ^{JSE} and $\Sigma^{\text{TRUE}} = \Sigma$, the true (population)

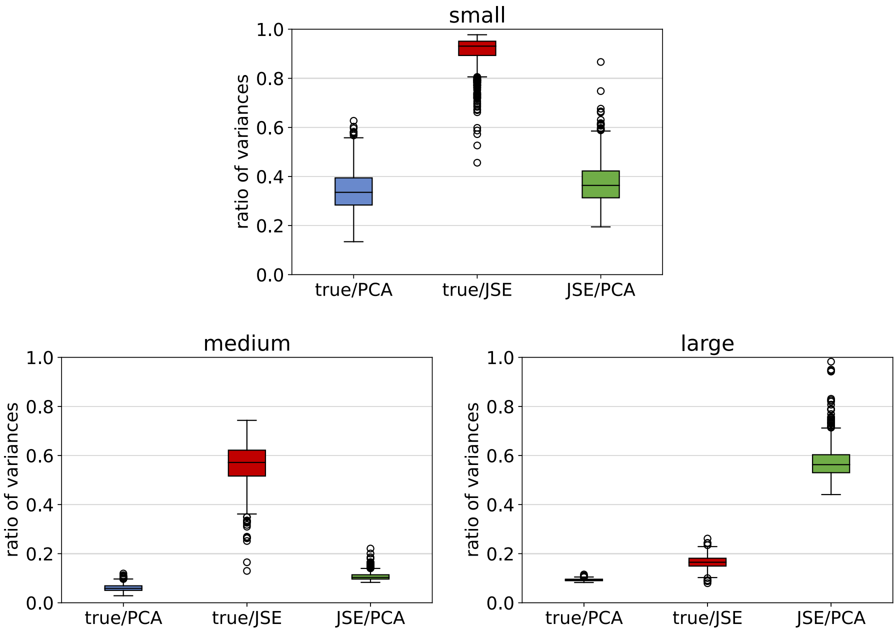

Fig. 2 Box plots for 10'000 simulations of ratios of variances of optimised and optimal portfolios, w^{PCA}, w^{JSE} and w^{TRUE}, for $p = 3'000$. The small, medium and large panels correspond to $\cos\Theta = 0.969$, 0.707 and 0.174. The expected return target is $m = 0.01$. Return data follow (4.1) with parameters specified in Table 2

covariance matrix. The portfolio w^{TRUE} and covariance matrix Σ^{TRUE} and fixed and known in the simulation independent of the simulation samples.

The blue and red box plots in Fig. 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, the improvement is best when the angle between b and C is small, and declines as this angle increases towards $\pi/2$. (In the limit where b is orthogonal to C, we expect no improvement.)

These results are displayed for $p = 3'000$; 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.

The asymptotic behaviour 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 towards a one-dimensional target improves on the leading sample eigenvector as an estimate of ground truth. We show that those guarantees persist when we shrink towards a target of dimension greater than one. This generalisation greatly enlarges the range of applications of JSE, which can now be used to build strategy-specific covariance matrices suitable for quadratic optimisation 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 optimisation. The connection between JSE and the variance of optimised portfolios is via the optimisation bias, which was formulated for minimum variance in earlier work and extended to take account of an arbitrary number of linear constraints in this article. The optimisation bias asymptotically controls the variance of optimised portfolios, and it tends to zero as the number of securities tends to infinity under JSE optimisation.

Also new in this article is a formula for the degree of improvement of JSE over the leading sample 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 optimisation in a multi-factor setting, which has been shown to be effective in numerical experiments. Another direction forward is to pursue the theoretical connections between JSE and concentration of measure in high-dimensional spheres, an understanding of which may provide new, deeper perspectives on these powerful and often counter-intuitive results.

Appendix: Proofs

A.1 Lemmas

We begin with some preliminary results needed for the subsequent proofs. The first lemma is the triangular strong law of large numbers (see Tao [69]).

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

The next lemma is known as the Kolmogorov strong law of large numbers.

Lemma A.2 *Suppose $X_1, X_2, \dots$ is a sequence of independent mean-zero random variables with finite variance and such that*

$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}[X_i]}{i^2} < \infty,$$

and for each p, *define the partial sum* $S_p = X_1 + \cdots + X_p$. *Then* S_p/p *converges almost surely to zero.*

Lemma A.3 *Let* $(z_i)_{i\in\mathbb{N}}$ *be a sequence of independent mean-zero random variables with uniformly bounded fourth moments and* $(b_{i,p})_{i=1,\ldots,p,\, p\in\mathbb{N}}$ *a collection of scalars satisfying* $\sup_{i,p} p|b_{i,p}|^2 < \infty$. *Then*

$$\frac{1}{\sqrt{p}}\sum_{i=1}^{p} b_{i,p} z_i \longrightarrow 0 \qquad \text{almost surely as } p \to \infty.$$

Proof Let $X_{i,p} = \sqrt{p} b_i z_i$ 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 A.1 with $\mu = 0$,

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

converges to zero almost surely. □

Lemma A.4 *Let* $(z_i)_{i\in\mathbb{N}}$ *be a sequence of independent mean-zero random variables with uniformly bounded fourth moments. Suppose*

$$\lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} E[z_i^2] = \delta^2.$$

Then almost surely,

$$\lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} z_i^2 = \delta^2.$$

Proof 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

$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}[X_i]}{i^2} < \infty$$

and the result follows by Lemma A.2. □

Lemma A.5 *Recall our* $p \times n$ *data matrix of returns*

$$Y = \beta X^\top + Z.$$

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

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

and

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

These limits hold in probability under Assumptions 4.1, 1)–3), *and hold almost surely if Assumption* 4.1, 2) *is replaced by Assumption* 4.1, 2*).

Proof The limits in probability follow from straightforward calculation using Assumption 4.1, 2) and Markov's inequality. The almost sure limits follow from Assumptions 4.1, 2*) and 4.1, 3) and Lemmas A.3 and A.4. □

The following is a version of Gurdogan and Kercheval [35, Proposition 5.2], which remains true with a similar proof under our slightly adapted hypotheses.

Proposition A.6 *Under Assumptions* 4.1, 1)–3), *let* $L = (L_p)_{p\in\mathbb{N}}$ *with* $L_p \subseteq \mathbb{R}^p$ *be a sequence of linear subspaces with constant dimension and independent of the random variables* z. *Then*:

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

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

A.2 Proof of Proposition 4.2

Proposition 4.2 *Under Assumptions* 4.1, 1)–3), *the limits*

$$\theta^{\mathrm{PCA}} = \lim_{p\to\infty} \angle(h, b) \qquad \text{and} \qquad \psi_\infty^2 = \lim_{p\to\infty} \psi_p^2$$

exist, and

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

Proof 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, λ, χ, Z, Y on p.

Since h, χ and Y are related by

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

we have by (4.4) that

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

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

$$\langle h, b\rangle = \frac{\eta X^\top \chi}{\lambda\sqrt{n}} + \frac{b^\top Z}{\sqrt{p}}\frac{\chi\sqrt{p}}{\lambda\sqrt{n}},$$
$$\frac{\lambda^2}{p} = \frac{\eta^2(X^\top\chi)^2}{np} + \frac{\chi^\top Z^\top Z\chi}{np} + 2(X^\top\chi)\frac{b^\top Z}{\sqrt{p}}\frac{\eta\chi}{n\sqrt{p}}.$$

Applying Lemma A.5, we deduce that $Z^\top Z/p$ tends to $\delta^2 I$ and $b^\top Z/\sqrt{p}$ to zero as $p \to \infty$. This means that λ^2/p is eventually bounded between zero and infinity, and

$$\langle h, b\rangle_\infty = \lim_{p\to\infty} \frac{\eta X^\top\chi}{\lambda\sqrt{n}}, \tag{A.1}$$

provided the limit in (A.1) exists.

Recall that ℓ_p^2 is the average of the nonzero sample eigenvalues less than λ^2. The proof of the following result is essentially identical to the proof of Goldberg et al. [32, Lemma A.2].

Lemma A.7 *Under Assumptions* 4.1, 1)–3) *and with the notation as above*, *we have the limits*

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

Applying Lemma A.7 to (A.1), we obtain

$$\begin{aligned}\langle h, b\rangle_\infty = \lim_{p\to\infty} \frac{\eta X^T\chi}{\lambda\sqrt{n}} &= \lim_{p\to\infty} \frac{\eta}{\sqrt{p}}\frac{\sqrt{p}}{\lambda}\frac{X^\top\chi}{\sqrt{n}} \\ &= \sigma B \frac{1}{\sqrt{\sigma^2 B^2|X|^2/n + \delta^2/n}}\frac{|X|}{\sqrt{n}} \\ &= \sqrt{\frac{\sigma^2 B^2 |X|^2}{\sigma^2 B^2|X|^2 + \delta^2}} \in (0, 1).\end{aligned}$$

By Lemma A.7,

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

converges to

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

and hence $\langle h, b\rangle_\infty = \psi_\infty$. This completes the proof of Proposition 4.2. □

A.3 Proof of Theorem 4.4

Theorem 4.4 *Suppose the angle $\angle(b, C)$ between b and C tends to a limit*

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

Then the limit

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

exists, and

$$\cos\Theta^h = \cos\theta^{\text{PCA}} \cos\Theta = \psi_\infty \cos\Theta. \tag{A.2}$$

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

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

and

$$0 < \cos\Theta^h < \cos\Theta.$$

Proof We apply Proposition A.6, 1) with $L = C$, noting that $\langle h, h_C\rangle = \cos\angle(h, C)$ and $\langle b, b_C\rangle = \cos\angle(b, C)$. Because we have $\langle h, b\rangle \to \psi_\infty$ from Proposition 4.2 and $\cos\angle(b, C) \to \cos\Theta$ by hypothesis, (A.2) follows immediately. ☐

A.4 Proof of Theorem 4.5

Theorem 4.5 *With the notation as above, suppose the limit*

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

exists. Then the limits

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

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

$$\cos^2\theta^{\text{JSE}} - \cos^2\theta^{\text{PCA}} = \frac{1}{\phi_\infty^2 + 1} \frac{\cos^2\Theta}{\phi_\infty^2 \sin^2\Theta + 1}.$$

If $\Theta = \pi/2$, then h^{JSE} converges to h^{PCA}, $\theta^{\text{JSE}} = \theta^{\text{PCA}}$ and there is no improvement, while if $\Theta = 0$, then h^{JSE} converges to b. In other cases, $\theta^{\text{JSE}} < \theta^{\text{PCA}}$ almost surely, with the improvement given by (4.12).

Proof The existence of the limit θ^{PCA} has already been established in Proposition 4.2. The JSE estimator h^{JSE} relative to the subspace C is an example of the "MAPS" estimator defined and studied in Gurdogan and Kercheval [35]. 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

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

and define the unit vector

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

The oracle h^o is the normalised 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 result is a simpler version of Gurdogan and Kercheval [35, proof of Theorem 5.1], for slightly adjusted assumptions.

Proposition A.8 *We have*

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

Next, let

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

Then $U = \mathrm{span}(h, C) = \mathrm{span}(u, C)$ and u is a unit vector orthogonal to C (assuming that h does not belong to C, otherwise set $u = 0$). Hence

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

and so

$$\langle h^o, b\rangle^2 = \left\langle \frac{b_U}{|b_U|}, b\right\rangle^2 = |b_U|^2 = |b_C|^2 + \langle u, b\rangle^2 = |b_C|^2 + \frac{(\langle h, b\rangle - \langle h_C, b\rangle)^2}{1 - |h_C|^2}.$$

All the terms on the right-hand side have previously been shown to have limits as $p \to \infty$, namely

$$\begin{aligned}
|b_C|^2 &\longrightarrow \cos^2\Theta,\\
|h_C|^2 &\longrightarrow \psi_\infty^2 \cos^2\Theta,\\
\langle h, b\rangle &\longrightarrow \psi_\infty = \cos\theta^{\mathrm{PCA}},\\
\langle h_C, b\rangle &\longrightarrow \psi_\infty \cos^2\Theta.
\end{aligned}$$

Therefore $\lim_{p\to\infty} \langle h^o, b\rangle^2$ exists, and by Proposition A.8,

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

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

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

we obtain in the limit that

$$\begin{aligned}\cos^2\theta^{\mathrm{JSE}} - \cos^2\theta^{\mathrm{PCA}} &= \cos^2\Theta + \frac{\psi^2(1-\cos^2\Theta)^2}{1-\psi^2\cos^2\Theta} - \psi^2 \\ &= (1-\psi^2)^2\frac{\cos^2\Theta}{1-\psi^2\cos^2\Theta} \\ &= \frac{1}{\phi^2+1}\,\frac{\cos^2\Theta}{\phi^2\sin^2\Theta+1}. \end{aligned} \tag{A.3}$$

This is positive when $\Theta < \pi/2$. In case $\Theta = \pi/2$, Theorem 4.4 implies that h_C tends to zero and h^{JSE} 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 (A.3) that $\theta^{\mathrm{JSE}} = 0$, and so h^{JSE} tends to b itself. □

A.5 Proof of Proposition 4.7

Proposition 4.7 *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

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

Proof From Proposition A.6, 3) with $L = C$, we have, in the limit as $p \to \infty$,

$$|h_C|^2 = \langle h, b\rangle_\infty^2\, |b_C|^2 = \psi_\infty^2 |b_C|^2. \tag{A.4}$$

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

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

where

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

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

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

This follows from (A.4) and the assumption that the angle between b and C is asymptotically strictly between 0 and $\pi/2$. □

A.6 Proof of Theorem 4.8

Theorem 4.8 *Let $v \in \mathbb{R}^p$ be a unit vector for each p and satisfying*

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

Recall that w^v denotes the unique vector in $\mathbb{R}^p$ minimising $w^\top \Sigma^v w$ subject to the constraint $C^\top w = a$. Then for n, k fixed, the true variance of the estimated portfolio w^v is

$$\mathcal{V}(w^v) := (w^v)^\top \Sigma w^v = \eta^2 |(C^\dagger)^\top a|^2 \mathcal{E}_p(v, C, a)^2 + O(1/p) \tag{A.5}$$

as $p \to \infty$. Furthermore,

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

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 (see Golub and Van Loan [33, Sect. 2.1.3]) shows that

$$w^v = \left(I + \frac{(v_C - v)v^\top}{1 + \kappa^2 - |v_C|^2} \right) (C^\dagger)^\top a. \tag{A.6}$$

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 4.1, 4) and 5) 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 the true variance, first notice that

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

For the second term, it follows from Assumption 4.1, 4) and $C^\top w^v = a$ that

$$|w^v|^2 \le O(1/p).$$

It remains to analyse the first term. Making use of (A.6) and recalling

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

we have

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

□

A.7 Proof of Theorem 4.9

Theorem 4.9 *Suppose that the angle between b and $\mathrm{col}(C)$ is asymptotically between 0 and $\pi/2$. In addition, assume (by passing to a subsequence if needed) that*

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

Then

$$\lim_{p\to\infty} \mathcal{E}_p(h^{\mathrm{JSE}}, C, a)^2 = 0. \tag{A.7}$$

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

$$\lim_{p\to\infty} \mathcal{E}_p(h, C, a)^2 > 0. \tag{A.8}$$

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

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

tends to zero as $p \to \infty$.

Proof By Proposition 4.7, we know that

$$\limsup |v_C| < 1$$

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

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

stays away from zero in both cases. For (A.7), it then suffices to show that the numerator

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

vanishes asymptotically. In light of Proposition A.8, 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 the straightforward identities

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

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

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

Turning to (A.8), note that Proposition A.6 applied to the subspace $L = \text{span}(\alpha)$ implies that 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

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

When $\mathcal{E}(h, C, a)$ is positive but $\mathcal{E}(h^{\text{JSE}}, C, a)$ tends to zero, (A.5) 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. □

A.8 Proof of Lemma 4.10

Lemma 4.10 *Assume that the limiting angle Θ is less than $\pi/2$. Suppose a does not belong to the orthogonal complement of the unit vector*

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

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

Proof We express the singular value decomposition of C as

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

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 that $(C^{\dagger})^{\top} = UZ^{-1}V$.

Assumptions 4.1, 4) and 5) 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 can assume that the inner product $U^{\top}b$ tends to a nonzero limit $(U^{\top}b)_\infty \in \mathbb{R}^k$ as $p \to \infty$. A short calculation shows that

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

and

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

Hence

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

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

Acknowledgements We are grateful to Jeongyoun Ahn, Stjepan Begušić, Haim Bar, Alex Shkolnik, Sungku Jung and Youhong Lee for support. We thank Richard Michaud and Alex Ulitsky for helpful comments on an early draft of this article, and two anonymous referees for their valuable comments which helped us to improve the quality of the paper. We should also like to thank the anonymous referees and the editors for their thorough review that led to a substantially improved manuscript.

Declarations

Competing interests The authors declare no competing interests.

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**, 760–766 (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. Aust. N. Z. J. Stat. **60**, 4–19 (2018)
3. Bai, Z.: Methodologies in spectral analysis of large dimensional random matrices: a review. Stat. Sin. **9**, 611–677 (1999)
4. Bai, Z., Silverstein, J.W.: Spectral Analysis of Large Dimensional Random Matrices, 2nd edn. Springer, Berlin (2010)
5. Ball, K.: An elementary introduction to modern convex geometry. In: Levy, S. (ed.) Flavors of Geometry, pp. 1–58. Cambridge University Press, Cambridge (1997)
6. Bianchi, S.W., Goldberg, L.R., Rosenberg, A.: The impact of estimation error on latent factor model forecasts of portfolio risk. J. Portf. Manag. **43**(5), 147–156 (2017)
7. Bodnar, T., Okhrin, Y., Parolya, N.: Optimal shrinkage-based portfolio selection in high dimensions. J. Bus. Econ. Stat. **41**, 140–156 (2021)
8. Bodnar, T., Parolya, N., Schmid, W.: Estimation of the global minimum variance portfolio in high dimensions. Eur. J. Oper. Res. **266**, 371–390 (2018)
9. Boyd, S., Johansson, K., Kahn, R., Schiele, P., Schmelzer, T.: Markowitz portfolio construction at seventy. J. Portf. Manag. **50**(8), 117–160 (2024)
10. Chamberlain, G., Rothschild, M.: Arbitrage, factor structure, and mean–variance analysis on large asset markets. Econometrica **51**, 1281–1304 (1983)
11. Clarke, R., De Silva, R., Thorley, S.: Minimum-variance portfolio composition. J. Portf. Manag. **37**(2), 31–45 (2011)
12. Connor, G.: A unified beta pricing theory. J. Econ. Theory **34**, 13–31 (1984)
13. Connor, G.: The three types of factor models: a comparison of their explanatory power. Financ. Anal. J. **51**(3), 42–46 (1995)
14. Connor, G., Korajczyk, R.: Risk and return in equilibrium APT. J. Financ. Econ. **21**, 255–289 (1988)
15. Connor, G., Korajczyk, R.: Factor models. In: Cont, R. (ed.) Encyclopedia of Quantitative Finance, pp. 669–674. Wiley, New York (2010)
16. Connor, G., Korajczyk, R.A.: Performance measurement with the arbitrage pricing theory. J. Financ. Econ. **15**, 373–394 (1986)
17. Ding, Y., Li, Y., Zheng, X.: High dimensional minimum variance portfolio estimation under statistical factor models. J. Econom. **222**, 502–515 (2021)
18. Donoho, D., Gavish, M., Johnstone, I.: Optimal shrinkage of eigenvalues in the spiked covariance model. Ann. Stat. **46**, 1742–1778 (2018)
19. Edelman, A., Rao, N.R.: Random matrix theory. Acta Numer. **14**, 233–297 (2005)
20. Efron, B., Morris, C.: Data analysis using Stein's estimator and its generalizations. J. Am. Stat. Assoc. **70**, 311–319 (1975)
21. Efron, B., Morris, C.: Stein's paradox in statistics. Sci. Am. **236**(5), 119–127 (1977)
22. El Karoui, N.: High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation. Ann. Stat. **38**, 3487–3566 (2010)
23. El Karoui, N.: On the realized risk of high-dimensional Markowitz portfolios. SIAM J. Financ. Math. **4**, 737–783 (2013)
24. Fan, J., Liao, Y., Mincheva, M.: High-dimensional covariance matrix estimation in approximate factor models. Ann. Stat. **39**, 3320–3356 (2011)
25. Fan, J., Liao, Y., Mincheva, M.: Large covariance estimation by thresholding principal orthogonal complements. J. R. Stat. Soc., Ser. B Stat. Methodol. **75**, 603–680 (2013)
26. Fourdrinier, D., Strawderman, W.E., Wells, M.T.: Shrinkage Estimation. Springer, Cham (2018)
27. Gabaix, X.: Power laws in economics and finance. Annu. Rev. Econ. **1**, 255–294 (2009)
28. Garlappi, L., Uppal, R., Wang, T.: Portfolio selection with parameter and model uncertainty: a multi-prior approach. Rev. Financ. Stud. **20**, 41–81 (2007)
29. Garvey, G., Kahn, R.N., Savi, R.: Practical applications of the dangers of diversification: managing multiple manager portfolios. J. Portf. Manag. **5**(2), 1–4 (2017)
30. Goldberg, L.R., Kercheval, A.N.: James–Stein for the leading eigenvector. Proc. Natl. Acad. Sci. **120**(2), e2207046120 (2023)
31. Goldberg, L.R., Papanicolaou, A., Shkolnik, A., Ulucam, S.: Better betas. J. Portf. Manag. **47**(1), 119–136 (2020)
32. Goldberg, L.R., Papanicolaou, A., Shkolnik, A.: The dispersion bias. SIAM J. Financ. Math. **13**, 521–550 (2022)
33. Golub, G.H., van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)

34. Green, R.C., Hollifield, B.: When will mean–variance efficient portfolios be well diversified? J. Finance **47**, 1785–1809 (1992)
35. Gurdogan, H., Kercheval, A.: Multi anchor point shrinkage for the sample covariance matrix. SIAM J. Financ. Math. **13**, 1112–1143 (2022)
36. Hall, P., Marron, J., Neeman, A.: Geometric representation of high dimension, low sample size data. J. R. Stat. Soc., Ser. B Stat. Methodol. **67**, 427–444 (2005)
37. Jagananthan, R., Ma, T.: Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Finance **58**, 1651–1683 (2003)
38. James, W., Stein, C.: Estimation with quadratic loss. In: Neyman, J. (ed.) Proc. Fourth Berkeley Symp. Math. Stat. Prob., pp. 361–397. University of California (1961)
39. Jobson, J., Korkie, B.: Estimation for Markowitz efficient portfolios. J. Am. Stat. Assoc. **75**, 544–554 (1980)
40. Johnstone, I.M.: On the distribution of the largest eigenvalue in principal components analysis. Ann. Stat. **29**, 295–327 (2001)
41. Johnstone, I.M., Lu, A.Y.: On consistency and sparsity for principal components analysis in high dimensions. J. Am. Stat. Assoc. **104**, 682–693 (2009)
42. Jung, S., Marron, J.: PCA consistency in high dimension, low sample size context. Ann. Stat. **37**, 4104–4130 (2009)
43. Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. J. Financ. Econ. **3**, 215–231 (1976)
44. Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Finance **10**, 603–621 (2003)
45. Ledoit, O., Wolf, M.: Honey, I shrunk the sample covariance matrix. J. Portf. Manag. **30**(4), 110–119 (2004)
46. Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. **88**, 365–411 (2004)
47. Ledoit, O., Wolf, M.: Nonlinear shrinkage estimation of large-dimensional covariance matrices. Ann. Stat. **40**, 1024–1060 (2012)
48. Ledoit, O., Wolf, M.: Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. Rev. Financ. Stud. **30**, 4349–4388 (2017)
49. Lee, J.H., Stefek, D.: Do risk factors eat alpha? J. Portf. Manag. **34**(4), 12–25 (2008)
50. Marchenko, V., Pastur, L.: Distribution of eigenvalues for some sets of random matrices. Mat. Sb. **114**, 507–536 (1967)
51. Markowitz, H.: Portfolio selection. J. Finance **7**, 77–92 (1952)
52. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments. Yale University Press, New Haven (1959)
53. Markowitz, H.M.: Topics in applied investment management: from a Bayesian viewpoint. J. Invest. **21**, 7–13 (2012)
54. Menchero, J., Lazanas, A.: Evaluating and comparing risk model performance. J. Portf. Manag. **50**(3), 90–110 (2024)
55. Menchero, J., Wang, J., Orr, D.: Improving risk forecasts for optimized portfolios. Financ. Anal. J. **68**(3), 40–50 (2012)
56. Michaud, R.O.: The Markowitz optimization enigma: is 'optimized' optimal? Financ. Anal. J. **45**(1), 31–43 (1989)
57. Paul, D.: Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. Stat. Sin. **17**, 1617–1642 (2007)
58. Paul, D., Aue, A.: Random matrix theory in statistics: a review. J. Stat. Plan. Inference **150**, 1–29 (2014)
59. Rosenberg, B.: Extra-market components of covariance in security returns. J. Financ. Quant. Anal. **9**, 263–274 (1974)
60. Rosenberg, B., McKibben, W.: The prediction of systematic and specific risk in common stocks. J. Financ. Quant. Anal. **8**, 317–333 (1973)
61. Ross, S.: The arbitrage theory of capital asset pricing. J. Econ. Theory **13**, 341–360 (1976)
62. Saxena, A., Stubbs, R.A.: An empirical case study of factor alignment problems using the user model. J. Invest. **21**, 25–43 (2012)
63. Sharpe, W.F.: A simplified model for portfolio analysis. Manag. Sci. **9**, 277–293 (1963)
64. Shen, D., Shen, H., Zhu, H., Marron, J.S.: The statistics and mathematics of high dimension low sample size asymptotics. Stat. Sin. **26**, 1747–1770 (2016)

65. Shkolnik, A.: James–Stein estimation of the first principal component. Stat **11**(1), e419 (2022)
66. Shkolnik, A., Kercheval, A., Gurdogan, H., Goldberg, L.R., Bar, H.: Portfolio selection revisited. Ann. Oper. Res. **346**, 137–155 (2025)
67. Spearman, C.: General intelligence, objectively determined and measured. Am. J. Psychol. **15**, 201–293 (1904)
68. Stein, C.: Some problems in multivariate analysis part I. Technical report no. 6, Office of Naval Research (1956). Available online at https://purl.stanford.edu/cm945fg5866
69. Tao, T.: 275a, notes 3: the weak and strong law of large numbers. https://terrytao.wordpress.com/2015/10/23/275a-notes-3-the-weak-and-strong-law-of-large-numbers/
70. Tao, T.: Topics in Random Matrix Theory. Graduate Studies in Mathematics, vol. 132. Am. Math. Soc., Providence (2012)
71. Wang, W., Fan, J.: Asymptotics of empirical eigenstructure for high dimensional spiked covariance. Ann. Stat. **45**, 1342–1374 (2017)
72. Warusawitharana, M.: Time-varying volatility and the power law distribution of stock returns. J. Empir. Finance **49**, 123–141 (2018)
73. Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955)
74. Wigner, E.: On the distribution of the roots of certain symmetric matrices. Ann. Math. **67**, 325–328 (1958)
75. Xidonas, P., Steuer, R., Hassapis, C.: Robust portfolio optimization: a categorized bibliographic review. Ann. Oper. Res. **292**, 533–553 (2020)