Portfolio Selection Revisited
in memory of Harry Markowitz

Alex Shkolnik,[*] Alec Kercheval,[†] Hubeyb Gurdogan,[‡] Lisa R. Goldberg,[§] Haim Bar[¶]

July 16, 2024

Abstract

Throughout his life, Harry Markowitz considered how best to estimate security return and risk for mean-variance optimization. We show how recent developments can increase accuracy of risk estimates of large portfolios, leading to superior portfolio selection.

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. 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, featured in business schools everywhere, is composed of "Markowitz portfolios," whose weights can be conveniently expressed with a closed-form formula. But Markowitz portfolios typically have short positions, and 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 too large relative to the number of observations. In a prescient comment in his 1952 paper, Markowitz'a considered alternatives:

> *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*

[*]Deparment of Statistics and Applied Probability, University of California, Santa Barbara, CA, 93106.

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

[‡]Department of Mathematics, University of California, Los Angeles, CA 90095

[§]Department of Economics and Consortium for Data Analytics in Risk, UC Berkeley, Berkeley, CA 94720 and Aperio Group LLC, a wholly-owned indirect subsidiary of BlackRock, Inc., Sausalito, CA 94965

[¶]Department of Statistics, University of Connecticut, Storrs CT 06269

[‖]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. We are grateful to Nick Gunther, Ron Kahn and Alex Ulitsky for insightful comments on an early draft of this article.

[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 & López de Prado (2013) describe an implementation of Markowitz's critical line algorithm that was developed more than 60 years after it was invented. Cottle & 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).

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 & Pastur (1967), which lay out the foundations of random matrix theory, a rich area of mathematics that has applications to high dimensional covariance matrix (or, σ_{ij}) estimation. It also preceded a major development in the work of Stein (1956) who convinced the statistics community that in high dimensions, better estimators than the sample mean (or μ_i) provably exist. A complication is the dynamic nature of financial markets. Observations from a volatile period may not be useful when the market is calm. What data from an irregular past should we use to forecast risk in an uncertain future? What additional "information" did Markowitz believe would make better estimates of means and covariances?

Factor models make the problem of estimating a high-dimensional return covariance matrix possible. Early developments include the market model (Sharpe 1963) introduced in the 1960s, and the arbitrage pricing theory (Ross 1976), which typically relies on principal component analysis, and the commercially successful Barra models (Rosenberg 1974). [4] These models reduce the dimension of the covariance matrix estimation problem to a manageable size and conform to the empirical fact that a few factors are adequate to explain correlation in security returns in developed public equity markets. However, the estimation error in factor-based covariance matrices can be selectively distorted and amplified by mean-variance optimization. It is this problem that can be addressed, to some extent, with insights from random matrix theory and the estimators introduced by Stein.

In what follows, we develop a factor-based covariance matrix tailored for use in portfolio construction. Results include a statistical perspective on the interaction between mean-variance optimization and the estimation of the means and covariances. In Section 2, we review the construction of Markowitz portfolios with mean-variance optimization and introduce a measure of estimated volatility distortion in factor-based estimates of their risk. In Section 3, we provide implementable recipes for estimates of means and variances when returns are generated by a factor model. The first is based on classical principal component analysis (PCA). The second James-Stein-Markowitz (JSM) recipe employs theoretically grounded shrinkage methods to reduce the distortion in PCA. A numerical illustration comparing PCA and JSM is in Section 4. Section 5 concludes and technical details follow in appendices.

2. The Accuracy of Optimized Portfolios

For a universe of p securities, we estimate a p-vector m of means and a $(p \times p)$-matrix Σ of covariances. These two estimates determine a Markowitz portfolio w via optimization, as the solution to

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

where e is a p-vector of ones and μ is a return target. If we knew the true means and covariances, the computed portfolio w would be "*mean-variance*" optimal. But in practice, the parameters m and Σ are estimated, and the resulting errors affect the accuracy of the optimized portfolio return and risk. As was aptly stated in Michaud (1989), mean-variance "*optimizers are, in a fundamental sense, estimation-error maximizers*".

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 may be quantified by an accuracy ratio, denoted by $\mathcal{A}$, and defined via the relation

$$\mathrm{EV} = \text{True Volatility (TV)} \times \mathcal{A},$$

so that

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

Since we don't know the true volatility, we don't know $\mathcal{A}$. But quantities such as $\mathcal{A}$ are of great interest to both investors and academics, with a vast literature spanning many disciplines; e.g, mathematical finance, physics,

[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.

economics, statistics and operations research. Ideally, $\mathcal{A}$ is very close to 1. Not only is that unlikely in actual use, but the opposite tends to be true in high dimensions. Unless the estimated covariance Σ is chosen with special care, mathematical analysis confirms, under reasonable assumptions, that

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

In other words, the estimated variance 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 and a good starting point is the work by Best & Grauer (1991). We explore the accuracy of Markowitz portfolios when security returns follow a factor model, 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 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.[5] We observe only the left side of (4) and seek to understand the breakdown on the right side. Here and elsewhere, the bold typeface indicates true parameter values, which we cannot observe and must estimate from data (e.g., $\mathrm{TV}^2 = w^\top \boldsymbol{\Sigma} w$, the true variance of portfolio w).

If we assume that the ϵ's are uncorrelated with factor returns f and pairwise uncorrelated with one another (see Appendix A), then Σ follows a sum of factor and specific components

$$\Sigma = \beta F \beta^\top + \Delta, \tag{5}$$

where F is a $(k \times k)$-matrix of factor returns, and Δ is a diagonal matrix of security specific variances. Both F and Δ as well as the factor sensitivity matrix β are estimated from data, composed of observations of (4). The constraint vector m that appears in (1) estimates $\boldsymbol{m}$, the expected value of (4), from the data as well.

Squaring (3) and substituting the estimated and true parameters as well as the optimized porfolio w,

$$\mathcal{A}^2 = \frac{w^\top \beta F \beta^\top w + w^\top \Delta w}{w^\top \boldsymbol{\beta} \boldsymbol{F} \boldsymbol{\beta}^\top w + w^\top \boldsymbol{\Delta} w}.$$

As the number of securities p grows, the four terms in the expression above behave in surprising ways, and only one of them matters once the dimension is sufficiently large. Their behavior elucidates how the errors in m and Σ (which determine w) are amplified in portfolios computed by a quadratic program, such as (1). Calculations reveal, under most reasonable scenarios, that all terms except for the factor component of the true variance, $w^\top \boldsymbol{\beta} \boldsymbol{F} \boldsymbol{\beta}^\top w$, decay as $1/p$ or faster, and a first-order approximation of $\mathcal{A}$ obeys the proportionality

$$\mathcal{A} \propto \frac{1}{\sqrt{p}\, \mathcal{M}_p(\beta, m) + 1} \tag{6}$$

where $\mathcal{M}_p(\beta, m)$ is a quantity bounded between zero and infinity that is key to understanding the accuracy of optimized portfolios – see Appendix B.[6] We emphasize that (6) implies $\mathcal{A}$ decays to zero at rate $1/\sqrt{p}$. That is, as investors grows 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.

The proportionality in (6) has several curious features, including its suggestion that the estimates F and Δ do not matter asymptotically. For large enough mean-variance portfolios, they indeed do not. New works[7] investigate the behaviour of the mysterious $\mathcal{M}_p(\beta, m)$ on the right side of (6) for large p and identify what they refer to as the "optimization biases." These biases are caused by the interactions of quadratic programs and the errors in their estimated parameters. The elimination of these biases leverages the high dimensional properties of random matrices to bring the $\mathcal{A}$ into proximity of the ideal $\mathcal{A} = 1$.[8] These solutions take the form of recipes derived from an intricate use of mathematics and data science; the very tools Markowitz brought to finance more than seven decades ago.

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

[6] $\mathcal{M}_p(\beta, m)$ remains bounded in $(0, \infty)$ as the dimension p grows for the vast majority of reasonable models (5).

[7] References include Goldberg et al. (2020), Goldberg et al. (2022), Gurdogan & Kercheval (2022), Goldberg & Kercheval (2023), Goldberg et al. (2024), Gurdogan & Shkolnik (2024*a*) and Gurdogan & Shkolnik (2024*b*)

[8] The driving force behind identification and correction of optimization biases is concentration of measure, as discussed in Talagrand (1996). Shkolnik (2022) and Goldberg & Kercheval (2023) first demonstrated the connection to James-Stein shrinkage, which is described, for example, in Efron & Morris (1977).

3. Markowitz Recipes for Increased Accuracy

We provide implementable recipes for estimation of means and covariance matrices that can be input to the mean-variance optimization program (1). Our recipes are based on PCA factor models and they guarantee that the estimated covariance matrix is well conditioned. The JSM recipe relies on James-Stein shrinkage for estimated means and eigenvectors of the PCA estimated covariance matrix.

We begin with a $p \times n$ data matrix R of excess returns, the columns of which constitute n observation of the left side of equation (4). Starting with only this ingredient we present the following recipes.

3.1. Recipes for PCA weights of an optimized portfolio. We use the sample average and principal component analysis from the return data to estimate the return constraint and a k-factor covariance matrix as inputs into the optimization (1).

Recipe for the PCA return constraint

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

Recipe for the PCA covariance model

1. With $\bar{r}$ defined as above, center the data to define

$$Y = R - \big[\bar{r}, \dots, \bar{r}\big] \tag{7}$$

where the matrix being subtracted has the sample mean $\bar{r}$ in its n identical columns.

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

$$S = \textstyle\sum_{(\jmath^2,h)} \jmath^2 hh^\top = HH^\top + N \tag{8}$$

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,

$$\Delta = \mathbf{diag}(N). \tag{9}$$

4. The PCA covariance is $\Sigma = HH^\top + \Delta$.[a]

[a]For factor analysis it is common to seek (F, β) satisfying $\beta F^{1/2} = H$ for a $p \times k$ matrix β and a $k \times k$ covariance matrix F of factor returns.

3.2. Recipes for James-Stein-Markowitz weights of an optimized portfolio. In the PCA recipes above, we set the return constraint gradient m to be the vector of sample averages $\bar{r}$. Since the 1950s with the work of Charles Stein, it has been known that vectors of sample averages may be improved upon. We use James-Stein shrinkage to estimate a constraint m for use in (1) that is provably better than the sample mean (in mean-squared error). Then we apply shrinkage to the factors (the eigenvectors) in the PCA covariance matrix.

Recipe for the JSM return constraint

1. The James-Stein recipe to improve the sample mean estimate $m = \bar{r}$ is given for any p-vector $M \neq m$ by

$$m_{\text{JS}} = \mathrm{c}\, m + (1 - \mathrm{c})\, M, \qquad \mathrm{c} = 1 - \frac{\nu^2}{|m - M|^2} \tag{10}$$

where $|m - M|^2 = (m - M)^\top (m - M)$,

(11)
$$\nu^2 = \frac{\mathbf{tr}\,(N)}{n_+ - k},$$

and n_+ is the number of nonzero eigenvalues of S in the PCA recipe. Setting $\nu^2 = \frac{\mathbf{tr}\,(N)(1+n_+/p)}{n_+-k+n_+/p}$ is theoretically a little better.

2. The shrinkage target M may be any p-vector, but a popular choice is the grand mean,

$$M = \frac{\langle m, \mathrm{e}\rangle}{\langle \mathrm{e}, \mathrm{e}\rangle}\,\mathrm{e} = \left(\textstyle\sum_{i=1}^{p} m_i/p\right)\mathrm{e}\,.$$

Recipe for the JSM covariance model

1. For any estimate Δ (e.g., (9)) that is not a scalar matrix, we replace (7) by

(12)
$$Y = \Delta^{-1/2}\big(R - \big[\bar{r}, \ldots, \bar{r}\big]\big)$$

where $\Delta^{-1/2}$ is diagonal with $\Delta_{ii}^{-1/2} = 1/\sqrt{\Delta_{ii}}$.

2. We recompute H following (8) but from the reweighted sample covariance S that uses (12).
3. Taking m_{JS} in (10) we assemble the matrix

(13)
$$A = \Delta^{-1/2}\begin{pmatrix} m_{\text{JS}} & \mathrm{e} \end{pmatrix}.$$

4. Computing the pseudo-inverse $A^+ = (A^\top A)^{-1} A^\top$, and taking ν^2 in (11) we define the variables

(14)
$$\begin{aligned} M &= AA^+H\,, \\ J &= (H - M)^\top (H - M)\,, \\ C &= I - \nu^2 J^{-1}\,. \end{aligned}$$

5. Our JSM estimator is analogous to (10) but with C a $k \times k$ matrix and M a $p \times k$ matrix.

(15)
$$H_{\text{JSM}} = HC + M(I - C)\,.$$

6. The basic JSM covariance model is $\Sigma = \Delta^{1/2}\big(H_{\text{JSM}} H_{\text{JSM}}^\top + I\big)\Delta^{1/2}$.[a]

[a]Factor analysis may be applied to find (β, F), as for PCA.

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 A. 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 A. From each data set, we construct PCA and JSM Markowitz portfolios from (1) with target annualized expected return 8.5%. PCA portfolios follow the recipe in Subsection 3.1, while JSM portfolios follow recipes in Subsection 3.2. We consider universes of size p ranging between 500 and 3,000 to shed light on problems commonly encountered in practice, and, we include $p = 100,000$ to highlight asymptotic effects.

True and estimated volatility of Markowitz portfolios are featured in Table 1. The OPT column shows oracle

p	OPT	PCA TV	PCA EV	$\mathcal{A}$	JSM TV	JSM EV	$\mathcal{A}$
500	6.33	8.66	6.71	0.77	7.93	10.0	1.26
1000	4.73	7.78	4.71	0.61	6.1	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	0.4	3.8	4.12	1.09
100000	0.5	6.29	0.47	0.07	0.81	0.83	1.02

Table 1. *Portfolio Volatility and Accuracy. 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.* $\mathcal{A}$*: Average accuracy ratio of estimated Markowitz portfolios.* $n = 125$, $\mu = 8.5$, *10000 simulations.* $n = 125$, $\mu = 8.5$, *400 simulations.*

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 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 $\mathcal{A} = 0.77$. 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 the volatility of the PCA Markowitz portfolio does not. The accuracy ratio of PCA plummets as p, while the 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	8.5	1.23	7.56	8.51	1.14
1000	8.5	7.04	8.5	1.21	7.59	8.51	1.13
2000	8.5	7.03	8.5	1.21	7.63	8.51	1.12
3000	8.5	6.92	8.5	1.23	7.51	8.51	1.14
100000	8.5	6.93	8.5	1.23	7.27	8.5	1.17

Table 2. *Portfolio Expected Excess Return 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 featured 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 $\mathcal{D} = ER/TR$ is greater for PCA than for JSM.

Risk-adjusted expected excess return or *Sharpe ratio* for Markowitz portfolios are featured 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 $\mathcal{D} = ESR/TSR$ explodes for PCA as p grows. There are two sources of the discrepancy between the relatively tame distortion for JSM and the explosion 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 accuracy 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.

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 estimated Markowitz portfolios. $n = 125$, $\mu = 8.5$, 400 simulations*

5. Harry Markowitz Was a Statistician

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 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.

A. Return Generating Process

We specify a seven-factor instance of the excess return generating process $r = \boldsymbol{\beta} f + \epsilon$, introduced in (4). Security excess returns are expressed in terms of factor returns f and specific returns ϵ, and security sensitivities to factors $\boldsymbol{\beta}$. Here, we explain how these inputs are calibrated in our experiment. With ϵ's uncorrelated with factor returns f and pairwise uncorrelated with one another, the covariance of r is given by

$$\boldsymbol{\Sigma} = \boldsymbol{\beta} \boldsymbol{F} \boldsymbol{\beta}^\top + \boldsymbol{\Delta}.$$

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. We specify the (7×7)-covariance matrix $\boldsymbol{F}$ in terms of factor volatilities $\boldsymbol{\sigma_f}$ and factor correlations $\boldsymbol{\rho_f}$. The factor volatilities are inspired by Barra models and, in annualized percentages, are

$$\boldsymbol{\sigma_f} = \begin{pmatrix} 16.0 & 4.0 & 2.0 & 20.0 & 15.0 & 10.0 & 5.0 \end{pmatrix}.$$

The style and market factor correlations are taken from (Fama & French 2105, Table 4), the correlations between the industries and the market are inspired by Barra models. The remaining correlations are set to zero. The factor correlation matrix is given by

$$\boldsymbol{\rho_f} = \begin{pmatrix} 1.00 & 0.28 & -0.30 & 0.16 & 0.08 & 0.04 & 0.02 \\ 0.28 & 1.00 & -0.11 & 0.00 & 0.00 & 0.00 & 0.00 \\ -0.30 & -0.11 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.16 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.08 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.04 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.02 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix}$$

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.

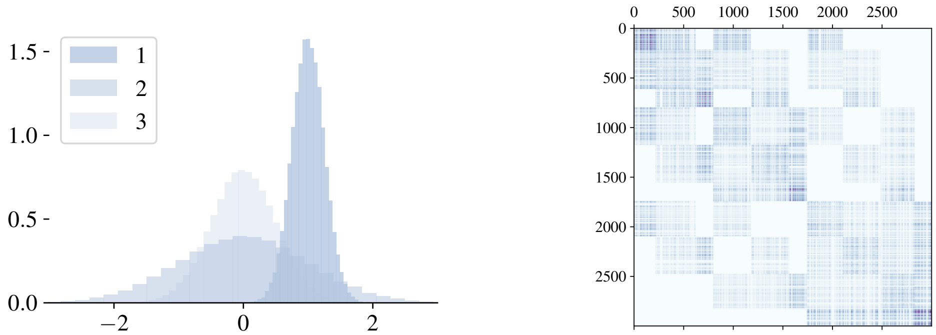

Figure 1. ***Left panel:*** *Histogram of the first three columns of $\boldsymbol{\beta}$ (market and two style factor sensitivities).* ***Right panel:*** *Industry membership visulization (i.e. entries of a matrix $\sum_c cc^\top$ where the sum is over the last four (industry) columns of $\boldsymbol{\beta}$ – white entries indicates 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 units).

Taking the expectation of r, we obtain the vector of security expected security returns,

$$\boldsymbol{m} = \boldsymbol{\beta}\boldsymbol{m_f} + \boldsymbol{m_s}. \tag{16}$$

where $\boldsymbol{m_f}$ and $\boldsymbol{m_s}$ are the expected factor and specific returns. We rely on (Fama & French 2105, Table 4) for guidance and following Ang (2023), we set the expected returns on industry factors to be zero. We set,

$$\boldsymbol{m_f} = \begin{pmatrix} 4.80 & 2.40 & 1.20 & 0.00 & 0.00 & 0.00 & 0.00 \end{pmatrix}.$$

The expected specific returns are obtained by the projection

$$\boldsymbol{m_s} = \frac{0.5}{100}\left(\boldsymbol{\Delta} - \boldsymbol{\beta}\boldsymbol{\beta}^{+}\boldsymbol{\Delta}\right), \tag{17}$$

where $\boldsymbol{\beta}^{+}$ is the pseudo-inverse of $\boldsymbol{\beta}$.[9] This means that the vector $\boldsymbol{m_s}$ of expected returns is orthogonal to the vectors of factor exposures, and securities with higher specific volatility have higher returns on average. In this way the expected return $\boldsymbol{m}$ decomposes into a factor return component $\boldsymbol{\beta m_f}$ and a specific return $\boldsymbol{m_s}$ 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.

Returns to factors f are drawn from a normal distribution with mean $\boldsymbol{m_f}$ and covariance matrix $\boldsymbol{F}$. Specific returns ϵ are uncorrelated with factor returns, and the components of ϵ are drawn from a joint normal distribution[10] with mean $\boldsymbol{m_s}$ and diagonal covariance matrix $\boldsymbol{\Delta}$. The returns are generated identically and independently over n dates keeping the model parameters $(\boldsymbol{m_f}, \boldsymbol{m_s}, \boldsymbol{F}, \boldsymbol{\beta}, \boldsymbol{\Delta})$ fixed.

B. Technical Details

We give a mathematical description of the term $\mathcal{M}_p(\beta, m)$ in (6). Let,

$$A = \begin{pmatrix} \mathrm{e} & m \end{pmatrix} \qquad \text{and} \qquad \langle U, V\rangle_\Delta = U^\top \Delta^{-1} V.$$

Then,

$$\mathcal{M}_p(\beta) = \langle \boldsymbol{\beta}, A - \beta\langle\beta,\beta\rangle_\Delta^{-1}\langle\beta, A\rangle_\Delta\rangle_\Delta \Big(\langle A, A - \beta\langle\beta,\beta\rangle_\Delta^{-1}\langle\beta, A\rangle_\Delta\rangle_\Delta\Big)^{-1}.$$

References

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

[9] 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 the no-arbitrage conclusion in Ross (1976).

[10] Normality of f and ϵ is not necessary.

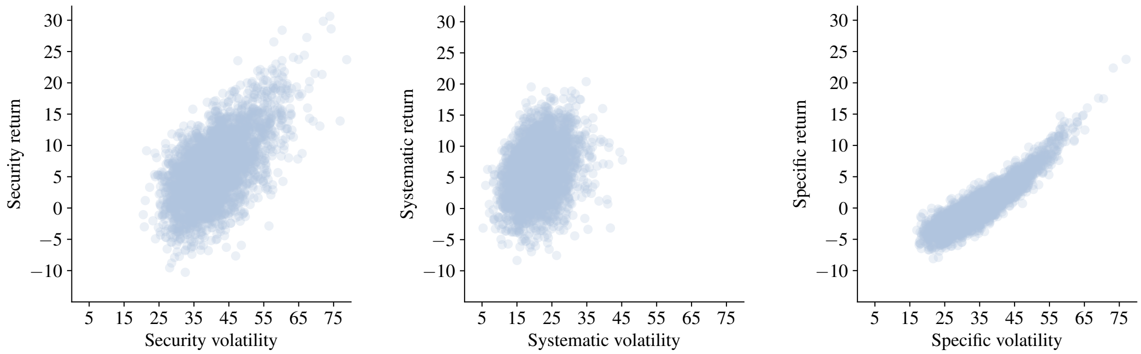

Figure 2. *Scatter plots of expected return versus volatility for $p = 3000$.* ***Left panel:*** *Total return ($\boldsymbol{m}$) vs. total volatility (square-roots of the diagonal of $\boldsymbol{\Sigma}$).* ***Center panel:*** *Systematic security return ($\boldsymbol{\beta m_f}$) vs systematic risks (square-roots of the diagonal of $\boldsymbol{\beta F \beta}^\top$).* ***Right panel:*** *Specific returns ($\boldsymbol{m_s}$) versus specific risks (square-roots of the diagonal of $\boldsymbol{\Delta}$).*

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

Best, M. J. & Grauer, R. R. (1991), '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.

Boyd, S., Johansson, K., Kahn, R., Schiele, P. & Schmelzer, T. (2024), 'Markowitz portfolio construction at seventy', *The Journal of Portfolio Management* **80**(8), 117–160.

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. (2010), Factor models, *in* 'Encyclopedia of Quantitative Finance', Wiley.

Cottle, R. W. & Infanger, G. (2010), Harry Markowitz and the early history of quadratic programming, *in* J. John B. Guerard, ed., 'Handbook of Portfolio Construction: Contemporary Applications of Markowitz Techniques', Springer.

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

Fama, E. F. & French, K. R. (2105), 'A five-factor asset pricing model', *Review of Financial Studies* **116**, 1–22.

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

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

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

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

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

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

Gurdogan, H. & Shkolnik, A. (2024*a*), Cdar working paper. The Quadratic Optimization Bias of Large Covariance Matrices.

Gurdogan, H. & Shkolnik, A. (2024*b*), Cdar working paper. The Markowtitz Optimization Enigma.

MacKenzie, D. (2006), *An Engine, Not a Camera: How Financial Models Shape Marketrs*, The MIT Press.

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

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

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

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

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

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

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

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

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

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

Talagrand, M. (1996), 'A new look at independence', *Annals of Probability* **24**(1), 1–34.

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