The Dispersion Bias

Lisa Goldberg[*] Alex Papanicolaou[†] Alex Shkolnik[‡]

November 4, 2017
This Version: March 8, 2020[§]

Abstract

We identify and correct excess dispersion in the leading eigenvector of a sample covariance matrix, when the number of variables vastly exceeds the number of observations. Our correction is data-driven, and it materially diminishes the substantial impact of estimation error on weights and risk forecasts of minimum variance portfolios. We quantify that impact with a novel metric, the *optimization bias*, which has a positive lower bound prior to correction and tends to zero almost surely after correction. The sample eigenvalues are used to correct excess dispersion in the leading eigenvector. However, the sample eigenvalues have no direct bearing on large minimum variance portfolios: correcting the sample eigenvalues to their population counterparts does nothing to diminish the optimization bias.

[*]Departments of Economics and Statistics and Consortium for Data Analytics in Risk, University of California, Berkeley, CA 94720 and Aperio Group, lrg@berkeley.edu.

[†]Intelligent Financial Machines, Palo Alto, CA and Consortium for Data Analytics in Risk, University of California, Berkeley, CA (apapanicolaou@berkeley.edu).

[‡]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA and Consortium for Data Analytics in Risk, University of California, Berkeley, CA (shkolnik@ucsb.edu).

[§]We thank the Center for Risk Management Research, the Consortium for Data Analytics in Risk, and the Coleman Fung Chair for financial support. We thank Marco Avellaneda, Bob Anderson, Kay Giesecke, Nick Gunther, Guy Miller, George Papanicolaou, Yu-Ting Tai, participants at the the 3rd Annual CDAR Symposium in Berkeley, participants at the Swissquote Conference 2017 on FinTech, and participants at the UC Santa Barbara Seminar in Statistics and Applied Probability for discussion and comments. We are grateful to Stephen Bianchi, whose incisive experiment showing that it is errors in eigenvectors, and not in eigenvalues, that corrupt large minimum variance portfolios, pointed us in a good direction. Conversations with Hubeyb Gurdogan, David He and Alec Kercheval led to numerous insights that improved this article.

1. Introduction

There are countless instances throughout the physical, social and data sciences where covariance matrices of large random vectors must be estimated from small samples. In this article, we show that the sampling error inherent in this process leads to excess dispersion of the leading eigenvector, and we provide a data-driven adjustment that corrects the bias. The motivation for our work comes from quantitative finance, where vast numbers of securities and non-stationarity make large, noisy covariance matrices the norm. These matrices are routinely used to construct portfolios with mean-variance optimization, which overweights securities whose volatilities and correlations with other securities are underforecast. The embedded sampling error tricks the optimizer into constructing distorted and highly inefficient portfolios. This practical problem is the starting point for the theory developed in this article.

Simulation in a one-factor PCA model reveals that errors in security weights and risk forecasts of the simplest mean-variance optimized portfolio, minimum variance, are driven by errors in the leading eigenvector and not in its associated eigenvalue (or variance). In this experiment, communicated to us by Stephen Bianchi, errors in weights and risk forecasts of an estimated minimum variance portfolios are not diminished when the estimated leading eigenvalue is replaced by its population counterpart. In contrast, replacing the estimated leading eigenvector with its counterpart (and leaving the estimated eigenvalue alone) substantially improves estimates of both weights and risk forecasts for a minimum variance portfolio. The strength of this experiment lies in well-known empirical facts: a single, positive (or market-like) factor drives substantial return and risk in equity markets, and that this factor determines, to a great extent, the weights of mean-variance optimized portfolios.

Further investigation identifies the specific source of the problem as excess dispersion in the entries of the estimated leading eigenvector. To develop some intuition for why this is the case, consider a market where correlations are driven by a single factor, and suppose all security exposures to that factor are identical. With probability one, a PCA estimate of the leading factor will have higher dispersion, or coefficient of variation, of its entries. Decreasing the dispersion mitigates the estimation error. A fresh perspective and some non-trivial analysis are required to mathematically articulate and verify these effects in a general setting, and we carry that out in this paper. We remove just the right amount of dispersion required to produce minimum variance portfolios with good properties. We do not correct all of the estimation error. Rather, we correct estimation error stemming from excess dispersion in the leading estimated eigenvector. This turns out to be sufficient to mitigate distortion and inefficiency in an optimized minimum variance portfolio.

We frame our results in the context of a single-factor model. This enables us to highlight our novel approach to covariance matrix estimation in a setting that incorporates the most salient features of equity markets and minimizes irrelevant complications. We show that shrinking an estimate of the leading eigenvector toward the unique, positive dispersionless vector on the sphere by a precisely defined amount materially improves the accuracy of the weights of minimum variance portfolios and their risk forecasts. Our analysis sheds light on previously unknown aspects of how sampling error corrupts an estimated covariance matrix, and it has deep connections to quadratic optimization.

1.1. Our contributions

We identify and correct excess dispersion in the leading eigenvector of a sample covariance matrix, when the number of variables vastly exceeds the number of observations. Our analysis leads to a number of surprising results, and also to a method that substantially improves the accuracy of weights and risk forecasts for estimated minimum variance portfolios.

The centerpiece of our results is the optimization bias $\mathscr{E}$, which drives both the misspecification of a minimum variance portfolio and errors in its risk forecasts. The optimization bias depends on the inner product between the true leading eigenvector b and an estimate of it, as well along inner products of the true and estimated eigenvector with the unique, positive, dispersionless vector z on the sphere. The first surprise is that $\mathscr{E}$ has no dependence on estimated eigenvalues. In other words, you can get the eigenvalue very wrong and still get the minimum variance portfolio and its risk forecast very right.

For the PCA estimate h of the leading eigenvector b, $\mathscr{E}_p(h)$ is bounded away from zero almost surely as $p \uparrow \infty$, so that errors in estimated minimum variance portfolio weights and risk forecasts have a hard lower bound. For the population eigenvector b, the optimization bias $\mathscr{E}_p(b)$ is zero of course. The second surprise is the existence of a vector h_{τ^*}, determined by the spherical law of cosines along the geodesic between h and z, for which $\mathscr{E}_p(h_{\tau^*}) = 0$. In other words, h_{τ^*} zeroes out an important source of estimation error in a minimum variance portfolio even though the fixed number of observations in our sample prevents h_{τ^*} from being a consistent estimator of the population eigenvector b.

The vector h_{τ^*} is defined explicitly in terms of the population eigenvector b. However, we obtain a data-driven estimate h_τ of h_{τ^*}, and show that the optimization bias $\mathscr{E}_p(h_\tau)$ tends to 0 almost surely as $p \uparrow \infty$.

Proofs of our results rely on delicate arguments concerning the asymptotic behavior of sample eigenvectors. The third surprise is that our arguments are constructed entirely with tools from classical probability theory: strong laws of large numbers and central limit theorems. This emphasizes unexpected parallels between the high p low n regime, where the number of variables vastly exceeds the number of observations, and classical statistics, where the number of observations vastly exceeds the number of variables.

1.2. Related literature

Sampling error has been an issue for investors since 1952, when Harry Markowitz transformed finance by framing portfolio construction as a trade-off between mean or expected return and its variance. Markowitz's mean-variance optimal portfolios form an efficient frontier, which is the basis of theoretical breakthroughs as fundamental as the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT), as well as practical innovations as impactful as Exchange Traded Funds (ETFs). Since we do not observe efficient portfolios, we estimate them from data, so sampling error permeates every aspect of finance. The seminal paper is Markowitz (1952). See Treynor (1962), Sharpe (1964), Lintner (1965*b*), Lintner (1965*a*) and Mossin (1966) for the Capital Asset Pricing Model and Ross (1976) for the Arbitrage Pricing Theory.

The impact of sampling error on efficient frontier portfolios has been investigated thoroughly in simulation and empirical settings. For example, see Jobson & Korkie (1980), Britten-Jones (1999), Bianchi, Goldberg & Rosenberg (2017) and the references therein. DeMiguel, Garlappi & Uppal (2007) compare a variety of methods for mitigating estimation error, benchmarking against the equally weighted portfolio in out-of-sample tests. They conclude that unreasonably long estimation windows are required for current methods to consistently outperform the benchmark. We briefly mention a few important references that do not overlap at all with out work. Michaud & Michaud (2008) recommends the use of bootstrap resampling. Lai, Xing & Chen (2011) reformulate the problem of finding the mean-variance efficient frontier as one of stochastic optimization with unknown moments. Goldfarb & Iyengar (2003) develop a robust optimization procedure to determine the efficient frontier by embedding a factor structure in the constraint set.

Early work on estimation error and the efficient frontier focused on Bayesian approaches. Vasicek (1973) and Frost & Savarino (1986) were perhaps the first to impose informative priors on the model parameters. Prior work analyzed diffuse priors and was shown to be inefficient (Frost & Savarino 1986). The latter, instead, presumes all stocks are identical and have the same correlations. Vasicek (1973) specifies a normal prior on the cross-sectional market betas (leading factor). More realistic priors incorporating multi-factor modeling are analyzed in Pástor (2000) (sample mean) and Gillen (2014) (sample covariance). Formulae for Bayes's estimates of the return mean and covariance matrix based on normal and inverted Wishart priors may be found in Lai & Xing (2008, Chapter 4, Section 4.4.1).

A related approach to the Bayesian framework is that of shrinkage or regularization of the sample covariance matrix. In the Bayesian setup, sample estimates are "shrunk" toward the prior (Lai & Xing 2008). Shrinkage methods have been proposed in contexts where little underlying structure is present (Bickel & Levina 2008) as well as those in which a factor or other correlation structure is presumed to exist (e.g. Ledoit & Wolf (2003), Ledoit & Wolf (2004), Fan, Liao & Mincheva (2013) and Bun, Bouchaud & Potters (2016)). Perhaps surprisingly, shrinkage methods turn out to be related to placing constraints on the portfolio weights in the Markowitz optimization. Jagannathan & Ma (2003) show that imposing a positivity constraint typically shrinks the large entries of the sample covariance downward. This is generalized and analyzed further in DeMiguel, Garlappi, Nogales & Uppal (2009).

Factor models mitigate the impact of sampling error on an estimated covariance matrix by reducing the number of required parameters. Investors typically rely on fundamental models, where the factors (or correlation drivers) are identified in advance. Financial practitioners typically use the fundamental factor models developed in Sharpe (1963) and Rosenberg (1974), in which factor exposures are specified from observable data and factor returns are estimated with cross-sectional regression. Finance academics favor the dual construction of factor models popularized in Fama & French (1992), in which factor returns are observed and exposures are estimated by time series regression. Latent factor models, in which both exposures and returns are extracted from are used everywhere in science. TIn a financial context, the strengths and shortcomings of fundamental and latent factor models are complementary. Fundamental models are intuitive but prone to miss emerging return sources. Latent models are prone to false positives and can be hard to interpret, but they have the capacity to identify new sources of return. Further details are in Connor (1995).

Principal component analysis (PCA) has been the dominant technique for extracting latent factors from observed security returns since Ross (1976). Its use in a high dimensional low sample size (HL) regime, where the number of variables vastly exceeds the number of observations, is justified by Chamberlain & Rothschild (1983), as the population eigenvectors approach the true factors under a mild set of assumptions. PCA is applied in the HL regime in the pioneering work by Connor & Korajczyk (1986) and Connor & Korajczyk (1988). In this regime, sample eigenvectors exhibit behavior that can be counterintuitive, as discussed in Hall, Marron & Neeman (2005). Recent analysis of the HL regime is in (Wang & Fan 2017).

In the HL regime, the largest eigenvalues of the covariance matrix grow linearly in tits dimension. This is not the traditional random matrix theory, in which the number of variables grows in proportion to the number of observations. The seminal paper in this HH regime is Marchenko & Pastur (1967), and an extensive treatment of the subject is Bai & Silverstein (2010). In the HH regime, consistency of principal component analysis (PCA) estimates can be established, as shown in Bai & Ng (2008). In the setting of Markowitz portfolios, the impact of eigenvalue bias and optimal corrections are investigated in in El Karoui et al. (2010) and El Karoui (2013). Donoho, Gavish & Johnstone (2018) consider eigenvalue corrections in "spiked" covariance matrices, which are similar to the covariance matrices we consider (in the HL) regime in this article. Onatski (2012) extends this framework to consider "weak" factors.

It appears that while eigenvector bias is acknowledged, direct bias corrections are made only to the eigenvalues corresponding to the principal components (e.g. Ledoit & Péché (2011) in the HH regime and and Wang & Fan (2017) in the HL regime). Several approaches to alter the sample eigenvectors indirectly do exist. For example, Ledoit & Wolf (2004) shrinks a sample covariance matrix toward a structured covariance matrix. However, these approaches are not focused on characterizing the bias inherent to the sample eigenvectors themselves. Some work on characterizing the behavior of sample eigenvectors may be found in Paul (2007) and Shen, Shen, Zhu & Marron (2016).

A stream of the portfolio construction literature considers the impact of the shape of the leading factor on the weights of Markowitz portfolios in general and minimum variance in particular. Green & Hollifield (1992) shows that the dispersion of the leading factor exposures drives the extreme positions in the portfolio composition. Minimum variance is identified in Markowitz (1952, footnote 9) as the efficient portfolio for which security expected returns are assumed to be equal. Since that distinguished beginning, minimum variance has played an important role in financial theory and practice. As just one of many illustrations of its theoretical importance, consider the place of minimum variance in the family of optimized portfolios that can be constructed without reference to expected value, which is notoriously difficult or even impossible to forecast. This family includes risk parity and maximum diversification; see Anderson, Bianchi & Goldberg (2012) and Clarke, De Silva & Thorley (2013) as well as references therein. The tens or even hundreds of billions of dollars that have been invested in ETFs on minimum variance since the financial crisis provide evidence of its practical importance. The empirical properties of minimum variance portfolios are studied in Clarke, De Silva & Thorley (2006), and Clarke, De Silva & Thorley (2011) provides simple formulas for the weights of minimum variance portfolios in a single-index model. Goldberg, Papanicolaou, Shkolnik & Ulucam (2019) shows the beneficial impact of beta shrinkage on minimum variance portfolios.

Bender, Lee, Stefek & Yao (2009), Bianchi et al. (2017), Ledoit & Wolf (2017), Wang & Fan (2017) and Goldberg et al. (2019), and many studies referenced in those articles use portfolio metrics such as variance or volatility forecast ratios, out-of-sample volatility and tracking error to assess the accuracy of a covariance matrix. Tracking error is the workhorse of the financial services industry and it is used, for example, to construct ETFs. By definition, tracking error is the width of the distribution of the return difference between a portfolio and its benchmark. Typically, the benchmark is taken to be a broad market index. Bianchi et al. (2017) and Goldberg et al. (2019) use tracking error to gauge the impact of sampling error on optimization by measuring the width of the distribution of the return difference between portfolios constructed with population and finite sample covariance matrices.

Finally, we note that while the notion of eigenvector shrinkage is new, market beta shrinkage is widely used by financial practitioners. The idea has its origins in Vasicek (1973) and Blume (1975). A detailed history is in Goldberg et al. (2019).

2. Problem formulation

Let $\mathrm{e} = (1, \ldots, 1)$ be the vector in $\mathbb{R}^p$ of all ones and denote by $|\cdot|$ the Euclidian norm so that $|\mathrm{e}| = \sqrt{p}$. Given a $p \times p$ covariance matrix $\mathbf{\Sigma} = \mathrm{Var}(Y)$ of returns $Y \in \mathbb{R}^p$ to p securities, we consider the following optimization problem.

$$\min_{w \in \mathbb{R}^p} w^\top \mathbf{\Sigma} w \quad \mathrm{e}^\top w = 1 \tag{1}$$

The solution minimizes the variance of the portfolio return over all fully invested portfolios. In practice, the matrix $\mathbf{\Sigma}$ must be estimated from security returns data, and there is a plethora of literature documenting the detrimental impact of estimation error on the portfolio weights computed via (1) and related optimization problems. Our choice of (1) is guided by the simplicity and practical importance of minimum variance and the fact that it provides an ideal setting to illustrate the delicate tradeoffs inherent in correcting estimation error in a covariance matrix.

2.1. The optimization bias

We adopt a framework in which the number of securities is large and the number of observed returns is small. This arises in many practical situations. One example concerns the estimation of equity alpha and risk models based on daily data. In such settings, typical estimation universes include hundreds or even thousands of securities, and market non-stationarity severely limits the available data history.

These considerations lead us to treat p as large, with the associated asymptotics $p \uparrow \infty$, and accept finite sample error in all estimates. We begin by illustrating a phenomenon we term the *optimization bias*. Our analysis focuses on a simple model.

For a vector $\beta \in \mathbb{R}^p$ and $\sigma, \delta \in (0, \infty)$, consider the covariance matrix

$$\mathbf{\Sigma} = \sigma^2 \beta\beta^\top + \delta^2 \mathbf{I}, \tag{2}$$

where the $\mathbf{I}$ denotes a $p \times p$ identity matrix. This covariance model is consistent with a market model which captures, in a remarkably simple manner, the systematic and specific risk we observe in equity markets. In practice, the betas (β) in (2) are often taken to be security sensitivities to a cap-weighted index. For many investors, beta is the main indicator, or even the only indicator, of systematic risk. The σ and δ denote the volatilities of the market and the specific (diversifiable) return.

For our analysis, we adopt a normalization to the unit sphere in $\mathbb{R}^p$, defining

$$b = \frac{\beta}{|\beta|} \quad \text{and} \quad z = \frac{\mathrm{e}}{\sqrt{p}}. \tag{3}$$

Letting $\langle x, y \rangle = x^\top y$, the projection of $x \in \mathbb{R}^p$ onto $y \in \mathbb{R}^p$, we define

$$\mathscr{E}(h) = \frac{\langle b, z \rangle - \langle b, h \rangle \langle h, z \rangle}{1 - \langle h, z \rangle^2} \qquad |h| = 1,\, h \in \mathbb{R}^p. \tag{4}$$

We refer to $\mathscr{E}$ as the *optimization bias*, since it is arises from the interaction of the optimization in (1) and the estimation error in the estimated covariance matrix. To see this, consider a portfolio $\hat{w}$ computed by solving (1) but after replacing the covariance $\boldsymbol{\Sigma}$ by an estimate $\hat{\boldsymbol{\Sigma}}$. In particular, the triplet (β, σ, δ) that leads to the $\boldsymbol{\Sigma}$ in (2) is estimated by some $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$ from which an estimate $\hat{\boldsymbol{\Sigma}}$ is then constructed. The true variance V^2 of this estimated portfolio $\hat{w}$ is given (under the mild assumptions on $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$ stated in Appendix B) by

$$\mathrm{V}^2 = \hat{w}^\top \boldsymbol{\Sigma} \hat{w} = \sigma^2 \mu^2(\beta)\,(1 + \mathrm{d}^2(\beta))\,\mathscr{E}^2(h) + o_p \tag{5}$$

where the remainder o_p has $o_p \asymp 1/p$, i.e., there are fixed constants $c, C \in \mathbb{R}$ such that $c/p \le o_p \le C/p$ for all p sufficiently large. In (5), the vector $h = \hat{\beta}/|\hat{\beta}| \in \mathbb{R}^p$ and the $\mu(\beta)$ and $\mathrm{d}(\beta)$, the mean and dispersion (or coefficient of variation) of β respectively, are defined by

$$\mu(\beta) = \frac{1}{p}\sum_{i=1}^{p} \beta_i \quad \text{and} \quad \mathrm{d}^2(\beta) = \frac{1}{p}\sum_{i=1}^{p} \Big(\frac{\beta_i}{\mu(\beta)} - 1 \Big)^2. \tag{6}$$

A remarkable observation is that the dependence of V^2 on the estimates of the volatilities σ and δ vanishes for p large. In fact, the sole estimated quantity that determines the true variance V^2 in (5) is h, the estimate of the normalized betas b defined in (3). This dependence occurs through $\mathscr{E}(h)$ and we note that $\mathscr{E}(b) = 0$.

Note, V^2 is the expected out-of-sample variance of a given estimated portfolio $\hat{w}$, i.e., $\mathrm{V}^2 = \mathrm{Var}(Y^\top x \,|\, x = \hat{w})$ as $Y^\top x$ is the actual return to any given portfolio x. Since $\hat{w}$ minimizes the in-sample variance (with respect to $\hat{\boldsymbol{\Sigma}}$), it is instructive to compare V^2 in (5) to this estimated variance $\hat{\mathrm{V}}^2$. The latter (under the mild assumptions on $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$ stated Appendix B) has

$$\hat{\mathrm{V}}^2 = \hat{w}^\top \hat{\boldsymbol{\Sigma}} \hat{w} \asymp 1/p. \tag{7}$$

This states that the in-sample variance of the portfolio $\hat{w}$ vanishes as p grows.

The asymptotic estimates supplied by (5) and (7) provide a first indication of how the optimization bias $\mathscr{E}$ is related to the investment process. In particular, we observe that the ratio of the true variance to the estimated variance satisfies

$$\mathrm{V}^2/\hat{\mathrm{V}}^2 \asymp p\mathscr{E}^2(h). \tag{8}$$

which explodes for large p unless $\mathscr{E}^2(h)$ tends to zero. In finite sample however, regardless of the estimation procedure, we expect $\mathscr{E}^2(h)$ to be bounded away from zero. Thus, the in-sample minimum variance will be severely underestimated, for large portfolios, relative to that encountered out-of-sample. This is because the optimization in (1) exploits the deviations of h from the true vector b to hedge out the perceived systematic risk, yielding a deceptively small portfolio variance. Additional portfolio metrics such as the tracking error, which we use to measure the impact of sampling error on portfolio weights and s asymptotically proportional to $\mathscr{E}^2(h)$, are also adversely affected.[1]

It may seem entirely impossible to remedy the dilemma posed by (8) since in our finite sample regime, we cannot expect h to be a consistent estimator (i.e., h cannot tend to b for which $\mathscr{E}(b) = 0$). Yet, this is precisely what we accomplish.

2.2. Model and assumptions

We consider a linear model for the excess return to p securities of the form

$$Y = \beta X + Z, \tag{9}$$

where the $X \in \mathbb{R}$ and $Z \in \mathbb{R}^p$ are random variables, while $\beta \in \mathbb{R}^p$ represents a constant parameter to be estimated from data that is generated from the model.

To accommodate a forthcoming asymptotic analysis we consider the sequences $\{\beta_i\}_{i \in \mathbb{N}}$ and $\{Z^i\}_{i \in \mathbb{N}}$ and write $\beta = (\beta_1, \ldots, \beta_p)^\top$ and $Z = (Z^1, \ldots, Z^p)^\top$ for the vectors in (9) (with dimension p implied from context). All random variables are defined on a common probability space equipped with an expectation E, variance VAR and covariance COV operators, all with respect to a probability measure P.

Assumption 2.1. *For finite $\sigma, \delta > 0$, we have $\mathrm{VAR}(X) = \sigma^2$ and every $\mathrm{VAR}(Z^i) = \delta^2$. Furthermore, $X \neq 0$ almost surely and every $\mathrm{COV}(X, Z^i) = \mathrm{E}(X) = \mathrm{E}(Z^i) = 0$.*

[1] See Goldberg et al. (2019) for details.

The generating process based on (9), under Assumption 2.1, is called a single-index or "market model."[2] The systematic component βX is the sole driver of correlation in the security return Y, and the specific component of return Z diversifies away in large portfolios. While it is common to include additional drivers of correlation, they are not relevant to minimum variance portfolios (see, Clarke et al. (2011) and Goldberg et al. (2019)). Model (9) also provides the simplest and most parsimonious means to capture the empirically observed systematic and specific return components.[3] It allows us to isolate the profound influence of the leading factor β on portfolio construction without the the distraction of less important effects.

Our analysis adopts an asymptotic regime wherein the number observations n of the return Y are finite (and fixed) while the number of securities p grows large. This corresponds to the high dimension and low sample size (HL) setting, which is relevant to modern applications involving large data sets. Our assumptions below are concerned with the applicability of the HL regime to financial data and the technical conditions that are required for our analysis in Section 3 and Section 4.

Let $\mu_p(\beta) = \mu(\beta)$ and $\mathrm{d}_p(\beta) = \mathrm{d}(\beta)$ denote the mean and dispersion of β as defined in (6) with the subscript p denoting the dependence on the dimension.

Assumption 2.2. *The sequence $\{\beta_i\}_{i\in\mathbb{N}}$ is such that the $\{\mu_p(\beta)\}_{p\in\mathbb{N}}$ and $\{\mathrm{d}_p(\beta)\}_{p\in\mathbb{N}}$ converge to the limits $\mu_\infty(\beta) \in (0,\infty)$ and $\mathrm{d}_\infty(\beta) \in (0,\infty)$ respectively as $p \uparrow \infty$.*

Assumption 2.2 imposes regularity on the sequence $\{\beta_i\}_{i\in\mathbb{N}}$ in order to simplify the statements of the theoretical results. The requirement that the limit $\mu_\infty(\beta)$ is positive is without loss of generality, i.e., the $\{\beta_i\}_{i\in\mathbb{N}}$ may always be negated to ensure the limit has a positive sign, while simultaneously negating the return X. This results in no change to the model (9) nor to the covariance matrix $\boldsymbol{\Sigma}$ in (2).

We require further assumptions on the $\{Z^i\}_{i\in\mathbb{N}}$ and on the temporal correlation of their realizations. Let $\mathrm{Z}_j = (\mathrm{Z}_j^1, \ldots, \mathrm{Z}_j^p) \in \mathbb{R}^p$ be the random variable equal in law to $Z \in \mathbb{R}^p$ so that $\mathrm{Y}_j = \beta \mathrm{X}_j + \mathrm{Z}_j$ for X_j the jth realization of X.

Assumption 2.3. *The random variables $\{Z^i\}_{i\in\mathbb{N}}$ are pairwise independent and identically distributed and moreover, $\mathrm{Cov}(\mathrm{Z}_j^i, \mathrm{Z}_k^i) = 0$ for all $i \in \mathbb{N}$ and every $j \neq k$.*

[2] The market model is also the standard, one-factor model that, under Assumption 2.1, yields to the theoretical requirements of estimation procedures such as principal component analysis (PCA) (Jolliffe, Trendafilov & Uddin 2003). Note, for example, that $\mathrm{VAR}(Y) = \boldsymbol{\Sigma}$ is of the form in (2).

[3] The market model as developed in Sharpe (1963) facilitates the efficient implementation of mean-variance portfolio construction (Markowitz 1952) via the critical line algorithm (Markowitz 1956).

Assumption 2.3 may be relaxed to various forms of weak dependence (replacing pairwise independence) across the securities and mixing conditions (allowing for correlation) in time. This is evident from the proofs of the main results in Appendix A. We do not pursue such extensions, focusing instead on introducing the concept of the dispersion bias and its relationship to minimum variance portfolios.

We discuss the realism of Assumption 2.3 in the context of the temporal correlation and non-stationarity of financial returns. For example, market microstructure is evident in the time-series of returns at horizons of fractions of a section. Our focus is primarily on daily and even lower frequencies for which temporal correlation may not be a concern.[4] Moreover, Assumption 2.3 removes the temporal correlation from the specific return only, leaving the market return X and ultimately the security return Y to be potentially correlated in time. With respect to stationarity, returns do exhibit volatility regimes, indicating that long histories may not be relevant to current forecasts. As a consequence, risk estimates that rely on historical returns are often based on short histories. The length of the applicable history varies with analysis date and data frequency, and it also depends on the application. This underscores the importance of the asymptotic regime (HL) that we adopt, i.e., when the number of securities p vastly exceeds n, the number of observations.

3. Dispersion bias

Let $\mathbf{Y}$ denote the $p \times n$ data matrix of realized security returns, i.e., the matrix whose jth column is $\mathrm{Y}_j \in \mathbb{R}^p$. We denote by s_p^2 the largest eigenvalue of

$$\mathbf{S} = \mathbf{Y}\mathbf{Y}^\top / n, \tag{10}$$

the sample covariance matrix of the returns. Since b in (3) is the eigenvector of $\mathbf{\Sigma}$ in (2) with the largest eigenvalue, a natural estimate of b is the corresponding sample eigenvector. To this end, we take the following definition for the estimate h of b.

$$h \in \mathbb{R}^p \ : \ \mathbf{S}h = \mathrm{s}_p^2 h, \quad |h| = 1, \quad \mu_p(h) \geq 0. \tag{11}$$

Note that the condition on $\mu_p(h) = \frac{1}{p}\sum_{i=1}^p h_i$ is without loss of generality, as an h with $\mu_p(h) < 0$ can always be negated preserving the remaining requirements. This convention is adopted for consistency with Assumption 2.2 (c.f., $\mu_\infty(\beta) > 0$).[5]

The relative gap between s_p^2 and the average of the remaining nonzero eigenvalues of $\mathbf{S}$, denoted by ℓ_p^2, plays an important role in our analysis.

[4]In a compendium of stylized facts about financial returns, Cont (2001) argues that temporal dependence is not an important consideration at a daily horizon.

[5]Note, h is not directly comparable to β in the sense that h estimates $b = \beta/|\beta|$.

Define,

$$\psi_p = \sqrt{\frac{\mathrm{s}_p^2 - \ell_p^2}{\mathrm{s}_p^2}}. \tag{12}$$

To highlight the dependence on p, we write $\langle x, y\rangle_p = \langle x, y\rangle = x^\top y$ for any $x, y \in \mathbb{R}^p$ and let $\langle x, y\rangle_\infty = \lim_{p\uparrow\infty}\langle x, y\rangle_p$ provided that the limit exists. The following result characterizes the bias in a p-dimensional sample eigenvector h that estimates its population counterpart b for p large with respect to the z in (3).

Theorem 3.1. *Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. Then,*

$$\langle h, z\rangle_\infty = \langle h, b\rangle_\infty \langle b, z\rangle_\infty \quad \text{and} \quad \langle h, b\rangle_\infty = \psi_\infty \in (0, 1) \tag{13}$$

almost surely and $\psi_\infty = \lim_{p\uparrow\infty} \psi_p$ is a nondegenerate random variable almost surely.

The proof of Theorem 3.1 is deferred to Appendix A.

Remark 3.2. *The (deterministic) limit $\langle b, z\rangle_\infty$ exists and is in $(0, 1)$ under Assumption 2.2. This is easily seen from the calculation in Appendix C which shows that*

$$\langle b, z\rangle_p^2 = \frac{1}{1 + \mathrm{d}_p^2(\beta)} \to \frac{1}{1 + \mathrm{d}_\infty^2(\beta)} \in (0, 1) \text{ as } p \uparrow \infty.$$

since $\mathrm{d}_p(\beta)$ is assumed to converge to $\mathrm{d}_\infty(\beta) \in (0, 1)$ as part of Assumption 2.2. This confirms the relation in (13) is not trivial (i.e., $\langle h, z\rangle_p$ and $\langle b, z\rangle_p$ converge to zero).

Remark 3.3. *The asymptotic angle $\langle h, b\rangle_\infty$ between the sample eigenvector h and its population counterpart b has been studied in Shen et al. (2016) and elsewhere. Our result differs in three respects from these prior works. First, the proof leverages the structure of the factor model in Section 2.2 and consequently uses different techniques. Second, our characterization of $\langle h, b\rangle_\infty$ is in terms of the limit of ψ_p which may be computed from the observed returns data* **Y**. *This facilitates the correction for the bias in Section 4. Third, an expression for $\langle h, b\rangle_\infty$ alone does not point to a correction, as bias has to be characterized with respect to some known vector. In our case, it is z.*

Remark 3.4. *Numerical evidence suggests this dispersion bias phenomenon continues to hold under a much weaker set of conditions than those of Assumptions 2.1–2.3.*

We refer to the systematic error identified by Theorem 3.1 as the dispersion bias (of a sample eigenvector) for the following reason. The dispersion $\mathrm{d}_p(h)$ of h has

$$\mathrm{d}_p^2(h) = \frac{1}{p}\sum_{i=1}^{p}\Big(\frac{h_i}{\mu_p(h)} - 1\Big)^2 = \frac{1 - \langle h, z\rangle_p^2}{\langle h, z\rangle_p^2} \tag{14}$$

by a calculation similar to that of Remark 3.2 (see Appendix C). Theorem 3.1 implies that $\langle b, z\rangle_p > \langle h, z\rangle_p$ with high probability (w.h.p.)[6] in p. Consequently, for p large, $\mathrm{d}_p(h)$ typically exceeds the dispersion $\mathrm{d}_p(b) = \mathrm{d}_p(\beta) > 0$ of b, since

$$\mathrm{d}_p^2(h) = \frac{1 - \langle h, z\rangle_p^2}{\langle h, z\rangle_p^2} > \frac{1 - \langle b, z\rangle_p^2}{\langle b, z\rangle_p^2} = \mathrm{d}_p^2(b) \quad \text{w.h.p. in } p. \tag{15}$$

More specifically, we have the following corollary of Theorem 3.1 which specifies (asymptotically) the amount by which h is overly dispersed relative to b.

Corollary 3.5. *Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. Then,*

$$\mathrm{d}_\infty^2(h) = \frac{1 - \psi_\infty^2}{\langle h, z\rangle_\infty^2} + \mathrm{d}_\infty^2(b) \tag{16}$$

almost surely where $\mathrm{d}_\infty^2(h) = \lim_{p\uparrow\infty} \mathrm{d}_p^2(h)$ and $\mathrm{d}_\infty(\beta) = \mathrm{d}_\infty(b)$.

Proof. This is a consequence of equations (13) and (14). ■

The characterization of the dispersion bias in the leading eigenvector of the sample return covariance matrix has significant implications for PCA estimates of optimized portfolios. In particular, recalling that the optimization bias $\mathscr{E}(h)$ in (4),

$$\mathscr{E}_p(h) = \frac{\langle b, z\rangle_p - \langle b, h\rangle_p \langle h, z\rangle_p}{1 - \langle h, z\rangle_p^2} \tag{17}$$

(we add the subscript p to highlight the dependence) is the primary driver of error in minimum variance portfolios, motivates the following corollary of Theorem 3.1. To see its ramifications for PCA estimated portfolios, recall from (8) that the ratio of the true to the estimated minimum variance satisfies $\mathrm{V}^2/\hat{\mathrm{V}}^2 \asymp p\mathscr{E}_p^2(h)$.

Corollary 3.6. *Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. Then,*

$$\mathscr{E}_\infty(h) = \frac{1 - \psi_\infty^2}{\mathrm{d}_\infty^2(h) \langle h, z\rangle_\infty \psi_\infty} \tag{18}$$

almost surely where $\mathscr{E}_\infty(h) = \lim_{p\uparrow\infty} \mathscr{E}_p(h)$ with $\mathscr{E}_\infty(h) > 0$ almost surely.

Proof. The result follows upon combining (13), (14) and (17). That $\mathscr{E}_\infty(h) > 0$ follows from (13) and that $\langle b, z\rangle_\infty > 0$ (hence $\langle h, z\rangle_\infty > 0$) per Remark 3.2. ■

[6]We say $A_p > B_p$ w.h.p. (in p) if for any $\epsilon > 0$ there is a p_ϵ such that $A_p > B_p$ on a set of probability $1 - \epsilon$ for all $p \geq p_\epsilon$. Our usage of this term is not standard, as w.h.p. typically states that an event $\mathcal{E}_p$ holds w.h.p (in p) if for all $\epsilon > 0$ there is a p_ϵ such that $\mathrm{P}(\mathcal{E}_p) > 1 - \epsilon$ for all $p \geq p_\epsilon$. The stronger statement we make is facilitated by Egoroff's theorem (Cohn 2013, Proposition 1.3.4).

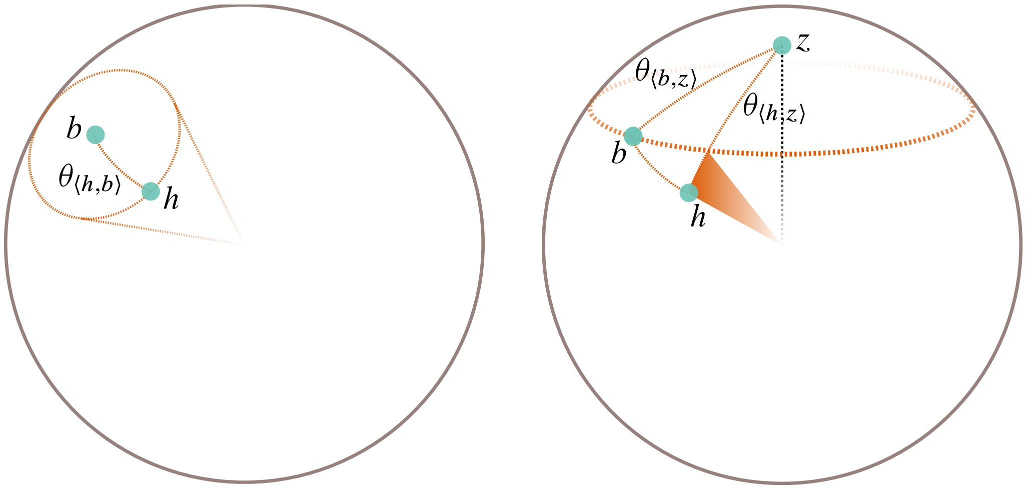

Figure 1. *An illustration of the sample and population eigenvectors h and b respectively. The angle $\theta_{\langle h,b\rangle}$ between h and b is also the length of the arc on the sphere between h and b. The left panel shows that the identification of the bias in h is not possible without a reference frame. The latter is provided by the vector z in the right panel, where the bias is illustrated by the shaded area. This bias is the amount by which $\theta_{\langle h,z\rangle}$ exceeds $\theta_{\langle b,z\rangle}$ and is due to the excess dispersion in h.*

In the remainder of this section, we discuss the the geometric interpretation of the dispersion bias suggested by the natural normalization of PCA estimates to the unit sphere. Figure 1 illustrates the vectors h, b and z on the unit sphere in $\mathbb{R}^p$, with angle between any x and y denoted by $\theta_{\langle x,y\rangle_p}$ so that $\langle x, y\rangle_p = \cos\theta_{\langle x,y\rangle_p}$. Since the population eigenvector b is unknown, there is no available direction with respect to which the sample eigenvector h is biased. In particular, even given the known estimate ψ_p of $\langle h, b\rangle_p = \cos\theta_{\langle h,b\rangle_p}$, the left panel of Figure 1 shows that h may be located anywhere on the cone around b of radius $\theta_{\langle h,b\rangle_p}$ (c.f., Remark 3.3). The right panel of Figure 1 illustrates the portion of the bias of h that is identifiable relative to the vector z. In particular, $\theta_{\langle h,z\rangle_p} > \theta_{\langle b,z\rangle_p}$ w.h.p. in p, which is equivalent to the statement in (15) in terms of the dispersions of h and b. This bias representation also points to a potential correction which is the topic explored in Section 4.

4. Bias correction

To correct the dispersion bias in the estimate h (of b) specified by the sample eigenvector in (11), we propose the following parametrized family of estimators.

$$h_t = \frac{h + tz}{|h + tz|}, \qquad t \in \mathbb{R}. \tag{19}$$

The optimization bias $\mathscr{E}_p(h)$ given in (17) that stems from the estimate h may then be replaced by $\mathscr{E}_p(h_t)$ upon replacing h with the estimator h_t in

(17). We have,

$$\mathscr{E}_p(h_t) = \mathscr{E}_p(h) - t\left(\frac{\langle h,b\rangle_p - \langle b,z\rangle_p\langle h,z\rangle_p}{1-\langle h,z\rangle_p^2}\right) \tag{20}$$

for any $t \in \mathbb{R}$ (see Appendix C). We propose a randomized choice τ_p for t in (19),

$$\tau_p = \frac{(1-\psi_p^2)\,\langle h,z\rangle_p}{\psi_p^2 - \langle h,z\rangle_p^2}. \tag{21}$$

We let h_τ be the estimator constructed with τ_p replacing t in (19) with subscript p in (21) inferred from the dimension of $h \in \mathbb{R}^p$. It is optimal in the following sense.

Theorem 4.1. *Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. Then,*

$$\mathscr{E}_\infty(h_\tau) = \lim_{p\uparrow\infty}\mathscr{E}_p(h_\tau) = 0 \quad \textit{almost surely.} \tag{22}$$

Moreover, the parameter τ_p in (21) may be computed from $p \times n$ data matrix $\mathbf{Y}$ only.

The proof (see Appendix A) is a consequence of Theorem 3.1 and fact that

$$\mathscr{E}_p(h_\tau) = \frac{\langle b,z\rangle_p\psi_p^2 - \langle h,z\rangle_p\langle h,b\rangle_p}{\psi_p^2 - \langle h,z\rangle_p^2}. \tag{23}$$

The second part of the result is trivial, but it crucially shows that the optimal parameter τ_p in (21) is computable directly from the observed quantities. In particular, it may be directly computed from the sample covariance matrix $\mathbf{S}$.[7] The first part of the result is remarkable in that even for only two observations ($n = 2$) of the return Y we are able to remove all of the optimization bias asymptotically.

The implications for the minimum variance portfolio are as follows. Recall that the true variance of an estimated portoflio $\hat{w}$ is $\mathrm{V}_p^2 = \hat{w}^\top\mathbf{\Sigma}\hat{w}$ and given in (5) by

$$\mathrm{V}_p^2 = \sigma^2\mu_p^2(\beta)\,(1+\mathrm{d}_p^2(\beta))\,\mathscr{E}_p^2(h) + o_p \tag{24}$$

for $o_p \downarrow 0$. Recall from Corollary 3.6 that the limit $\mathscr{E}_\infty^2(h) > 0$ almost surely for the plain PCA estimate h, and under our assumptions, the V_p^2 remains bounded away from zero almost surely. In other words, the expected out-of-sample variance is strictly positive and potentially large (see (18)). On

[7]The vector h is an eigenvector of $\mathbf{S}$ and ψ_p is a function of the eigenvalues of $\mathbf{S}$.

the other hand, replacing h with the estimator h_τ ensures the expected out-of-sample variance tends to zero.

Next, we partially address the rate of convergence of the corrected optimization bias $\mathscr{E}_p^2(h_\tau)$ to zero. This has important implications for the ratio of the true to the estimated portfolio variance in (8) which behaves as $p\mathscr{E}_p^2(h_\tau)$ (see Appendix B).

Conjecture 1. *Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. In addition, suppose that every Z^i of Assumption 2.1 has a finite 6th moment. Then,*

$$\sup_p \mathrm{E}\left(p\mathscr{E}_p^2(h_\tau)\right) < \infty. \tag{25}$$

Remark 4.2. *We provide numerical support for (25) in Section 5.*

Remark 4.3. *In contrast to (25), we have $\limsup_{p\uparrow\infty} p\mathscr{E}_p^2(h_\tau) = \infty$ almost surely. This stems from a variant of the law of iterated logarithms, the fact that the scaled sum $Q_p = \frac{Z^1+\cdots+Z^p}{\sqrt{p}}$ for i.i.d. random variables $\{Z^i\}_{i\in\mathbb{N}}$ has $\limsup_{p\uparrow\infty} Q_p = \infty$ almost surely. The random walk oscillations are too erratic to avoid path-by-path entirely, but given sufficient finite moments of Z^1 (as above), they cancel in expectation.*

Remark 4.4. *In view of Corollary 3.6, $\sup_p \mathrm{E}\left(c_p\mathscr{E}_p^2(h)\right) = \infty$ for any $c_p \to \infty$.*

In the remainder of this section we explore some of the features of the estimator h_τ as compared with h, the (unadjusted) sample eigenvector. Theorem 3.1 implies that h is adversely affected by a dispersion bias since $\mathrm{d}_p^2(h) > \mathrm{d}_p^2(b)$ w.h.p. per (15), i.e., the dispersion of h is larger than that of b, the population eigenvector. But,

$$\mathrm{d}_p^2(h_t) = \frac{1-\langle h_t, z\rangle_p^2}{\langle h_t, z\rangle_p^2} = \frac{1-\langle h, z\rangle_p^2}{(\langle h, z\rangle_p + t)^2} < \mathrm{d}_p^2(h) \qquad t > 0 \tag{26}$$

for h_t in (19). In other words, for a positive parameter value, the estimator h_t has the effect of decreasing the dispersion of h. The left panel of Figure 2 illustrates the placement of a h_t relative to the b and the dispersionless vector z. Observe, that since $\psi_\infty^2 - \langle h, z\rangle_\infty^2 = \psi_\infty^2(1 - \langle b, z\rangle_\infty^2) > 0$, under the assumptions of Theorem 3.1, the optimal parameter τ_p given in (21) has $\tau_\infty = \lim_{p\uparrow\infty} \tau_p > 0$ almost surely.

To obtain some intuition for the precise value of the paramter τ_p and its effect on the optimization bias $\mathscr{E}_p$, we make the following two observations.

(1) As noted previously, $\mathscr{E}_p(b) = 0$ which implies that h being a consistent estimator (i.e., $\langle h, b\rangle_p \to 0$) is sufficient for removing the optimization bias asymptotically. However, this is not possible in finite sample (when n is fixed). Consistency turns out, surprisingly, to not be necessary. A remarkable property of the optimization bias is that it has a root that is distinct

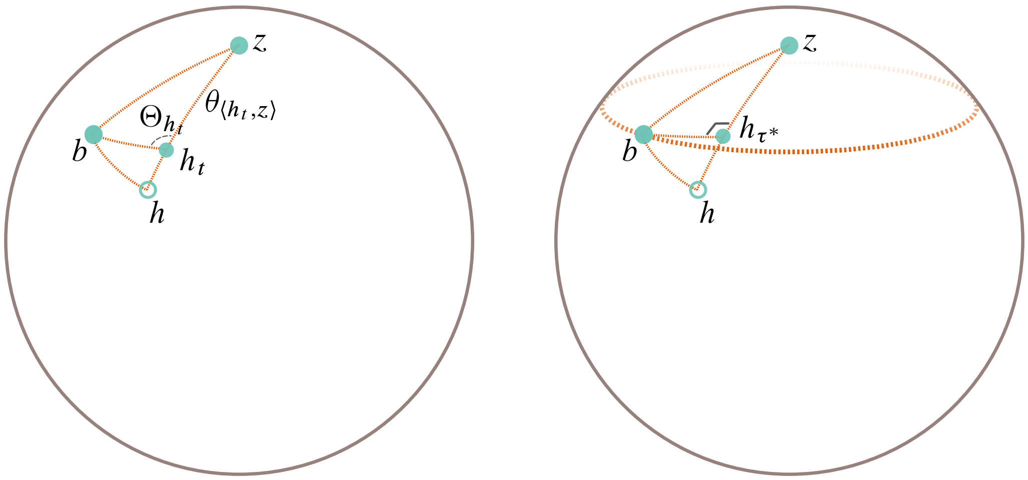

Figure 2. *An illustration of the placement of the estimator h_t for a particular $t > 0$ (left panel), which shrinks the dispersion $\mathrm{d}(h_t)$ of h_t relative to that of h. Equivalently, we shrink the angle $\theta_{\langle h_t,z\rangle}$ relative to $\theta_{\langle h,z\rangle}$ by taking $t > 0$ (see also Figure 1). The spherical angle Θ_{h_t} controls the optimization bias $\mathscr{E}_p(h_t)$ per formula (28). The right panel illustrates the choice h_{τ^*} at which $\Theta_{h_{\tau^*}} = 90$ degrees. The equi-mean contour of b (points x with $|x| = 1$ and $\langle x, z\rangle = \langle b, z\rangle$) is further from z than the optimal point h_{τ^*}. This means that the distance to the point b is not minimized at the intersection of its equi-mean contour and $\{h_t\}_{t\in\mathbb{R}}$.*

from the unknown vector b. It is easy to verify, via (20), that $\mathscr{E}_p(h_{\tau^*}) = 0$ for h_t as in (19) but with $t = \tau^*$ given by

$$\tau_p^* = \frac{\langle b, z\rangle_p - \langle h, b\rangle_p \langle h, z\rangle_p}{\langle h, b\rangle_p - \langle h, z\rangle_p \langle b, z\rangle_p} \tag{27}$$

where the suppressed subscript p in h_{τ^*} is inferred from the dimension of $h \in \mathbb{R}^p$. However, τ_p^* cannot be constructed in practice since b is not known. Theorem 4.1 states that τ_p approximates τ_p^* for p large, and τ_p is implementable from the observed data. Indeed, it is not difficult to check that $|\tau_p - \tau_p^*| \to 0$ as $p \uparrow \infty$ almost surely.

(2) The geometry of the (finite p) optimal point τ_p^* in (27) is best illustrated with the spherical law of cosines (Banerjee 2004). Recalling that $\theta_{\langle x,y\rangle_p}$ denotes the angle between x and y in $\mathbb{R}^p$, we can write the optimization bias of h_t for any $t \in \mathbb{R}$ as

$$\mathscr{E}_p(h_t) = \frac{\langle b, z\rangle_p - \langle h_t, b\rangle_p \langle h_t, z\rangle_p}{1 - \langle h_t, z\rangle_p^2} = \left(\frac{\sin\theta_{\langle h_t, b\rangle_p}}{\sin\theta_{\langle h_t, z\rangle_p}}\right)\cos\Theta_{h_t} \tag{28}$$

where Θ_{h_t} denotes the spherical angle between the arcs emanating from h_t, i.e., the arcs from h_t to z and h_t to b. Figure 2 illustrates the spherical angle Θ_{h_t} and we note that when $\Theta_{h_t} = 90$ degrees, for such $t \in \mathbb{R}$, we have

$\mathscr{E}_p(h_t) = 0$ per (28). This occurs precisely for $t = \tau_p^*$ in (27). It may be verified that τ_p^* is also the maximizer of $\langle h_t, b\rangle_p$ over $t \in \mathbb{R}$ and equivalently, τ_p^* is the minimizer of $\theta_{\langle h_t, b\rangle_p}$ over $t \in \mathbb{R}$, i.e., $t = \tau_p^*$ minimizes the arc length between b and h_t on the unit sphere in $\mathbb{R}^p$. The random parameter τ_p approximates this minimizer asymptotically as $p \uparrow \infty$.

Our findings are particularly interesting from the perspective of the interplay between the geometry of the optimization bias and the optimality of parametrized family of estimators $\{h_t\}_{t\in\mathbb{R}}$ in (19). The analysis of the minimum variance in Section 2.1 motivated this family of estimators. But, conversely, viewing the $\{h_t\}_{t\in\mathbb{R}}$ as a family of (dispersion) shrinkage estimators, the optimal choice of h_t is naturally the one that minimizes its "distance" to b, the unknown. This h_{τ^*} coincides with the root of the optimization bias $\mathscr{E}_p$ and so yields "optimal" minimum variance portfolios. The fact that the h_{τ^*} may be arbitrarily well approximated in finite sample, only from the observed data, and simply by considering more variables, is striking.

5. Numerical study

We present results of two experiments that illustrate the impact of the dispersion bias correction on an optimized minimum variance portfolio and corroborate the theoretical results of Section 4.[8]

We generate observations $Y_j \in \mathbb{R}^p$ of returns for securities $j = 1, \dots, n$ from model (9) so that

$$Y_j = \beta X_j + Z_j \tag{29}$$

for unobserved factor and specific returns $X_j^\top$ and $Z_j = (Z_j^1, \dots, Z_j^p)^\top$ respectively, which are mean-zero and normally distributed. We require the generating process to obey Assumptions 2.1 and 2.3, so every $\mathrm{Var}(Y_j) = \mathbf{\Sigma} = \sigma^2\beta\beta^\top + \delta^2\mathbf{I}$ as in (2). We further take i.i.d. $\{X_j\}_{j=1}^n$ with $\mathrm{Var}(X_1) = \sigma^2 = (0.16)^2$ and i.i.d. $\{Z_j\}_{j=1}^n$ with $\mathrm{Var}(Z_j) = \delta^2\mathbf{I} = (0.5)^2\mathbf{I}$. The vector $\beta \in \mathbb{R}^p$ is constructed to have mean $\mu(\beta) = 1$ and dispersion $\mathrm{d}(\beta) = 0.5$, and is held constant over the observations.[9]

We extract a PCA estimate h from the $p \times n$ data matrix $\mathbf{Y} = (Y_1, \dots, Y_n)$, which per (11) is the positive-mean eigenvector of the sample covariance $\mathbf{S}$ in (10). The estimator $h_t \propto h + tz$ is formed via (19) for constant $t \in \mathbb{R}_+$. Similarly, we form the estimators h_{τ^*} and h_τ via (19) with the random (dispersion) shrinkage parameters τ_p^* and τ_p in (27) and (21). The latter relies

[8] For complementary simulations calibrated to the US equity market, see Goldberg et al. (2019).

[9] More precisely, (independently of all other variables) some $\{\eta_i\}_{i=1}^p$ are drawn independently from the Normal distribution of mean one and variance one. The transformation $\beta_i = c\eta_i/\mu(\eta) + (1-c)$ is then applied with $c = 0.5/\mathrm{d}(\eta)$ so that $\mathrm{d}(\beta) = 0.5$ for $\beta = (\beta_1, \dots, \beta_p)^\top$.

on the eigenvalues of $\mathbf{S}$. We refer to h_{τ^*} as the *exact* estimator since it carries no optimization bias. We call h_τ the *blind* estimator as it is unable to observe β, and the family h_t, for constant $t \in \mathbb{R}_+$, the *parametric* estimator, to which the PCA estimator $h_0 = h$ belongs.

5.1. Market model estimation

We use each of the estimates h_t, h_{τ^*} and h_τ as the basis of an estimate $\hat{\mathbf{\Sigma}}$ of the covariance matrix

$$\mathbf{\Sigma} = \sigma^2 \beta\beta^\top + \delta^2 \mathbf{I},$$

as in (2). The structural form of $\mathbf{\Sigma}$ is implied by (9) and Assumptions 2.1 and 2.3. It follows that our task amounts to specifying estimates $\hat{\sigma}^2$, $\hat{\beta}$ and $\hat{\delta}^2$ of σ^2, β and δ^2.

To motivate our model calibration, we rely on Lemma A.2, which shows specific variance δ^2 is equal to the average of the eigenvalues of $\mathbf{\Sigma}$ after excluding the leading eigenvalue scaled by n/p. We adapt this recipe to the sample covariance matrix by taking $\ell_p^2 = (\mathrm{Tr}(\mathbf{S}) - \mathrm{s}_p^2)/(n-1)$, the average of the eigenvalues of $\mathbf{S}$ excluding zeros and s_p^2, the largest value. Then $\hat{\delta}^2 = (n/p)\ell_p^2$. As it is only that product of σ^2 and $|\beta|$ that can be identified and not their individual values (since the factor returns X_j are not observed), and since the product $\sigma^2|\beta|^2$ is equal to the leading eigenvalue of $\mathbf{\Sigma}$ minus the specific variance scaled by $1/n$ as shown in Lemma A.2, take $\hat{\sigma}^2|\hat{\beta}|^2 = p(\mathrm{s}_p^2/p - \hat{\delta}^2/n) = \mathrm{s}_p^2 - \ell_p^2$. As the scale $|\hat{\beta}|$ cannot be identified, we are free to assume it is 1 and to set $\hat{\sigma}^2|\hat{\beta}|^2 = \hat{\sigma}^2 = \mathrm{s}_p^2 - \hat{\delta}^2$. Finally we set $\hat{\beta}$ to be h_t, h_{τ^*} or h_τ.[10] Given a choice of $\hat{\beta}$, we estimate the covariance matrix $\mathbf{\Sigma}$

$$\hat{\mathbf{\Sigma}} = \hat{\sigma}^2 \mathbf{H}_t + \hat{\delta}^2 \mathbf{I}; \qquad \mathbf{H}_t = \hat{\beta}\hat{\beta}^\top. \tag{30}$$

We remark that the values $\hat{\sigma}^2$ and $\hat{\delta}^2$ are identical for all our estimators $\hat{\mathbf{\Sigma}}$; it is only the estimate of $\hat{\beta}$ that changes. Further, the leading eigenvalue is the same for all our estimators $\mathbf{\Sigma}$, and it is equal to s_p^2, the leading eigenvalue if $\hat{\mathbf{\Sigma}}$.

5.2. Dispersion bias identification

Table 1 provides support for Theorem 3.1 with estimates of the means and standard deviations of $\langle h, z\rangle_p$, $\langle h, b\rangle_p$ and ψ_p over 10^6 simulations with $n = 50$ observations and numbers of securities p ranging from 500 to 8000. The value of $\langle b, z\rangle$ multiplied by the point estimate $\mathrm{E}(\langle h, b\rangle_p)$ is equal to the point estimate $\mathrm{E}(\langle h, z\rangle_p)$ to four decimal places for each p. is well within the 99% confidence intervals around $\mathrm{E}(\langle h, z\rangle_p)$ for each p. Further, the sample means of the values $\langle h, b\rangle_p$ and ψ_p, which are asymptotically equal, get

[10] A different choice of scale for $|\hat{\beta}|$ would lead to estimates of $\hat{\beta}$ that look more like security market betas. For example, we could normalize $\hat{\beta}$ so that its mean entry is equal to 1. However, this choice would not affect the numerical results presented here.

closer as p increases from 500 to 8000. The 99% confidence intervals around $\mathrm{E}(\psi_p)$ and $\mathrm{E}(\langle h, b\rangle_p)$, however, do not overlap for the values of p we considered.

p	$\mathrm{E}(\langle h,z\rangle_p)$	$\mathrm{E}(\langle h,b\rangle_p)$	$\mathrm{E}(\psi_p)$	$\mathrm{SD}(\langle h,z\rangle_p)$	$\mathrm{SD}(\langle h,b\rangle_p)$	$\mathrm{SD}(\psi_p)$
500	0.8287	0.9265	0.9290	0.01513	0.01467	0.01305
1000	0.8292	0.9271	0.9283	0.01358	0.01399	0.01318
2000	0.8295	0.9274	0.9280	0.01227	0.01366	0.01325
4000	0.8296	0.9275	0.9278	0.01235	0.01349	0.01328
8000	0.8297	0.9276	0.9277	0.01213	0.01340	0.01330

Table 1. *Sample means and standard deviations for $\langle h,z\rangle_p$, $\langle h,b\rangle_p$ and ψ_p with $\langle b,z\rangle \approx 0.89442$ (corresponding to $\mathrm{d}^2(b) = 0.5$) over 10^6 simulations. For each value of p, we use $n = 50$. Estimates for $\mathrm{E}(\langle h,z\rangle_p)$ have 99% confidence intervals of $\pm 2.58\,\mathrm{SD}(\langle h,z\rangle_p) \times 10^{-3}$ and analogously for $\mathrm{E}(\langle h,b\rangle_p)$ and $\mathrm{E}(\psi_p)$.*

5.3. Minimum variance portfolio volatility

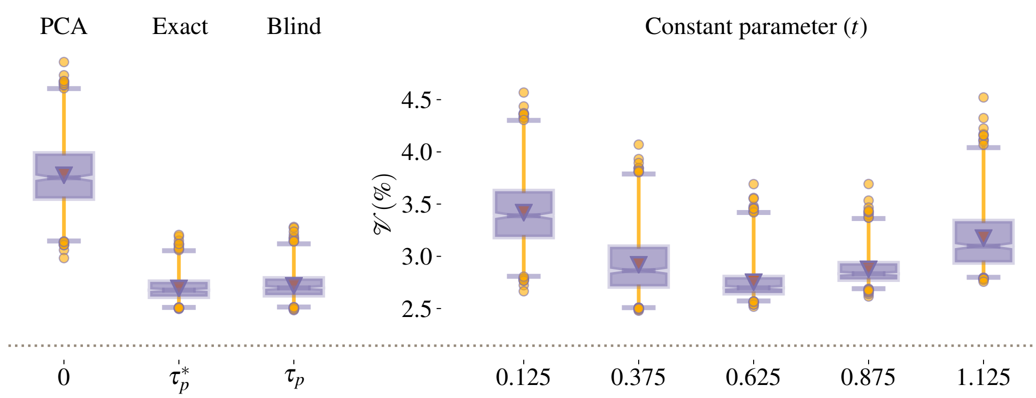

Figure 3. *Boxplots of the true volatility* V *of the estimated minimum variance portfolio $\hat{w}$ constructed with the parametric, exact and blind estimators. The boxplots for five constant parameter estimators h_t are shown on the right of the* V*-axis. The minimizer (over t) of the average volatility is approximately* 0.625*. The left of the* V*-axis shows boxplots for the* PCA, exact *and* blind *estimators. The dashed line marks the optimal minimum volatility $\mathscr{M}$. For each boxplot, we perform 10^3 simulations, each consisting of $n = 50$ observations of $p = 500$ securities. Each boxplot shows the interquartile range, the means marked with triangles, and the outliers that lie below* 1% *and above* 99% *of the distribution.*

We compare the performance of the exact and blind estimators to PCA and investigate the behavior of the parametric estimator. For each estimator we form a covariance estimate $\hat{\boldsymbol{\Sigma}}$ and compute the estimated, minimum

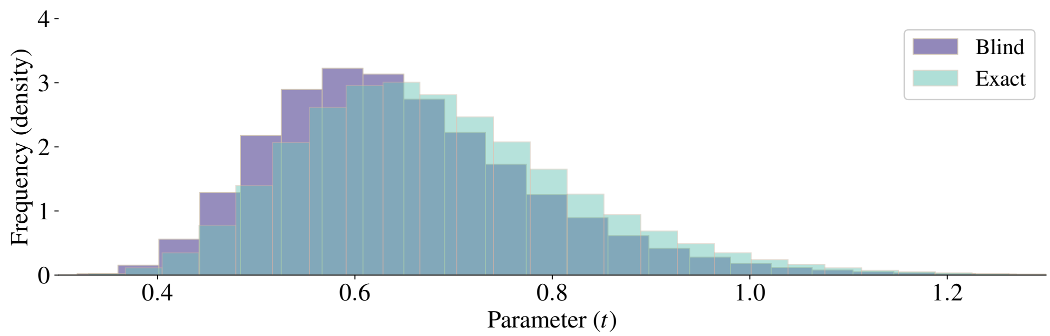

Figure 4. *Histograms of* 10^5 *simulations for the* exact *and* blind *(dispersion) shrinkage parameters* τ_p^* *and* τ_p *respectively (see formulas* (21) *and* (27)*). The sample means and variances are approximately* 0.685 *and* 0.021 *for* τ_p^* *and* 0.649 *and* 0.019 *for* τ_p*. For each, we take* $n = 50$ *observations and* $p = 500$ *securities.*

variance portfolio $\hat{w}$ by solving (1) after replacing $\mathbf{\Sigma}$ by $\hat{\mathbf{\Sigma}}$. We take the true volatility V (the square root of $\mathrm{V}^2 = \hat{w}^\top \mathbf{\Sigma} \hat{w}$ defined in (5)) of the estimated minimum variance portfolio as our performance metric. This metric, which may be regarded as the out-of-sample volatility, emphasizes the practical utility of the experiments we conduct.

Figure 3 shows distribution boxplots of V for each of the estimators. Even for the moderate number of securities ($p = 500$) and small sample size ($n = 50$) used in experiment, the exact and blind estimators materially outperformed PCA. For instance, relative to the PCA estimator, we observe a reduction in median V of more that 25% along with a reduction of more than 50% in the interquartile range for the exact and blind estimators. The (horizontal) dashed line in Figure 3 marks the value of the optimal minimum volatility $\mathscr{M}$, the square root of the minimum value $w^\top \mathbf{\Sigma} w$ attained in optimization (1). All estimators produce portfolios with higher volatility than $\mathscr{M}$ (approximately 2.144), indicating a higher level of risk than optimal. However, the median volatility produced by the exact and blind estimators are both within 25% of the optimum $\mathscr{M}$, while the PCA estimator yields a median volatility that exceeds the optimum by 75%. Figure 3 displays results for five parametric estimators (on the right of the V-axis). The best parametric estimator achieves similar performance gains to the blind and exact estimators. The estimator h_{t^*} corresponds to the value $t^* = 0.625$ that necessarily depends on the unknown β. It approximately minimizes the median volatility (a function of $\mathbf{\Sigma}$) over the nonrandom parameter choices. However, this value is never accessible in practice since $\mathbf{\Sigma}$ is not known in such settings. Remarkably, it underperforms (in term of the mean and the variance) the blind estimator, which relies only on the observed data $\mathbf{Y}$. The distributions of the random, dispersion-shrinkage parameters τ_p^* and τ_p are shown in Figure 4. These histograms support the theoretical finding that τ_p approximates well the exact parameter τ_p^*, which benefits from the knowledge of β in the model.

	$\mathrm{E}(\mathrm{V})$	$\mathrm{E}(\hat{\mathrm{V}})$	$\mathrm{E}(\mathrm{V}/\hat{\mathrm{V}})$	$\mathrm{SD}(\mathrm{V})$	$\mathrm{SD}(\hat{\mathrm{V}})$
PCA	3.770	1.739	2.178	0.313	0.070
Exact	2.691	2.690	1.000	0.109	0.086
Blind	2.715	2.646	1.029	0.124	0.159
Param	2.750	2.621	1.053	0.159	0.121

Table 2. *Sample means and standard deviations for the true and estimated volatilities* V *and* $\hat{\mathrm{V}}$ *and their ratio* $\mathrm{V}/\hat{\mathrm{V}}$ *over* 10^6 *simulations. Estimates for* $\mathrm{E}(\mathrm{V})$ *have* 99% *confidence intervals of* $\pm 2.58\,\mathrm{SD}(\mathrm{V}) \times 10^{-3}$ *and analogously for* $\mathrm{E}(\hat{\mathrm{V}})$*. Param denotes the best parametric estimator* $h_{0.625}$.

Table 2 supplements Figure 3 and reports the sample means and variances of the true volatility V for PCA, exact, blind and best parametric (Param) estimators. It also reports the same statistics for the estimated volatility $\hat{\mathrm{V}}$ where $\hat{\mathrm{V}}^2 = \hat{w}^\top \hat{\boldsymbol{\Sigma}} \hat{w}$ was defined in (7). The blind estimator outperforms PCA and the best parametric estimator (h_{t^*} with $t^* = 0.625$) in terms of mean and variance. We find a variance reduction of factors 6.22 and 1.68 relative to PCA and the best parametric estimator respectively. Conversely, the blind estimator exhibits a variance for the estimated volatility $\hat{\mathrm{V}}$ that is larger than both the PCA and h_{t^*}. This is an advantage as the higher number allows for a larger level of uncertainty to be taken into account in practice. Table 2 also reports statistics for the ratio of the true to the estimated volatility $\mathrm{V}/\hat{\mathrm{V}}$. This reports how much the forecast volatility deviates from the true volatility. The exact, blind and best parametric estimators show a desirably small level of deviation. On the other hand the volatility forecast produced by PCA is a factor larger than two away from the true volatility (c.f., (8) in Section 2.1).

5.4. Asymptotics of the optimization bias

We experimentally confirm the statements of Theorems 3.1 and 4.1 and Conjecture 1 by simulating models of increasing size, taking p as large as 8,000. For every p,[11] we generate a $\beta \in \mathbb{R}^p$ and draw $n = 50$ i.i.d. observations of the returns obeying (29) as described at the outset. The subscript p highlights the dependence on size.

We study the optimization bias $\mathscr{E}_p(h)$ for the PCA estimate and its corrected counterpart $\mathscr{E}_p(h_\tau)$ that is produced by the blind estimator (the exact estimator has no optimization bias). The error $\mathscr{E}_p$ was shown to be closely related to true volatility of the estimated minimum variance portfolio investigated in Section 5.3. Indeed, the error $\mathscr{E}_p$ is the sole component of the

[11] Following footnote 9, we generate a sequence $\eta_1, \eta_2, \ldots$ and take (increasing) subsets $\{\eta_i\}_{i=1}^p$ for each size p to produce the vector $\beta \in \mathbb{R}^p$ with $\mu_p(\beta) = 1.0$ and $\mathrm{d}_p(\beta) = 0.5$.

asymptotic description of the true volatility (see V_p in (5)) that may be manipulated in an estimation context. Moreover, by (8), the ratio of the true to the estimated variance $\mathrm{V}_p^2/\hat{\mathrm{V}}_p^2$ is proportional to $p\mathscr{E}_p^2$.

Figure 5 considers the moderate deviations (on the scale $\sqrt{p}$) of the corrected error $\mathscr{E}_p(h_\tau)$. It confirms that these deviations do not grow in p and further suggests a convergence (in law) of the rescaled variable $\sqrt{p}\mathscr{E}_p(h_\tau)$ to some nondegenerate limit of mean zero and finite variance. These results confirm Theorem 4.1 (not almost surely, but as convergence in law) that states that the corrected optimization bias $\mathscr{E}_p(h_\tau)$ vanishes as the size of the portfolio increases. It also indirectly supports Conjecture 1, which posits that $\sqrt{p}\mathscr{E}_p(h_\tau)$ has a bounded (in p) second moment. Tables 3 and 4 provide further support. Table 3 (the first two columns) illustrates that the mean and standard deviation of of $\mathscr{E}_p(h_\tau)$ both tend to zero as p grows. The rate at which the mean tends to zero appears to be linear while the standard deviation looks to converge at the rate $\sqrt{p}$ (c.f., Figure 5). Table 4 supplements these statistics with those for $p\mathscr{E}_p^2(h_\tau)$. The first column of Table 4 provides evidence for Conjecture 1 by showing that the mean of $p\mathscr{E}_p^2(h_\tau)$ does not grow in p. It further demonstrates (second column) that the standard deviation likely remains bounded as well (at least the Gaussian setting that we adopt).

Figure 6 draws a comparison between the optimization bias produced by the PCA and blind estimators. The distribution of the optimization bias $\mathscr{E}_p(h)$ scaled by $\sqrt{p}$ and the mean of that of the blind estimator are illustrated in Figure 6. As predicted, the mean and standard deviation of the distribution of $\sqrt{p}\mathscr{E}_p(h)$ both grow at rate $\sqrt{p}$, This growth results in a scale difference of tenfold as compared to Figure 5, which displays the same quantities for the blind estimator. Table 3 records (fifth column) the ratio of the means of $\mathscr{E}_p(h))$ and $\mathscr{E}_p(h_\tau)$ showing an improvement of a factor in the hundreds over PCA for large values of p. Columns three and four of Table 3 confirm the convergence of the optimization bias of PCA to a nondegenerate positive limit (c.f. Corollary 3.6). Table 4 further confirms the linear blow up of the mean and standard deviation of $p\mathscr{E}_p^2(h)$ as the portfolio size grows.

We point out that the theoretical results we confirm here in a setting of i.i.d. Gaussian observations hold much more broadly. Neither the i.i.d. not the Gaussian requirements are needed to obtain qualitatively similar results. We refer to reader to extensive simulations in Goldberg et al. (2019) that test our conclusions even for more complex models of security returns and under several market calibrations.

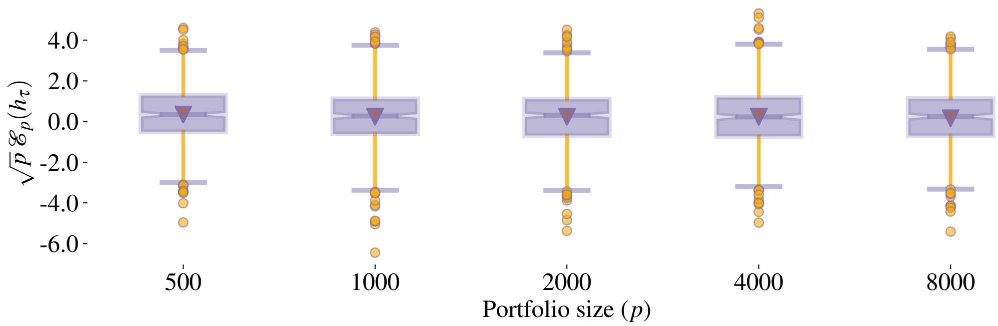

Figure 5. *Boxplots of* $\sqrt{p}\mathscr{E}_p(h_\tau)$ *vs growing portfolio size p. Each* blind *estimator corrected optimization bias* $\mathscr{E}_p(h_\tau)$ *is constructed from* (23) *by using* 50 *observations.* 1000 *simulations are used to construct each boxplot that displays the median, the interquartile range and outliers below* 1% *and above* 99% *are shown. Sample means for each value of p are marked with a (upside down) triangle.*

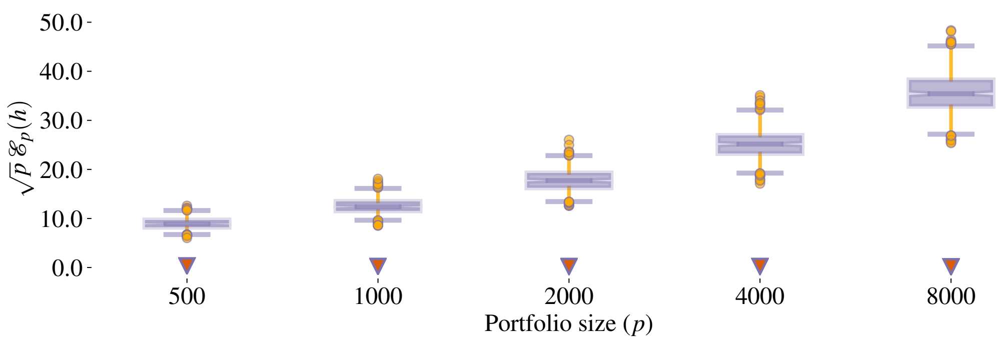

Figure 6. *Boxplots of* $\sqrt{p}\mathscr{E}_p(h)$ *vs growing portfolio size p. The optimization bias* $\mathscr{E}_p(h)$ *of* PCA *is constructed from* (17) *by using* 50 *observations. Sample means for each corrected bias* $\sqrt{p}\mathscr{E}_p(h_\tau)$ *is depicted with a (upside down) triangle marker.* 1000 *simulations are used to construct each boxplot that displays the median, the interquartile range and outliers below* 1% *and above* 99% *are shown.*

p	$\mathrm{E}(\mathscr{E}_p(h_\tau))$	$\mathrm{SD}(\mathscr{E}_p(h_\tau))$	$\mathrm{E}(\mathscr{E}_p(h))$	$\mathrm{SD}(\mathscr{E}_p(h))$	$\frac{\mathrm{E}(\mathscr{E}_p(h))}{\mathrm{E}(\mathscr{E}_p(h_\tau))}$
500	0.0185	0.0611	0.4004	0.0485	21.62
1000	0.0093	0.0436	0.3987	0.0464	42.77
2000	0.0047	0.0310	0.3980	0.0453	84.26
4000	0.0024	0.0220	0.3976	0.0448	166.4
8000	0.0012	0.0156	0.3973	0.0445	338.3

Table 3. *Sample statistics for the optimization bias $\mathscr{E}_p$ produced by the* PCA *estimator (h) and the* blind *estimator (h_τ) versus growing portfolio size p. Estimates of the ratio $\mathrm{E}(\mathscr{E}_p(h))/\mathrm{E}(\mathscr{E}_p(h_\tau))$ measure the improvement of the* blind *estimator relative to* PCA. *Each sample estimate for an expectation* (E) *and a standard deviation* (SD) *is computed using 10^6 simulations. Every estimate of $\mathrm{E}(\mathscr{E}_p(h_\tau))$ has the 99% confidence interval $\pm 2.58\,\mathrm{SD}(\mathscr{E}_p(h_\tau)) \times 10^{-3}$ and analogously for the* PCA *estimator. For each value of p we use $n = 50$ observations.*

A. Proofs

Recall the $p \times n$ data matrix $\mathbf{Y} = \mathbf{Y}_{p\times n}$ of n excess returns to p securities. According to (9), the jth observation (jth column of $\mathbf{Y}$) is $\mathrm{Y}_j = \beta \mathrm{X}_j + \mathrm{Z}_j$, and

$$\mathbf{Y} = \beta \mathrm{X}^\top + \mathbf{Z} \tag{31}$$

for $\beta \in \mathbb{R}^p$, a row vector $\mathrm{X}^\top = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$ of realized market return X and $\mathbf{Z} = \mathbf{Z}_{p\times n}$, the $p \times n$ matrix with jth column Z_j, the jth realized specific return.

Under Assumption 2.2 for all p sufficiently large, we have $\mu_p(\beta) > 0$ and

$$|\beta|^2 = p\mu_p^2(\beta)\,(1 + \mathrm{d}_p^2(\beta)) \tag{32}$$

as in Appendix C, for the mean $\mu(\beta) = \mu_p(\beta)$ and dispersion $\mathrm{d}(\beta) = \mathrm{d}_p(\beta)$ in (6).

A.1. Proof of Theorem 3.1

Recall the eigenvector h of the $(p \times p)$ sample covariance matrix $\mathbf{S} = \mathbf{Y}\mathbf{Y}^\top/n$ with the eigenvalue s_p^2 as in (11). We consider the the singular value decomposition of $\mathbf{Y}$ and the $\chi_p \in \mathbb{R}^n$ with $|\chi_p| = 1$ such that h and χ_p form the left and right singular vectors of $\mathbf{Y}/\sqrt{n}$ respectively with singular

p	$p\mathrm{E}(\mathscr{E}_p^2(h_\tau))$	$p\mathrm{SD}(\mathscr{E}_p^2(h_\tau))$	$p\mathrm{E}(\mathscr{E}_p^2(h))$	$p\mathrm{SD}(\mathscr{E}_p^2(h))$
500	2.036	3.013	81.32	19.74
1000	1.990	2.991	161.2	37.66
2000	1.964	2.971	321.0	73.42
4000	1.959	2.992	640.5	145.1
8000	1.953	2.992	1279.	287.9

Table 4. *Sample statistics for the scaled square of the optimization bias $p\mathscr{E}_p^2$ produced by the* PCA *estimator (h) and the* blind *estimator (h_τ) versus growing portfolio size p. Each sample estimate for an expectation* (E) *and a standard deviation* (SD) *are computed using 10^6 simulations. Every estimate of $p\mathrm{E}(\mathscr{E}_p^2(h_\tau))$ has the 99% confidence interval $\pm 2.58\, p\mathrm{SD}(\mathscr{E}_p^2(h_\tau)) \times 10^{-3}$ and analogously for the* PCA *estimator. For each value of p we use $n = 50$ observations.*

value $\mathrm{s}_p \geq 0$.[12] Then, by (31),

$$\text{(33)} \qquad h\mathrm{s}_p = \mathbf{Y}\chi_p/\sqrt{n} = \frac{\beta \mathrm{X}^\top \chi_p + \mathbf{Z}\chi_p}{\sqrt{n}}.$$

Taking a dot product of both sides with b and z yields the following identities.

$$\text{(34)} \qquad \langle h, b\rangle_p = h^\top b = \left(\frac{|\beta|\mathrm{X}^\top \chi_p}{\mathrm{s}_p\sqrt{n}}\right) + \left(\frac{\beta^\top \mathbf{Z}}{\sqrt{p}|\beta|}\right)\left(\frac{\chi_p\sqrt{p}}{\mathrm{s}_p\sqrt{n}}\right)$$

$$\text{(35)} \qquad \langle h, z\rangle_p = h^\top z = \langle b, z\rangle_p \left(\frac{|\beta|\mathrm{X}^\top \chi_p}{\mathrm{s}_p\sqrt{n}}\right) + \left(\frac{\mathrm{e}^\top \mathbf{Z}}{\sqrt{p}|\mathrm{e}|}\right)\left(\frac{\chi_p\sqrt{p}}{\mathrm{s}_p\sqrt{n}}\right)$$

Taking the dot product of both sides of (33) with $h\mathrm{s}_p$ and dividing by p yields

$$\text{(36)} \qquad \mathrm{s}_p^2/p = \frac{|\beta|^2(\mathrm{X}^\top\chi_p)^2}{np} + \frac{\chi_p^\top \mathbf{Z}^\top \mathbf{Z}\chi_p}{np} + 2(\mathrm{X}^\top\chi_p)\left(\frac{\beta^\top \mathbf{Z}}{\sqrt{p}|\beta|}\right)\left(\frac{\chi_p|\beta|}{n\sqrt{p}}\right).$$

The next result facilitates limit ($p \uparrow \infty$) computations in (34), (35) and (36).

Lemma A.1. *Let $\{\eta_i\}_{i\in\mathbb{N}} \subseteq \mathbb{R}$ be a sequence with $\mu_p(\eta) = \frac{1}{p}\sum_{i=1}^p \eta_i$ satisfying* $\liminf_{p\uparrow\infty} \mu_p(\eta) > 0$*. For $\{\mathrm{Z}^k\}_{k\in\mathbb{N}}$ a collection of mean-zero, pairwise indepenent and identically distributed (real) random variables with $\mathrm{Var}(\mathrm{Z}^1) < \infty$, writing $\eta = (\eta_1, \ldots, \eta_p)^\top$ and $\mathrm{Z} = (\mathrm{Z}^1, \ldots, \mathrm{Z}^p)^\top$, we have $\frac{\eta^\top \mathrm{Z}}{\sqrt{p}|\eta|} \to 0$ almost surely as $p \uparrow \infty$.*

[12] By convention, the singular values of a real matrix $\mathbf{A}$ are taken as the nonnegative square roots of the (nonnegative) eigenvalues of $\mathbf{A}^\top\mathbf{A}$. The largest such value a satisfies $a^2 = \sup_{|x|=1} \mathbf{A}^\top\mathbf{A}$.

PROOF.Since for all p large enough, $\mu_p(\eta) > 0$, the dispersion $\mathrm{d}_p(\eta)$ is well defined (see Appendix C) and then, for $W_k = \eta_k Z^k / \big(\mu_p(\eta)\sqrt{1+\mathrm{d}_p^2(\eta)}\big)$, we have

$$\frac{\eta^\top Z}{\sqrt{p}|\eta|} = \frac{1}{p}\sum_{k=1}^{p} W_k.$$

The result now follows by the SLLN of Chandra & Goswami (1992, Theorem 6) provided $\sup_p \frac{1}{p}\sum_{k=1}^{p} \mathrm{E}(W_k^2) < \infty$. As $\mathrm{E}(W_k^2) = \eta_k^2 \mathrm{VAR}(Z^1)/(\mu_p^2(\eta)(1+\mathrm{d}_p^2(\eta))$, we have $\frac{1}{p}\sum_{k=1}^{p} \mathrm{E}(W_k^2) = \mathrm{VAR}(Z^1)|\eta|^2/(\mu_p^2(\eta)(1+\mathrm{d}_p^2(\eta)) = \mathrm{VAR}(Z^1) < \infty$. ■

By applying Lemma A.1 to each $\big(\frac{\beta^\top \mathbf{Z}}{\sqrt{p}|\beta|}\big)_j$ and $\big(\frac{\mathrm{e}^\top \mathbf{Z}}{\sqrt{p}|\mathrm{e}|}\big)_j$ for $1 \le j \le n$ in (34) and (35) we have, under Assumptions 2.2 and 2.3 (only $\mu_\infty(\beta) > 0$ and the pairwise independence of the $\{Z^i\}_{i\in\mathbb{N}}$ is required here) that (34) and (35) reduce to

$$\langle h, b\rangle_\infty = \lim_{p\uparrow\infty}\left(\frac{|\beta|\mathrm{X}^\top \chi_p}{\mathrm{s}_p\sqrt{n}}\right) \tag{37}$$

$$\langle h, z\rangle_\infty = \langle b, z\rangle_\infty \lim_{p\uparrow\infty}\left(\frac{|\beta|\mathrm{X}^\top \chi_p}{\mathrm{s}_p\sqrt{n}}\right) \tag{38}$$

provided the limit on the rights side exist almost surely (for the existence of $\langle b, z\rangle_\infty$, see Remark 3.2) and that $\sup_p \mathrm{s}_p/\sqrt{p} < \infty$ almost surely (see Lemma A.2 below).

Define the (nondegenerate) random variable σ_{X}^2 (which is well-defined by Assumption 2.2 and is strictly positive almost surely by Assumption 2.1) as

$$\sigma_{\mathrm{X}}^2 = \frac{|\mathrm{X}|^2}{n}\mu_\infty^2(\beta)\,(1+\mathrm{d}_\infty^2(\beta)). \tag{39}$$

With $\mathrm{TR}(\mathbf{S})$ denoting the matrix trace[13] of $\mathbf{S}$, we have $\ell_p^2 = (\mathrm{TR}(\mathbf{S}) - \mathrm{s}_p^2)/(n-1)$ as in the definition of ψ_p in (12) after taking $p \ge n \ge 2$.

Lemma A.2. *Suppose that Assumptions 2.2 and 2.3 hold. Then, almost surely, we have* $\lim_{p\uparrow\infty} \mathrm{s}_p/\sqrt{p} = \sqrt{\sigma_{\mathrm{X}}^2 + \delta^2/n}$, $\lim_{p\uparrow\infty}\chi_p \to \mathrm{X}/|\mathrm{X}|$ *and* $\lim_{p\uparrow\infty} \ell_p^2/p \to \delta^2/n$.

PROOF.Let $\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$. By definition (see, footnote 12) we have that $\mathrm{s}_p^2/p = g_p(\chi_p) = \sup_{x\in\mathbb{S}^{n-1}} g_p(x)$ where $g_p(x) = (x^\top \mathbf{Y}^\top \mathbf{Y} x)/(np)$ so that

$$g_p(x) = \frac{|\beta|^2(\mathrm{X}^\top x)^2}{np} + \frac{x^\top \mathbf{Z}^\top \mathbf{Z} x}{np} + 2(\mathrm{X}^\top x)\left(\frac{\beta^\top \mathbf{Z}}{\sqrt{p}|\beta|}\right)\left(\frac{x|\beta|}{n\sqrt{p}}\right). \tag{40}$$

[13] The sum of the diagonal elements of $\mathbf{S}$, and equivalently the sum of its eigenvalues.

To conclude the first two claims (pertaining to $\lim_{p\uparrow\infty} \mathrm{s}_p/p$ and $\lim_{p\uparrow\infty} \chi_p$), it suffices to show that the functions g_p converge uniformly to g_∞ on $\mathbb{S}^{n-1}$ almost surely, where

$$g_\infty(x) = \sigma_{\mathrm{X}}^2 \left(\frac{\mathrm{X}^\top x}{|\mathrm{X}|} \right)^2 + \frac{\delta^2}{n}. \tag{41}$$

If so, then almost surely,

$$\lim_{p\uparrow\infty} \mathrm{s}_p^2/p = \lim_{p\uparrow\infty} \sup_{x\in\mathbb{S}^{n-1}} g_p(x) = \sup_{x\in\mathbb{S}^{n-1}} g_\infty(x) = \sigma_{\mathrm{X}}^2 + \delta^2/n,$$

since the convergence of g_p to g_∞ must be uniform. Taking (positive) square roots, we have $\lim_{p\uparrow\infty} \mathrm{s}_p^2/p = \sqrt{\sigma_{\mathrm{X}}^2 + \delta^2/n}$.

Almost surely, $\sigma_{\mathrm{X}}^2 + \delta^2/n = \sup_{x\in\mathbb{S}^{n-1}} g_\infty(x)$ is attained only at $\pm\mathrm{X}/|\mathrm{X}|$, so the almost sure convergence

$$\lim_{p\uparrow\infty} \chi_p = \frac{\mathrm{X}}{|\mathrm{X}|}$$

follows from the uniform convergence of the sequence g_p to the continuous limit g_∞ and the fact that left-hand side of (38) must be positive almost surely for p sufficiently large whether or not the limit on the right side of this equation exists.

We proceed to show the almost sure convergence of g_p to g_∞, which follows from the bound $|g_p(x) - g_\infty(x)| \leq |\gamma_1(x)| + |\gamma_2(x)| + |\gamma_3(x)|$ where

$$\begin{aligned}
|\gamma_1(x)| &= (\mathrm{X}^\top x)^2 \left| \frac{|\beta|^2}{np} - \frac{\sigma_{\mathrm{X}}^2}{|\mathrm{X}|^2} \right|, \\
|\gamma_2(x)| &= \left| \frac{x^\top \mathbf{Z}^\top \mathbf{Z} x}{np} - \frac{\delta^2}{n} \right|, \\
|\gamma_3(x)| &= 2(\mathrm{X}^\top x) \left| \left(\frac{\beta^\top \mathbf{Z}}{\sqrt{p}|\beta|} \right) \left(\frac{x|\beta|}{n\sqrt{p}} \right) \right|.
\end{aligned}$$

Using that $|\beta|^2 = p\mu_p^2(\beta)(1 + \mathrm{d}_p^2(\beta))$ by (32) for all p sufficiently large (under Assumption 2.2), we have $\sup_{x\in\mathbb{S}^{n-1}} \gamma_1(x) \to 0$ by the definition of σ_{X}^2 in (39). For the term $\gamma_2(x)$, observe that $x^\top \mathbf{Z}^\top \mathbf{Z} x = \sum_{j=1}^n \sum_{k=1}^n x_j \mathrm{Z}_j^\top \mathrm{Z}_k x_k$ so that

$$\begin{aligned}
|\gamma_2(x)| &\leq \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n |x_j x_k| |\mathrm{Z}_j^\top \mathrm{Z}_k/p - \delta^2 \mathbf{1}_{\{k=j\}}| \\
&\leq n \max_{1\leq j,k\leq n} |\mathrm{Z}_j^\top \mathrm{Z}_k/p - \delta^2 \mathbf{1}_{\{k=j\}}|
\end{aligned}$$

where $\mathbf{1}_{\mathscr{A}}$ denotes the indicator of $\mathscr{A}$. Per Assumption 2.1, $\mathrm{E}((\mathrm{Z}_j^i)^2) = \delta^2 < \infty$. Thus under Assumption 2.3, every $\mathrm{Z}_j^\top \mathrm{Z}_j/p = \frac{1}{p}\sum_{i=1}^p (\mathrm{Z}_j^i)^2 \to \delta^2$ by the SLLN (Etemadi 1981, Theorem 1). Similarly, $\mathrm{E}(\mathrm{Z}_j^i \mathrm{Z}_k^i) = 0$ for $j \neq k$ by Assumption 2.3, so $\mathrm{Z}_j^\top \mathrm{Z}_k / p = \frac{1}{p} \sum_{i=1}^p \mathrm{Z}_j^i \mathrm{Z}_k^i \to 0$ for every $j \neq k$. Thus, $\sup_{x \in \mathbb{S}^{n-1}} |\gamma_2(x)| \to 0$.

The required convergence of g_p follows by applying Cauchy-Schwartz to $|\gamma_3(x)|$,

$$|\gamma_3(x)| \leq 2|\mathrm{X}|^2 \left| \frac{\beta^\top \mathbf{Z}}{\sqrt{p}|\beta|} \right| \left(\frac{|\beta|}{n\sqrt{p}} \right), \tag{42}$$

so that $\sup_{x \in \mathbb{S}^{n-1}} |\gamma_3(x)| \to 0$ almost surely by Lemma A.1 and the fact that $\sup_p |\beta|^2/p = \sup_p \mu_p^2(\beta)(1 + \mathrm{d}_p^2(\beta)) < \infty$ under Assumptions 2.2 and 2.3.

Finally, observe that since $n\mathbf{S} = \mathbf{Y}\mathbf{Y}^\top = \mathrm{X}^\top \mathrm{X} \beta\beta^\top + \mathbf{Z}\mathbf{Z}^\top + \beta \mathrm{X}^\top \mathbf{Z}^\top + \mathbf{Z}\mathrm{X}\beta^\top$, the almost sure limit as $p \uparrow \infty$ of the trace of $\mathbf{S}$ is given by

$$\lim_{p \uparrow \infty} \frac{\mathrm{Tr}(\mathbf{S})}{p} = \lim_{p \uparrow \infty} \left(\frac{|\beta|^2 |\mathrm{X}|^2}{pn} + \frac{\mathrm{tr}(\mathbf{Z}^\top \mathbf{Z})}{pn} + 2\left(\frac{|\beta|}{\sqrt{pn}} \right) \frac{\beta^\top \mathbf{Z} \mathrm{X}}{|\beta|\sqrt{p}} \right) = \sigma_{\mathrm{X}}^2 + \delta^2 \tag{43}$$

where we applied (32) and (39) to the first term to obtain σ_{X}^2, an argument identical to that for $x^\top \mathbf{Z}^\top \mathbf{Z} x$ above to the second term to obtain δ^2, and Lemma A.1 to remove the third term. From (43) and the fact that $\mathrm{s}_p^2/p \to \sigma_{\mathrm{X}}^2 + \delta^2/n$ almost surely, we deduce that

$$\ell_p^2/p = (\mathrm{Tr}(\mathbf{S})/p - \mathrm{s}_p^2/p)/(n-1) \to \delta^2/n$$

almost surely. ■

We now complete the proof of the main result. Applying (32) and Lemma A.2 to the right sides of (37) and (38) we obtain that $\langle h, z \rangle_\infty = \langle b, z \rangle_\infty \langle h, b \rangle_\infty$ and

$$\langle h, b \rangle_\infty = \frac{\sigma_{\mathrm{X}}^2}{\sigma_{\mathrm{X}}^2 + \delta^2} \in (0, 1) \tag{44}$$

almost surely. Applying Lemma A.2 to $\psi_p^2 = (\mathrm{s}_p^2 - \ell_p^2)/\mathrm{s}_p^2$ in (12), we obtain that the limit ψ_∞ equals the right side of (44) almost surely. This concludes the proof.

B. Asymptotic estimates

Let $\hat{w} \in \mathbb{R}^p$ denote the solution to $\min_{\mathrm{e}^\top w = 1} w^\top \hat{\mathbf{\Sigma}} w$, the optimization in (1) with $\hat{\mathbf{\Sigma}}$ replacing $\mathbf{\Sigma} = \sigma^2 \beta\beta^\top + \delta^2 \mathbf{I}$ so that the (β, σ, δ) is replaced by $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$. With $b = \beta/|\beta|$ and $z = \mathrm{e}/\sqrt{p}$ as in (3), we have $\mathbf{\Sigma}$ in (2) taking the form

$$\mathbf{\Sigma} = p\sigma_b^2 \, bb^\top + \delta^2 \mathbf{I}$$

where by (32) we have $\sigma_b^2 = \sigma^2|\beta|^2/p = \sigma^2\mu^2(\beta)(1 + \mathrm{d}^2(\beta))$ (c.f., σ_X^2 of (39)). Define $\kappa = \hat{\delta}/(\hat{\sigma}\sqrt{p})$ and $h = \hat{\beta}/|\hat{\beta}|$, the estimate of b (not necessarily the sample eigenvector h that appears in (11)). Analogously to σ_b^2, we let $\sigma_h^2 = \hat{\sigma}^2|\hat{\beta}|/p$.

The solution $\hat{w}$ (which has a closed-form) may be written as

$$\hat{w} = \frac{1}{\sqrt{p}}\left(\frac{z\rho - h}{\rho - \langle h, z\rangle}\right) \quad \text{where} \quad \rho = \frac{1+\kappa^2}{\langle h, z\rangle}.$$

The true variance $\mathrm{V}^2 = \hat{w}^\top \boldsymbol{\Sigma} \hat{w} = p\sigma_b^2(b^\top \hat{w})^2 + \delta^2|\hat{w}|^2$ and the estimated variance $\hat{\mathrm{V}}^2 = \hat{w}^\top \hat{\boldsymbol{\Sigma}} \hat{w} = p\sigma_h^2(h^\top \hat{w})^2 + \hat{\delta}^2|\hat{w}|^2$ depend on the asymptotics of the following.

$$(p\sigma_b^2)(b^\top \hat{w})^2 = \sigma_b^2\left(\frac{(1+\kappa^2)\langle b, z\rangle - \langle h, b\rangle\langle h, z\rangle}{\kappa^2 + 1 - \langle h, z\rangle^2}\right)^2$$

$$(p\sigma_h^2)(h^\top \hat{w})^2 = \sigma_h^2\left(\frac{\kappa^2\langle h, z\rangle}{\kappa^2 + 1 - \langle h, z\rangle^2}\right)^2$$

$$|\hat{w}|^2 = \frac{1}{p}\left(\frac{1+\kappa^2}{\kappa^2 + 1 - \langle h, z\rangle^2} - \frac{\kappa^2\langle h, z\rangle^2}{(\kappa^2 + 1 - \langle h, z\rangle^2)^2}\right)$$

From the expressions given above, the asymptotics of V^2 and $\hat{\mathrm{V}}^2$ are immediate. We add a subscript p below to highlight the dependence on p.

Assumption B.1. *The estimates $\hat{\sigma}^2$ and $\hat{\delta}^2$ are bounded away from zero and $\langle h, z\rangle_p$ is nonnegative and bounded away from one in $p \in \mathbb{N}$.*

Note that these assumptions are satisfied for the PCA estimates we analyze (see Theorem 3.1). Let $o_p \asymp 1/p$, recalling (per Section (2.1)) this notation means there are fixed constants $c, C \in \mathbb{R}$ such that $c/p \le o_p \le C/p$ for all p sufficiently large. Under Assumption B.1, for some constant $K > 0$ and $\mathscr{E}_p$ in (4) we have

$$\mathrm{V}_p^2 = \sigma_b^2\mathscr{E}_p^2(h) + o_p$$
$$\hat{\mathrm{V}}_p^2 = \hat{\delta}^2|\hat{w}|^2 + K\kappa_p^4$$

confirming the asymptotics stated in (5) and (7).

C. Auxiliary calculations

Consider any $\eta \in \mathbb{R}^p$ with (Euclidean) length $|\eta|$, mean $\mu(\eta) = \frac{1}{p}\sum_{i=1}^p \eta_i$ and dispersion $\mathrm{d}^2(\eta) = \frac{1}{p}\sum_{i=1}^p(\eta_i/\mu_p(\eta) - 1)^2$ defined for $\mu(\eta) > 0$, we have

$$|\eta|^2 = \sum_{i=1}^p \eta_i^2 = p\mu_p^2(\eta) + \sum_{i=1}^p(\eta_i - \mu(\eta))^2$$

$$= p\mu_p^2(\eta)(1 + \mathrm{d}_p^2(\eta)) \qquad \mu(\eta) > 0. \tag{45}$$

Next, recall the vectors $\mathrm{e} = (1, \dots, 1)^\top \in \mathbb{R}^p$ and $z = \mathrm{e}/\sqrt{p}$ and for η above with $|\eta| > 0$, define $h = \eta/|\eta|$. We have $\langle h, z\rangle = h^\top z = \frac{p\mu_p(\eta)}{\sqrt{p}|\eta|}$ and by (32),

$$\langle h, z\rangle^2 = \frac{1}{1 + \mathrm{d}^2(\eta)} \quad \text{and} \quad \mathrm{d}^2(\eta) = \frac{1 - \langle h, z\rangle^2}{\langle h, z\rangle^2}.$$

This calculation justifies equations (14) and (15).

We proceed to justify (20). For any $q \in \mathbb{R}^p$ with $|q| = 1$ let $\mathscr{E}(q) = \frac{\langle b,z\rangle - \langle q,z\rangle\langle q,b\rangle}{1 - \langle q,z\rangle^2}$. Then for any $h \in \mathbb{R}^p$ with $|h| = 1$ and $t \in \mathbb{R}$, defining $h_t = \frac{h+tz}{|h+tz|}$, we have

$$\begin{aligned}
\mathscr{E}_{h_t} &= \frac{\langle b, z\rangle |h + tz|^2 - \langle h + tz, b\rangle \langle h + tz, z\rangle}{|h + tz|^2 - \langle h + tz, z\rangle^2} \\
&= \frac{\langle b, z\rangle (1 + 2t\langle h, z\rangle + t^2) - (\langle h, b\rangle + t\langle z, b\rangle)(\langle h, z\rangle + t)}{1 + 2t\langle h, z\rangle + t^2 - (\langle h, z\rangle + t)^2} \\
&= \frac{\langle b, z\rangle + t\langle b, z\rangle\langle h, z\rangle - \langle h, b\rangle\langle h, z\rangle - t\langle h, b\rangle}{1 - \langle h, z\rangle^2} \\
&= \mathscr{E}_h - t\,\frac{\langle h, b\rangle - \langle b, z\rangle\langle h, z\rangle}{1 - \langle h, z\rangle^2}.
\end{aligned}$$

To justify (26) we check that,

$$\begin{aligned}
\mathrm{d}_p^2(h_t) = \frac{1 - \langle h_t, z\rangle^2}{\langle h_t, z\rangle^2} &= \frac{|h + tz|^2 - (\langle h, z\rangle + t)^2}{(\langle h, z\rangle + t)^2} \\
&= \frac{1 - \langle h, z\rangle^2}{(\langle h, z\rangle + t)^2}.
\end{aligned}$$

References

Anderson, R. M., Bianchi, S. & Goldberg, L. R. (2012), 'Will my risk parity portfolio outperform', *Financial Analysts Journal* **68**(6), 75–93.

Bai, J. & Ng, S. (2008), 'Large dimensional factor analysis', *Foundations and Trends in Econometrics* **3**(2), 89–163.

Bai, Z. & Silverstein, J. W. (2010), *Spectral analysis of large random matrices*, Springer Series in Statistics, second edn, Springer.

Banerjee, S. (2004), 'Revisiting spherical trigonometry with orthogonal projectors', *The College Mathematics Journal* **35**(5), 375–381.

Bender, J., Lee, J.-H., Stefek, D. & Yao, J. (2009), 'Forecast risk bias in optimized portfolios'.

Bianchi, S. W., Goldberg, L. R. & Rosenberg, A. (2017), 'The impact of estimation error on latent factor models of portfolio risk', *The Journal of Portfolio Management* **43**(5), 145–156.

Bickel, P. J. & Levina, E. (2008), 'Covariance regularization by thresholding', *The Annals of Statistics* pp. 2577–2604.

Blume, M. E. (1975), 'Betas and their regression tendencies', *The Journal of Finance* **30**(3), 785–795.

Britten-Jones, M. (1999), 'The sampling error in estimates of mean-variance efficient portfolio weights', *The Journal of Finance* **54**(2), 655–671.

Bun, J., Bouchaud, J. & Potters, M. (2016), 'Cleaning correlation matrices', *Risk Magazine* .

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica* **51**(5), 1281–1304.

Chandra, T. K. & Goswami, A. (1992), 'Cesaro uniform integrability and the strong law of large numbers'.

Clarke, R., De Silva, H. & Thorley, S. (2006), 'Minimum-variance portfolios in the us equity market', *The Journal of Portfolio Management* **33**(1), 10–24.

Clarke, R., De Silva, H. & Thorley, S. (2011), 'Minimum-variance portfolio composition', *Journal of Portfolio Management* **2**(37), 31–45.

Clarke, R., De Silva, H. & Thorley, S. (2013), 'Risk parity, maximum diversification, and minimum variance: An analytic perspective', *Journal of Portfolio Management* **39**(3), 39–53.

Cohn, D. L. (2013), *Measure theory*, Springer.

Connor, G. (1995), 'The three types of factor models: A comparison of their explanatory power', *Financial Analysts Journal* **51**(3), 42–46.

Connor, G. & Korajczyk, R. A. (1986), 'Performance measurement with the arbitrage pricing theory: A new framework for analysis', *Journal of financial economics* **15**, 373–394.

Connor, G. & Korajczyk, R. A. (1988), 'Risk and return in equilibrium apt: Application of a new test methodology', *Journal of financial economics* **21**, 255–289.

Cont, R. (2001), 'Empirical properties of asset returns: stylized facts and statistical issues', *Quantitative Finance* **1**, 223–236.

DeMiguel, V., Garlappi, L., Nogales, F. J. & Uppal, R. (2009), 'A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms', *Management Science* **55**(5), 798–812.

DeMiguel, V., Garlappi, L. & Uppal, R. (2007), 'Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?', *The review of Financial studies* **22**(5), 1915–1953.

Donoho, D., Gavish & Johnstone, I. (2018), 'Optimal shrinkage of eigenvalues in the spiked covariance model', *Annals of Statistics* **46**(4), 1742–1778.

El Karoui, N. (2013), 'On the realized risk of high-dimensional markowitz portfolios', *SIAM Journal on Financial Mathematics* **4**(1), 737–783.

El Karoui, N. et al. (2010), 'High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: risk underestimation', *The Annals of Statistics* **38**(6), 3487–3566.

Etemadi, N. (1981), 'An elementary proof of the strong law of large numbers', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **55**(1), 119–122.

Fama, E. F. & French, K. R. (1992), 'The cross section of expected stock returns', *The Journal of Finance* **47**(2), 427–465.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(4), 603–680.

Frost, P. A. & Savarino, J. E. (1986), 'An empirical bayes approach to efficient portfolio selection', *Journal of Financial and Quantitative Analysis* **21**(3), 293–305.

Gillen, B. J. (2014), 'An empirical bayesian approach to stein-optimal covariance matrix estimation', *Journal of Empirical Finance* **29**, 402–420.

Goldberg, L. R., Papanicolaou, A., Shkolnik, A. & Ulucam, S. (2019), Better betas. CDAR working paper.

Goldfarb, D. & Iyengar, G. (2003), 'Robust portfolio selection problems', *Mathematics of operations research* **28**(1), 1–38.

Green, R. C. & Hollifield, B. (1992), 'When will mean-variance efficient portfolios be well diversified?', *The Journal of Finance* **47**(5), 1785–1809.

Hall, P., Marron, J. S. & Neeman, A. (2005), 'Geometric representation of high dimension, low sample size data', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(3), 427–444.

Jagannathan, R. & Ma, T. (2003), 'Risk reduction in large portfolios: Why imposing the wrong constraints helps', *The Journal of Finance* **58**(4), 1651–1683.

Jobson, J. D. & Korkie, B. (1980), 'Estimation for markowitz efficient portfolios', *Journal of the American Statistical Association* **75**(371), 544–554.

Jolliffe, I. T., Trendafilov, N. T. & Uddin, M. (2003), 'A modified principal component technique based on the lasso', *Journal of computational and Graphical Statistics* **12**(3), 531–547.

Lai, T. L. & Xing, H. (2008), *Statistical models and methods for financial markets*, Springer.

Lai, T. L., Xing, H. & Chen, Z. (2011), 'Mean-variance portfolio optimization when means and covariances are unknown', *The Annals of Applied Statistics* pp. 798–823.

Ledoit, O. & Péché, S. (2011), 'Eigenvectors of some large sample covariance matix ensembles', *Probability Theory and Related Fields* **151**(1–2), 233–264.

Ledoit, O. & Wolf, M. (2003), 'Improved estimation of the covariance matrix of stock returns with an application to portfolio selection', *Journal of empirical finance* **10**(5), 603–621.

Ledoit, O. & Wolf, M. (2004), 'Honey, I shrunk the sample covariance matrix', *The Journal of Portfolio Management* **30**, 110–119.

Ledoit, O. & Wolf, M. (2017), 'Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks', *The Review of Financial Studies* **30**(12), 4349–4388.

Lintner, J. (1965*a*), 'Security prices, risk and maximal gains from diversification', *Journal of Finance* **4**(587–615).

Lintner, J. (1965*b*), 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets', *Review of Economics and Statistics* **73**, 13–37.

Marchenko, V. A. & Pastur, L. A. (1967), 'Distribution of eigenvalues for some sets of random matrices', *Mathematics of the USSR-Sbornik* **1**(4), 457.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77–91.

Markowitz, H. (1956), 'The optimization of a quadratic function subject to linear constraints', *Naval Research Logistics Quarterly* **3**, 111–133.

Michaud, R. O. & Michaud, R. O. (2008), *Efficient asset management: a practical guide to stock portfolio optimization and asset allocation*, Oxford University Press.

Mossin, J. (1966), 'Equilibrium in a capital asset market', *Econometrica* **34**(4), 768–783.

Onatski, A. (2012), 'Asymptotics of the principal components estimator of large factor models with weakly influential factors', *Journal of Econometrics* **168**(2), 244–258.

Pástor, L. (2000), 'Portfolio selection and asset pricing models', *The Journal of Finance* **55**(1), 179–223.

Paul, D. (2007), 'Asymptotics of sample eigenstructure for a large dimensional spiked covariance model', *Statistica Sinica* pp. 1617–1642.

Rosenberg, B. (1974), 'Extra-market components of covariance in security returns', *Journal of Financial and Quantitative Analysis* **9**(2), 263–274.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of economic theory* **13**(3), 341–360.

Sharpe, W. (1963), 'A simplified model for portfolio analysis', *Management Science* **9**(2), 277–293.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance* **19**(3), 425–442.

Shen, D., Shen, H., Zhu, H. & Marron, S. (2016), 'The statistics and mathematics of high dimensional low sample size asympotics', *Statistica Sinica* **26**(4), 1747–1770.

Treynor, J. L. (1962), Toward a theory of market value of risky assets. Presented to the MIT Finance Faculty Seminar.

Vasicek, O. A. (1973), 'A note on using cross-sectional information in bayesian estimation of security betas', *The Journal of Finance* **28**(5), 1233–1239.

Wang, W. & Fan, J. (2017), 'Asymptotics of empirical eigenstructure for high dimensional spiked covariance', *The Annals of Statistics* **45**(3), 1342–1374.