

THE DISPERSION BIAS

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

November 4, 2017

This Version: January 2, 2022[§]

SIAM Journal on Financial Mathematics, forthcoming

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. Our analysis sheds light on previously unknown aspects of how estimation error corrupts an estimated covariance matrix and is transmitted to portfolios via quadratic optimization.

*Department of Economics and Consortium for Data Analytics in Risk, University of California, Berkeley, CA 94720 and Aperio Group at BlackRock, lrg@berkeley.edu.

[†]Intelligent Financial Machines, Palo Alto, CA and Consortium for Data Analytics in Risk, University of California, Berkeley, CA alex.papanic@gmail.com.

[‡]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 Bob Anderson, Marco Avellaneda, Gérard Ben Arous, Kay Giesecke, Nick Gunther, Guy Miller, George Papanicolaou, Peter Sarnak, Yu-Ting Tai, participants at the the 3rd Annual CDAR Symposium in Berkeley, participants at the Swissquote Conference 2017 on FinTech, participants at the fall 2017 UC Santa Barbara Seminar in Statistics and Applied Probability and participants in the Special Session on Aspects of High Dimensional Covariance Estimation for Equity Risk at the 2019 SIAM Conference on Financial Mathematics 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, Alec Kercheval and Simge Uluçam led to numerous insights that improved this article, which benefitted from a careful review by two anonymous referees.

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. Here, dispersion refers to the standard deviation of the entries of the eigenvector divided by their mean. Dispersion is also known as coefficient of variation. To develop an understanding of why there might be excess dispersion in an estimated eigenvector, consider a market where correlations are driven by a single factor, and suppose all security exposures to that factor are identical. With high probability, 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 as introduced in Section 2. 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. Section 2 also introduces what we call the “optimiza-

tion bias”, the quantifiable error that is key to understanding the interplay between the model estimation error and the optimizer that computes a minimum variance portfolio. Section 3 demonstrates that it is the excess dispersion of the leading eigenvector of a sample covariance matrix that must be removed to address the impact of the optimization bias. In Section 4, we show that shrinking that excess dispersion (i.e., the dispersion bias) 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. We illustrate our results numerically in Section 5 with simulation experiments that corroborate our theoretical findings and study the behavior of the estimators we propose.

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 \mathcal{E} , which is an important driver of 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 as inner products of the true and estimated eigenvector with the unique, positive, dispersionless vector z on the sphere. The first surprise is that \mathcal{E} allows us to correct substantial errors in minimum variance portfolio weights and risk forecasts while having no direct dependence on estimated eigenvalues. 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 , $\mathcal{E}_p(h)$ 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 $\mathcal{E}_p(b)$ 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 $\mathcal{E}_p(h_{\tau^*}) = 0$. 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 $\mathcal{E}_p(h_{\tau})$ 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. 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 tradeoff 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 our 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 ad-

mal choice for an investor who believes the expected returns of all securities are equal. 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 (2020) 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. (2020), 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. (2020) 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.

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. (2020). Generalizations of the eigenvector shrinkage method developed in this article to include multiple anchor points and order information are in Gurdogan & Kercheval (2021). Shkolnik (2021) demonstrates parallels between our eigenvector shrinkage method and James-Stein shrinkage for a collection of averages toward their grand mean. That article also provides elegant alternative proofs of some of the results presented here.

2. PROBLEM FORMULATION

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

$$(1) \quad \begin{aligned} \min_{w \in \mathbb{R}^p} \quad & w^\top \Sigma w \\ & e^\top w = 1 \end{aligned}$$

The solution minimizes the variance of the portfolio return over all fully invested portfolios. In practice, the matrix Σ must be estimated from security returns data, and

there exists 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 trade-offs inherent in correcting estimation error in a returns 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 situations. One example concerns the estimation of equity 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

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

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 $\beta = (\beta_1, \dots, \beta_p)$ 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

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

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

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

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

$$(5) \quad \mathcal{V}^2 = \hat{w}^\top \Sigma \hat{w} \asymp \mathcal{E}^2(h)$$

where $h = \hat{\beta}/|\hat{\beta}|$ is the normalized estimate $\hat{\beta}$. Here, the (Vinogradov) asymptotic notation $a_p \asymp f(p)$ refers to the existence of two constants $c, C > 0$ such that for all p sufficiently large, $c|f(p)| \leq |a_p| \leq C|f(p)|$. Sharper asymptotics that

reveal the leading constant in (5) depends only on the true parameters σ and β may be found in Appendix B. It is remarkable that the dependence of \mathcal{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 \mathcal{V}^2 is h , the estimate of the normalized betas b defined in (3). This dependence occurs through $\mathcal{E}(h)$ and we further note that $\mathcal{E}(b) = 0$.

Note, \mathcal{V}^2 is the expected out-of-sample variance of the estimated, minimum variance portfolio \hat{w} . That is, for a return vector Y , independent of the data generating process that yielded \hat{w} , the out-of-sample portfolio return is $Y^\top \hat{w}$, and consequently, $\mathcal{V}^2 = \text{Var}(Y^\top \hat{w})$. Since \hat{w} minimizes the in-sample variance (with respect to $\hat{\Sigma}$), it is instructive to compare \mathcal{V}^2 in (5) to this estimated variance $\hat{\mathcal{V}}^2$. The latter (under the mild assumptions on $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$ stated Appendix B) obeys the large p asymptotics,

$$(6) \quad \hat{\mathcal{V}}^2 = \hat{w}^\top \hat{\Sigma} \hat{w} \asymp 1/p.$$

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

The asymptotic estimates supplied by (5) and (6) provide a first indication of how the optimization bias $\mathcal{E}(h)$ is related to the investment process. In particular, we observe that the ratio of the true variance to the estimated variance satisfies

$$(7) \quad \mathcal{V}^2 / \hat{\mathcal{V}}^2 \asymp 1 + p \mathcal{E}^2(h)$$

which explodes for large p unless $\mathcal{E}^2(h)$ vanishes. In finite sample however, regardless of the estimation procedure, we expect $\mathcal{E}^2(h)$ 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. More refined asymptotics of the variance forecast ratio in (7) are supplied in Appendix B. Additional portfolio metrics such as tracking error, which may be used to measure the impact of sampling error on portfolio weights and is asymptotically driven by $\mathcal{E}^2(h)$, are also adversely affected.¹

It may seem entirely impossible to remedy the dilemma posed by (7) since in our finite sample regime, we cannot expect h to be a consistent estimator (i.e., h does not tend to b for which $\mathcal{E}(b) = 0$). Yet, this is precisely what we accomplish.

2.2. Model and assumptions. For random $X \in \mathbb{R}$ and $Z \in \mathbb{R}^p$, we consider a linear model for the excess return to p securities of the form

$$(8) \quad Y = \beta X + Z,$$

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

To accommodate a forthcoming asymptotic analysis we consider the infinite sequences $\{\beta_i\}_{i \in \mathbb{N}}$ and $\{Z^i\}_{i \in \mathbb{N}}$ and write $\beta = (\beta_1, \dots, \beta_p)^\top$ and $Z = (Z^1, \dots, Z^p)^\top$ for the vectors in (8) (with the 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, with respect to a probability measure P .

¹See Goldberg et al. (2020) for details.

Assumption 2.1. For constants $\delta, \sigma > 0$, $\text{Var}(X) = \sigma^2$ and every $\text{Var}(Z^i) = \delta^2$. Furthermore, $X \neq 0$ almost surely and every $\text{Cov}(X, Z^i) = \text{E}(X) = \text{E}(Z^i) = 0$.

The generating process based on (8), 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. (2020)). Model (8) 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) framework, which is increasingly relevant for 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 $d_p(\beta) = d(\beta)$ denote the mean and dispersion of the vector $\beta \in \mathbb{R}^p$ with the subscript p denoting the dependence on the dimension.

$$(9) \quad \mu(\beta) = \frac{1}{p} \sum_{i=1}^p \beta_i \quad \text{and} \quad d^2(\beta) = \frac{1}{p} \sum_{i=1}^p \left(\frac{\beta_i}{\mu(\beta)} - 1 \right)^2$$

Assumption 2.2. The sequence $\{\beta_i\}_{i \in \mathbb{N}}$ is such that the $\{\mu_p(\beta)\}_{p \in \mathbb{N}}$ and $\{d_p(\beta)\}_{p \in \mathbb{N}}$ converge to the limits $\mu_\infty(\beta) \in (0, \infty)$ and $d_\infty(\beta) \in (0, \infty)$ 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. For conclusions that address only the efficiency of a given estimator of β , it suffices to work over subsequences that achieve the best or worst case asymptotics of the optimization bias (i.e., subsequences attaining $\liminf_{p \uparrow \infty} \mathcal{E}_p$ and $\limsup_{p \uparrow \infty} \mathcal{E}_p$). In this manner, the regularity conditions may be removed entirely (see Remark 4.2 for further detail). 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 in (8) nor to the covariance matrix Σ in (2).

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

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

²The 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 $\text{Var}(Y) = \Sigma$ is of the form in (2).

³The 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 accomodate 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, however, 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 observed returns.

3. DISPERSION BIAS

Let Y denote the $p \times n$ data matrix of realized security returns, i.e., the matrix whose j th column is $Y_j \in \mathbb{R}^p$. We denote by \mathfrak{s}_p^2 the largest eigenvalue of

$$(10) \quad S = YY^\top/n,$$

the sample covariance matrix of the returns. Since b in (3) is the eigenvector of Σ 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 .

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

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 \mathfrak{s}_p^2 and the average of the remaining nonzero eigenvalues of S , denoted by ℓ_p^2 , plays an important role in our analysis. Define,

$$(12) \quad \psi_p = \sqrt{\frac{\mathfrak{s}_p^2 - \ell_p^2}{\mathfrak{s}_p^2}}.$$

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

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

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,*

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

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*

$$\langle b, z \rangle_p^2 = \frac{1}{1 + d_p^2(\beta)} \rightarrow \frac{1}{1 + d_\infty^2(\beta)} \in (0, 1) \quad \text{as } p \uparrow \infty.$$

since $d_p(\beta)$ is assumed to converge to $d_\infty(\beta) \in (0, 1)$ as part of Assumption 2.2. This confirms the relation in (13) is not trivial (i.e., $\langle h, z \rangle_p$ and $\langle b, z \rangle_p$ both converge to zero). The relevant part of Assumption 2.2 is $\mu_\infty(\beta) \neq 0$ which prevents the trivial case.

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 must be characterized with respect to some known vector. In our case, it is z , the vector in the expression for the optimization bias $\mathcal{E}(h)$ in (4) and related to the constraint in (1).*

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 $d_p(h)$ of h has

$$(14) \quad d_p^2(h) = \frac{1}{p} \sum_{i=1}^p \left(\frac{h_i}{\mu_p(h)} - 1 \right)^2 = \frac{1 - \langle h, z \rangle_p^2}{\langle h, z \rangle_p^2}$$

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,

⁶We 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 $P(\mathcal{E}_p) > 1 - \epsilon$ for all $p \geq p_\epsilon$. The stronger statement we make is facilitated by Egoroff's theorem (Cohn 2013, Proposition 1.3.4).

the $d_p(h)$ typically exceeds the dispersion $d_p(b) = d_p(\beta) > 0$ of b , since

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

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,*

$$(16) \quad d_\infty^2(h) = d_\infty^2(b) + \frac{1 - \psi_\infty^2}{\langle h, z \rangle_\infty^2}$$

almost surely where $d_\infty^2(h) = \lim_{p \uparrow \infty} d_p^2(h)$ and $d_\infty(\beta) = d_\infty(b)$.

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 $\mathcal{E}(h)$ in (4),

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

(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 (7) that the ratio of the true to the estimated minimum variance satisfies $\mathcal{V}^2/\hat{\mathcal{V}}^2 \asymp 1 + p\mathcal{E}_p^2(h)$.

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

$$(18) \quad \mathcal{E}_\infty(h) = \frac{1 - \psi_\infty^2}{d_\infty^2(h) \langle h, z \rangle_\infty \psi_\infty}$$

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

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

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 $\theta_{\langle x, y \rangle} = \theta_{\langle x, y \rangle_p}$ so that $\langle x, y \rangle = \cos \theta_{\langle x, y \rangle}$. 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 $\theta_{\langle h, b \rangle}$ (c.f., Remark 3.3). The right panel of Figure 1 illustrates the portion of the bias of h that is identifiable relative to

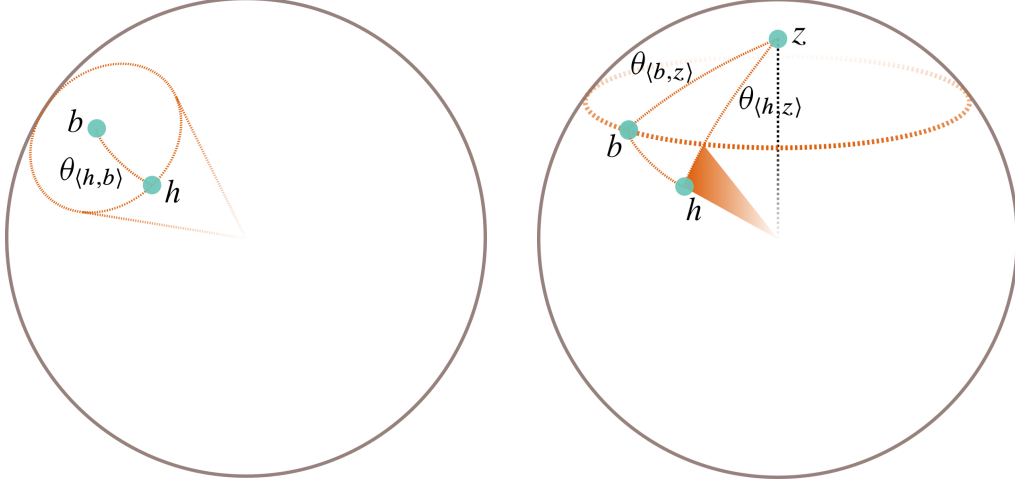


Figure 1. An illustration of the sample and population eigenvectors h and b respectively. The angle $\theta_{(h,b)}$ between h and b is also the length of the arc on the sphere of unit radius between h and b . The left panel shows that the identification of the bias in h is not possible without a reference 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 $\theta_{(h,z)}$ exceeds $\theta_{(b,z)}$ and is due to the excess dispersion in h .

the vector z . In particular, $\theta_{(h,z)_p} > \theta_{(b,z)_p}$ 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 now also points to a potential correction which is the problem explored in Section 4.

There is by now an extensive literature on the bias in sample eigenvalues and the corresponding shrinkage estimators that correct for these errors (e.g., see Donoho et al. (2018)). In our setting, for a result in this direction, see Lemma A.2 of Appendix A. This lemma suggests eigenvalue bias corrections to the estimates of the volatility parameters (σ, δ) in the population covariance $\Sigma = \sigma^2 \beta \beta^\top + \delta^2 I$ in (2). Due to the fact that minimum variance portfolios are driven exclusively by the sample eigenvector bias (see Section 2.1), these estimates are not essential in our context. However, we wish to emphasize the remarkable fact that the bias in the sample eigenvectors may be used to correct the bias in the sample eigenvalues. To this end, see Table 1 which relies on Theorem 3.1 as well as several of the calculations found in Appendix C.

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.

$$(19) \quad h_t = \frac{h + tz}{|h + tz|}, \quad t \in \mathbb{R}.$$

	finite p	limit p
$\hat{\sigma}^2$	$s_p^2 \mu_p^2(h)$	$\mu_\infty^2(\beta) X ^2/n$
$\hat{\delta}^2$	$n \ell_p^2/p$	δ^2

Table 1. The market and specific return variance estimators $\hat{\sigma}^2$ and $\hat{\delta}^2$. Here, $X = (X_1, \dots, X_n)^\top$ is the vector of n realizations of the market return X in (8). Note, $|X|^2/n$ is an unbiased estimator of $\sigma^2 = \text{Var}(X)$ and w.l.o.g. $\mu_\infty(\beta) = 1$ as the σ and β are unidentifiable (i.e. $\beta X = (c\beta)(X/c)$ for any $c \neq 0$). The estimates $\hat{\sigma}^2$ and $\hat{\delta}^2$ are compatible with the normalization $\mu(\hat{\beta}) = 1$ for the estimate $\hat{\beta}$ of β .

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

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

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

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

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,

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

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

Remark 4.2. As conclusions concerning the ultimate performance of the PCA and corrected estimators h and h_τ would involve only best and worst case analyses, one could aim to prove only that $B = \liminf_{p \uparrow \infty} \mathcal{E}(h) > 0$ and $W = \limsup_{p \uparrow \infty} \mathcal{E}(h_\tau) = 0$. This leads to further relaxation of our assumptions. For example, Assumption 2.2 may be significantly relaxed by considering a subsequence that achieves either B or W and then a further subsequence along which both $\mu_p(\beta)$ and $d_p(\beta)$ do converge. This relaxes Assumption 2.2 to one requiring only the boundedness of $\mu_p(\beta)$ and $d_p(\beta)$.

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

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

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 S .⁷ The first part of

⁷The vector h is an eigenvector of S and ψ_p is a function of the eigenvalues of S .

the result is remarkable in that even for only two observations ($n = 2$) of the return vector 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 the estimated portoflio \hat{w} is $\mathcal{V}_p^2 = \hat{w}^\top \Sigma \hat{w}$ with (5) refined as

$$(24) \quad \mathcal{V}_p^2 = \sigma^2 \mu_p^2(\beta) (1 + d_p^2(\beta)) \mathcal{E}_p^2(h) + o_p$$

for $o_p \downarrow 0$ (see Appendix B). From Corollary 3.6, we have that $\mathcal{E}_\infty^2(h) > 0$ almost surely for the plain PCA estimate h , and under our hypotheses, \mathcal{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 the other hand, replacing h with the corrected estimator h_τ ensures the expected out-of-sample variance vanishes.

We see that the variance forecast ratio $\mathcal{V}^2/\hat{\mathcal{V}}^2$ in (7), of the true to the estimated portfolio variance, diverges for the PCA estimator. Since $\mathcal{V}^2/\hat{\mathcal{V}}^2 \asymp 1 + p \mathcal{E}^2(h_\tau)$ supplies the asymptotics of the corrected variance forecast ratio, it is the moderate deviations scale that becomes relevant. To this end, we study the behavior of $\sqrt{p} \mathcal{E}(h_\tau)$ numerically in Section 5. Here, we propose the following bound without proof.

Conjecture. Fix $n \geq 2$ and suppose Assumptions 2.1, 2.2 and 2.3 hold. In addition, suppose that $\sup_i E((Z^i)^8) < \infty$ and that $E(|X|^q) < \infty$ for $q = \pm 4$ where $X = (X_1, \dots, X_n)^\top$ is the vector of n realizations of X , with X and Z in (8). Then,

$$(25) \quad \sup_p E(p \mathcal{E}_p^2(h_\tau)) < \infty.$$

Remark 4.3. In contrast to (25), we must have $\limsup_{p \uparrow \infty} p \mathcal{E}_p^2(h_\tau) = \infty$ almost surely. This stems from a variant of the law of iterated logarithms, the fact that the scaled sum $W_p = \frac{Z^1 + \dots + Z^p}{\sqrt{p}}$ for i.i.d. random variables $\{Z^i\}_{i \in \mathbb{N}}$ has $\limsup_{p \uparrow \infty} W_p = \infty$ 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 out in expectation.

Remark 4.4. In view of Corollary 3.6, $\sup_p E(c_p \mathcal{E}_p^2(h)) = \infty$ for any $c_p \rightarrow \infty$.

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 excess dispersion since $d_p^2(h) > d_p^2(b)$ w.h.p. per (15), i.e., the dispersion of h is larger than that of b , the population eigenvector. But,

$$(26) \quad d_p^2(h_t) = \frac{1 - \langle h_t, z \rangle_p^2}{\langle h_t, z \rangle_p^2} = \frac{1 - \langle h, z \rangle_p^2}{(\langle h, z \rangle_p + t)^2} < d_p^2(h) \quad t > 0$$

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 parameter τ_p and its effect on the optimization bias \mathcal{E}_p , we make the following two observations.

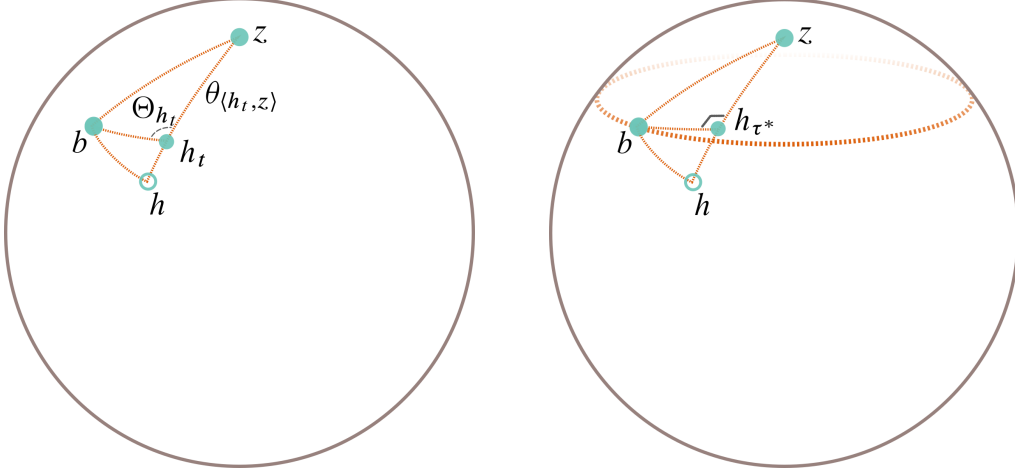


Figure 2. An illustration of the estimator h_t for $t > 0$ (left panel), which shrinks the dispersion $d(h_t)$ of h_t relative to that of h . Equivalently, we shrink the angle $\theta_{\langle h_t, z \rangle}$ relative to $\theta_{\langle h, z \rangle}$ by taking $t > 0$ (see also Figure 1). The spherical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

(1) As noted previously, $\mathcal{E}_p(b) = 0$ which implies that h being a consistent estimator (i.e., $\langle h, b \rangle_p \rightarrow 0$) is sufficient for removing the optimization bias asymptotically. However, this may not be 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 from the unknown vector b . It is very easy to verify, via (20), that $\mathcal{E}_p(h_{\tau^*}) = 0$ for h_t as in (19) but with $t = \tau^*$ given by

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

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 $|\tau_p - \tau_p^*| \rightarrow 0$ 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

$$(28) \quad \mathcal{E}(h_t) = \frac{\langle b, z \rangle - \langle h_t, b \rangle \langle h_t, z \rangle}{1 - \langle h_t, z \rangle^2} = \left(\frac{\sin \theta_{\langle h_t, b \rangle}}{\sin \theta_{\langle h_t, z \rangle}} \right) \cos \Theta_{h_t}$$

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 $\mathcal{E}(h_t) = 0$ per (28). This

occurs precisely for $t = \tau_p^*$ in (27). It may be verified that τ_p^* is also the maximizer of $\langle h_t, b \rangle$ over $t \in \mathbb{R}$ and equivalently, τ_p^* is the minimizer of $\theta_{\langle h_t, b \rangle}$ over $t \in \mathbb{R}$, i.e., the $t = \tau_p^*$ 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 (and Appendix B) 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 $\mathcal{E}(\cdot)$ 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 simulation experiments that illustrate the impact on a minimum variance portfolio of the dispersion bias correction, corroborating the theory of Section 4.⁸ Section 5.2 investigates the behavior of the portfolio volatility for the PCA and several corrected estimators in the case of small n and moderate p . Sections 5.3 and 5.4 study the large p asymptotics of the dispersion and optimization biases.

We generate a return observation $Y_j \in \mathbb{R}^p$ to p securities from model (8) so that

$$(29) \quad Y_j = \beta X_j + Z_j$$

for unobserved factor and specific returns $X_j \in \mathbb{R}$ and $Z_j = (Z_j^1, \dots, Z_j^p)^\top$. The Y_j for $j = 1, \dots, n$ form the columns of the $p \times n$ data matrix Y . We extract a PCA estimate h from the sample covariance $S = YY^\top/n$ in (10) ensuring it is positively oriented (i.e., $\mu(h) \geq 0$) per (11). The estimator $h_t \propto h + tZ$ is formed via (19) for constant $t \in \mathbb{R}_+$. Similarly, the estimators h_{τ^*} and h_τ follow (19) but with the random (dispersion) shrinkage parameters τ_p^* and τ_p in (27) and (21). The latter relies on the eigenvalues of S . We refer to h_{τ^*} as the *exact* estimator as it carries no optimization bias. We call h_τ the *blind* estimator as it is unable to observe β , and the family $\{h_t\}_{t \in \mathbb{R}_+}$, the *parametric* estimator, to which the PCA estimator $h = h_0$ belongs.

5.1. Market model calibration and estimation. We require the generating process to obey Assumptions 2.1 and 2.3 and take the $\{X_j\}$ and $\{Z_j\}$ in (29) to be i.i.d. Normal with mean zero, $\text{Var}(X_1) = \sigma^2 = (0.16)^2$ and $\text{Var}(Z_1) = \delta^2 I = (0.5)^2 I$. The unknown vector $\beta \in \mathbb{R}^p$ is constructed to have mean $\mu(\beta) = 1$ and dispersion

⁸For complementary simulations calibrated to the US equity market, see Goldberg et al. (2020).

$d(\beta) = 0.5$, and is held constant over the observations.⁹ Then, $\text{Var}(Y_j) = \Sigma$ with

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

as in (2). Our task amounts to specifying the estimates $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$ of (β, σ, δ) .

We use each of the estimates h_t, h_{τ^*} and h_τ as the basis of an estimate $\hat{\Sigma}$ of the covariance matrix. Given a choice of $\hat{\beta}$, we estimate the covariance matrix Σ by

$$(30) \quad \hat{\Sigma} = \hat{\sigma}^2 \hat{\beta} \hat{\beta}^\top + \hat{\delta}^2 \mathbf{I}$$

We take the estimates $\hat{\sigma}^2 = \mathcal{J}_p^2 \mu_p^2(h)$ and $\hat{\delta}^2 = n \ell_p^2 / p = (\text{tr}(S) - \mathcal{J}_p^2) \left(\frac{n/p}{n-1} \right)$ directly from Table 1 regardless of the choice $\hat{\beta}$ of the estimator used for β . These estimates are compatible with the assumption that $\mu(\hat{\beta}) = 1$ which is without loss of generality as the σ and β are undentifiable (i.e., $\mu(\beta) = 1$ is not exploited but serves solely as a convention). Thus $\hat{\beta}$ is either $\frac{h_t}{\mu(h_t)}$ (parametric), $\frac{h_\tau}{\mu(h_\tau)}$ (blind) or $\frac{h_{\tau^*}}{\mu(h_{\tau^*})}$ (exact).

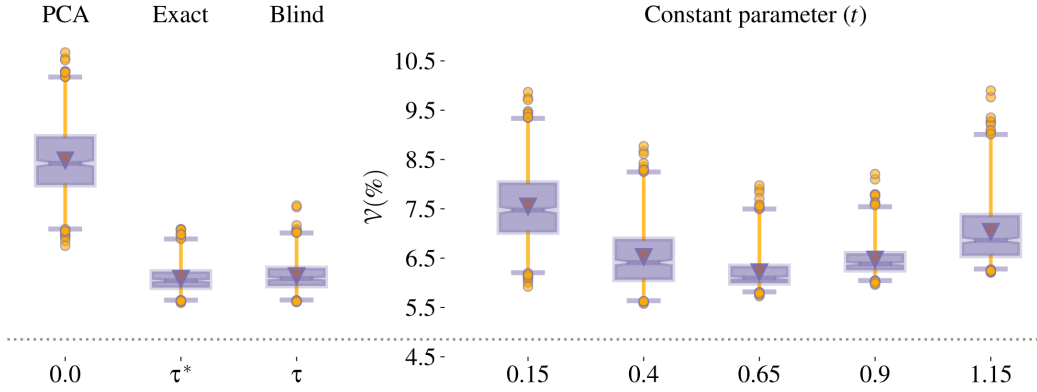


Figure 3. Boxplots of the true volatility \mathcal{V} of the estimated minimum variance portfolio \hat{w} constructed with the parametric, exact and blind estimators. The boxplots for five constant parameter estimators h_t are shown on the right of the \mathcal{V} -axis. The minimizer (over t) of the average volatility is approximately 0.5. The left of the \mathcal{V} -axis shows boxplots for the PCA, exact and blind estimators. The dashed line marks the optimal minimum volatility \mathcal{V} , the square-root of the minimum value of (1). 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.

5.2. Minimum variance portfolio volatility. 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 the $\hat{\Sigma}$ in (30) and compute the estimated, minimum variance portfolio \hat{w} by solving (1) after replacing Σ by $\hat{\Sigma}$. We take the true

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 $c = 0.5/d(\eta)$ so that $d(\beta) = 0.5$ and $\mu(\beta) = 1$ for $\beta = (\beta_1, \dots, \beta_p)^\top$.

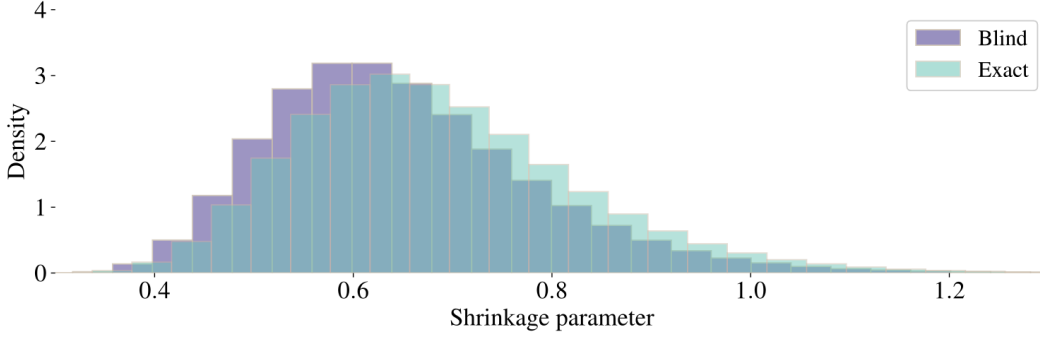


Figure 4. Histograms of 10^6 simulations for the exact and blind (dispersion) shrinkage parameters τ_p^* and τ_p (see formulas (21) and (27)). The sample means and standard deviations are approximately 0.684 and 0.144 for τ_p^* and 0.650 and 0.137 for τ_p . For each, we take $n = 50$ observations and $p = 500$ securities.

volatility \mathcal{V} (square root of $\mathcal{V}^2 = \hat{w}^\top \Sigma \hat{w}$) 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 boxplots of \mathcal{V} for each of the estimators. Even for the moderate number of securities ($p = 500$) and small sample size ($n = 50$) used in our experiment, the exact and blind estimators materially outperformed PCA. For instance, relative to the PCA estimator, we observe a reduction in median \mathcal{V} of more than 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 \mathcal{V}^* , the square root of the minimum value of $w^\top \Sigma w$ attained in optimization problem (1). All estimators produce portfolios with higher volatility than \mathcal{V}^* (approximately 4.853), indicating a higher level of risk than optimal. However, the median volatility produced by the exact and blind estimators are both within 26% of the optimum \mathcal{V}^* , while the PCA estimator yields a median volatility that exceeds the optimum by 74%. Figure 3 also displays results for five parametric estimators (on the right of the \mathcal{V} -axis). The best parametric estimator h_t achieves similar performance gains to the blind and exact estimators. This estimator corresponds to the value $t = 0.65$ that necessarily depends on the unknown β . It approximately minimizes the median volatility (a function of Σ) over the nonrandom parameter choices. However, this value is never accessible in practice since Σ is not known in such settings. Remarkably, it underperforms (in term of the mean and the variance) relative to the blind estimator, which relies only on the observed data Y . The distributions of the random, exact and blind dispersion-shrinkage parameters τ^* and τ are shown in Figure 4. These histograms support the theoretical finding that the τ approximates well the exact parameter τ^* , which benefits from the knowledge of β in the model.

Table 2 supplements Figure 3 and reports the sample means and variances of the true volatility \mathcal{V} for the PCA, exact, blind and best parametric (Param) estimators. It also reports the same statistics for the estimated volatility $\hat{\mathcal{V}}$ where $\hat{\mathcal{V}}^2 = \hat{w}^\top \hat{\Sigma} \hat{w}$ was

	$E(\mathcal{V})$	$E(\hat{\mathcal{V}})$	$E(\mathcal{V}/\hat{\mathcal{V}})$	$SD(\mathcal{V})$	$SD(\hat{\mathcal{V}})$	$SD(\mathcal{V}/\hat{\mathcal{V}})$
PCA	8.490	3.944	2.163	0.701	0.157	0.262
Exact	6.091	6.054	1.006	0.262	0.182	0.017
Blind	6.149	5.958	1.035	0.297	0.347	0.069
Param	6.225	5.982	1.045	0.355	0.280	0.103

Table 2. Sample means and standard deviations for the true and estimated volatilities \mathcal{V} and $\hat{\mathcal{V}}$ and their ratio $\mathcal{V}/\hat{\mathcal{V}}$ computed using 10^6 simulations. The estimates for $E(\mathcal{V})$ have Normal 99% confidence intervals of $\pm 2.58 SD(\mathcal{V}) \times 10^{-3}$ and analogously for $E(\hat{\mathcal{V}})$ and $E(\mathcal{V}/\hat{\mathcal{V}})$. Param denotes the best parametric estimator $h_{0.65}$.

defined in (6) and shown to be of order $1/p$. The blind estimator outperforms PCA and the best parametric estimator $h_{0.65}$ on the \mathcal{V} -metric in terms of both mean and variance. We find a variance reduction of factors 5.57 and 1.43 relative to PCA and the best parametric estimator respectively. Conversely, the blind estimator exhibits a variance for the estimated volatility $\hat{\mathcal{V}}$ that is larger than both the PCA and $h_{0.65}$. 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 $\mathcal{V}/\hat{\mathcal{V}}$. 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., (7) in Section 2.1).

5.3. Dispersion bias identification. Table 3 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 computed with 10^6 simulations. We fix $n = 50$ observations and numbers of securities p ranging from 500 to 8000. The value of $\langle b, z \rangle$ multiplied by the point estimate $E(\langle h, b \rangle_p)$ is equal to the point estimate $E(\langle h, z \rangle_p)$ to four decimal places for each p . This result is well within the Normal 99% confidence intervals around $E(\langle h, z \rangle_p)$ for each value of p . Further, the sample means of the values $\langle h, b \rangle_p$ and ψ_p , which are asymptotically equal, get closer as p increases from 500 to 8000. The 99% confidence intervals around $E(\psi_p)$ and $E(\langle h, b \rangle_p)$, however, do not overlap for the values of p we considered.

5.4. Asymptotics of the optimization bias. We experimentally confirm the statements of Theorems 3.1 and 4.1 and support the conjecture of Section 4 by simulating models of increasing size, taking p as large as 8000. For every p ,¹⁰ 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 the size of the portfolio. We study the optimization bias $\mathcal{E}_p(h)$ for the PCA estimate and its corrected counterpart $\mathcal{E}_p(h_\tau)$ that is produced by the blind estimator (the exact estimator has no optimization bias). The error \mathcal{E}_p was shown to be closely related to true volatility of the estimated minimum variance portfolio investigated in Section 5.2. Indeed, the er-

¹⁰Following footnote 9, we generate a sequence η_1, η_2, \dots 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 $d_p(\beta) = 0.5$.

p	$E(\langle h, z \rangle_p)$	$E(\langle h, b \rangle_p)$	$E(\psi_p)$	$SD(\langle h, z \rangle_p)$	$SD(\langle h, b \rangle_p)$	$SD(\psi_p)$
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 3. 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 $d^2(b) = 0.5$) computed with 10^6 simulations. For each value of p , we use $n = 50$. Estimates for $E(\langle h, z \rangle_p)$ have 99% confidence intervals of $\pm 2.58 SD(\langle h, z \rangle_p) \times 10^{-3}$ and analogously for $E(\langle h, b \rangle_p)$ and $E(\psi_p)$.

ror \mathcal{E}_p is the sole component of the asymptotic description of the true volatility (see $\mathcal{V} = \mathcal{V}_p$ in (5) and Appendix B) that may be manipulated in an estimation context. Moreover, per (7), the ratio of the true to the estimated variance $\mathcal{V}_p^2 / \hat{\mathcal{V}}_p^2$ is proportional to $p \mathcal{E}_p^2$.

Figure 5 displays the moderate deviations (on the scale \sqrt{p}) of the corrected optimization bias $\mathcal{E}_p(h_\tau)$. It confirms that these deviations do not grow in p and further suggests a convergence (in law) of the rescaled variable $\sqrt{p} \mathcal{E}_p(h_\tau)$ to some nondegenerate limit (e.g., Normally distributed) of zero mean and finite variance. By illustrating convergence in law, these results support Theorem 4.1, which states that $\mathcal{E}_p(h_\tau)$ vanishes almost surely as the size of the portfolio grows. They also provide evidence for the conjecture of Section 4, which posits that $\sqrt{p} \mathcal{E}_p(h_\tau)$ has a second moment that is bounded in p . Tables 4 and 5 supply further support. Table 4 (the first two columns) illustrates that the mean and standard deviation of $\mathcal{E}_p(h_\tau)$ 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 5 supplements these statistics with those for $p \mathcal{E}_p^2(h_\tau)$. The first column of Table 5 provides evidence for our conjecture by showing that the mean of $p \mathcal{E}_p^2(h_\tau)$ 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 in the experiments).

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

The theoretical results we confirm here in the setting of i.i.d. Gaussian observations hold much more broadly. Neither the i.i.d. nor the Gaussian requirements are

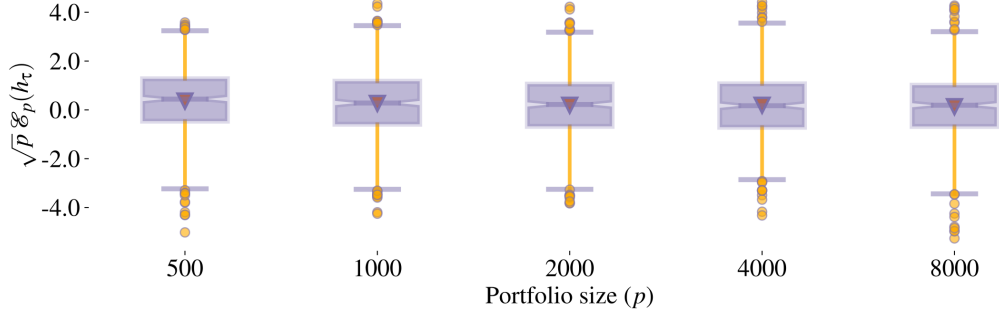


Figure 5. Boxplots of $\sqrt{p}\mathcal{E}_p(h_\tau)$ vs growing portfolio size p . Each blind estimator corrected optimization bias $\mathcal{E}_p(h_\tau)$ is constructed from (23) by using $n = 50$ observations. 10^3 simulations are used to construct each boxplot that displays the median, the interquartile range and outliers below 1% and above 99% are shown. Sample means for each value of p are marked with a (upside down) triangle.

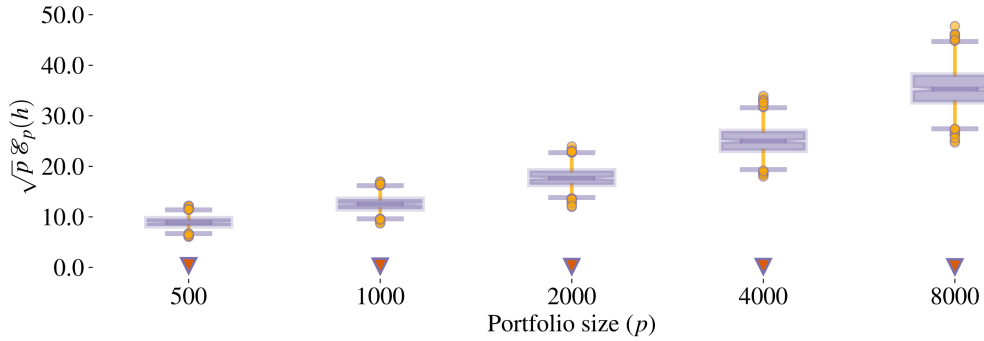


Figure 6. Boxplots of $\sqrt{p}\mathcal{E}_p(h)$ vs growing portfolio size p . The optimization bias $\mathcal{E}_p(h)$ of PCA is constructed from (17) by using $n = 50$ observations. Sample means for each corrected bias $\sqrt{p}\mathcal{E}_p(h_\tau)$ is depicted with a (upside down) triangle marker. 10^3 simulations are used to construct each boxplot that displays the median, the interquartile range and outliers below 1% and above 99% are shown.

needed to obtain qualitatively similar results. We refer to reader to extensive simulations in Goldberg et al. (2020) that test our conclusions for more complex models of security returns under more realistic market calibrations.

p	$E(\mathcal{E}_p(h_\tau))$	$SD(\mathcal{E}_p(h_\tau))$	$E(\mathcal{E}_p(h))$	$SD(\mathcal{E}_p(h))$	$\frac{E(\mathcal{E}_p(h))}{E(\mathcal{E}_p(h_\tau))}$
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 4. Sample statistics for the optimization bias \mathcal{E}_p produced by the PCA estimator (h) and the blind estimator (h_τ) versus growing portfolio size p . Estimates of the ratio $E(\mathcal{E}_p(h))/E(\mathcal{E}_p(h_\tau))$ 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 $E(\mathcal{E}_p(h_\tau))$ has the Normal 99% confidence interval $\pm 2.58 SD(\mathcal{E}_p(h_\tau)) \times 10^{-3}$ and analogously for the PCA estimator h . For each value of p we use $n = 50$ observations.

p	$pE(\mathcal{E}_p^2(h_\tau))$	$pSD(\mathcal{E}_p^2(h_\tau))$	$pE(\mathcal{E}_p^2(h))$	$pSD(\mathcal{E}_p^2(h))$
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 5. Sample statistics for the scaled square of the optimization bias $p\mathcal{E}_p^2$ generated 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 sample estimate of $pE(\mathcal{E}_p^2(h_\tau))$ has the Normal 99% confidence interval $\pm 2.58 pSD(\mathcal{E}_p^2(h_\tau)) \times 10^{-3}$ and analogously for the PCA estimator h . For each value of p we use $n = 50$ observations.

A. PROOFS

Recall the $p \times n$ data matrix $Y = Y_{p \times n}$ of n excess returns to p securities. According to (8), the j th observation (j th column of Y) is $Y_j = \beta X_j + Z_j$, and

$$(31) \quad Y = \beta X^\top + Z$$

for $\beta \in \mathbb{R}^p$, a row vector $X^\top = (X_1, \dots, X_n)$ of realized market return X and $Z = Z_{p \times n}$, the $p \times n$ matrix with j th column Z_j , the j th realized specific return.

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

$$(32) \quad |\beta|^2 = p\mu^2(\beta)(1 + d^2(\beta))$$

as in Appendix C, for the mean $\mu(\beta) = \mu_p(\beta)$ and dispersion $d(\beta) = d_p(\beta)$ in (9).

PROOF OF THEOREM 3.1. Recall the eigenvector h of the sample covariance matrix $S = YY^\top/n$ with the eigenvalue \mathcal{J}_p^2 as in (11). We consider the the singular value decomposition of Y and $\chi_p \in \mathbb{R}^n$ with $|\chi_p| = 1$ such that h and χ_p form the left and right singular vectors of Y/\sqrt{n} respectively with singular value $\mathcal{J}_p \geq 0$.¹¹ By (31),

$$(33) \quad h\mathcal{J}_p = Y\chi_p/\sqrt{n} = \frac{\beta X^\top \chi_p + Z\chi_p}{\sqrt{n}}.$$

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

$$(34) \quad \langle h, b \rangle_p = h^\top b = \left(\frac{|\beta|X^\top \chi_p}{\mathcal{J}_p \sqrt{n}} \right) + \left(\frac{\beta^\top Z}{\sqrt{p}|\beta|} \right) \left(\frac{\chi_p \sqrt{p}}{\mathcal{J}_p \sqrt{n}} \right)$$

$$(35) \quad \langle h, z \rangle_p = h^\top z = \langle b, z \rangle_p \left(\frac{|\beta|X^\top \chi_p}{\mathcal{J}_p \sqrt{n}} \right) + \left(\frac{e^\top Z}{\sqrt{p}|e|} \right) \left(\frac{\chi_p \sqrt{p}}{\mathcal{J}_p \sqrt{n}} \right)$$

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

$$(36) \quad \mathcal{J}_p^2/p = \frac{|\beta|^2(X^\top \chi_p)^2}{np} + \frac{\chi_p^\top Z^\top Z \chi_p}{np} + 2(X^\top \chi_p) \left(\frac{\beta^\top Z}{\sqrt{p}|\beta|} \right) \left(\frac{\chi_p |\beta|}{n\sqrt{p}} \right).$$

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 $c < \frac{1}{p} \sum_{i=1}^p \eta_i^2 < C$ for fixed constants $c, C > 0$ and all p sufficiently large. For mean-zero, pairwise independent and (real) random variables $\{Z^i\}_{i \in \mathbb{N}}$ with $\sup_i \text{Var}(Z^i) < \infty$, writing $\eta = (\eta_1, \dots, \eta_p)^\top$ and $Z = (Z^1, \dots, Z^p)^\top$ for any p , we have that $\frac{\eta^\top Z}{\sqrt{p}|\eta|} \rightarrow 0$ almost surely as $p \uparrow \infty$.*

PROOF. Letting $W_i = \eta_i Z^i$, we have

$$\frac{\eta^\top Z}{\sqrt{p}|\eta|} = \left(\frac{\sqrt{p}}{|\eta|} \right) \frac{1}{p} \sum_{i=1}^p W_i.$$

¹¹ By convention, the singular values of a real matrix A are taken as the nonnegative square roots of the (nonnegative) eigenvalues of $A^\top A$. The largest such value a satisfies $a^2 = \sup_{|x|=1} A^\top A$.

The result now follows by the SLLN of Chandra & Goswami (1992, Theorem 6) provided $\sup_p \frac{1}{p} \sum_{i=1}^p \mathbb{E}(W_i^2) < \infty$ (taking $\phi(t) = t^2$ in that reference) and since $\sup_p \sqrt{p}/|\eta| < 1/\sqrt{c} < \infty$ by our assumptions. Since $\mathbb{E}(W_i^2) = \eta_i^2 \text{Var}(Z^i)$ we have $\sup_p \frac{1}{p} \sum_{i=1}^p \mathbb{E}(W_i^2) = \sup_i \text{Var}(Z^i) \sup_p |\eta|^2/p \leq C \sup_i \text{Var}(Z^i) < \infty$. ■

By applying Lemma A.1 to each $(\frac{\beta^\top Z}{\sqrt{p}|\beta|})_j$ and $(\frac{e^\top Z}{\sqrt{p}|e|})_j$ for $1 \leq j \leq n$ in (34) and (35) we have, under Assumptions 2.2 and 2.3, that (34) and (35) reduce to

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

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

provided the limit on the right side of (37) exists almost surely (for the existence of $\langle b, z \rangle_\infty$, see Remark 3.2) and that $\sup_p \sqrt{p}/\mathfrak{s}_p < \infty$ almost surely (see Lemma A.2).

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

$$(39) \quad v_X^2 = \frac{|X|^2}{n} \mu_\infty^2(\beta) (1 + d_\infty^2(\beta)).$$

With $\text{tr}(S)$ denoting the matrix trace¹² of S , we have $\ell_p^2 = (\text{tr}(S) - \mathfrak{s}_p^2)/(n-1)$ as in the definition of ψ_p in (12) after taking $p \geq n \geq 2$.

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

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

$$(40) \quad \langle h, b \rangle_\infty = \sqrt{\frac{nv_X^2}{nv_X^2 + \delta^2}} \in (0, 1)$$

almost surely. Applying Lemma A.2 to $\psi_p^2 = (\mathfrak{s}_p^2 - \ell_p^2)/\mathfrak{s}_p^2$ per (12), we obtain that the limit ψ_∞ equals the right side of (40) almost surely. ■

PROOF OF LEMMA A.2. Let $S^{n-1} = \{x \in \mathbb{R}^n : |x| = 1\}$ and define

$$(41) \quad g_p(x) = \frac{|\beta|^2 (X^\top x)^2}{np} + \frac{x^\top Z^\top Z x}{np} + 2(X^\top x) \left(\frac{\beta^\top Z}{\sqrt{p}|\beta|} \right) \left(\frac{x|\beta|}{n\sqrt{p}} \right).$$

By definition we have that $\mathfrak{s}_p^2/p = g_p(\chi_p) = \sup_{x \in S^{n-1}} g_p(x)$ where the last equality follows due to the fact that $g_p(x) = (x^\top Y^\top Y x)/(np)$ (see, footnote 11).

¹²The sum of the diagonal elements of S , and equivalently the sum of its eigenvalues.

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

$$(42) \quad g_\infty(x) = v_X^2 \left(\frac{X^\top x}{|X|} \right)^2 + \frac{\delta^2}{n}.$$

If so, then $\lim_{p \uparrow \infty} \mathcal{J}_p^2/p = \sup_{x \in \mathbb{S}^{n-1}} g_\infty(x) = v_X^2 + \delta^2/n$ since by (36) we have $\mathcal{J}_p^2/p = g_p(\chi_p) \rightarrow \sup_{x \in \mathbb{S}^{n-1}} g_\infty(x)$ attained by $X/|X|$ almost surely. Moreover,

$$|g_\infty(\chi_p) - g_\infty(X/|X|)| \leq |g_\infty(\chi_p) - g_p(\chi_p)| + |g_p(\chi_p) - g_\infty(X/|X|)|$$

which converges to zero as $p \uparrow \infty$ almost surely (as $g_p \rightarrow g_\infty$ uniformly on \mathbb{S}^{n-1}). Consequently, since g_∞ is continuous $g_\infty(\lim_{k \uparrow \infty} \chi_{p_k}) = g_\infty(X/|X|)$ for any convergent subsequence $\{\chi_{p_k}\}$ with $p_k \uparrow \infty$. Also, since g_∞ has exactly two maximizers on \mathbb{S}^{n-1} , namely $X/|X|$ and $-X/|X|$, every convergent subsequence of $\{\chi_p\}_{p \in \mathbb{N}}$ must converge to $X/|X|$ in view of (38) and the choices of sign for h and b (i.e., w.l.o.g. we have $\langle h, z \rangle_p \geq 0$ per (11) and $\langle b, z \rangle_p > 0$ for all p sufficiently large under Assumption 2.2). Since every convergent subsequence of the bounded sequence $\{\chi_p\}_{p \in \mathbb{N}}$ converges to $X/|X|$ almost surely, the same conclusion for χ_p holds.

We proceed to show the almost sure, uniform (on \mathbb{S}^{n-1}) convergence of g_p , which follows from the bound $|g_p(x) - g_\infty(x)| \leq |\gamma_p^1(x)| + |\gamma_p^2(x)| + |\gamma_p^3(x)|$ where

$$\begin{aligned} |\gamma_p^1(x)| &= (X^\top x)^2 \left| \frac{|\beta|^2}{np} - \frac{v_X^2}{|X|^2} \right|, \\ |\gamma_p^2(x)| &= \left| \frac{x^\top Z^\top Z x}{np} - \frac{\delta^2}{n} \right|, \\ |\gamma_p^3(x)| &= 2(X^\top x) \left| \left(\frac{\beta^\top Z}{\sqrt{p}|\beta|} \right) \left(\frac{x|\beta|}{n\sqrt{p}} \right) \right|. \end{aligned}$$

Using the identity $|\beta|^2 = p\mu_p^2(\beta)(1 + d_p^2(\beta))$ in (32), under Assumption 2.2, we deduce that $\sup_{x \in \mathbb{S}^{n-1}} |\gamma_p^1(x)| \rightarrow 0$ by the definition of v_X^2 in (39). For the second term $|\gamma_p^2(x)|$, observe that $x^\top Z^\top Z x = \sum_{j=1}^n \sum_{k=1}^n x_j Z_j^\top Z_k x_k$, and consequently

$$\begin{aligned} |\gamma_p^2(x)| &\leq \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n |x_j x_k| |Z_j^\top Z_k / p - \delta^2 1_{\{k=j\}}| \\ &\leq n \max_{1 \leq j, k \leq n} |Z_j^\top Z_k / p - \delta^2 1_{\{k=j\}}| \end{aligned}$$

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

The required convergence of g_p follows by applying Cauchy-Schwartz to $|\gamma_p^3(x)|$,

$$(43) \quad |\gamma_p^3(x)| \leq 2|X|^2 \left| \frac{\beta^\top Z}{\sqrt{p}|\beta|} \right| \left(\frac{|\beta|}{n\sqrt{p}} \right),$$

so that $\sup_{x \in \mathbb{S}^{n-1}} |\gamma_p^3(x)| \rightarrow 0$ by Lemma A.1, the continuity of the norm $|\cdot|$ and that $\sup_p |\beta|^2/p = \sup_p \mu_p^2(\beta)(1 + d_p^2(\beta)) < \infty$ under Assumptions 2.2 and 2.3.

Finally, observe that since $nS = YY^\top = X^\top X \beta \beta^\top + ZZ^\top + \beta X^\top Z^\top + ZX \beta^\top$, the almost sure limit as $p \uparrow \infty$ of the trace of S is given by

$$(44) \quad \lim_{p \uparrow \infty} \frac{\text{tr}(S)}{p} = \lim_{p \uparrow \infty} \left(\frac{|\beta|^2 |X|^2}{pn} + \frac{\text{tr}(Z^\top Z)}{pn} + 2 \left(\frac{|\beta|}{\sqrt{pn}} \right) \frac{\beta^\top ZX}{|\beta| \sqrt{p}} \right) = v_X^2 + \delta^2$$

where we applied (32) with Assumption 2.2 and (39) to the first term to obtain v_X^2 , an argument identical to that for $x^\top Z^\top Z x$ above to the second term to obtain δ^2 , and Lemma A.1 to remove the third term. From (44) and that $s_p^2/p \rightarrow v_X^2 + \delta^2/n$ almost surely, we deduce that $\ell_p^2/p = (\text{tr}(S)/p - s_p^2/p)/(n-1) \rightarrow \delta^2/n$ almost surely. ■

PROOF OF THEOREM 4.1. From Appendix C, we have

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

for $h_t = \frac{h+tz}{|h+tz|}$ and $t \in \mathbb{R}$. For $\tau_p^* = \frac{\langle b, z \rangle_p - \langle h, b \rangle_p \langle h, z \rangle_p}{\langle h, b \rangle_p - \langle b, z \rangle_p \langle h, z \rangle_p}$ we have $\mathcal{E}_p(h_{\tau^*}) = 0$ (the subscript p on τ^* in h_{τ^*} is suppressed) by definition of $\mathcal{E}_p(h)$ in (17). Then,

$$(45) \quad \mathcal{E}_p(h_t) = (t_p^* - t) \left(\frac{\langle h, b \rangle_p - \langle b, z \rangle_p \langle h, z \rangle_p}{1 - \langle h, z \rangle_p^2} \right) \quad t \in \mathbb{R}.$$

For $\tau_p = \frac{(1-\psi_p^2)\langle h, z \rangle_p}{\psi_p^2 - \langle h, z \rangle_p^2}$ in (21), we have by a tedious calculation that

$$\begin{aligned} t_p^* - \tau_p &= \frac{\langle h, z \rangle_p^2 (\langle h, b \rangle_p \langle h, z \rangle_p - \langle b, z \rangle_p) + \langle h, z \rangle_p (\langle h, z \rangle_p \langle b, z \rangle_p - \langle h, b \rangle_p)}{(\langle h, b \rangle_p - \langle h, z \rangle_p \langle b, z \rangle_p)(\psi_p^2 - \langle h, z \rangle_p^2)} \\ &\quad + \frac{(\langle b, z \rangle_p - \langle h, b \rangle_p \langle h, z \rangle_p) \psi_p^2 + \langle h, z \rangle_p \psi_p^2 (\langle h, b \rangle_p - \langle h, z \rangle_p \langle b, z \rangle_p)}{(\langle h, b \rangle_p - \langle h, z \rangle_p \langle b, z \rangle_p)(\psi_p^2 - \langle h, z \rangle_p^2)} \\ &= \frac{\langle h, z \rangle_p^2 (\langle h, b \rangle_p \langle h, z \rangle_p - \psi_p^2 \langle b, z \rangle_p) + \langle b, z \rangle_p \psi_p^2 - \langle h, z \rangle_p \langle h, b \rangle_p}{(\langle h, b \rangle_p - \langle h, z \rangle_p \langle b, z \rangle_p)(\psi_p^2 - \langle h, z \rangle_p^2)}. \end{aligned}$$

Then continuing with (45), we obtain

$$\begin{aligned} \mathcal{E}_p(h_\tau) &= (t_p^* - \tau_p) \frac{\langle h, b \rangle_p - \langle b, z \rangle_p \langle h, z \rangle_p}{1 - \langle h, z \rangle_p^2} \\ &= \frac{\langle h, z \rangle_p^2 (\langle h, b \rangle_p \langle h, z \rangle_p - \psi_p^2 \langle b, z \rangle_p) + (\langle b, z \rangle_p \psi_p^2 - \langle h, z \rangle_p \langle h, b \rangle_p)}{(1 - \langle h, z \rangle_p^2)(\psi_p^2 - \langle h, z \rangle_p^2)} \\ (46) \quad &= \frac{\langle b, z \rangle_p \psi_p^2 - \langle h, z \rangle_p \langle h, b \rangle_p}{\psi_p^2 - \langle h, z \rangle_p^2}. \end{aligned}$$

Taking $p \uparrow \infty$, the claim now follows by Theorem 3.1 under its assumptions. \blacksquare

B. ASYMPTOTIC VARIANCE OF MINIMUM VARIANCE PORTFOLIOS

We develop asymptotic expressions for the variance of a minimum variance portfolio, which supply refined versions of the asymptotic expressions given in Section 2.1.

Setting $v^2 = |\beta|^2/p$, the covariance matrix $\Sigma = \sigma^2 \beta \beta^\top + \delta^2 \mathbf{I}$ takes the form

$$\Sigma = \sigma^2 p v^2 b b^\top + \delta^2 \mathbf{I},$$

where as always, $b = \beta/|\beta|$. The estimated model $\hat{\Sigma}$ employs the estimates $(\hat{\beta}, \hat{\sigma}, \hat{\delta})$ of the parameters (β, σ, δ) and may be written in the form $\hat{\Sigma} = \hat{\sigma}^2 p \hat{v} h h^\top + \hat{\delta}^2 \mathbf{I}$ where $q = \hat{\beta}/|\hat{\beta}|$ and $\hat{v}^2 = |\hat{\beta}|^2/p$. Here, the estimates $(h, \hat{\sigma}, \hat{\delta})$ need not involve PCA but may result from any estimation procedure obeying Assumption B.1 below.

Define $\hat{\kappa} = \hat{\delta}/(\hat{\sigma} \hat{v})$ and, as before, we let $\hat{w} \in \mathbb{R}^p$ denote the minimizer of $\min_{e^\top w=1} w^\top \hat{\Sigma} w$, the optimization problem (1) but with $\hat{\Sigma}$ replacing Σ . As in (3), set $z = e/\sqrt{p}$ for $e = (1, \dots, 1)$. The solution \hat{w} (in closed-form) may be written as

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

The true variance $\mathcal{V}^2 = \hat{w}^\top \Sigma \hat{w} = \sigma^2 p v^2 (\hat{w}^\top b)^2 + \delta^2 |\hat{w}|^2$ and the estimated variance $\hat{\mathcal{V}}^2 = \hat{w}^\top \hat{\Sigma} \hat{w} = \hat{\sigma}^2 p \hat{v}^2 (\hat{w}^\top h)^2 + \hat{\delta}^2 |\hat{w}|^2$ of the portfolio \hat{w} (both introduced in Section 2.1) are governed by the asymptotics of the following. First, for $\mathcal{E}(h)$ in (4),

$$\begin{aligned} p(\hat{w}^\top b)^2 &= \left(\frac{(\hat{\kappa}^2/p + 1) \langle b, z \rangle - \langle h, b \rangle \langle h, z \rangle}{\hat{\kappa}^2/p + 1 - \langle h, z \rangle^2} \right)^2 \\ &= \left(\mathcal{E}(h) + \frac{1}{p} \left(\frac{\hat{\kappa}^2 \langle b, z \rangle}{1 - \langle h, z \rangle^2} \right) \right)^2 \left(\frac{1}{1 + o_p} \right)^2 \end{aligned}$$

where $o_p = \frac{\hat{\kappa}^2/p}{1 - \langle h, z \rangle^2}$ and $\left(\frac{1}{1 + o_p} \right)^2 = \sum_{k \geq 1} k o_p^{k-1}$ when the series converges. Next,

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

and the series $\frac{1}{1 + o_p} = \sum_{k \geq 1} o_p^{k-1}$ is useful when convergent. Finally,

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

From the expressions given above, the asymptotics of \mathcal{V}^2 and $\hat{\mathcal{V}}^2$ are immediate provided the following assumptions on the estimates $(h, \hat{\sigma}, \hat{\delta})$.

Assumption B.1. *The estimates $\hat{\sigma}$ and $\hat{\delta}$ (which possibly depend on p) are bounded above and away from zero and the estimate h satisfies $\sup_p \langle h, z \rangle_p^2 < 1$.*

Note that these assumptions are satisfied for the PCA estimates we analyze (see Theorem 3.1 and Table 1). Under Assumption B.1, for p sufficiently large,

$$\begin{aligned}\mathcal{V}^2 &= \sigma^2 \nu^2 \mathcal{E}^2(h) + \frac{2}{p} \left(\mathcal{E}^2(h) + \frac{\mathcal{E}(h) \hat{\kappa}^2 \langle b, z \rangle + \delta^2/2}{1 - \langle h, z \rangle^2} \right) + O(1/p^2) \\ \hat{\mathcal{V}}^2 &= \frac{\hat{\delta}^2/p}{1 - \langle h, z \rangle^2} + O(1/p^2)\end{aligned}$$

confirming the asymptotics stated in (5) and (6) using Vinogradov's notation.

The variance forecast ratio $\mathcal{V}^2/\hat{\mathcal{V}}^2$ asymptotics in (7) may now be refined as

$$\mathcal{V}^2/\hat{\mathcal{V}}^2 \sim A + Bp\mathcal{E}^2(h)$$

where $A = \frac{\delta^2 + \mathcal{E}(h)\hat{\kappa}^2 \langle b, z \rangle}{\hat{\delta}^2} + \frac{2(1 - \langle h, z \rangle^2)}{\hat{\delta}^2} \mathcal{E}^2(h)$ and $B = \frac{\sigma^2 \nu^2 (1 - \langle h, z \rangle^2)}{\hat{\delta}^2}$.

C. AUXILIARY CALCULATIONS

We present several calculations useful in the proof of Theorem 4.1, the estimators of Table 1, and equations stated without derivation (i.e., (14), (15), (20) and (26)).

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 $d^2(\eta) = \frac{1}{p} \sum_{i=1}^p (\eta_i / \mu(\eta) - 1)^2$ defined for $\mu(\eta) \neq 0$, we have

$$\begin{aligned}|\eta|^2 &= \sum_{i=1}^p \eta_i^2 = p\mu^2(\eta) + \sum_{i=1}^p (\eta_i - \mu(\eta))^2 \\ (47) \quad &= p\mu^2(\eta) (1 + d^2(\eta)) \quad (\mu(\eta) \neq 0).\end{aligned}$$

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

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

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

Table 1 involves the following calculation (note, the estimate $\hat{\delta}^2 = n\ell_p^2/p$ is justified on the basis of Lemma A.2). The estimate $\hat{\sigma}^2 = \mathcal{J}_p^2 \mu_p^2(h) = \mathcal{J}_p^2 \langle h, z \rangle_p^2/p$ uses the identities for η above. The limit follows from Lemma A.2 and Theorem 3.1.

We proceed to justify (20). For $\mathcal{E}(h)$ as in (4) and any $h \in \mathbb{R}^p$ with $|h| = 1$,

letting $h_t = \frac{h+tz}{|h+tz|}$ for $t \in \mathbb{R}$, we have

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

To justify (26) we check that,

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

REFERENCES

- Anderson, R. M., Bianchi, S. & Goldberg, L. R. (2012), ‘Will my risk parity portfolio outperform’, *Financial Analysts Journal* 68(6), 75–93.
- Aoshima, M., Shen, D., Shen, H., Yata, K., Zhou, Y.-H. & Marron, J. S. (2018), ‘A survey of high dimension low sample size asymptotics’, *Australian & New Zealand Journal of Statistics* 60(1), 4–19.
- 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’, *MSCI Barra Research Insight*.
- 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', *Sankhya: The Indian Journal of Statistics, Series A* pp. 215–231.
- 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. (2020), ‘Better betas’, *The Journal of Portfolio Management* 47(1), 119–136.
- 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.
- Gurdogan, H. & Kercheval, A. (2021), Multi anchor point shrinkage for the sample covariance matrix. CDAR working paper.
- 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.

- 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, J. S. (2016), 'The statistics and mathematics of high dimensional low sample size asymptotics', *Statistica Sinica* 26(4), 1747–1770.
- Shkonik, A. (2021), 'James-Stein estimation of the first principal component', *Stat* .
- 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.