Portfolio Selection Revisited

In memory of Harry M. Markowitz

Annals of Operations Research – forthcoming.

Alex Shkolnik[1*], Alec Kercheval[2], Hubeyb Gurdogan[3], Lisa R. Goldberg[4,5*], Haim Bar[6]

[1]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA, 93106, USA .
[2]Department of Mathematics, Florida State University, FL, 32306, USA .
[3]Department of Mathematics, University of California, Los Angeles, CA, 90095, USA .
[4]Department of Economics and Consortium for Data Analytics in Risk, University of California, Berkeley, CA, 94720, USA .
[5]Aperio, BlackRock, Inc., CA, 94965, USA .
[6]Department of Statistics, University of Connecticut, CT, 06269, USA .

*Corresponding author(s). E-mail(s): shkolnik@ucsb.edu; lrg@berkeley.edu;
Contributing authors: akercheval@fsu.edu; hgurdogan@math.ucla.edu; haim.bar@uconn.edu;

Abstract

In 1952, Harry Markowitz formulated portfolio selection as a trade-off between expected, or mean, return and variance. This launched a massive research effort devoted to finding suitable inputs to mean-variance optimization. The estimation problem is high dimensional and a factor model is at the core of many attempts. A factor model can reduce the number of parameters that need to be estimated to a manageable size, but these parameters may incorporate substantial, hidden estimation error. Recent analysis elucidates the nature of this error, identifies a mechanism by which it can corrupt optimization and provides a method for its mitigation. We explore this analysis here by illustrating how to improve the volatility ratio of large optimized portfolios, leading to superior portfolio selection.*

*We certify that we have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Keywords: mean-variance portfolio optimization, high dimensional covariance estimation, principal component analysis, factor models, volatility ratio, James-Stein shrinkage.

1 In the Beginning ...

Harry Markowitz launched modern finance when he was an economics graduate student in the early 1950s. By framing portfolio construction as an optimization that trades off expected return against risk, Markowitz brought mathematics, computing and data science to bear on investing, even though computing and data science had scarcely been invented. In a seminal article that Myron Scholes described as "the big bang,"[1] Markowitz (1952) introduced the concept of an *efficient portfolio*, which minimizes risk for a prescribed level of expected return, subject to constraints. Hiding in this simple formulation are two profound ideas that, prior to Markowitz, had not been explicitly central to finance or economics. The first is a portfolio level perspective, which leads to high dimensional analysis of statistical problems that involve many variables. The second is a quantitative notion of risk, which Markowitz had encountered in engineering and operations research.[2] Markowitz characterized risk as variance of portfolio return, and he mused about how to construct efficient portfolios.

The name "Markowitz" is sometimes attached to a portfolio that is completely determined by the mean-variance tradeoff under full investment. The efficient frontier, as featured in business schools everywhere, is composed of "Markowitz portfolios," whose weights can be conveniently expressed with a closed formula. Markowitz portfolios typically have short positions, but there were no securities lending desks in the 1950s. Harry Markowitz, was more interested in long-only portfolios, the kind that were available to investors, but the weights of a long-only portfolio required mathematical recipes that did not exist. So, Markowitz (1956) developed the critical line algorithm to incorporate position limits into mean-variance optimization, a development that was roughly coincident with the introduction of Fortran.[3] Well into his 90s, Markowitz wrote code.

The inputs to mean-variance optimization include a vector of expected returns and a matrix of return covariances. These inputs are never observable. A massive research effort dedicated to finding suitable estimates followed Markowitz' portfolio selection article, and continues today in industry and the academy. Why is this problem difficult? One obvious contributor is "dimension." As Markowitz realized early on, methods from classical statistics are not adequate when the number of securities, or variables, is large. In a prescient comment in his 1952 paper, Markowitz'a considered alternatives:

---

[1]Scholes offered his comments at the March 2024 *Journal of Investment Management* conference in Markowitz's honor held in San Diego.

[2]A discussion of how Markowitz brought ideas from engineering to finance is in MacKenzie (2006).

[3]Bailey and López de Prado (2013) describe an implementation of Markowitz's 1956 critical line algorithm. Cottle and Infanger (2010) provides a history of Markowitz's contributions to quadratic programming. New algorithms descending from the critical line algorithm are described in Boyd et al. (2024). A compendium of Markowitz's early ideas is in Markowitz (1959).

Perhaps there are ways, by combining statistical techniques and the judgment of experts, to form reasonable probability beliefs (μ_i, σ_{ij}). ... One suggestion as to tentative μ_i, σ_{ij} is to use the observed μ_i, σ_{ij} for some period of the past. I believe that better methods, which take into account more information, can be found. I believe that what is needed is essentially a "probabilistic" reformulation of security analysis. I will not pursue this subject here, for this is "another story." It is a story of which I have read only the first page of the first chapter.

This query preceded works by Wigner (1955) and Marcenko and Pastur (1967), which lay out the foundations of random matrix theory, a rich area of mathematics that informs high dimensional covariance matrix (or, σ_{ij}) estimation. It also preceded a major development in the work of Stein (1956) and James and Stein (1961) who convinced the statistics community that in high dimensions, better estimators than the sample mean (or μ_i) provably exist. For Markowitz, the dimension played a key role in the "law of the average covariance," which he often used to point out the "do's and don'ts of large portfolios" (Markowitz, 1959, Chapter 5). He used many thousands of securities for his numerical illustrations. In what follows, we use numerics to illustrate the "probabilistic" properties of large mean-variance optimized portfolios. We also combine insights from random matrix theory and James-Stein estimation to show how higher dimensions yield the additional "information" needed to improve the estimates (μ_i, σ_{ij}).

A complication is the dynamic and noisy nature of financial markets. Observations from a volatile period may not be useful when the market is calm. What limited data from an irregular past should we use to forecast risk in an uncertain future? Markowitz, as a self-described Bayesian (Markowitz, 2010), believed in the use of historical averages that are corrected for uncertainty in accordance with Bayesian principles (e.g., Markowitz and Xu (1994), Markowitz and Usmen (1996) and Markowitz (2012)). But mean-variance optimization in the presence of noise proved to be a challenging problem. As Markowitz concluded, even with a Bayesian approach "the investor is still too optimistic for his or her own best interest" (Markowitz and Usmen, 2003). This finding alludes to the now well-known observation that estimated mean-variance optimized portfolios tend to severely underestimate the true risk. Higher dimensions amplify the problem, and this so-called "Markowitz optimization enigma" has led to an active area of research in the past several decades. A small sampling of this literature covering various approaches includes Klein and Bawa (1976), Jobson and Korkie (1980), Best and Grauer (1991), Michaud and Ma (2001), Pafka and Kondor (2003), Ledoit and Wolf (2004), Lai et al. (2011), Fan et al. (2012), Bun et al. (2017), Ledoit and Wolf (2017), Bodnar et al. (2022) and Blanchet et al. (2022). We highlight Jobson et al. (1979) and Jorion (1986) for their use of James-Stein estimation for Markowitz problems, and a related thread of literature on "beta adjustments" which relies on Stein-type estimators (Elton et al., 2009, Chap. 7).

In this article, we provide easy to implement "James-Stein-Markowitz" (JSM) recipes for the estimates (μ_i, σ_{ij}). They incorporate the Bayes' rule instincts espoused by Markowitz regarding the use of historical averages in the form of James-Stein (JS) estimation. The JS estimator, which combines a vector of historical means with some "shrinkage" target, may be derived by applying Bayes' rule (e.g., Efron (1978)). In this context, a Bayesian prior is replaced by a "shrinkage" target, which stands in for the "information" that Markowitz sought to correct the historical mean. Our contribution is the realization that for mean-variance portfolios, that shrinkage target can usually be obtained from the constraints of the Markowitz optimization itself. We show that this allows for optimized portfolio selection that has superior performance to portfolios that do not result from this "shrinkage" rule. This principle takes firmer hold as more securities are added to the portfolio to increase dimension.

The JSM recipe is applied not just to the vector of historical means of the security returns, but more importantly, to the eigenvectors of their sample covariance matrix. These sample eigenvectors are high dimensional, and they govern security return correlations. In other words, they are risk drivers from which a factor model may be constructed. Early developments of such models includes the market model of Sharpe (1963) and the multi-factor models in the arbitrage pricing theory of Ross (1976). Beyond their theoretical underpinnings, factor models facilitate the problem of estimating a high-dimensional return covariance matrix.[4] They reduce the dimension to a manageable size, produce robust covariance matrices, and conform to the empirical fact that a few factors are adequate to explain correlation in security returns in developed public equity markets. Factor models in finance typically rely on either principal component analysis (PCA) or the commercially successful Barra models (Rosenberg, 1974). Blin et al. (2022) covers many of the historical developments of multi-factor models in finance, and the use of PCA for empirical work is grounded in the pioneering work of Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986). The latter had a large influence on the recent advances in high dimensional factor models and PCA (e.g. Bai and Ng (2008), Fan et al. (2013), Bai and Ng (2023) and Fan et al. (2023)). We adopt a PCA framework in which sample eigenvectors are used as risk factors to construct Markowtitz portfolios.

In Section 2, we review the construction of Markowitz portfolios with mean-variance optimization and introduce a measure of the volatility ratio of an estimated portfolio. Here, we allude to the literature showing that in situations endemic to financial markets, volatility ratio of large portfolios optimized with PCA is low. An asymptotic description of the volatility ratio is provided by formula (7), which indicates, perhaps counterintuitively, that in a large enough universe, the volatility ratio of risk forecasts of optimized portfolios may not depend on estimates of factor or specific variances. Rather, it is errors in means and factor exposures that distort mean-variance portfolios. The recipes for mean and covariance estimates that correct this distortion for PCA are in Section 3. A numerical illustration comparing PCA and JSM is in Section 4. In Section 5, we return to Markowitz the statistician, who had a deep interest in generating the best possible inputs to mean-variance analysis Appendix A covers factor-model portfolio construction and summarizes the derivation of formula (7). The calibration of a seven-factor return generating process used in our numerical results is specified in Appendix B. Appendix C contains technical details related to our numerical recipes.

[4]Factor modeling originated with an inquiry into the determinants of human intelligence in Spearman (1904). Spearman's g factor for intelligence is equivalent from a modeling viewpoint to the market factor in finance.

2 The Volatility Ratio of Markowitz Portfolios

For a universe of p securities, we estimate a p-vector $\mu = (\mu_i)$ of means and a $(p \times p)$-matrix $\Sigma = (\sigma_{ij})$ of covariances. These two estimates determine a Markowitz portfolio w via optimization, as the solution to

$$\begin{aligned} &\min_w \; w^\top \Sigma w \\ &\text{subject to:} \\ &\mu^\top w \geq \alpha \\ &\mathrm{e}^\top w = 1, \end{aligned} \tag{1}$$

where e is a p-vector of ones and α is a return target. Assume μ does not have identical entries. If we knew the true means $\boldsymbol{\mu}$ and covariances $\boldsymbol{\Sigma}$, the computed portfolio w would be "*mean-variance*" optimal. But in practice, the parameters μ and Σ are estimates, and the resulting errors affect the accuracy of the optimized portfolio return and risk. As aptly stated in Michaud (1989), mean-variance "*optimizers are, in a fundamental sense, estimation-error maximizers*".

We can characterize this mathematically in terms of volatility (or risk). The estimated variance of w is,

$$(\mathrm{EV})^2 = w^\top \Sigma w \tag{2}$$

and the square-root yields the estimated portfolio volatility EV. Its relationship to the true volatility TV, and the true variance $(\mathrm{TV})^2 = w^\top \boldsymbol{\Sigma} w$, may be quantified by the volatility ratio, denoted by $\mathcal{V}$, and defined via the relation

$$\mathrm{EV} = \mathrm{TV} \times \mathcal{V},$$

so that

$$\mathcal{V} = \frac{\mathrm{EV}}{\mathrm{TV}} = \frac{\text{Estimated Volatility}}{\text{True Volatility}}\,. \tag{3}$$

Since we don't know the true volatility, we don't know $\mathcal{V}$. But quantities such as $\mathcal{V}$ are of great interest to both investors and academics, with a vast literature spanning many disciplines; e.g, mathematical finance, physics, economics, statistics and operations research. Ideally, $\mathcal{V}$ is close to 1. Not only is that unlikely in actual use, but the opposite tends to be true in high dimensions, as we discuss in this section. Unless the estimated mean μ and covariance Σ are chosen in special way, as described below and in Appendix A, under reasonable assumptions:

The ratio $\mathcal{V}$ tends to zero as the number of securities p grows to infinity.

In other words, the estimated volatility EV may be severely understated relative to the truth for a portfolio optimized from a large universe of securities. There is a rich literature on the cause of this phenomenon as discussed in the introduction.

We explore the volatility ratio of Markowitz portfolios when security returns follow a factor model,[5] the industry standard for a security return generating process.

Suppose that security returns in excess of the riskless rate are generated by the process

$$r = \boldsymbol{\beta} f + \epsilon, \tag{4}$$

where f denotes a random k-vector of returns to k risk factors, ϵ denotes a random p-vector of security specific returns, and $\boldsymbol{\beta}$ is an unknown non-random $(p \times k)$-matrix of sensitivities of the securities to the factors.[6] We observe only the p-vector r on the left side of (4). We seek estimates of the unknown parameter $\boldsymbol{\beta}$ and the covariance matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Delta}$ that form the nonrandom parameters of the latent components of return f and ϵ.

If we assume that the entries of ϵ are uncorrelated with f and pairwise uncorrelated with one another, the true covariance matrix $\boldsymbol{\Sigma}$ decomposes into a sum of factor and specific risk components (see Appendix B). Assuming the same factor-structure for the estimated model, for estimates (β, Φ, Δ) of the true parameters $(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Delta})$, we let

$$\Sigma = \beta \Phi \beta^\top + \Delta. \tag{5}$$

Squaring (3) and substituting the estimated and true parameters as well as the optimized portfolio w, we obtain

$$\mathcal{V}^2 = \frac{w^\top \beta \Phi \beta^\top w + w^\top \Delta w}{w^\top \beta \boldsymbol{\Phi} \boldsymbol{\beta}^\top w + w^\top \boldsymbol{\Delta} w}. \tag{6}$$

Under empirically reasonable assumptions, unless the estimate β is chosen so that the optimization bias, defined below, tends to zero (e.g., per the JSM recipe in the next section), all terms except for the factor component of the true variance, $w^\top \boldsymbol{\beta} \boldsymbol{\Phi} \boldsymbol{\beta}^\top w$, decay as $1/p$ or faster. In such cases, a first-order approximation of $\mathcal{V}$ obeys the proportionality,

$$\mathcal{V} \propto \frac{1}{\sqrt{p}\, \mathcal{M}_p(\beta) + 1} \tag{7}$$

where $\mathcal{M}_p(\beta)$ is bounded between zero and infinity, and is called the *optimization bias* (see Appendix A for further details and assumptions). This systematic bias is the key to understanding the accuracy of mean-variance optimized portfolios. We emphasize that (7) implies the volatility ratio $\mathcal{V}$ of the portfolio w decays to zero at rate $1/\sqrt{p}$. That is, as investors grow their portfolios (perhaps with the aim of diversifying), their risk estimates become less and less accurate. These portfolios mislead the investors into seeing much less risk on paper than there is in reality.

[5] An additional source of error in practice is mis-specification of the number k of factors. We do not address this source of error here.

[6] See Connor (1995) and Connor and Korajczyk (2010) for discussions of different factor model architectures used in finance.

The optimization bias $\mathcal{M}_p(\beta)$ has no dependence on the estimate Φ, and depends on β only through col(β), the column space of β, or equivalently, the span of the estimated factor exposures. The estimates μ and Δ do effect the value of $\mathcal{M}_p(\beta)$ but cannot prevent $\mathcal{V}$ from decaying to zero.[7] This turns out to imply, perhaps unexpectedly, that the optimization bias $\mathcal{M}_p(\beta)$ may be removed only by changing the estimate β. Errors in the column space of β can be understood in high dimensions due to the influence of the law of large numbers, in the following way. A high dimensional noise vector will be nearly orthogonal to any fixed set of reference directions. This noise may be found in an estimate col(β), constructed by PCA for example, in the form of an additive perturbation of the true column space col($\boldsymbol{\beta}$). Its effect in high dimensions is to therefore push col(β) away from the chosen reference directions. Shrinkage back toward those directions can undo this effect. A quantitative version of this idea is described in the JSM recipes below, which leverages the constraint vectors e and μ in (1) to construct the reference directions.

New work and work in progress[8] investigates the behaviour of $\mathcal{M}_p(\beta)$ for large p and identifies optimization biases caused by the interactions between quadratic programs and errors in estimated parameters. The elimination of these biases makes use of the high dimensional properties of random matrices to bring $\mathcal{V}$ closer to the ideal $\mathcal{V} = 1$. In the next section, we provide recipes for optimization bias-free inputs to mean-variance programs, yielding special estimates β and μ that send the optimization bias $\mathcal{M}_p(\beta)$ to zero as p tends to infinity. They are derived from an intricate use of mathematics and data science; the very tools Markowitz brought to finance in the 1950s.

3 Recipes for an Improved Volatility Ratio

We provide implementable recipes to estimate the means and covariances as inputs to the mean-variance optimization program (1). The first recipe adopts the "usual" estimators based on maximum likelihood (ML). Here, the estimate $\mu = (\mu_i)$ for the target return constraint in (1) is computed as the usual sample average. Similarly, a maximum likelihood estimator, in the form of principal component analysis,[9] is applied to produce a factor-model covariance estimate $\Sigma = (\sigma_{ij})$. These estimates will be denoted by μ_{ML} and Σ_{PCA}. The second recipe will leverage James-Stein shrinkage to improve the usual estimators. The estimate μ_{ML} of the mean security return is replaced by the well-known James-Stein estimator μ_{JS}, which improves upon the vector of sample averages when $p > 2$ (see Stein (1956), James and Stein (1961), Efron and Morris (1975) and Efron and Morris (1977)). The estimate Σ_{PCA} is replaced by a more recent Stein-type estimators that improve sample eigenvectors (i.e., principal component (factor) loadings – see Shkolnik (2022) and Goldberg and Kercheval (2023)). The new covariance estimate is denoted by Σ_{JSM}. The corresponding James-Stein-Markowitz estimator is

[7] We remark that when Δ is a scalar matrix the dependence of $\mathcal{M}_p(\beta)$ on it vanishes in a cancellation.

[8] Including Goldberg et al. (2020), Goldberg et al. (2022), Gurdogan and Kercheval (2022), Goldberg and Kercheval (2023), Goldberg et al. (2024), andGurdogan and Shkolnik (2024)

[9] e.g., Tipping and Bishop (1999) give an interpretation of PCA as maximum likelihood for factor analysis under suitable conditions.

designed specifically to remove optimization biases from the weights of portfolios constructed with PCA. The outputs of both recipes, first $(\mu_{\text{ML}}, \Sigma_{\text{PCA}})$ and second $(\mu_{\text{JS}}, \Sigma_{\text{JSM}})$, are used in (1) to compute the portfolio weights. We test the volatility ratio of these weights in the next section.

We begin with a $p \times n$ data matrix R of excess returns, the columns of which hold n observation of the left side of equation (4). Starting with only this ingredient, our recipes output the estimates of the means and covariances.

Recipe for the return constraint

1. Let $\bar{r}$ be the p-vector average of the n columns of R.
2. Let $\mu = \bar{r}$ for the target return constraint in (1).

Recipe for the covariance model

1. With $\bar{r}$ as above, let $\bar{R}$ be the $(p \times n)$ matrix with $\bar{r}$ in every column, to center the data, i.e,

$$Y = R - \bar{R}. \tag{8}$$

2. For the centered sample covariance matrix $S = YY^\top/n$, write its spectral decomposition as

$$S = \textstyle\sum_{(\jmath^2,h)} \jmath^2 hh^\top = HH^\top + N \tag{9}$$

where the sum is over all eigenvalue/eigenvector pairs $(\jmath^2, h)$ of S, H is a $p \times k$ matrix with every column of the form $\jmath h$ sourced from the k largest eigenvalues $\jmath^2$, and $N = S - HH^\top$.
3. The specific risk estimate Δ in (5) sets all the off-diagonal elements of N to zero, i.e.,

$$\Delta = \mathbf{diag}(N). \tag{10}$$

4. The PCA covariance matrix is $\Sigma_{\text{PCA}} = HH^\top + \Delta$.

Recipe 1 Principal component analysis (PCA) recipes for means and covariances.

Recipe 1 computes the sample average $\mu_{\text{ML}} = \bar{r}$ of the columns of R for the constraint μ and a PCA covariance matrix Σ_{PCA} given return data R as input. The estimate $\Sigma_{\text{PCA}} = HH^\top + \Delta$ may now be expressed, if desired, in terms of a factor model estimate $\beta\Phi\beta^\top + \Delta$, by finding (β, Φ) satisfying $\beta\Phi^{1/2} = H$ for a $(p \times k)$ matrix β and a $(k \times k)$ covariance matrix Φ of factor returns. However, the individual terms β and Φ are separately unidentifiable, so are not unique. However, there is no need to find separate estimates for β and Φ, because the optimal portfolio only depends on the sum of $HH^\top$ and Δ. Step 2 of Recipe 1 may be computationally expensive and we

Recipe for the return constraint

1. The James-Stein estimate μ_{JS} adjusts the sample mean $\bar{r}$ by a shrinkage parameter,

$$\mathrm{c} = 1 - \nu^2 J^{-1}, \qquad J = (\bar{r} - m)^\top (\bar{r} - m), \tag{11}$$

where ν^2 is the variance of noise and some p-vector $m \neq \bar{r}$, a shrinkage target.

2. Let $\mu = \mu_{\text{JS}}$ for the target return constraint in (1) be computed as,

$$\mu_{\text{JS}} = \bar{r}\,\mathrm{c} + m(1 - \mathrm{c}). \tag{12}$$

3. The noise variance ν^2, given the $S = HH^\top + N$ in the PCA recipe, is computed via

$$\nu^2 = \frac{\operatorname{trace}(N)}{n_+ - k}, \tag{13}$$

where n_+ is the number of nonzero eigenvalues of the sample covariance S.

4. The shrinkage target m may be any p-vector, but a popular choice is (the grand mean),

$$m = \mathrm{e}\,(\mathrm{e}^\top \mathrm{e})^{-1}\,\bar{r} = \mathrm{e}^\top \bar{r}/p. \tag{14}$$

Recipe 2 James-Stein-Markowitz (JSM) recipe for means.

defer to Appendix C for a faster procedure to compute the eigenvector matrix H. An improved estimate Δ is also stated there. An efficient computation of the Markowitz portfolio weights, given Σ_{PCA}, is described in Appendix A.

Recipe 2 uses the classic James-Stein estimator μ_{JS} to improve the sample average μ_{ML} in the sense of expected mean-square error when $p > 2$ with ν^2 known. For a discussion of James-Stein in the asymptotic case of p growing to infinity, see Casella and Hwang (1982). For these asymptotics, step 3 supplies a consistent estimate ν^2 of the variance of the noise in the vector of averages μ_{ML} under reasonable assumptions. The vector of ones used in step 4 implements a shrinkage toward the grand mean (the average of averages in $\mu_{\text{ML}} = \bar{r}$) as popularized by Efron and Morris. However, this estimate alone cannot correct for the optimization biases described in Section 2. To this end, the covariance estimate Σ_{JSM} applies shrinkage to the eigenvectors H (the principal component loadings) in the estimate Σ_{PCA} to address its decay in volatility ratio $\mathcal{V}$ per formula (7) as p grows. For Σ_{JSM}, the optimization bias $\mathcal{M}_p(\beta)$ for the Markowitz portfolio problem (1) tends to zero (see Appendix A for more detail). Our numerical results show that the rate of decay of $\mathcal{M}_p(\beta)$ is sufficiently fast so that $\mathcal{V}$ remains bounded above zero when Σ_{JSM} is used. The mitigation of the optimization bias is achieved by rotating the column space of $\bar{H} = \Delta^{1/2} H$ to obtain H_{JSM}, defined in (18). This bears a striking resemblance to classic James-Stein shrinkage formula (12). When Δ is a scalar matrix the re-weighting by $\Delta^{-1/2}$ in all steps of the recipe

RECIPE FOR THE COVARIANCE MODEL

1. For any estimate Δ (e.g., (10)), centering and weighting the data, we set

$$Y = \Delta^{-1/2}\big(R - \bar{R}\big) \tag{15}$$

where $\Delta^{-1/2}$ is diagonal with $\Delta_{ii}^{-1/2} = 1/\sqrt{\Delta_{ii}}$ and $\bar{R}$ is the matrix in (8).
2. Recompute H following (9) but from the re-weighted sample covariance S that uses (15). Set,

$$\bar{H} = \Delta^{1/2} H\,. \tag{16}$$

3. The JSM estimator of the weighted eigenvectors $\bar{H}$ computes a $(k \times k)$-matrix valued shrinkage parameter,

$$C = I - \nu^2 J^{-1}\,, \qquad J = (\bar{H} - M)^\top \Delta^{-1} (\bar{H} - M)\,, \tag{17}$$

where ν^2 is the variance of the noise and $M \neq \bar{H}$ is a $(p \times k)$-matrix shrinkage target.[a]
4. The JSM estimator is analogous to (12) but with matrix valued C and M.

$$H_{\text{JSM}} = \bar{H}C + M(I - C) \tag{18}$$

5. The variance ν^2 is computed per (13) but with N from the reweighted sample covariance S.
6. A shrinkage target M analogous to (14) uses a $(p \times 2)$-matrix $A = \begin{pmatrix} \mu & \mathrm{e} \end{pmatrix}$ as

$$M = A(A^\top \Delta^{-1} A)^{-1} A^\top \Delta^{-1} \bar{H}\,. \tag{19}$$

7. The basic JSM covariance model is $\Sigma_{\text{JSM}} = H_{\text{JSM}} H_{\text{JSM}}^\top + \Delta$.

[a] Here, $\neq$ is in the sense that the columns spaces of the two matrices are not identical.

Recipe 2 (continued) James-Stein-Markowitz (JSM) recipe for covariances.

may be omitted. The vector shrinkage target m used to shrink the sample mean $\bar{r}$ is replace by the matrix M to shrink the (re-weighted) sample eigenvectors $\bar{H}$. This M is constructed from the constraint vectors, $\mu = \mu_{\text{JS}}$ and e, in order to address the bias $\mathcal{M}_p(\beta)$ resulting from optimization (1). Our earlier comments regarding PCA and the factor model estimate $\beta\Phi\beta^\top + \Delta$ equally apply to the estimate Σ_{JSM}.

Prior work that applies James-Stein ideas to covariance estimation and optimized portfolio construction includes Jobson and Korkie (1981) and Jorion (1986) as well as Ledoit and Wolf (2003) and Ledoit and Wolf (2004). The novelty of the JSM recipes, however, lies in the application of these ideas directly to the principal component loadings and the accompanying theoretical analysis of the optimization bias $\mathcal{M}_p(\beta)$ and the volatility ratio $\mathcal{V}$ in formula (7).

4 Numerical Illustration

We look at risk, excess return and Sharpe ratio of Markowitz portfolios optimized with PCA and JSM, when data are generated by a seven-factor instance of the return generating process (4). The model is based on excess return to the market, two style factors and four industry factors. A specification of the return generating process is in Appendix B. The shrinkage target for the JSM return constraint is the grand mean.

In 400 fictional universes, we simulate six months of daily data, $n = 125$ observations, of p security returns that follow the generating process detailed in Appendix B. From each data set, we construct PCA and JSM Markowitz portfolios from (1) with target annualized expected return 8.5%. Recipes for the construction of PCA and JSM portfolios are in Section 3. We consider universes of size p ranging between 500 and 3,000 to shed light on problems commonly encountered in practice, and, we include the unrealistic value $p = 100,000$ to highlight asymptotic effects.

p	OPT	PCA TV	PCA EV	$\mathcal{V}$	JSM TV	JSM EV	$\mathcal{V}$
500	6.33	8.66	6.71	0.77	7.93	10.0	1.26
1000	4.73	7.78	4.71	0.61	6.10	7.17	1.17
2000	3.34	6.89	3.31	0.48	4.43	4.96	1.12
3000	2.78	6.68	2.68	0.40	3.80	4.12	1.09
100000	0.50	6.29	0.47	0.07	0.81	0.83	1.02

Table 1 Portfolio Volatility (annual %) and Volatility Ratio. OPT: True volatility of Markowitz portfolios optimized with true parameters. TV: Average true volatility of estimated Markowitz portfolios. EV: Average estimated volatility of estimated Markowitz portfolios. $\mathcal{V}$: Average volatility ratio of estimated Markowitz portfolios. $n = 125$, $\mu = 8.5$, 400 simulations.

True and estimated volatility of Markowitz portfolios are shown in Table 1. The OPT column shows oracle values: the true volatility of the true Markowitz portfolio optimized with true means and covariance matrix for each p. Under empirically sound assumptions about the calibration of the return generating process for large p, theory predicts these values tend to zero as p tends to infinity, since factor return tends to be hedged and specific return tends to diversify away. This limiting behavior is suggested by the volatility of 0.50% for the optimal Markowitz portfolio estimated from a universe of $p = 100,000$ securities. The oracle values serve as a benchmark against which we can assess portfolios optimized with estimated parameters.

The remainder of Table 1 concerns true and estimated volatility of Markowitz portfolios optimized with PCA and JSM estimates. For $p = 500$, PCA and JSM Markowitz portfolios have similar average true volatilities of 8.66% and 7.93%. Unlike JSM, however, the estimated volatility PCA is underforecast at an average of 6.71%, leading to an average $\mathcal{V} = 0.77$. An asset manager sees a $p = 500$ Markowitz portfolio estimated with PCA as 23% less risky than it is, and less risky than the JSM analog whose volatility is overforecast. As p grows, the average true volatility of the Markowitz portfolio estimated with JSM diminishes toward 0 as it does for the oracle, but the volatility of the PCA Markowitz portfolio does not. The distortion of PCA plummets as p, while that of JSM gets closer to 1.

p	OPT	PCA TR	PCA ER	$\mathcal{D}$	JSM TR	JSM ER	$\mathcal{D}$
500	8.5	6.96	8.50	1.23	7.56	8.51	1.14
1000	8.5	7.04	8.50	1.21	7.59	8.51	1.13
2000	8.5	7.03	8.50	1.21	7.63	8.51	1.12
3000	8.5	6.92	8.50	1.23	7.51	8.51	1.14
100000	8.5	6.93	8.50	1.23	7.27	8.50	1.17

Table 2 Portfolio Expected Excess Return (annual %) and Distortion: OPT: True returns of Markowitz portfolios optimized with true parameters. TR: Average true return of estimated Markowitz portfolios. ER: Average estimated return of estimated Markowitz portfolios. $\mathcal{D}$: Average ratios of estimated to true return for estimated Markowitz portfolios. $n = 125$, $\mu = 8.5$, 400 simulations.

True and estimated expected excess returns of optimized Markowitz portfolios are shown in Table 2. For all values of p, estimated expected returns of PCA and JSM are 8.5% because the optimizer targets that value with estimated security returns. These estimates are equal to the expected excess return of the oracle, noted in OPT, since they are made with true returns. The unobservable truth is higher for JSM than PCA due to the JSM return constraint detailed in Section 3. This means that average distortion $\mathcal{D} = \text{ER}/\text{TR}$ is greater for PCA than for JSM.

p	OPT	PCA TSR	PCA ESR	$\mathcal{D}$	JSM TSR	JSM ESR	$\mathcal{D}$
500	1.34	0.81	1.72	2.13	0.95	1.12	1.17
1000	1.8	0.91	2.35	2.59	1.24	1.52	1.22
2000	2.54	1.03	3.33	3.25	1.73	2.21	1.28
3000	3.06	1.04	4.08	3.91	1.98	2.62	1.32
100000	16.98	1.11	22.99	20.71	8.97	12.78	1.42

Table 3 Portfolio Sharpe Ratio and Distortion: OPT: True Sharpe Ratio of Markowitz portfolios optimized with true parameters. TSR: Average true Sharpe Ratio of estimated Markowitz portfolios. ESR: Average estimated Sharpe ratios of estimated Markowitz portfolios optimized with PCA or JSM. $\mathcal{D}$: Average ratio of estimated to true Sharpe Ratio for estimated Markowitz portfolios. $n = 125$, $\mu = 8.5$, 400 simulations.

Risk-adjusted expected excess returns or *Sharpe ratios* for Markowitz portfolios are shown in Table 3. Estimated Sharpe ratio exceeds true Sharpe ratio on average for Markowitz portfolios optimized with PCA and JSM for all values of p considered, but is close to 1 for the latter. The Sharpe ratio distortion $\mathcal{D} = \text{ESR}/\text{TSR}$ explodes for PCA as p grows. There are two sources of the discrepancy between the relatively tame distortion for JSM and the explosion one observes with PCA. The first is that return distortion is greater for PCA than for JSM, as shown in Table 2. The second, more potent source is the plummeting volatility ratio of PCA volatility estimates, shown in Table 1.

For an asset manager tracking their risk-adjusted excess return, disappointment in our fictional universes is rife when Markowitz portfolios are constructed with PCA.

5 Harry Markowitz Was a Statistician

Prompted by questions posed in Markowitz (1952) we apply tools from high dimensional statistics to estimate inputs appropriate for use in mean-variance optimization. We illustrate how shrinkage techniques used in the James-Stein-Markowitiz recipe for means and covariance matrices correct optimization biases that, left unchecked, corrupt optimized quantities. Markowitz looked holistically at problems in a way that allowed theoreticians to build on his work and practitioners to use it. He explored widely outside his fields of expertise. This may help explain why he was so effective at solving big problems that require deep understanding of many subjects.

Harry fans sometimes ask whether their hero was an economist, a computer scientist or a mathematician. Let's add "statistician" to the list. His early inquiries about the importance of risk in portfolio selection and the suitability of classical statistics for estimating inputs to mean-variance optimization launched vast bodies of research. His late-in-life crusade to clarify the assumptions on data required for a mean-variance optimized portfolio to be the best choice is ongoing. As we strive to develop better inputs to optimization, we are inspired by Markowitz's stubborn insistence on getting the right answer.

Acknowledgements. We are grateful to John Blin, Kay Giesecke, Nick Gunther, Ron Kahn, Ananth Madhavan and Alex Ulitsky for insightful comments on an early draft of this article.

A Factor Models & Markowitz Portfolios

Provided the $(p \times p)$ covariance Σ is invertible, the optimized (mean-variance) portfolio solving (1) has the form,

$$w = \gamma_{\mathrm{e}} \Sigma^{-1} \mathrm{e} + \gamma_{\mu} \Sigma^{-1} \mu \tag{20}$$

a combination of a global minimum variance portfolio and the characteristic portfolio of μ, weighted by their shadow prices γ_{e} and γ_{μ}.[10] Letting $\phi(x, y) = x^{\top} \Sigma^{-1} y$, these two portfolios are given by $\Sigma^{-1}\mathrm{e}/\phi(\mathrm{e}, \mathrm{e})$ and $\Sigma^{-1}\mu/\,\phi(\mu, \mu)$.

When the return to the global minimum variance portfolio, $\mu^{\top}\Sigma^{-1}\mathrm{e}/\phi(\mathrm{e}, \mathrm{e})$, exceeds the target return α,

$$\begin{aligned} \gamma_{\mathrm{e}} &= \frac{1}{\phi(\mathrm{e}, \mathrm{e})} \\ \gamma_{\mu} &= 0 \end{aligned} \qquad \Big(\frac{\phi(\mu, \mathrm{e})}{\phi(\mathrm{e}, \mathrm{e})} \geq \alpha \Big). \tag{21}$$

[10] Shadow prices are the values of the Lagrange multipliers of the constrained optimization problem at its optimal solution. See Grinold and Kahn (1999) for an exposition of the algebra of efficient frontiers.

Otherwise, the shadow prices are expressed in terms of all the inputs to (1) as follows.

$$\begin{aligned} \gamma_{\mathrm{e}} &= \frac{\phi(\mu,\mu) - \alpha\,\phi(\mathrm{e},\mu)}{\phi(\mathrm{e},\mathrm{e})\phi(\mu,\mu) - \phi(\mathrm{e},\mu)^2} \\ \gamma_{\mu} &= \frac{\alpha\,\phi(\mathrm{e},\mathrm{e}) - \phi(\mu,\mathrm{e})}{\phi(\mathrm{e},\mathrm{e})\,\phi(\mu,\mu) - \phi(\mathrm{e},\mu)^2} \end{aligned} \qquad \left(\frac{\phi(\mu,\mathrm{e})}{\phi(\mathrm{e},\mathrm{e})} < \alpha\right). \tag{22}$$

Formula (20) is costly to evaluate when the matrix Σ is very large. But given a factor-structure, (20) may be computed efficiently via the Woodbury matrix identity. This is accomplished by adopting a factor model for the security returns; a choice with additional advantages (e.g., reducing the number of estimated parameters).

The excess return generating process $r = \boldsymbol{\beta} f + \epsilon$ introduced in (4) is expressed in terms of factor returns $f \in \mathbb{R}^k$ specific returns $\epsilon \in \mathbb{R}^p$, and (true) security sensitivities to factors $\boldsymbol{\beta} \in \mathbb{R}^{p\times q}$. In a factor model, the entries of ϵ are uncorrelated with factor returns f and pairwise uncorrelated with one another. Then, the covariance of r is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}^\top + \boldsymbol{\Delta} \tag{23}$$

where $\boldsymbol{\Phi}$ is the covariance matrix of the factor return f and $\boldsymbol{\Delta}$ is a covariance matrix of the specific return ϵ. The matrix $\boldsymbol{\Delta}$ is further assumed to be diagonal and invertible. Taking the expectation of r, we obtain the decomposition

$$\boldsymbol{\mu} = \boldsymbol{\beta}\boldsymbol{\mu_f} + \boldsymbol{\mu_s}\,, \tag{24}$$

of the (true) expected security returns $\boldsymbol{\mu}$, where $\boldsymbol{\mu_f} \in \mathbb{R}^k$ and $\boldsymbol{\mu_s} \in \mathbb{R}^p$ are the expected factor and specific returns.

The estimated covariance matrix Σ discards the bold lettering notation, and is written as

$$\Sigma = \beta\Phi\beta^\top + \Delta \tag{25}$$

where the estimates (β, Φ, Δ) have the same dimensions and properties as the true (population) parameters $(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Delta})$. Implicit in this is the knowledge of the true number of factors k, which we assume is granted. Our assumptions (in particular the invertibility of Δ) ensure Σ^{-1} exists and hence the Markowitz portfolio in (20) is well-defined.

Given the factor-structure in (25) and Φ invertible, Σ^{-1} is computed efficiently via the Woodbury identity as,

$$\Sigma^{-1} = \Delta^{-1} - \Delta^{-1}\beta\,(\Phi^{-1} + Q)^{-1}\beta^\top\Delta^{-1}, \qquad Q = \beta^\top\Delta^{-1}\beta\,, \tag{26}$$

by leveraging the fact that the matrix $(\Phi^{-1}+Q)$ is a $(k\times k)$ matrix (with the number of factors k being small or moderate) may be efficiently inverted (or, used in computations that perform this inversion implicitly).

We analyze the asymptotic behavior of the volatility ratio $\mathcal{V}$ in (3) of the estimated mean-variance portfolio in (20). The following conditions on $(\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{\Delta})$, their estimates and the estimate μ of $\boldsymbol{\mu}$ in (24) are required for our results.

Regularity conditions. *$\beta^\top \Delta^{-1}\beta/p$ and $\boldsymbol{\beta}^\top \Delta^{-1}\boldsymbol{\beta}/p$ (the estimate Δ of $\boldsymbol{\Delta}$ used for both) converge to invertible (limit) $k \times k$ matrices, and the diagonal entries of Δ and $\boldsymbol{\Delta}$ remain bounded in $(0, \infty)$ as p tends to infinity. The matrix $\boldsymbol{\Phi}$ and its estimate Φ do not depend on p and are both invertible. The vector μ does not vanish,* e *and μ are not collinear and neither* e *nor μ belongs to the column space of $\Delta^{-1/2}\boldsymbol{\beta}$ in the limit where p diverges.*[11]

With these conditions in place, the sequence of portfolios (20), may be shown[12] to have $w^\top \boldsymbol{\Delta} w$, $w^\top \Delta w$ and $w^\top \beta \Phi \beta^\top w$ decaying to zero at the rate $1/p$ as p grows, and (6) may be simplified as follows.[13]

$$\mathcal{V}^2 = \frac{(\mathrm{EV})^2}{(\mathrm{TV})^2} = \frac{w^\top \beta \Phi \beta^\top w + w^\top \Delta w}{w^\top \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}^\top w + w^\top \boldsymbol{\Delta} w} = \left(\frac{C_p}{p}\right) \frac{1}{w^\top \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}^\top w} \tag{27}$$

for some sequence C_p that is bounded in $(0, \infty)$. To analyze the remaining term $w^\top \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}^\top w$, we define

$$\psi(\beta, \zeta) = \Delta^{-1}(\zeta - \beta Q^{-1}\beta^\top \Delta^{-1}\zeta) \qquad\qquad (\zeta \in \mathbb{R}^p)$$

and, using (26), it may be shown that our characteristic portfolios satisfy $\frac{\Sigma^{-1}\zeta}{\phi(\zeta,\zeta)} = \frac{\psi(\beta,\zeta)}{\zeta^\top \psi(\beta,\zeta)} + c_p/p$ for some sequence c_p bounded in $\mathbb{R}$, taking $\zeta = \mathrm{e}$ for the global minimum variance or $\zeta = \mu$. Moreover, the quantities

$$\frac{\boldsymbol{\beta}^\top \psi(\beta, \mathrm{e})}{\mathrm{e}^\top \psi(\beta, \mathrm{e})} \qquad \text{and} \qquad \frac{\boldsymbol{\beta}^\top \psi(\beta, \mu)}{\mu^\top \psi(\beta, \mu)} \tag{28}$$

are vectors in $\mathbb{R}^k$ that remain bounded as p grows. These quantities are "optimization biases" that result from mismatches between β and $\boldsymbol{\beta}$ and their interplay with the optimization constraint vectors e and μ.

Now, for $|\cdot|$ the Euclidean norm and an o_p that vanishes as p tends to infinity,

$$w^\top \boldsymbol{\beta}^\top \boldsymbol{\Phi}\boldsymbol{\beta}^\top w = |\boldsymbol{\Phi}^{1/2}\boldsymbol{\beta}^\top w| = |\boldsymbol{\Phi}^{1/2}(\gamma_{\mathrm{e}}\boldsymbol{\beta}^\top \psi(\mathrm{e}) + \gamma_\mu \boldsymbol{\beta}^\top \psi(\mu))| + o_p\,.$$

Under our conditions, $\Gamma_{\mathrm{e}} = \boldsymbol{\Phi}^{1/2}\gamma_{\mathrm{e}}\mathrm{e}^\top \psi(\beta, \mathrm{e})$ and $\Gamma_\mu = \boldsymbol{\Phi}^{1/2}\gamma_\mu \mu^\top \psi(\beta, \mu)$ are both bounded in $\mathbb{R}^k$ and the former is bounded away from zero irrespective of the cases, (21) or (22). We arrive at $\mathcal{M}_p(\beta, \mu)$ in (7) in the form,

$$\mathcal{M}_p(\beta, \mu) = \left|\Gamma_{\mathrm{e}} \frac{\boldsymbol{\beta}^\top \psi(\beta, \mathrm{e})}{\mathrm{e}^\top \psi(\beta, \mathrm{e})} + \Gamma_\mu \frac{\boldsymbol{\beta}^\top \psi(\beta, \mu)}{\mu^\top \psi(\beta, \mu)}\right| \tag{29}$$

---

[11] The matrix $\Delta^{-1/2}$ is diagonal with $\Delta^{-1/2} = 1/\sqrt{\Delta_{ii}}$

[12] These calculations are analogous to Theorem 2.3 and Lemma A.1 of Gurdogan and Shkolnik (2024).

[13] Here, we use that $\frac{1}{x+\delta} = \frac{1}{x}\big(1 - \frac{\delta/x}{1+\delta/x}\big)$.

where we omit the constant argument e and leave (β, μ) as the pair of estimates that are "levers" that may be adjusted to set the optimization biases to zero. The estimates Φ and Δ do not have this capacity on their own.

The term $\frac{\boldsymbol{\beta}^\top \psi(\beta,\zeta)}{\zeta^\top \psi(\beta,\zeta)}$ which appears in (29) with $\zeta = \mathrm{e}$ and $\zeta = \mu$, is studied in Gurdogan and Shkolnik (2024) in terms of what they call the "*quadratic optimization bias*". Each such bias is a k-vector with each component corresponding to a risk factor in the covariance model. Each may be set to zero by an orthogonal projection of the column space of the $(p \times (k+1))$-matrix $(\beta\ \zeta)$ onto the column space of the true factor loadings $\boldsymbol{\beta}$. Theorem 5.1 of that paper states the conditions under which this orthogonal projection may be estimated from the observed data, and supplies the estimator that accomplishes this in the setting of PCA, so that $\beta = H$ per Recipe 1. Recipe 2 is an extension that estimator that combines the vectors $\zeta = \mathrm{e}$ and $\zeta = \mu$ in an orthogonal projection in (19). The resulting estimator may then be put into the James-Stein form H_{JSM} in (18). These theoretical results establish that $\mathcal{M}_p(H_{\text{JSM}})$ tends to zero almost surely as p diverges provided the estimate Δ is either a scalar matrix or independent of the data. The numerical results of Section 4 suggest the dependent case of (10) has the same property. It is also remarkable that the proofs can be modified for $\zeta = \mu$, a sample mean (or the corresponding James-Stein estimator μ_{JS} of Recipe 2) that depends on the observed data.

B Security Return Generating Process

We specify a seven-factor instance of the excess return generating process $r = \boldsymbol{\beta} f + \epsilon$, introduced in (4), in terms of the true mean $\boldsymbol{\mu}$ and true covariance $\boldsymbol{\Sigma}$. The seven factors include excess return to the market, two styles, which we might think of as size and value, and membership in four industries. The dimension dependent components $(\boldsymbol{\beta}, \boldsymbol{\Delta})$ of $\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}^\top + \boldsymbol{\Delta}$ are generated for the largest value of p (i.e., $100,000$) used in the numerical results first, and subsets of these are taken to produce returns for a smaller number of securities.

Market	Size	Value	Industry 1	Industry 2	Industry 3	Industry 4
16.0	4.0	2.0	20.0	15.0	10.0	5.0

Table 4 Volatilities of the factor returns $f = (f_1, \ldots, f_k)$ in percent annualized.

The (7×7)-covariance matrix $\boldsymbol{\Phi}$ of the factor returns $f = (f_1, \ldots, f_7)$ is specified in terms of the factor volatilities and their correlations. The factor volatilities are calibrated as in Bayraktar et al. (2014, Table J4) and are presented in Table 4. Table 5 presents the correlations between the seven factor returns. The style and market factor correlations are taken from Fama and French (2015, Table 4). The correlations between the industries and the market are relatively small, since we think of industry factors as residual to the market as in Menchero et al. (2011). The remaining correlations are set to zero.

Market	Size	Value	Industry 1	Industry 2	Industry 3	Industry 4
1.00	0.28	−0.30	0.16	0.08	0.04	0.02
	1.00	−0.11	0.00	0.00	0.00	0.00
		1.00	0.00	0.00	0.00	0.00
			1.00	0.00	0.00	0.00
				1.00	0.00	0.00
					1.00	0.00
						1.00

Table 5 Correlation matrix of the factor returns $f = (f_1, \ldots, f_k)$. Blank entries omitted due to symmetry.

The $(p \times 7)$-matrix of $\boldsymbol{\beta}$ sensitivities to factors is summarized in Figure 1. Its left panel shows histograms of the first three columns of $\boldsymbol{\beta}$, the entries of which are drawn independently from $N(1, 0.25^2)$, $N(0, 1)$ and $N(0, 0.5^2)$ respectively. The industry factor sensitivities are generated as follows. Each security selects two (of four) industries for membership (with replacement). Independently generating two numbers uniformly in $(0, 1)$, we assign each as a sensitivity to the two industries. If only one industry was selected, the sensitivity equals their sum. An illustration of common memberships to industries for each pair of securities is illustrated in the right panel of Figure 1.

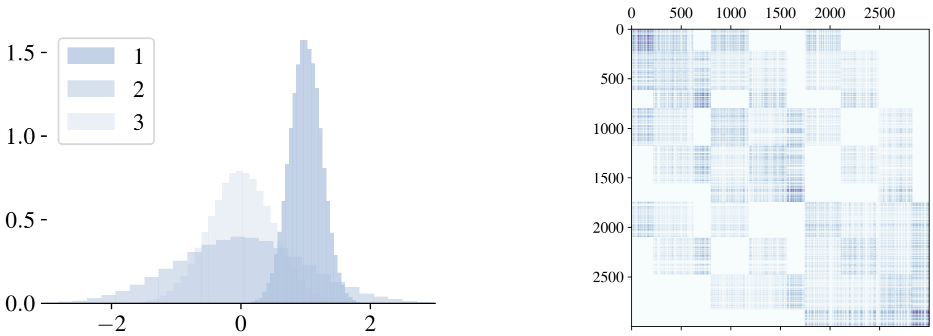

Figure 1 **Left panel:** Histogram of the first three columns of $\boldsymbol{\beta}$ (market and two style factor sensitivities). **Right panel:** Industry membership visualization (i.e. entries of a matrix $\sum_c cc^\top$ where the sum is over the last four (industry) columns of $\boldsymbol{\beta}$ – white entries indicate no industry in common between two securities.

The square roots of the diagonal entries of $\boldsymbol{\Delta}$, the specific volatilities, are drawn from $15 + 100 \times \text{Beta}(4, 16)$, and they range from 15% to 77% (annualized). See third panel of Figure 2 for illustration.

To calibrate the expected returns $\boldsymbol{\mu} = \boldsymbol{\beta}\boldsymbol{\mu_f} + \boldsymbol{\mu_s}$ per (24), we rely on (Fama and French, 2015, Table 4) for guidance on $\boldsymbol{\mu_f}$ and following Ang (2023), we set the expected returns on industry factors to be zero. See Table 6.

Market	Size	Value	Industry 1	Industry 2	Industry 3	Industry 4
4.80	2.40	1.20	0.00	0.00	0.00	0.00

Table 6 Expectations $\boldsymbol{\mu_f}$ of the factor returns $f = (f_1, \ldots, f_k)$ in percent annualized.

The expected specific returns are obtained by the projection,

$$\boldsymbol{\mu_s} = \frac{0.5}{100}\big(\boldsymbol{\Delta} - \boldsymbol{\beta}\boldsymbol{\beta}^{+}\boldsymbol{\Delta}\big), \tag{30}$$

where $\boldsymbol{\beta}^{+}$ is the pseudo-inverse of $\boldsymbol{\Delta}$.[14] This results in a vector $\boldsymbol{\mu_s}$ orthogonal to the risk factor exposures (i.e., the columns of $\boldsymbol{\beta}$), and such that securities with a higher specific volatility have higher returns on average. In this way, $\boldsymbol{\mu}$ decomposes into a factor return component $\boldsymbol{\beta}\boldsymbol{\mu_f}$ and a specific return $\boldsymbol{\mu_s}$ which are orthogonal. Scatter plots of the expected returns against the volatilities for each component of return and the sum are shown in Figure 2.

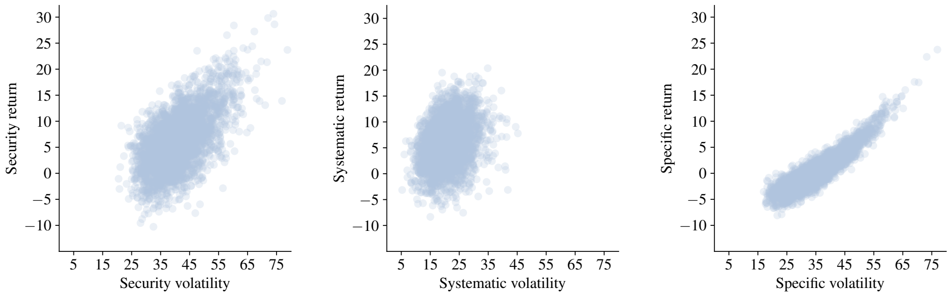

Figure 2 Scatter plots of various components of return versus volatilities for a representative sample of $p = 3000$ securities. **Left panel:** Total return ($\boldsymbol{\mu}$) vs. total volatility (square-roots of the diagonal entries of $\boldsymbol{\Sigma}$). **Center panel:** Systematic security return ($\boldsymbol{\beta}\boldsymbol{\mu_f}$) versus systematic risks (square-roots of the diagonal entries of $\boldsymbol{\beta}\boldsymbol{\Phi}\boldsymbol{\beta}^{\top}$). **Right panel:** Specific returns ($\boldsymbol{\mu_s}$) versus specific risks (square-roots of the diagonal entries of $\boldsymbol{\Delta}$).

Lastly, returns to factors f are drawn from a normal distribution with mean $\boldsymbol{\mu_f}$ and covariance matrix $\boldsymbol{\Phi}$. The specific returns ϵ are uncorrelated with factor returns, and the components of ϵ are drawn from a joint normal distribution[15] with mean $\boldsymbol{\mu_s}$ and covariance matrix $\boldsymbol{\Delta}$. The returns are generated identically and independently over the n dates keeping all model parameters $(\boldsymbol{\mu_f}, \boldsymbol{\mu_s}, \boldsymbol{\Phi}, \boldsymbol{\beta}, \boldsymbol{\Delta})$ fixed.

[14] A discussion of factor premia versus specific return alpha in the context of multiple managers is in Garvey et al. (2017). Non-zero specific return alpha is inconsistent with the no-arbitrage conclusion in Ross (1976).

[15] Normality of f and ϵ is not required for the shrinkage methods described in this article to be effective.

C Technical Supplement on the Recipes

Recipe 1 addresses the unidentifiability issues of factor analysis by computing orthogonal factors (i.e., principal component loadings). In the PCA covariance model recipe we note that $H^\top H = \mathcal{S}^2$, the $(k \times k)$ diagonal matrix of the largest k sample eigenvalues. To put $HH^\top$ into the form $\beta \Phi \beta^\top$ as required by the estimate (25), we set,

$$\Phi = H^\top H = \mathcal{S}^2 \tag{31}$$

and then the columns of β become orthonormal principal component loadings (i.e. $\beta^\top \beta$ is a $(k \times k)$ identity matrix).

Moreover, in finite sample regime in which the dimension p tends to infinity, we may perform a bias correction on the eigenvalues in (31) and formula (13). To improve the estimate ν^2 in (13) for $S = HH^\top + N$, we may take,

$$\nu^2 = \frac{\text{trace}\,(N)}{n_+ - (1 + n_+/p)\, k}\,. \tag{32}$$

See page 1355 of Wang and Fan (2017) for this estimator. It may be used to improve the estimate (10) by renormalizing its average to that of ν^2.

Letting $\mathcal{S}^{-2}$ be the inverse of $\mathcal{S}^2$, we may compute $\Psi^2 = I - \nu^2 \mathcal{S}^{-2}$ to improve (31) further by computing,

$$\Phi = \mathcal{S}^2\, \Psi^2\,. \tag{33}$$

This adjusts the diagonal entries of $\mathcal{S}^2$ downward to reduce their biases. These biases do not impact $\mathcal{M}_p(\beta)$ in (7) however. The computation of these eigenvalues/eigenvectors may be challenging for large p. Instead, letting

$$L = Y^\top Y/p$$

for Y in either (8) or (15), we compute the $(n \times k)$ matrix of eigenvectors W corresponding to the k largest eigenvalues of the $(n \times n)$-matrix L. Then, $H = YV/\sqrt{n}$ which is highly efficient for n much smaller than p.

Recipe 2 which computes Σ_{JSM} can also follow the orthogonality conventions of the PCA estimate. Here, compute β_{JSM} by taking for its columns the k eigenvectors of the matrix H_{JSM} from the JSM covariance model recipe. Then.

$$\Sigma_{\text{JSM}} = \beta_{\text{JSM}} \Phi \beta_{\text{JSM}}^\top + \Delta \tag{34}$$

with Φ in (33) is a covariance model that has improved factor variances. We use these improved PCA and JSM models in the numerical results of Section 4. They lead to improved results relative to the plain recipes of Section 3.

Lastly, the computation of the $(k \times k)$ shrinkage parameters matrix C in the JSM recipe should not invert the matrix J in (17) numerically. Instead, we apply the

Woodbury identity (as in (26)) to the right side of,

$$J = (\bar{H} - M)^\top \Delta^{-1} (\bar{H} - M) = \mathcal{S}^2 - \bar{H}^\top \Delta^{-1} M\,.$$

References

Ang, A.: A Sharper Lens—Factors & Sectors. https://www.blackrock.com/us/individual/investment-ideas/what-is-factor-investing/factor-commentary/andrews-angle/factors-and-sectors (2023)

Bun, J., Bouchaud, J.-P., Potters, M.: Cleaning large correlation matrices: tools from random matrix theory. Physics Reports **666**, 1–109 (2017)

Blanchet, J., Chen, L., Zhou, X.Y.: Distributionally robust mean-variance portfolio selection with wasserstein distances. Management Science **68**(9), 6382–6410 (2022)

Best, M.J., Grauer, R.R.: On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. The review of financial studies **4**(2), 315–342 (1991)

Blin, J., Guerard, J., Mark, A.: A History of Commercially Available Risk Models, pp. 1–39. Springer, New York, NY (2022)

Boyd, S., Johansson, K., Kahn, R., Schiele, P., Schmelzer, T.: Markowitz portfolio construction at seventy. The Journal of Portfolio Management **80**(8), 117–160 (2024)

Bailey, D.H., López de Prado, M.: An open-source implementation of the critical-line algorithm for portfolio optimization. algorithms **6**, 169–196 (2013)

Bayraktar, M.K., Mashtaler, I., Meng, N., Radchenko, S.: Barra US Total Market Equity Model for Long-Term Investors (2014)

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

Bai, J., Ng, S.: Approximate factor models with weaker loadings. Journal of Econometrics **235**(2), 1893–1916 (2023)

Bodnar, T., Okhrin, Y., Parolya, N.: Optimal shrinkage-based portfolio selection in high dimensions. Journal of Business & Economic Statistics **41**(1), 140–156 (2022)

Casella, G., Hwang, J.T.: Limit expressions for the risk of james-stein estimators. Canadian Journal of Statistics **10**(4), 305–309 (1982)

Cottle, R.W., Infanger, G.: Harry Markowitz and the Early History of Quadratic Programming. Springer, New York, NY (2010)

Connor, G., Korajczyk, R.A.: Performance measurement with the arbitrage pricing theory. Journal of Financial Economics **15**, 373–394 (1986)

Connor, G., Korajczyk, R.: Factor Models. Wiley, Hoboken, NJ (2010)

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

Chamberlain, G., Rothschild, M.: Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica: Journal of the Econometric Society, 1281–1304 (1983)

Efron, B.: Controversies in the foundations of statistics. The American Mathematical Monthly **85**(4), 231–246 (1978)

Elton, E.J., Gruber, M.J., Brown, S.J., Goetzmann, W.N.: Modern Portfolio Theory and Investment Analysis. John Wiley & Sons, Hoboken, NJ (2009)

Efron, B., Morris, C.: Data analysis using stein's estimator and its generalizations. Journal of the American Statistical Association **70**(350), 311–319 (1975)

Efron, B., Morris, C.: Stein's paradox in statistics. Scientific American **236**(5), 119–127 (1977)

Fama, E.F., French, K.R.: A five-factor asset pricing model. Review of Financial Studies **116**, 1–22 (2015)

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

Fan, J., Masini, R.P., Medeiros, M.C.: Bridging factor and sparse models. The Annals of Statistics **51**(4), 1692–1717 (2023)

Fan, J., Zhang, J., Yu, K.: Vast portfolio selection with gross-exposure constraints. Journal of the American Statistical Association **107**(498), 592–606 (2012)

Goldberg, L.R., Gurdogan, H., Kercheval, A.: Portfolio optimization via strategy-specific eigenvector shrinkage (2024)

Grinold, R.C., Kahn, R.N.: Active Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk, 2nd edn. McGraw Hill, New York, NY (1999)

Gurdogan, H., Kercheval, A.: Multi anchor point shrinkage for the sample covariance matrix. SIAM Journal on Financial Mathematics **13**(3), 1112–1143 (2022)

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

Garvey, G., Kahn, R.N., Savi, R.: The dangers of diversification: Managing multiple manager portfolios. The Journal of Portfolio Management **43**(2), 13–23 (2017)

Goldberg, L.R., Papanicolau, A., Shkolnik, A.: The dispersion bias. SIAM Journal of Financial Mathematics **13**(2), 521–550 (2022)

Goldberg, L.R., Papacinicolau, A., Shkolnik, A., Ulucam, S.: Better betas. The Journal of Portfolio Management **47**(1), 119–136 (2020)

Gurdogan, H., Shkolnik, A.: The Quadratic Optimization Bias of Large Covariance Matrices. Working paper (2024)

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

Jobson, J.D., Korkie, B.M.: Performance hypothesis testing with the sharpe and treynor measures. The Journal of Finance **36**(4), 889–908 (1981)

Jobson, D., Korkie, R., Ratti, V.: Improved estimation for markowitz portfolios using james-stein estimators, 1979 (1979)

Jorion, P.: Bayes-stein estimation for portfolio analysis. The Journal of Financial and Quantitative Analysis **21**(3), 279–292 (1986)

James, W., Stein, C.: Estimation with quadratic loss. In: Proc. Fourth Berkeley Symp. Math. Stat. Prob., pp. 361–397 (1961)

Klein, R.W., Bawa, V.S.: The effect of estimation risk on optimal portfolio choice. Journal of Financial Economics **3**, 215–231 (1976)

Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock retursn with an application to portfolio selection. J. of Empirical Finance **10**, 603–621 (2003)

Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis **88**, 365–411 (2004)

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

Lai, T.L., Xing, H., Chen, Z., *et al.*: Mean–variance portfolio optimization when means and covariances are unknown. The Annals of Applied Statistics **5**(2A), 798–823 (2011)

MacKenzie, D.: An Engine, Not a Camera: How Financial Models Shape Marketrs.

The MIT Press, Cambridge, MA (2006)

Markowitz, H.: Portfolio selection. The Journal of Finance **7**(1), 77–92 (1952)

Markowitz, H.M.: The optimization of a quadratic function subject to linear constraints. Naval Research Logistics Quarterly **III**, 77–91 (1956)

Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments. Yale University Press, ??? (1959)

Markowitz, H.M.: God, ants and thomas bayes. The American Economist **55**(2), 5–13 (2010)

Markowitz, H.M.: Topics in applied investment management: From a bayesian viewpoint. The Journal of Investing **21**(1), 7–13 (2012)

Michaud, R.O.: The Markowitz optimization enigma: Is 'optimized' optimal? Financial Analysts Journal **45**(1), 31–43 (1989)

Michaud, R.O., Ma, T.: Efficient asset management: a practical guide to stock portfolio optimization and asset allocation. Oxford University Press (2001)

Menchero, J., Orr, D.t.J., Wang, J.: The Barra US Equity Model (USE4): Methodology notes (2011)

Marcenko, V., Pastur, L.: Distribution of eigenvalues for some sets of random matrices. Matematicheskii Sbornik **114**(4), 507–536 (1967)

Markowitz, H.M., Usmen, N.: The likelihood of various stock market return distributions, part 1: Principles of inference. Journal of Risk and Uncertainty **13**, 207–219 (1996)

Markowitz, H.M., Usmen, N.: Resampled frontiers versus diffuse bayes: An experiment. Journal Of Investment Management **1**(4), 9–25 (2003)

Markowitz, H.M., Xu, G.L.: Data mining corrections. Journal of Portfolio Management **21**(1), 60 (1994)

Pafka, S., Kondor, I.: Noisy covariance matrices and portfolio optimization ii. Physica A: Statistical Mechanics and its Applications **319**, 487–494 (2003) https://doi.org/10.1016/S0378-4371(02)01499-1

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

Ross, S.: The arbitrage theory of capital asset pricing. Journal of Economic Theory **13**, 341–360 (1976)

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

Shkolnik, A.: James-Stein estimation of the first principal component. Stat **11**(1) (2022)

Spearman, C.: General intelligence, objectively determined and measured. American Journal of Psychology **15**(201–293) (1904)

Stein, C.: Inadmissibility of the usual estimator for the mean of a multivariate distribution. In: Proc. Third Berkeley Symp. Math. Stat. Prob., pp. 197–206 (1956)

Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology **61**(3), 611–622 (1999)

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

Wigner, E.: Characteristic vectors of bordered matrices with infinite dimensions. Annals of Mathematics **62**, 548–564 (1955)