Financial Analysts Journal

ISSN: 0015-198X (Print) 1938-3312 (Online) Journal homepage: www.tandfonline.com/journals/ufaj20

Enhanced Portfolio Optimization

Lasse Heje Pedersen, Abhilash Babu & Ari Levine

To cite this article: Lasse Heje Pedersen, Abhilash Babu & Ari Levine (2021) Enhanced Portfolio Optimization, Financial Analysts Journal, 77:2, 124-151, DOI: 10.1080/0015198X.2020.1854543

To link to this article: https://doi.org/10.1080/0015198X.2020.1854543

Published online: 19 Feb 2021.

Financial Analysts Journal | A Publication of **CFA Institute**
https://doi.org/10.1080/0015198X.2020.1854543

OPEN ACCESS

Enhanced Portfolio Optimization

Lasse Heje Pedersen, Abhilash Babu, CFA, and Ari Levine

Lasse Heje Pedersen is a principal at AQR Capital Management and a professor at Copenhagen Business School, Frederiksberg, Denmark. Abhilash Babu, CFA, is a vice president at AQR Capital Management, Greenwich, Connecticut. Ari Levine is a principal at AQR Capital Management, Greenwich, Connecticut.

Portfolio optimization should provide large benefits for investors, but standard mean–variance optimization (MVO) works so poorly in practice that optimization is often abandoned. Many of the approaches developed to address this issue are surrounded by mystique regarding how, why, and whether they really work. So, we sought to simplify, unify, and demystify optimization. We identified the portfolios that cause problems in standard MVO, and we present here a simple "enhanced portfolio optimization" method. Applying this method to industry momentum and time-series momentum across equities and global asset classes, we found significant alpha beyond the market, the 1/*N* portfolio, and standard asset pricing factors.

Disclosure: The authors report no conflicts of interest. AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described here. The views expressed here are those of the authors and not necessarily those of AQR. Lasse Heje Pedersen gratefully acknowledges support from Center for Financial Frictions (Grant No. DNRF102).

PL Credits: 2.0

Investors seek to construct portfolios that optimally trade off risk and expected return. A standard tool to achieve this goal is mean–variance optimization (Markowitz 1952), but mean–variance optimization (MVO) often produces large and unintuitive bets that perform poorly in practice (Michaud 1989). Indeed, finding optimization methods that beat the simple 1/*N* portfolio that allocates capital (or risk) equally across securities has proven surprisingly difficult (DeMiguel, Garlappi, and Uppal 2009). Perhaps as a result, many investors skip optimization altogether. Similarly, standard academic factors that bet on such characteristics as value (high book-to-market ratio minus low book-to-market ratio, or HML), size (small minus big, or SMB), and momentum (up minus down, or UMD) are constructed without the use of optimization or, in fact, the use of any volatility or correlation information (e.g., the factor models of Fama and French 1993, 2015). Theoretically, optimization should be a big help, but the practical failure of standard MVO raises several questions: Why does standard optimization perform so poorly? Is there a better way to use the information contained in estimated risks, correlations, and expected returns? If so, how much does this method improve performance?

In the study reported here, we sought to demystify optimization by addressing these questions. In short, we show (1) where the problem with standard optimization arises, (2) how to fix it in a simple way, (3) how the fix explains and unifies a number of enhanced optimization methods in the literature, and (4) that the fix works surprisingly well. Specifically, we show the following:

1. It is well-known that the problems with standard MVO arise because of noise in the estimation of risk and expected return,[1] but our contribution is to identify the "problem portfolios" that cause trouble for MVO.
2. Our fix is an "enhanced portfolio optimization" (EPO) method designed to downweight these problem portfolios. We provide a simple closed-form solution that makes EPO as simple to implement as standard MVO.

We thank Michele Aghassi, Stephen Brown (the editor), Ben Davis, Victor DeMiguel, Antti Ilmanen, Roni Israelov, Bryan Kelly, Lorenzo Garlappi, Ernst Schaumburg, and Raman Uppal for helpful comments and Matthew Silverman and Jusvin Dhillon for excellent research assistance.

3. The method unifies a broad range of existing methods to enhance portfolio optimization; it shows what these methods have in common and how they can be implemented in a simple way.
4. The EPO method improves industry momentum and time-series momentum performance in an economically and statistically significant way relative to standard benchmarks. For example, the EPO time-series momentum portfolio in global equities, bonds, currencies, and commodities shows a large improvement in Sharpe ratio and statistically significant alpha relative to equal-notional-weighted and equal-volatility-weighted time-series momentum portfolios. Similarly, in equities, we found large performance improvements relative to standard factors when we applied the EPO method to optimize industry momentum. These findings mean that the EPO method can be a powerful tool both for investment practice and for constructing strong academic factors.

To understand the poor performance of standard MVO, consider how optimization works in practice. An investor first identifies the securities that she likes and dislikes—or, said differently, estimates the securities' expected returns. Then, she estimates the securities' risks (volatilities and correlations). All these estimates naturally have measurement errors, which can lead MVO to take large unintuitive bets that work poorly in practice.

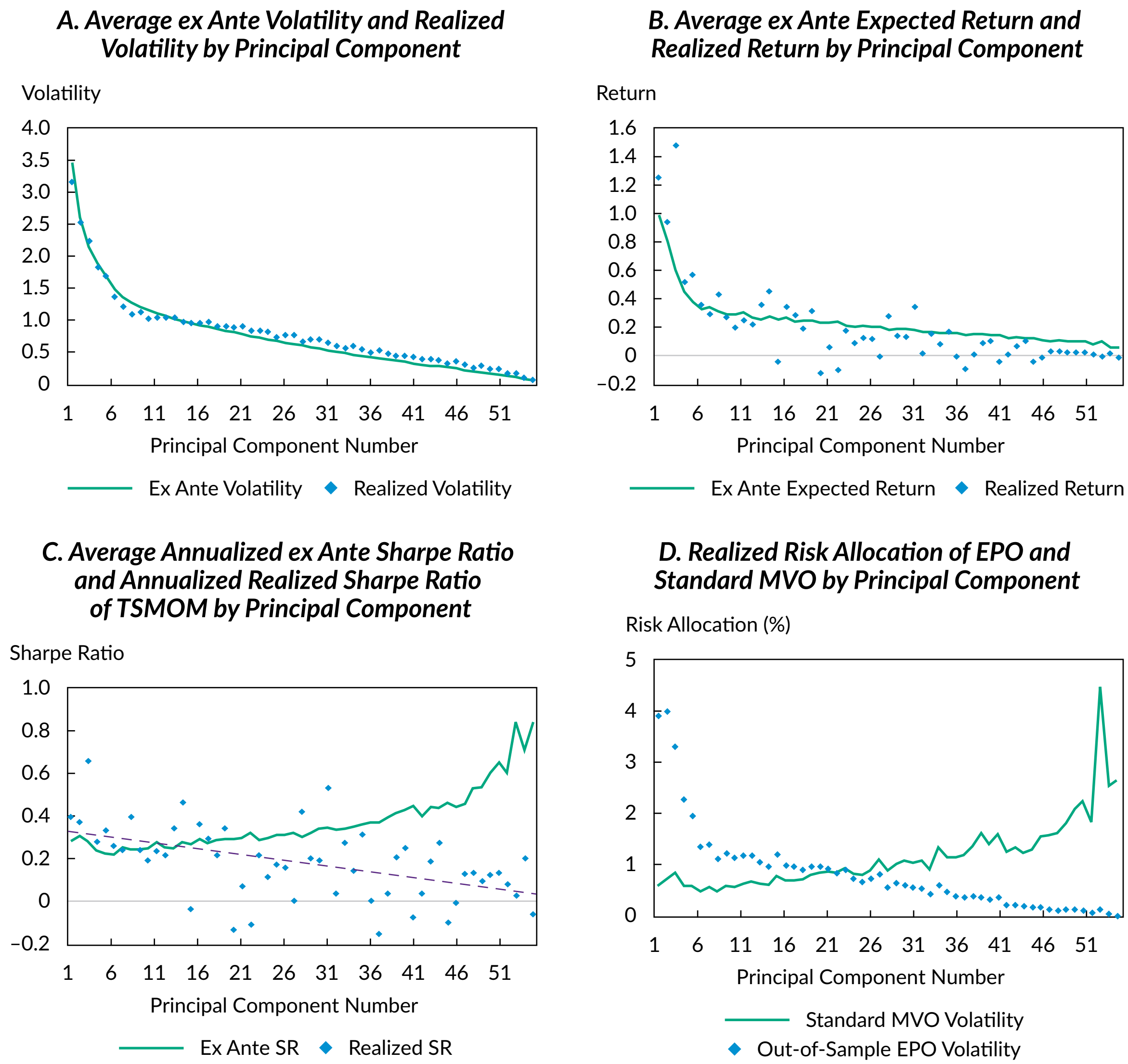
What are the problem portfolios that plague standard MVO? We show how to find them in a simple way. To do this, we transform the standard optimization problem into the space of principal components; that is, we work with long–short portfolios that are uncorrelated with each other and are ranked by their importance—namely, their variance. Working with principal components greatly simplifies the diagnosis of the problems with standard MVO because principal components are by definition uncorrelated, which means, in turn, that the risk that MVO takes in each principal component is simply proportional to its Sharpe ratio. The least important principal components are exactly the portfolios that cause trouble for standard MVO. Indeed, these portfolios have the lowest estimated risk, and as a result, their risks tend to be slightly underestimated, as shown in Panel A of **Figure 1**. (In Figure 1, the least important principal components are those with the highest numbers. Figure 1 is explained in detail in the subsection "Identifying Problem Portfolios" in the section "EPO in Practice.") Furthermore, although expected returns decrease with the principal-component number, the expected returns of the least important principal components are nevertheless too high relative to their realized returns, as can be seen in Panel B of Figure 1. As a result, from the perspective of standard MVO, these problem portfolios have large estimated Sharpe ratios, as shown in Panel C of Figure 1. These bets perform poorly in practice, as can be seen from their low realized Sharpe ratios in Panel C. MVO takes large risks in these portfolios, as shown in Panel D of Figure 1.

Having identified the problem portfolios, we show how to address the problem. In the simplest form, the solution is to reduce the estimated Sharpe ratios of the least important principal components—to make their ex ante Sharpe ratios more consistent with the realized Sharpe ratios seen in Panel C of Figure 1. Reducing estimated Sharpe ratios of the least important principal components can be achieved by increasing their estimated volatilities. Furthermore, we show that increasing the ex ante *volatilities* of the problem portfolios is exactly the same as shrinking *correlations* of the original assets toward zero! Thus, correlation shrinkage directly reduces the estimated Sharpe ratios of the problem portfolios.

This method is what we call the "simple EPO." The simple EPO first shrinks all correlations toward zero and then computes the standard MVO portfolio. The two key insights are (1) correlation shrinkage can fix errors in both risk and expected return and (2) this can be achieved by choosing the shrinkage parameter to maximize the portfolio's Sharpe ratio (out of sample). This approach contrasts with the existing literature cited below that chooses correlation shrinkage to maximize the fit of the correlation (or variance–covariance) matrix. Tuning to maximize Sharpe ratios yields a much larger shrinkage parameter, which empirically provides a large improvement in performance and is motivated by the theory that we develop. Indeed, recall that shrinkage correlations of the original assets corresponds to increasing the ex ante volatilities of the problem portfolios, which further corresponds to shrinking their Sharpe ratios, so this shrinkage addresses errors in *both* the risk model and expected returns.

This insight—that tuning correlation shrinkage to maximize risk-adjusted returns has more power than correlation shrinkage to reduce errors in risk alone—has deep theoretical foundations based on Bayesian estimation and robust optimization. Indeed, we solve a new form of robust optimization and show that

Figure 1. Understanding Problem Portfolios, 1985–2018



Notes: This figure shows that the least important principal components (those with high numbers) are the "problem portfolios" because the ex ante risk model underestimates their realized risk (Panel A), the ex ante expected return overestimates their realized returns (Panel B), and the ex ante Sharpe ratios are higher than those of the more intuitive factors (low-numbered principal components), whereas the reverse is true for realized Sharpe ratios (Panel C). Therefore, standard MVO invests too heavily in problem portfolios but EPO does not (Panel D). The sample consisted of monthly data for 55 global equities, bonds, commodities, and currencies.

uncertainty about expected returns leads endogenously to shrinkage of correlations, even when correlations are known without error. Furthermore, we show that the solution to this robust optimization equals the solution to the seminal model of Black and Litterman (1992) and methods used in machine learning. In addition to unifying these approaches, a key contribution of this article is to explain *why* these methods work—namely, because they shrink correlations, which fixes the problem portfolios.

To see how the simple EPO works in practice, consider a shrinkage parameter $w \in [0,1]$. First, we replace the off-diagonal correlation Ω_{ij} between any pair of assets i and j with $(1-w)\,\Omega_{ij}$. Then, we use this modified variance–covariance matrix to perform MVO. That is it!

Note how easy it is to do. When the EPO parameter is $w = 0$, there is no shrinkage, so our method yields the standard MVO. When $w = 1$, then all correlations are set to zero and the solution is essentially the same as not optimizing (similar to the use of standard Fama–French factors and even more similar to the signal-weighted portfolios considered in Asness, Moskowitz, and Pedersen 2013). With any shrinkage $w \in (0,1)$, we get somewhere in between standard MVO and no optimizing but in a way that works surprisingly well.[2]

How much shrinkage is needed? The simple answer is that this matter is an empirical question. We empirically choose out-of-sample w as follows: Each time period, we estimate what choice of w would have produced the highest Sharpe ratio in the time period up until today; then, we use this estimate in the next time period. In several applications, $w = 75\%$ worked well. Our theory provides some intuition for this finding. First, shrinking correlations means increasing the risk of unimportant principal components. To "fix" the correlation matrix (i.e., to fix errors in the risk model alone), we typically need to shrink the correlation matrix only about 5%–10%. So, why do we need a much larger shrinkage, around 75%? As explained previously, we show theoretically that errors in the estimates of expected returns also make correlation shrinkage useful, and these errors may be much larger than the errors in the correlation matrix itself. We found strong optimization improvements when we used a surprisingly large amount of shrinkage (surprisingly large from the perspective of what is needed to fix the correlation matrix from a pure risk perspective).

We also develop here a general form of EPO that allows the investor to control how close the solution stays to an "anchor portfolio." For example, an investor benchmarked to a certain stock index may desire to control how much his optimized portfolio deviates from this benchmark; that is, he is using the benchmark as an anchor. Or an investor may have a heuristic way to construct a portfolio—for example, splitting her money equally among good stocks ($1/N$)—and may wish for that optimized portfolio to stay close to the anchor.

Empirically, we applied our EPO method to optimize momentum portfolios using several realistic datasets. We show that EPO produces significant performance gains relative to standard benchmarks in the literature. When applied to a universe of global equity indexes, bonds, currencies, and commodities, the EPO time-series momentum portfolio substantially outperformed several benchmarks that are known to be difficult to beat. Indeed, EPO outperformed $1/N$ portfolios, equal-notional-weighted time-series momentum factors, equal-volatility-weighted time-series momentum factors, standard MVO, and MVO methods with enhanced risk models.

Furthermore, in the context of equity industry portfolios, the EPO industry momentum portfolio significantly outperformed the market portfolio, $1/N$ portfolios, standard MVO, MVO with an enhanced risk model, and standard industry momentum. The out-of-sample EPO industry momentum portfolio had significant alpha relative to the Fama–French five-factor model augmented with a standard industry momentum factor.

Related Literature

Our study is related to several approaches in the literature; indeed, one of our theoretical contributions is to unify and demystify these seemingly different frameworks.[3] First, some papers have focused on using shrinkage to improve the variance–covariance estimate (Ledoit and Wolf 2003; Elton, Gruber, and Spitzer 2006), factor models (Fan, Fan, and Lv 2008), or random matrix theory (e.g., Ledoit and Wolf 2004, 2012, 2017; El Karoui 2008; Bun, Bouchaud, and Potters 2017). We found that the EPO solution significantly outperforms such approaches because EPO uses a much larger shrinkage to account for noise in estimates of both risk and expected returns (as discussed).

Second, Black and Litterman (1992) pioneered the focus on noise in expected returns. Despite the fame of their paper, it remains mysterious to many readers, who find it difficult to apply and find where the result comes from difficult to understand, including what is being assumed and what the parameters mean. Although the EPO solution is seemingly different from Black and Litterman, we show that it is, in fact, equivalent to Black and Litterman. But EPO is simpler to apply and more transparent in how and why it works. Indeed, the EPO solution is given as a new expression, which shows how correlation shrinkage can help address uncertainty in expected returns.[4] Furthermore, we demystify the whole approach by proposing an easy and transparent method (the simple EPO) and by illustrating how it fixes the problem portfolios.

Third, we link our approach to the literature on robust optimization (see the survey by Fabozzi, Huang, and Zhou 2010 and references in it) by showing how to solve a problem with a general "ellipsoidal uncertainty" set on the mean and by showing, perhaps surprisingly, the exact equivalence between this form of robust optimization and the Bayesian estimator. Garlappi, Uppal, and Wang (2007) discussed robustness based on ambiguity aversion and uncovered a connection between their approach and shrinkage estimators. Raponi, Uppal, and Zaffaroni (2020) found strong results for robust portfolio optimization.

Fourth, Britten-Jones (1999) showed that standard MVO can be seen as the regression coefficient when a constant is being regressed on realized returns. Machine learning has many ways to regularize regressions, and Ao, Li, and Zheng (2019) found that a so-called LASSO regression significantly improves performance. These papers assumed that assets have constant expected returns, whereas we allow signals to vary over time. Furthermore, we show that the EPO can be viewed as a "ridge regression," another form of regularization used in machine learning. To generate the most general form of EPO, we must consider the regression of expected returns on the variance–covariance matrix, which is related to the elastic net regression of Kozak, Nagel, and Santosh (2020).

Fifth, our empirical results extend and enhance standard factor models—in particular, industry momentum (Moskowitz and Grinblatt 1999) and time-series momentum (Moskowitz, Ooi, Pedersen 2012). See Baltas (2015), Yang, Qian, and Belton (2019), and Baltas and Kosowski (2020) for other enhancements of time-series momentum based on risk-parity methods.

Finally, Clarke, de Silva, and Thorley (2006), studying the performance of the minimum-variance portfolio, showed the power of risk modeling when using principal components and Bayesian shrinkage even in the absence of return predictors.

Identifying the Problem with Standard Optimization

We first lay out the standard framework for portfolio choice. Then, we show how to identify problem portfolios. Appendix A contains a summary of our notation.

Standard Mean–Variance Optimization.

Consider an investor's problem of choosing a portfolio of n risky assets and a risk-free security. The risk-free return is r^f, and the risky assets have excess returns given by $\mathbf{r} = (r^1, ..., r^n)'$. The investor receives a signal, $\mathbf{s}$, about the assets (such as their past momentum) and, using this signal, computes the vector of the risky assets' conditional expected excess returns, $\boldsymbol{\alpha} = \text{E}(\mathbf{r}|\mathbf{s})$. For now, assume that the investor ignores potential noise in the signal. Furthermore, rather than considering an abstract signal, assume for simplicity that the signal is already scaled to be the conditional expected excess return—that is, $\boldsymbol{\alpha} = \mathbf{s}$. Similarly, the investor computes a risk model—that is, the conditional variance–covariance matrix of excess returns, $\boldsymbol{\Sigma} = \text{var}(\mathbf{r}|\mathbf{s})$.

The investor starts with a wealth of W_0 and chooses a portfolio $\mathbf{x} = \left(x^1, ..., x^n\right)'$. Specifically, x^i is the fraction of capital invested in security i; expressed differently, the investor buys $x^i W_0$ dollars worth of security i. Given this portfolio choice, the investor's future wealth is

$$W = W_0\left(1 + r^f + \mathbf{x}'\mathbf{r}\right).$$

The investor seeks to maximize mean–variance utility over final wealth with absolute risk aversion $\bar{\gamma} = \gamma / W_0$:

$$\text{E}(W|\mathbf{s}) - \frac{\bar{\gamma}}{2}\text{var}(W|\mathbf{s}) = W_0\left(1 + r^f + \mathbf{x}'\mathbf{s} - \frac{\gamma}{2}\mathbf{x}'\Sigma\mathbf{x}\right). \tag{1}$$

Hence, to pick the investor's optimal portfolio $\mathbf{x}$, the investor optimizes as follows:

$$\max_{\mathbf{x}}\left(\mathbf{x}'\mathbf{s} - \frac{\gamma}{2}\mathbf{x}'\Sigma\mathbf{x}\right). \tag{2}$$

Based on the first-order condition, $0 = \mathbf{s} - \gamma\boldsymbol{\Sigma}\mathbf{x}$, we get the standard mean–variance-optimal portfolio:

$$\mathbf{x}^{MVO} = \frac{1}{\gamma}\boldsymbol{\Sigma}^{-1}\mathbf{s}. \tag{3}$$

This portfolio has the highest possible Sharpe ratio among all portfolios if the expected excess return and risk are measured correctly, but the MVO portfolio is sensitive to measurement errors.

Problem Portfolios. We first show here how the problem portfolios for standard MVO can be identified by using principal components of the correlation matrix. To understand, note that the variance–covariance matrix, $\boldsymbol{\Sigma} = \boldsymbol{\sigma\Omega\sigma}$, can be decomposed into the correlation matrix, $\boldsymbol{\Omega}$, and the diagonal matrix of asset volatilities,

$$\sigma = \text{diag}\left(\sqrt{\Sigma^{11}}, \ldots, \sqrt{\Sigma^{nn}}\right). \tag{4}$$

Focusing on the correlation matrix essentially means that we first scale all the original assets to have equal volatility (but we could also use the variance–covariance matrix itself).

For background on principal components, note that the first principal component maximizes the function $\mathbf{h}'\mathbf{\Omega}\mathbf{h}$ subject to $\mathbf{h}'\mathbf{h} = 1$. In other words, it maximizes the variance $\mathbf{h}'\mathbf{\Omega}\mathbf{h}$ of any portfolio $\mathbf{h}$ (in the space of assets that have been scaled to unit volatility, given that we are working with the correlation matrix instead of the variance–covariance matrix). Hence, the first principal component is the most risky portfolio (for a given sum of squared weights). The second principal component maximizes the same function $\mathbf{h}'\Omega\mathbf{h}$ subject to being independent of the first, and so on. The last principal components are exactly those portfolios that potentially give trouble to the standard mean–variance optimization. These portfolios have, by definition, the smallest possible variance among all portfolios (relative to their sum of squared portfolio weights) but not necessarily a small magnitude of estimated expected returns. In other words, for these portfolios, the noise can easily swamp the signal, and what is worse, standard MVO tends to take large leveraged bets on these noise-driven portfolios. These points are illustrated in Figure 1 as discussed in the introduction and explained in detail in the section "EPO in Practice."

To identify the principal components, we consider the eigendecomposition of the correlation matrix,

$$\mathbf{\Omega} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}, \tag{5}$$

where $\mathbf{P}$ is a matrix whose columns are the principal components (also called eigenvectors) and $\mathbf{D}$ is a diagonal matrix of the variances of each principal component (also called eigenvalues).

Each principal component is scaled such that the sum of square weights is 1—that is, $\mathbf{P}\mathbf{P}' = \mathbf{I}$, so that $\mathbf{P}^{-1} = \mathbf{P}'$. The principal-component (PC) portfolios have realized excess returns $\mathbf{P}'\boldsymbol{\sigma}^{-1}\mathbf{r}$, with expected excess returns of $\mathbf{s}^{\mathbf{P}} = \mathbf{P}'\boldsymbol{\sigma}^{-1}\mathbf{s}$ and variance given by $\mathbf{D}$. Because $\mathbf{D}$ is diagonal, these PC portfolios are uncorrelated (by construction). The portfolio optimization problem can be written as

$$\mathbf{x}'\mathbf{s} - \frac{\gamma}{2}\mathbf{x}'\mathbf{\Sigma}\mathbf{x} = (\mathbf{P}'\boldsymbol{\sigma}\mathbf{x})'\mathbf{s}^{\mathbf{P}} - \frac{\gamma}{2}(\mathbf{P}'\boldsymbol{\sigma}\mathbf{x})'\mathbf{D}(\mathbf{P}'\boldsymbol{\sigma}\mathbf{x}) = \mathbf{z}'\mathbf{s}^{\mathbf{P}} - \frac{\gamma}{2}\mathbf{z}'\mathbf{D}\mathbf{z}, \tag{6}$$

where $\mathbf{z} = \mathbf{P}'\sigma\mathbf{x}$ is the overall portfolio weight measured in terms of the principal-component portfolios. Thus, the optimal portfolio weight, $\mathbf{z}$, for the principal components is

$$\mathbf{z}^{MVO} = \frac{1}{\gamma}\mathbf{D}^{-1}\mathbf{s}^{\mathbf{P}}. \tag{7}$$

Given that all principal components are uncorrelated (that is, $\mathbf{D}^{-1}$ is also a diagonal matrix calculated by simply replacing each diagonal element in $\mathbf{D}$ with its reciprocal), this solution means that the risk taken in principal-component portfolio i is proportional to its Sharpe ratio:

$$\underbrace{z_i^{MVO}}_{\substack{\text{notional}\\ \text{position}\\ \text{in portfolio } i}} = \underbrace{\frac{1}{\gamma}\underbrace{\frac{s_i^P}{\sqrt{D_i}}}_{\substack{\text{Sharpe}\\ \text{ratio of}\\ \text{portfolio } i}}}_{\substack{\text{desired}\\ \text{volatillity for}\\ \text{portfolio } i}} \underbrace{\frac{1}{\sqrt{D_i}}}_{\substack{\text{leverage}\\ \text{needed to}\\ \text{achieve a}\\ \text{volatility of 1}\\ \text{for portfolio } i}}. \tag{8}$$

The least important principal components are those with the lowest volatilities, $\sqrt{D_i}$. Any error in the estimation of the risk model is likely to lead to an underestimation of the volatilities of these portfolios (because they have been chosen as the lowest-risk portfolios). Furthermore, any noise in the estimation of expected return s_i^P will probably be large relative to the risk. Hence, as seen in Equation 8, estimation noise has two problematic effects for the least important principal components: (1) The optimizer may have a large desired volatility for such a problem portfolio because of a large (absolute value of the) Sharpe ratio (because of noise in the estimate of expected return, which is large relative to the low risk). (2) The low estimated risk, $\sqrt{D_i}$, leads the optimizer to apply high leverage to these portfolios to achieve a given level of risk. Furthermore, these two problems exacerbate each other.

Addressing the Problem: Enhanced Portfolio Optimization

Now that the problem with MVO has been identified, the solution is straightforward: Increase the estimated risk of the problem portfolios, which can be achieved by shrinking the estimated correlations of the assets, leading to the simple EPO as shown in the next subsection. The simple EPO underlies most of our empirical analysis, so readers who want to immediately apply these insights can go directly to the empirical section, "EPO in Practice," after reading the next subsection, "Shrinking Correlations: The Simple EPO." Readers who are interested in why this simple EPO approach works well, how to anchor the EPO portfolio to a benchmark, and how various optimization techniques are connected should continue with all the following subsections.

Shrinking Correlations: The Simple EPO. As discussed previously, principal components (PCs) can be viewed as portfolios that are ordered by their degree of troublesomeness for portfolio optimization. In essence, the problem is that the estimated variances are likely to be too low for the safest portfolios (and too high for the riskiest ones). An easy fix is to shrink the estimated variances toward their average.[5] The average variance of these PC portfolios is 1 (because they are the principal components of the correlation matrix, which has 1s along the diagonal). Hence, we can use the modified risks of the PCs:

$$\tilde{\mathbf{D}} = (1-\theta)\mathbf{D} + \theta\mathbf{I}, \tag{9}$$

where $\theta \in [0,1]$ is the degree of shrinkage, **I** is the identity matrix, and the tilde (~) over the **D** means that it has been adjusted to account for estimated error. The corresponding correlation matrix for the original assets is

$$\tilde{\Omega} = \mathbf{P}\tilde{\mathbf{D}}\mathbf{P}' = \mathbf{P}\left[(1-\theta)\mathbf{D} + \theta\mathbf{I}\right]\mathbf{P}' = (1-\theta)\Omega + \theta\mathbf{I}. \tag{10}$$

Hence, one can see that the adjusted correlation matrix is simply original matrix $\boldsymbol{\Omega}$ shrunk toward the identity matrix. In other words, we have shown the following:

> *Observation:* Adjusting the volatilities of PC portfolios corresponds to adjusting the correlations of the original assets. Specifically, increasing the volatility of the problem portfolios while lowering the volatility of the important PC portfolios is the same as multiplying all the correlations of the original assets by $1 - \theta$.

The variance–covariance matrix with the shrunk correlations is $\tilde{\Sigma} = \sigma\tilde{\Omega}\sigma$, which we can use as an input in portfolio optimization. The result is a simple enhanced portfolio optimization:

$$\mathbf{EPO}^s = \frac{1}{\gamma}\tilde{\Sigma}^{-1}\mathbf{s}. \tag{11}$$

This shrinkage is not only helpful in addressing misspecification of the variances; it also addresses misspecification in expected returns because it implicitly shrinks the Sharpe ratios of the PC portfolios, as we discuss further in the next subsections. That is, the simple EPO uses enough correlation information to improve diversification relative to an unoptimized portfolio but not too much correlation information in order to avoid the problems of standard MVO.

Anchoring Expected Returns: A Bayesian Approach. We next address that the investor's signal, **s**, is observed with noise. This section considers a Bayesian approach following Black and Litterman (1992) but with a different way of expressing the solution (and different notation). We first describe the assumptions and then provide some intuition. The investor observes a vector of signals $\mathbf{s} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, which is the true (unobserved) expected return vector, $\boldsymbol{\mu}$, plus the noise term, $\boldsymbol{\varepsilon}$, that captures measurement errors about expected returns. The noise is normally distributed with a mean of zero and a covariance of $\boldsymbol{\Lambda}$.

The investor must try to estimate true expected return $\boldsymbol{\mu}$ on the basis of noisy signal **s**. Although standard MVO estimates the true expected return simply as the signal **s** that contains measurement errors, we consider a Bayesian investor who updates his "prior beliefs" about $\boldsymbol{\mu}$ to make a better estimate of true expected returns by using the observed signal—that is, $E(\boldsymbol{\mu}|\mathbf{s})$. The investor's prior beliefs about the assets' true expected return vector, $\boldsymbol{\mu}$, is given by

$$\mu = \gamma\Sigma\mathbf{a} + \eta, \tag{12}$$

where $\gamma\Sigma\mathbf{a}$ is constant and $\boldsymbol{\eta}$ represents random fluctuations in investment opportunities. Specifically, $\boldsymbol{\eta}$ is normally distributed with mean zero and a covariance of $\tau\Sigma$ for some constant τ.[6] The first term in Equation 12, $\gamma\Sigma\mathbf{a}$, is the unconditional average return, which is written (without loss of generality) as a product of risk aversion γ (defined in the subsection "Standard Mean–Variance Optimization" in the section "Identifying the Problem with Standard Optimization"); the variance–covariance matrix of returns, $\boldsymbol{\Sigma}$ (also defined in the subsection "Standard Mean–Variance Optimization"); and an anchor portfolio, **a**. Writing the average return in this way means that the anchor is the investor's "typical portfolio."

Intuitively, this model means that the investor is aware that the signal is estimated with error and has a framework for the nature of this error.

This framework involves some standard parameters (the risk, $\boldsymbol{\Sigma}$; the signal about expected returns, **s**; and risk aversion, γ), and other more mysterious parameters $\boldsymbol{\Lambda}$, τ, and anchor portfolio **a**. The mysterious parameters can be explained as follows: The anchor portfolio is basically the investor's typical portfolio or strategic asset allocation, τ indicates the variation in the investor's optimal portfolio, and $\boldsymbol{\Lambda}$ is the amount of measurement error. We need not worry too much about these parameters, however, because we show in the upcoming subsection "Putting Optimization to Work" how the simple EPO makes all these mysterious parameters disappear!

We also consider an anchored EPO, which makes all the mysterious parameters disappear except the anchor, because having an anchor can be useful in practice—for example, to control how much an optimized portfolio deviates from a benchmark. Indeed, we can think of the anchor as the investor's benchmark, strategic asset allocation, or typical investment strategy. To understand the anchor, note that when expected returns are at their average value (i.e., $\boldsymbol{\eta} = 0$), the optimal portfolio is the anchor (i.e., $\mathbf{x} = (1/\gamma)\Sigma^{-1}\mu = \mathbf{a}$).[7]

To solve the model, we first compute the investor's view on expected returns based on her signal and prior—namely, $\mathrm{E}(\mu|\mathbf{s})$. Given that the investor maximizes her mean–variance utility (as defined in the section "Identifying the Problem with Standard Optimization"), the solution to the enhanced portfolio optimization problem is then $(1/\gamma)\Sigma^{-1}\mathrm{E}(\mu|\mathbf{s})$. The following proposition summarizes the result, and all proofs can be found in Appendix B.

Proposition 1. In this Bayesian model, given the observed signal, the investor's expected return is

$$\mathrm{E}(\mu|\mathbf{s}) = \Sigma(\tau\Sigma + \Lambda)^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}), \tag{13}$$

and the solution to the enhanced portfolio optimization problem is

$$\mathbf{x} = \frac{1}{\gamma}(\tau\Sigma + \Lambda)^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}). \tag{14}$$

Interestingly, the optimal portfolio, Equation 14, looks like the solution to an MVO when *both* the mean and variance have been modified, even though here, we have only assumed that the mean contains errors. That is, errors in expected returns alone lead to the shrinkage of correlations, even when correlations are assumed to be known without error.

Anchoring Expected Returns: Robust Optimization.

An alternative approach to address noise in expected returns is to use robust optimization. Robust optimization aims to improve upon standard MVO by explicitly modeling uncertainty around expected returns as a part of the optimization problem. Specifically, we want to choose the portfolio that gives the highest utility even if the expected return is the worst possible, within some uncertainty region:

$$\max_{\mathbf{x}}\min_{\mu}\left[(\mathbf{x}-\mathbf{a})'\mu - \frac{\gamma}{2}\mathbf{x}'\,\Sigma\mathbf{x}\right] \text{ subject to}$$
$$\mu \in \left\{\bar{\mu} \mid (\bar{\mu}-\mathbf{s})'\Lambda^{-1}(\bar{\mu}-\mathbf{s}) \le c^2\right\}. \tag{15}$$

This specification means that we seek to be robust to measurement error in the signal **s** about expected returns. In other words, the true expected return, $\boldsymbol{\mu}$, can deviate from **s**, and we wish to ensure good performance even for the worst possible $\boldsymbol{\mu}$. The parameters $\boldsymbol{\Lambda}$ and c control how much the true expected return can deviate from the signal—that is, the amount of measurement error in the signal **s**. But these mysterious parameters disappear in the simple EPO as explained in the next subsection. Finally, we interpret the anchor portfolio, **a**, as a benchmark portfolio that we wish to outperform (or are afraid of underperforming)—for example, the market portfolio.[8] The solution is given in the following proposition.

Proposition 2. The solution to the robust portfolio optimization problem is

$$\mathbf{x} = \frac{1}{\gamma}(\tau\Sigma + \Lambda)^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}), \tag{16}$$

where τ depends on c and the set of solutions for $c \in (0,\infty)$ equals the set of solutions for $\tau \in (0,\infty)$.

This result shows how robust optimization can be done via shrinkage of the mean and variance–covariance matrices. Surprisingly, the optimal portfolio (Equation 16) is exactly the same as the solution in the previous subsection! This result provides a new link between robust optimization and Bayesian optimization. What is the intuition behind this link? Both methods capture the ideas that signal **s** contains imperfect information about the conditional expected returns, that the amount of noise in the signal is related to $\boldsymbol{\Lambda}$, and that there exists an anchor portfolio, **a**, that one might not want to deviate too much from.

Putting Optimization to Work: Simple EPO and Anchored EPO. We have discussed that estimation errors occur in both the variance–covariance matrix and in expected returns. Hence, we first fix the problem with the variance–covariance matrix by using simple shrinkage as shown in the subsection "Shrinking Correlations: The Simple EPO" (or using the random matrix theory discussed in Appendix A), giving rise to enhanced risk estimate $\tilde{\Sigma}$, and second, we enhance expected returns as described in the two sections on anchoring returns, leading to the general EPO solution:

$$EPO = \frac{1}{\gamma}(\tau\tilde{\Sigma} + \Lambda)^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}). \tag{17}$$

The general EPO solution in Equation 17 depends on several parameters, some of which are straightforward to estimate, but others are tricky. So, we will provide some guidance. Let us start with the easier ones: The variance–covariance matrix, $\tilde{\Sigma}$, can be estimated in the standard way based on the sample counterpart, possibly enhanced with shrinkage as discussed in "Shrinking Correlations: The Simple EPO." The signal about expected returns, **s**, is the investor's favorite predictor of returns. (To be clear, predicting returns is never easy, but an investor would probably not be interested in portfolio optimization if he did not have some predictors to optimize.) The trickier parameters are the anchor, **a**, the risk aversion, γ, the magnitude of shocks to expected returns, τ, and the uncertainty matrix, Λ.

Starting with the uncertainty matrix, a natural assumption is that the noise in the measurement of expected returns is independent across assets; that is, $\boldsymbol{\Lambda} = \lambda\mathbf{V}$, where λ is a constant and $\mathbf{V}$ is the diagonal matrix of variances or, equivalently, the matrix of squared volatilities, $\mathbf{V} = \sigma^2$. The independence across assets arises, for example, from the common practice of estimating signals about returns in a way that is unrelated to the estimation of risk.[9] Under this assumption, the EPO solution can be written as

$$EPO(w) = \Sigma_w^{-1}\left[(1-w)\frac{1}{\gamma}\mathbf{s} + w\mathbf{V}\mathbf{a}\right], \tag{18}$$

where $\boldsymbol{\Sigma}_w$ is a shrunk variance–covariance matrix corresponding to a shrunk correlation matrix, $\tilde{\boldsymbol{\Omega}}$:

$$\begin{aligned}\Sigma_w &= (1-w)\tilde{\Sigma} + w\mathbf{V} \\ &= \sigma\left[(1-w)\tilde{\Omega} + w\mathbf{I}\right]\sigma,\end{aligned} \tag{19}$$

and we denote $w = \lambda/(\tau+\lambda) \in [0,1]$ as the "EPO shrinkage parameter." A benefit of Equation 18 is that two of the tricky parameters (λ and τ) disappear, so we need to keep track of only their relative magnitude via w.

The EPO shrinkage parameter. The shrinkage parameter, w, plays a key role in our empirical implementation. We see from Equations 18 and 19 that the EPO shrinkage parameter controls the shrinkage of both (1) expected returns toward the anchor and (2) the correlations toward zero. For example, a shrinkage of $w = 0$ gives the standard MVO solution; a shrinkage of $w = 100\%$ yields the anchor portfolio. For the empirical implementation, we chose the shrinkage parameter in a pragmatic way—namely, as the value that yields the best risk-adjusted returns—and we show how to choose w both in-sample and out-of-sample.

Because w becomes an empirical choice variable, we write the EPO solution as a function of w—that is, $EPO(w)$. On the one hand, intuitively, the optimal shrinkage parameter is larger when measurement errors are larger—because of, for example, poor data quality, illiquidity, or using a weak return predictor (i.e., w is increasing in λ). On the other hand, the shrinkage is smaller when true expected returns fluctuate more (i.e., w is decreasing in τ).

Simple EPO. A particularly simple expression arises if we choose the anchor portfolio as $\mathbf{a} = (1/\gamma)\mathbf{V}^{-1}\mathbf{s}$.[10] In this case, we recover the simple EPO already discussed in the subsection "Shrinking Correlations: The Simple EPO":

$$EPO^s(w) = \frac{1}{\gamma}\Sigma_w^{-1}\mathbf{s}, \tag{20}$$

where $\boldsymbol{\Sigma}_w$ is the shrunk variance–covariance matrix from Equation 19.

Remarkably, Equation 20 is the same as the solution to a standard MVO except that the correlations (or variance–covariance matrix) have been shrunk. So, surprisingly, errors in the estimation of mean and variance make it helpful to shrink the correlations; that is, these correlations are shrunk beyond what is justified by the errors in the variance alone because errors in the estimates of expected returns also make correlation shrinkage useful. Furthermore, the simple EPO solution given in Equation 20 is linear in the risk tolerance, so performance statistics such as the Sharpe ratio do not depend on risk aversion γ.

Therefore, this expression is straightforward to implement—for example, by setting $\gamma = 1$ or any other number that corresponds to a desirable level of risk.

Anchored EPO. Some investors prefer their portfolios to be tied to an anchor, so it is useful to consider a practical implementation of an anchored EPO. For example, an investor might have a signal, **s**, about the assets' expected returns based on their momentum and an anchor, **a**, based on the 1/*N* portfolio or based on a benchmark portfolio. In this case, we know all the inputs in Equation 18 except *w*, which we choose empirically, and risk aversion γ. So, the last question is how to choose the risk aversion. The risk aversion γ can be chosen based on the investor's preferences (typically, a number between 1 and 10).[11] But using Equation 18 with a γ based on risk aversion requires that the signal be measured in the right "units." Specifically, the signal must not only predict returns; it should also be scaled so that, for instance, $s_i = 2\%$ means that asset *i* has an expected return of 2%.

Suppose, instead, that our signal is proportional to expected returns but we do not really know the scale. For example, an asset's past momentum predicts that it will outperform in the future, but we do not know by how much. Or, as another example, suppose the signal is a relative ranking of securities based on their valuations. In these cases, the risk aversion γ can be chosen based on the insight that the investor apparently likes the risk level inherent in the anchor portfolio. Note that the EPO solution (Equation 18) is essentially a mixture of the anchor portfolio and the portfolio $\Sigma_w^{-1}(1/\gamma)\mathbf{s}$, so we can pick γ to equalize the variance of these portfolios:

$$\gamma = \frac{\sqrt{\mathbf{s}'\Sigma_w^{-1}\tilde{\Sigma}\Sigma_w^{-1}\mathbf{s}}}{\sqrt{\mathbf{a}'\tilde{\Sigma}\mathbf{a}}}, \tag{21}$$

which yields[12]

$$EPO^a(w) = \Sigma_w^{-1}\left[(1-w)\frac{\sqrt{\mathbf{a}'\,\tilde{\Sigma}\mathbf{a}}}{\sqrt{\mathbf{s}'\,\Sigma_w^{-1}\tilde{\Sigma}\Sigma_w^{-1}\mathbf{s}}}\mathbf{s} + w\mathbf{V}\mathbf{a}\right]. \tag{22}$$

Equation 22 is our *anchored EPO* solution for anchor **a** based on shrinkage parameter *w*, where risk aversion is chosen endogenously.

A Unified Approach to Optimization.

In summary, we have derived a general enhanced portfolio optimization method (Equation 17) and two straightforward ways to implement this method—the simple EPO (Equation 20) and the anchored EPO (Equation 22), which are used in our empirical implementations. The reader has already seen that the EPO method is related to several other approaches to portfolio optimization and, as described in the following proposition, the method has, in fact, even broader links to the literature.

Proposition 3. The EPO solution (Equation 17) is equal to

1. ***standard MVO*** when the estimate of variance has no noise, so $\tilde{\Sigma} = \Sigma$, and the signal of expected returns has no noise, so $\mathbf{\Lambda} = 0$;
2. the anchor when $\tau = 0$ as in ***reverse MVO***;
3. the ***Bayesian estimator*** from "Anchoring Expected Returns: A Bayesian Approach," which is equivalent to Black–Litterman (1992) when the anchor portfolio is the market portfolio, the signal is their "view portfolios," and we assume that the variance–covariance matrix is estimated without error;
4. the solution to ***robust optimization*** with ellipsoidal uncertainty set as defined in "Anchoring Expected Returns: Robust Optimization"; and
5. ***a generalized ridge regression*** (a form of regularization used in machine learning) of expected returns on the variance–covariance matrix.[13]

Proposition 3 shows how the EPO method helps unify seemingly unrelated strands of literature. Regarding Parts 1 and 2, EPO obviously contains as special cases the standard MVO and the anchor, which is trivial in itself, but by nesting these approaches, we get an enhanced version of things we already know. Furthermore, when using expected returns that imply the anchor is the optimal portfolio (Part 2), we get what is called "reverse MVO" in the context of an optimization that includes a set of constraints. The reason is that the optimal portfolio is taken as given and optimization is performed with the "implied expected returns," $E(\mu|\mathbf{s}) = \tilde{\Sigma}\mathbf{a}$, which is the expected return that makes the anchor portfolio optimal in the absence of constraints.

Regarding Part 3, we see that the Bayesian estimator from "Anchoring Expected Returns: A Bayesian Approach" is connected to the Black–Litterman (1992) formula, which is not surprising given the similar Bayesian structure. Despite this connection, we note that our empirical implementation is very different from previous applications of Black–Litterman: Black and Litterman always took the anchor portfolio to be the market portfolio. They considered certain "view portfolios" rather than maintaining our direct focus on a signal about expected returns. They considered a relatively small set of assets. And they ignored noise in the variance–covariance matrix. In contrast, we focus on different anchors, including where the anchor essentially disappears in the simple EPO, we use the shrinkage parameter as the key tuning variable, we consider noise in estimates of both risk and expected return, and we consider a number of datasets with many more assets.

Regarding Parts 4 and 5 of Proposition 3, note the interesting—and far from obvious—aspect that the Bayesian estimator corresponds to both robust optimization (as derived in "Anchoring Expected Returns: Robust Optimization") and regularization methods used in other strands of statistics and machine learning (not discussed previously here; the proof in Appendix B, however, describes ridge regressions and other regularizations).

EPO in Practice: Empirical Results

In this section, we describe the data and methodology for our empirical study and discuss results for application of the method to global asset classes and equity portfolios.

Data and Methodology.

For our empirical implementation, we constructed optimized industry momentum and time-series momentum portfolios for 11 samples that differed in terms of their test assets and methodology, as summarized in **Table 1**. The data used, the number of test assets, the methods used, the start date of the data, and the start date of our backtests are provided in Table 1. The first three samples—Global 1, Global 2, and Global 3—consist of equity indexes, bond futures, commodities, and currencies (foreign exchange, or FX); the Equity 1 through Equity 8 samples consist of equity portfolios, as we describe in detail next. The samples consist of various datasets and methodologies in order to examine the robustness of the EPO method.

Test assets and data. Our data for Global 1–Global 3 in Table 1 consist of 55 liquid futures and forward contracts described in Moskowitz et al. (2012). Specifically, in addition to every equity, commodity, and bond futures contract used in Moskowitz et al., we used the nine currency pairs in Moskowitz et al. that involve the US dollar. We excluded non-USD cross-currency pairs to ensure that the variance–covariance matrix would be of full rank.[14] For each instrument, we constructed a return series by computing the daily excess return of the most liquid contract at each point in time and then compounded daily returns to a cumulative return index from which we could compute returns at any horizon. The data start in 1970 and extend through 2018. Following Moskowitz et al. (2012), we started the backtest in 1985, at which time we had data for a broad set of instruments. Furthermore, having the earlier data allowed us to choose an initial out-of-sample EPO shrinkage parameter without shortening the time series relative to the Moskowitz et al. study.

The samples for Equity 1 through Equity 7 in Table 1 are the 49 value-weighted US equity industry portfolios from Kenneth French's website.[15] As noted, for Equity 8, we split each industry portfolio into two components, for a total of 2 × 49 = 98 test assets. Specifically, using the CRSP data on the underlying stocks, we computed a "high-momentum" and "low-momentum" portfolio within each of the 49 industry portfolios. Each low-momentum portfolio return is a value-weighted average of the half of the stocks in that industry with the lowest past 12-month returns, and the construction is similar for the high-momentum portfolio. To calculate excess returns of all the equity portfolios, we subtracted the one-month US T-bill rate, also sourced from French's website. The equity portfolio data begin in 1927 and end in 2018. To ensure enough data from which to select an initial out-of-sample EPO shrinkage parameter using only past information, we evaluated EPO performance for a sample period beginning 15 years after data were first available (as we did for shrunk Global 1–Global 3), so all equity backtests ran from 1942 to 2018.

Benchmark factors. We used monthly returns from French's website to evaluate the returns of optimized equity industry momentum portfolios relative to the Fama–French (2015) five-factor model. We also evaluated optimized time-series momentum portfolios relative to the time-series momentum benchmarks described in Appendix A.

Optimization methods. Table 1 shows the optimization methods we considered. To demonstrate the robustness of our results, we considered various optimization methods, various signals about expected returns, and various ways to estimate risk. Specifically, we used the simple EPO method from Equation 20 in the sample of global assets and in

Table 1. Samples and Summary Statistics: Backtests Ending 2018

Portfolio	Dataset	Number of Assets	Optimization Method	Risk Model	Return Signal	Start of Data (1 January each year)	Start of Backtest (1 January each year)
Global 1	Global equities, bonds, FX, and commodities	55	EPO^s	Exponentially weighted daily volatilities (60-day center of mass) and 3-day overlapping correlations (150-day center of mass)	TSMOM	1970	1985
Global 2	Global equities, bonds, FX, and commodities	55	EPO^s	Risk model from Global 1, where correlations are shrunk 5%	TSMOM	1970	1985
Global 3	Global equities, bonds, FX, and commodities	55	EPO^s	Risk model from Global 1, enhanced via random matrix theory	TSMOM	1970	1985
Equity 1	49 industry portfolios	49	EPO^s	60 months (equal weighted), 5% shrunk	XSMOM	1927	1942
Equity 2	49 industry portfolios	49	EPO^s	40 days (equal weighted), 5% shrunk	XSMOM	1927	1942
Equity 3	49 industry portfolios	49	EPO^s	120 days (equal weighted), 5% shrunk	XSMOM	1927	1942
Equity 4	49 industry portfolios	49	EPO^s	120 days (equal weighted), 5% shrunk	XSMOM*σ	1927	1942
Equity 5	49 industry portfolios	49	EPO^s	120 days (equal weighted), 5% shrunk	XSMOM*σ^2	1927	1942
Equity 6	49 industry portfolios	49	EPO^a with anchor = $1/N$	60 months (equal weighted), 5% shrunk	XSMOM	1927	1942
Equity 7	49 industry portfolios	49	EPO^a with anchor = $1/\sigma$	60 months (equal weighted), 5% shrunk	XSMOM	1927	1942
Equity 8	Each industry split in 2 portfolios based on past 12-month return	98	EPO^s	60 months (equal weighted), 5% shrunk	XSMOM	1927	1942

Notes: TSMOM is time-series momentum; XSMOM is cross-sectional momentum. All backtests began 15 years after the earliest initial data were available, so we always had at least 15 years of data to select an out-of-sample EPO shrinkage parameter.

Equity 1 through Equity 5 and Equity 8; we used the anchored EPO method from Equation 22 in Equity 6 and Equity 7. In Equity 6, the anchor portfolio is the $1/N$ portfolio that gives equal notional weight to each industry portfolio. In Equity 7, the anchor portfolio is the $1/\sigma$ portfolio that assigns equal ex ante volatility to each industry portfolio. Specifically, this portfolio has a notional weight in industry i given by $(\sigma_t^i)^{-1}/\Sigma_j(\sigma_t^j)^{-1}$, where σ_t^i is the estimated volatility of industry i at time t.

Risk models. Table 1 further shows how we estimated risk—again considering various methods to demonstrate robustness. For Global 1, we used a method similar to that of commercial risk models. The volatility of each instrument was estimated by using exponentially weighted daily returns with a 60-day center of mass. The correlations, $\tilde{\mathbf{\Omega}}^{\text{Global 1}}$, were estimated by using exponentially weighted 3-day overlapping returns with a 150-day center of mass.[16] We used three-day returns, $r_{i,t}^{3d} = \Sigma_{k=0}^{2} r_{t-k}^{i}$, to mitigate the effects of asynchronous trading among global assets, which affects correlations but not volatilities. For Global 2, we used the same risk model as Global 1, except that all nondiagonal correlations were shrunk 5% toward zero: $\tilde{\mathbf{\Omega}}^{\text{Global 2}} = 0.95\tilde{\mathbf{\Omega}}^{\text{Global 1}} + 0.05\mathbf{I}$. For Global 3, we started with the risk model of Global 1 and then enhanced the model by using random matrix theory, $\tilde{\mathbf{\Omega}}^{\text{Global 3}} = RIE(\tilde{\mathbf{\Omega}}^{\text{Global 1}})$, where RIE stands for rotationally invariant estimator (see Bun, Bouchaud, and Potters 2016) as described in Appendix A, with n set to the number of securities available at each given point in time and $T = 300$, which is twice the center of mass of 150 days. We then combined each of these correlation matrices with the diagonal matrix, $\boldsymbol{\sigma}$, of volatility estimates to arrive at variance–covariance matrix $\tilde{\Sigma} = \sigma\tilde{\Omega}\sigma$.

For the Equity 1 through Equity 8 samples, we started with the standard equal-weighted estimates of variances and covariances for 60 months, for 40 days, and for 120 days of data.[17] We then shrank all off-diagonal correlations (or, equivalently, covariances) 5% toward zero.

Signals about expected returns. Finally, we needed a signal about expected returns in each sample. To have a simple signal that we knew correlates with future returns, we decided on past 12-month returns,[18] which we used as our signal throughout this analysis. Note, however, that our EPO method is general and can be used to optimize any predictor of future returns (or combination of predictors), not just those predictors based on past returns.

For Global 1–Global 3, we used TSMOM signals, meaning that the signal of expected return for each instrument was related to its past 12-month excess return. Specifically, the signal about the expected return of instrument i in month t was

$$s_t^i = 0.1 \times \sigma_t^i \times \text{sign}(r_{t-12,t}^i). \tag{23}$$

Equation 23 means that each instrument had a positive expected excess return when the sign of the past 12-month excess return was positive (otherwise, expected excess return was negative) and that the monthly Sharpe ratio for each asset was constant and equal to 0.1. The assumption of a constant Sharpe ratio is consistent with the implicit assumption of Moskowitz et al. (2012) because they used a constant volatility target for each asset. The scaling of 0.1 is consistent with the average realized Sharpe ratios reported by Moskowitz et al. and, more recently, by Babu, Levine, Ooi, Pedersen, and Stamelos (2020),[19] but this choice is inconsequential for the Sharpe ratio of the final EPO portfolio. To ensure that the fully shrunk EPO portfolio—$EPO^s(w = 100\%)$—exactly matched the TSMOM strategy of Moskowitz et al., we used a risk aversion coefficient of $\gamma_t = n_t/40\%$, where n_t is the number of instruments at time t.[20]

For Equity 1 through 3 and Equity 6 through 8, we considered a simple version of cross-sectional momentum (XSMOM), meaning that the signal of each instrument depended on its past 12-month relative outperformance (i.e., its return minus the average return across all instruments):

$$s_t^i = XSMOM_t^i := c_t\left(r_{t-12,t}^i - \frac{1}{n}\sum_{j=1,\ldots,n} r_{t-12,t}^j\right), \tag{24}$$

where the scaling factor, c_t, was chosen such that the positive and negative signals would sum to 1.0; that is,

$$\sum_i s_t^i 1_{\{s_t^i>0\}} = \sum_i \left|s_t^i\right| 1_{\{s_t^i<0\}} = 1. \tag{25}$$

For Equity 4, each industry's signal of expected returns is its past 12-month outperformance multiplied by its volatility,

$$s_t^i = \sigma_t^i \times XSMOM_t^i. \tag{26}$$

This choice of multiplying by volatility is similar in spirit to the scaling of TSMOM as defined in Equation 23. To see again that this choice is natural, consider the implications for the fully shrunk EPO portfolio. This portfolio has a notional weight of each industry i given by

$$EPO^s(w=100\%)^i = \frac{1}{\gamma}\frac{s_t^i}{(\sigma_t^i)^2} = \frac{XSMOM_t^i}{\gamma\sigma_t^i}, \tag{27}$$

which is proportional to the Sharpe ratio of its outperformance—an intuitive scaling. Furthermore, the fully shrunk EPO's risk weight in industry i is $\sigma_t^i a_t^i = XSMOM_t^i/\gamma$, implying that when the absolute values of these risk weights are summed over all instruments, this total risk weight is constant over time (because of the definition of c_t). So, Equation 27 is also an expression of an intuitive scaling if we believe that the investment opportunity set is not varying much over time.

Finally, for Equity 5, we let $s_t^i = (\sigma_t^i)^2 \times XSMOM_t^i$, which implies that the fully shrunk EPO portfolio, $EPO^s(w=100\%)^i = XSMOM_t^i/\gamma$, is proportional to each industry's outperformance.

Global Asset Classes: Beating Time-Series Momentum.

In this subsection, we consider the performance of EPO versus benchmarks, how to identify problem portfolios, the alphas of EPO, and leverage and turnover.

Performance of EPO vs. benchmark portfolios. Turning to our empirical results, we first consider the performance of optimized TSMOM portfolios relative to key benchmarks for global assets, such as long-only portfolios and standard TSMOM factor portfolios, as shown in **Table 2**.

The first portfolio that we consider is the $1/N$ portfolio that invests an equal notional exposure across all assets. This portfolio delivered a Sharpe ratio of 0.44, arising from the equity risk premium and similar risk premiums in other asset classes. The $1/\sigma$ portfolio targeted an equal amount of standalone volatility in each asset; for example, the notional exposure in asset i was $\left(\sigma_t^i\right)^{-1}/\Sigma_j\left(\sigma_t^j\right)^{-1}$. This portfolio delivered a higher Sharpe ratio of 0.76. The Sharpe ratios are even higher for the standard time-series momentum factors. The risk-weighted TSMOM factor already had a high Sharpe ratio of 1.09 because it does several things that an optimizer hopes to achieve: It takes into account expected returns by trading on TSMOM, and it takes into account volatility differences across assets and over time by scaling positions accordingly. Said differently, although the $1/N$ portfolio is normally difficult to beat, we also consider benchmarks such as TSMOM that already beat $1/N$ hands down—so these benchmarks set a high bar.

Table 2. Gross Sharpe Ratios of Optimized TSMOM Portfolios: Global 1–Global 3, 1985–2018

	Global 1	Global 2 (Shrunk)	Global 3 (RMT)
Portfolio			
Long only: $1/N$	0.44	0.44	0.44
Long only: $1/\sigma$	0.76	0.76	0.76
TSMOM: equal notional weight	0.74	0.74	0.74
TSMOM: equal volatility weight	1.09	1.09	1.09
EPO^s: out-of-sample	1.24	1.24	1.23
$EPO^s(w)$: shrinkage parameter w			
0% (naive MVO)	0.87	1.08	1.02
10%	1.15	1.18	1.19
25%	1.24	1.26	1.26
50%	1.31	1.31	1.32
75%	1.32	1.31	1.32
90%	1.26	1.26	1.26
99%	1.13	1.13	1.13
100% (anchor)	1.09	1.09	1.09

Notes: Global 1–Global 3 samples are described in Table 1. The long-only $1/N$ portfolio invests with equal notional exposure across all assets; the $1/\sigma$ portfolio invests with equal volatility weight in each asset; the TSMOM strategy invests with equal notional exposure in each asset; the TSMOM strategy invests with equal volatility weight in each asset and a range of optimized portfolios. The optimized portfolios consist of the simple out-of-sample EPO and a range of in-sample EPO portfolios that differ on the basis of EPO shrinkage parameter w. The out-of-sample EPO uses only past data to choose w. For Global 1, the correlation matrix was estimated in the standard way; Global 2 had a 5% shrunk correlation matrix; Global 3 used a cleaned correlation matrix based on random matrix theory (RMT).

Nevertheless, the out-of-sample EPO significantly outperformed TSMOM—by 14%—delivering, as Table 2 shows, a Sharpe ratio of 1.24 in Global 1 and Global 2 and 1.23 in Global 3. Recall that these samples differ in their estimation of the risk model. Global 2 shrinks correlations initially by 5%, and Global 3 uses random matrix theory (RMT). This performance of the EPO TSMOM portfolio is remarkably strong.

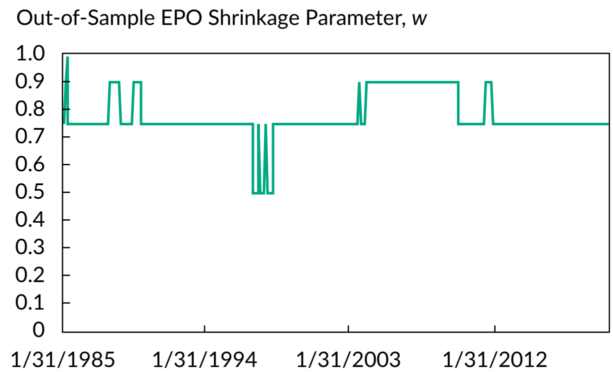
In our tests, the EPO portfolio relied on a single parameter—namely, the EPO shrinkage parameter, *w*. The out-of-sample EPO chose this parameter in an expanding fashion, using only data available before each month to decide on the parameter to use next month. Also informative is the performance of EPO when a constant *w* was used. The unshrunk EPO with $w = 0$ corresponds to standard MVO, and Table 2 shows that MVO performs worse than equal-volatility-weighted TSMOM. In other words, standard MVO does not work here. The fully shrunk EPO with $w = 1$ means that we invested in the anchor portfolio, which is the TSMOM factor by construction. With shrinkage factors in between zero and 100%, we can see that performance improves. It peaks at an even higher level than the out-of-sample EPO (which we will now call "OOS EPO"), but of course, picking the in-sample highest *w* is not implementable in real time.

Figure 2 shows the evolution of the OOS EPO shrinkage parameter over time for the Global 2 sample. Note that at least 15 years of data are required to select an initial OOS EPO shrinkage parameter. Over time, the shrinkage parameter used by the OOS EPO method approaches the optimal in-sample value, but initially OOS EPO used a lower shrinkage, and some time had to pass for the out-of-sample process to settle on the optimal shrinkage parameter. This aspect explains why the performance of OOS EPO is a bit below the in-sample maximum Sharpe ratio in Table 2.

Figure 3 shows how the realized Sharpe ratios of optimized portfolios also vary with the choice of EPO shrinkage parameter. The EPO performance is strong for a wide range of shrinkage parameters, reflecting the robustness of the process. Furthermore, the enhancements of the correlation matrix in Global 2 and Global 3 improve the performance relative to Global 1 in the case of $w = 0$ (the left side of the graph), which corresponds to standard MVO, but have almost no effect on the peak of the curve. In other words, improving the correlation matrix is important for standard MVO but has little effect when we subsequently shrink the correlation by a large factor.

Identifying problem portfolios. We have shown that standard MVO performs poorly whereas EPO performs strongly, so an interesting question is, what is the source of this difference? For simplicity, we illustrate problem portfolios for the sample in Global 1.

Figure 2. EPO Shrinkage Parameter over Time: Global 2, 1985–2018



Notes: We empirically chose *w* out-of-sample as follows: For each time period, we estimated what choice of *w* (within a finite grid of possible values) would have produced the highest EPO portfolio Sharpe ratio in the time period up until that date. Then, we used this estimate in the next time period.

Following the ideas in "Shrinking Correlations: The Simple EPO," we uncovered the problem portfolios as follows. Each month *t*, we first estimated the volatilities and correlation matrix of global assets $\boldsymbol{\Omega}_t$ as described in "Data and Methodology." We then computed the eigendecomposition of the correlation matrix,

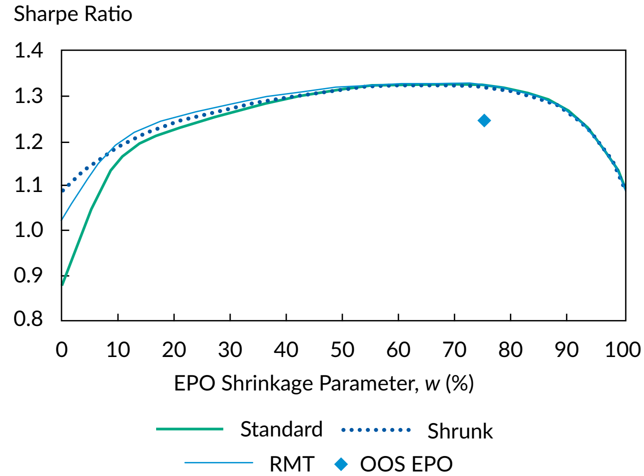
$$\Omega_t = \mathbf{P}_t \mathbf{D}_t \mathbf{P}_t^{-1}, \tag{28}$$

where $\mathbf{P}_t = \left(\mathbf{P}_t^1, \ldots, \mathbf{P}_t^{n_t}\right)$ is the matrix in which each column is a PC portfolio.

We then studied the expected returns, ex ante volatilities, and ex ante Sharpe ratios of these PC portfolios (which were rebalanced monthly). We were comparing these ex ante data with the realized data. To compute these statistics, we considered the assets rescaled to have unit volatility, $\sigma_t^{-1}\mathbf{r}_{t+1}$, with an ex ante variance–covariance matrix equal to the correlation matrix (recall that σ_t is the diagonal matrix of volatilities). Similarly, PC portfolio *i* has a return $\left(\mathbf{P}_t^i\right)' \sigma_t^{-1}\mathbf{r}_{t+1}$.

Therefore, based on this time series, we can compute the realized average excess return, volatility,

Figure 3. Performance of Optimized TSMOM (Global) Portfolios, 1985–2018



Notes: Global 1–Global 3 data were used. The three correlation matrices are the standard sample correlation matrix ("Standard" in the figure), a correlation matrix with 5% correlation shrinkage applied ("Shrunk"), and a cleaned correlation matrix based on RMT. A shrinkage of 0% is standard mean–variance optimization, 100% shrinkage is the anchor portfolio, and in between is EPO.

and Sharpe ratio. The ex ante expected return is $(\mathbf{P}_t^i)'\sigma_t^{-1}\mathbf{s}_t$, where signal s_t about the expected return is given in Equation 23. The ex ante volatility of PC portfolio *i* is given by its corresponding eigenvalue, $\sqrt{D_t^i}$, and the ex ante Sharpe is the ratio of expected return and ex ante volatility.[21]

Figure 1 plots the results. Looking back at it, first consider Panel A, showing the volatilities of the principal-component portfolios. By construction, PC#1 had the highest ex ante volatility and PC#55 had the lowest ex ante volatility. Looking at the realized volatilities of these portfolios, we see that the realized returns are also decreasing in the PC number with volatility levels that roughly match their average ex ante counterparts, indicating that the risk model works reasonably well. However, we do see systematic errors: The least important PCs (those with the highest numbers) have higher realized volatilities than their average ex ante volatility. The reason is that these portfolios have been chosen as those with the lowest ex ante volatility, so errors in the risk model may lead to underestimation of the ex ante risk of these portfolios. This low level of estimated risk leads the optimizer to apply excess leverage to these noise portfolios to achieve a given level of risk.

Now consider PC returns, plotted in Panel B of Figure 1. Naturally, realized returns are noisy; expected returns are smoother simply because realized performance always has an element of chance. Nevertheless, we see that both expected and realized returns tend to be lower for the less important PCs. Furthermore, we see that realized returns approach zero faster than the expected returns do. Said differently, the expected returns appear to be too high for the least important PCs, which adds to the problem identified in Panel A. Any noise in the expected returns of the actual assets leads to nonzero expected returns of the unimportant PCs, and because the optimizer can always choose the sign of the portfolio to make a nonzero expected return into a positive expected return, the optimizer wants to take a large position in these noise PCs.

Panel C of Figure 1 illustrates how the problems with risk and expected return interact by looking at the corresponding Sharpe ratios. We see a dramatic difference between ex ante and realized Sharpe ratios: Realized Sharpe ratios *decrease* with the PC number, whereas ex ante Sharpe ratios *increase*. Realized Sharpe ratios decrease because the important low-numbered PCs are more likely to be driven by true economic factors whereas the high-numbered PCs are unintuitive long–short factors. Said differently, the low-numbered PCs have larger signal-to-noise ratios than the high-numbered PCs. The ex ante Sharpe ratios are high for the unimportant PCs because their risk is underestimated, and their expected return is overestimated, especially relative to their level of risk.

To see the implications of this discrepancy, note in Panel D of Figure 1 the relative importance of each PC for the MVO and EPO portfolios. Specifically, we plotted the realized risk for each PC of the standard

MVO portfolio and the OOS EPO portfolio (where both portfolios were scaled to realize 10% volatility over the full sample to focus on differences in relative risks across PC portfolios). We see that the erroneous pattern in ex ante Sharpe ratios leads standard MVO to take large amounts of risk in the unimportant PC portfolios, which turns out, ex post, to be largely betting on noise in past data. Furthermore, the notional weights on the unimportant PC portfolios are even larger because these portfolios need to be leveraged as a result of their low risk per notional amount (not shown in Figure 1). This large risk exposure to "problem portfolios" highlights why standard mean–variance optimization techniques often perform poorly out-of-sample. In contrast, the EPO method accommodates this problem. Indeed, EPO shrinkage corresponds to reducing the ex ante Sharpe ratio of unimportant PC portfolios, which leads, in turn, to much smaller amounts of realized risk in the unimportant PC portfolios, as Panel D shows.

The alpha of EPO. Having shown the underlying cause of EPO's economically significant performance improvements, we now report the alphas of EPO over passive market exposures and other known factors. **Table 3** shows the alphas of OOS EPO for TSMOM (using the Global 2 sample) in relation to several benchmarks (with all variables standardized ex post to 10% volatility for comparability of coefficients). In column 1, we simply controlled for the volatility-adjusted TSMOM factor. The resulting improvement in Sharpe ratio that we saw in Table 2 translates into a statistically significant alpha and a large information ratio despite the high R^2. Column 2 reflects a further adjustment for the

Table 3. Alpha of Out-of-Sample EPO for TSMOM: Global 2, 1985–2018

	Dependent Variable				
	EPO	EPO	EPO	TSMOM	TSMOM
Alpha	2.48%	2.17%	2.11%	−0.43%	−0.34%
	(3.36)	(2.92)	(2.93)	(−0.57)	(−0.45)
Long only (1/σ)		0.06	0.07		−0.02
		(2.77)	(3.57)		(−0.86)
TSMOM	0.91	0.90			
	(44.82)	(43.77)			
TSMOM(COM)			0.53		
			(26.02)		
TSMOM(EQ)			0.30		
			(14.99)		
TSMOM(FI)			0.34		
			(16.77)		
TSMOM(FX)			0.32		
			(15.69)		
EPO				0.91	0.92
				(44.82)	(43.77)
Information ratio	0.60	0.53	0.54	−0.10	−0.08
R^2	83%	84%	85%	83%	83%

Notes: The alphas are for OOS EPO (using the Global 2 sample described in Table 1) when controlling for a volatility-scaled long-only portfolio diversified across all instruments and volatility-scaled TSMOM portfolios diversified across all instruments or all instruments within each asset class. Also reported are alphas of the volatility-scaled TSMOM portfolios in relation to OOS EPO. All variables were ex post standardized to an annualized full-sample volatility of 10% to make the alphas comparable. The scaling did not affect the *t*-statistics, which are reported in parentheses.

volatility-adjusted long-only portfolio (called 1/σ), which also had good performance, to see whether EPO simply benefits from being more long passive market exposures. Table 3 shows that the alpha remains statistically significant. Column 3 reflects controls for volatility-adjusted TSMOM strategies in each of the four asset classes to see whether EPO statically exploits a different asset allocation strategy. This test is stringent because we are now controlling for five high-performance volatility-adjusted strategies that already implicitly do part of the job that we hoped an optimizer would do. Nevertheless, the alpha of EPO remains statistically significant. The last two columns of Table 3 turn things around, regressing the volatility-adjusted TSMOM strategy on the EPO portfolio. We found an insignificant alpha, which is consistent with the dominant performance of EPO.

Leverage and turnover. Finally, to show that EPO produces realistic and implementable portfolios, we considered the turnover and gross leverage profiles of EPO portfolios. **Table 4** shows leverage and turnover statistics for the benchmark portfolios, the OOS EPO portfolio, and the EPO with various constant shrinkage parameters. We focus on the sample from Global 2 with a 5% shrunk correlation matrix. Furthermore, for comparability, gross leverage statistics are shown for portfolios ex post scaled to 10% annualized volatility, and annualized turnover statistics are reported as a percentage of average gross leverage. The lower EPO shrinkage parameters exhibit larger turnover and more gross leverage. For example, the standard MVO portfolio arising from an EPO shrinkage parameter of zero has substantially more turnover and leverage than the anchor portfolio arising from an EPO shrinkage parameter of 100%. Nevertheless, an EPO shrinkage parameter of 90% would yield turnover and leverage similar to the anchor, with a substantial improvement in performance, as shown in Table 2. The OOS EPO has a larger turnover and leverage than the anchor, but they remain of the same order of magnitude. In summary, when the EPO shrinkage parameter is chosen appropriately, EPO yields implementable portfolios with realistic leverage and turnover profiles, as well as substantial performance improvements over the standard TSMOM factors in the literature. Although the EPO method shown here abstracts from transaction costs, modeling transaction costs explicitly as a part of the optimization can potentially reduce turnover. Gârleanu and Pedersen (2013, 2016) derived the optimal portfolio in light of transaction costs but without taking estimation uncertainty into account, so their model could be combined with our enhancements in future research.

Table 4. Leverage and Turnover of Optimized TSMOM Portfolios: Global 2, 1985–2018

	Gross Leverage per 10% Volatility	Annualized Turnover as % of Avg. Gross Leverage
Portfolio		
Long only: 1/*N*	135%	26%
Long only: 1/σ	267	43
TSMOM: equal notional weight	167	153
TSMOM: equal risk	358	163
EPO^s: out-of-sample	457	254
EPO^s*(w): shrinkage parameter w*		
0% (naive MVO)	991%	546%
10%	767	480
25%	649	417
50%	551	339
75%	479	263
90%	424	208
99%	368	166
100% (anchor)	358	163

Notes: For comparability, all statistics are reported for portfolios that were ex post scaled to an annualized full-sample volatility of 10%. Annualized turnover is reported as a percentage of each portfolio's average gross leverage.

Results for Equity Portfolios: Beating Industry Momentum, the Market, 1/*N*, and Standard Factors.

We have seen that EPO substantially improves the performance of time-series momentum predictors applied to a universe of global assets. We next consider the performance of EPO for equity portfolios and study the robustness of the performance to a range of choices on optimization, risk estimation, and signals about expected returns.

EPO performance vs. benchmarks. **Table 5** reports the Sharpe ratios of the OOS EPO portfolio, a range of EPO portfolios with various constant shrinkage parameters, and three benchmark portfolios. The benchmark portfolios are the 1/*N* portfolio, a standard industry momentum (INDMOM) portfolio

Table 5. Realized Gross Sharpe Ratios of Optimized Equity Portfolios, 1942–2018

	Equity 1	Equity 2	Equity 3	Equity 4	Equity 5	Equity 6	Equity 7	Equity 8
Portfolio								
1/*N*	0.59	0.59	0.59	0.59	0.59	0.59	0.59	0.57
INDMOM	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.67
MVO (no correlation shrinkage)	0.19	−0.02	0.92	0.84	0.47	0.21	0.21	0.01
EPO: out-of-sample	**0.79**	**0.72**	**0.96**	**0.99**	**0.66**	**0.83**	**0.90**	**0.90**
EPO(w): in-sample with shrinkage of w								
0% (MVO with 5% correlation shrinkage)	0.56	0.82	0.97	0.96	0.66	0.50	0.51	0.60
10%	0.68	0.89	**0.98**	0.99	0.71	0.59	0.60	0.80
25%	0.75	0.92	0.98	**0.99**	**0.72**	0.66	0.67	0.91
50%	0.79	**0.93**	0.96	0.97	0.71	0.72	0.75	**0.98**
75%	**0.80**	0.91	0.93	0.94	0.69	**0.85**	**0.91**	0.98
90%	0.79	0.88	0.89	0.92	0.67	0.83	0.90	0.94
99%	0.73	0.77	0.77	0.91	0.65	0.60	0.63	0.86
100% (anchor)	0.71	0.73	0.73	0.91	0.63	0.59	0.62	0.81

Notes: The Equity 1 through Equity 8 samples are described in Table 1. The long-only 1/*N* portfolio invests with equal notional exposure across all industries; the INDMOM portfolio is long industries that outperformed over the past 12 months and short industries that underperformed; the standard MVO portfolio is without correlation shrinkage. The optimized portfolios are the out-of-sample and in-sample EPO portfolios. The optimal in-sample EPO portfolio and the out-of-sample EPO portfolio are shown in bold in each column. We considered in-sample EPO portfolios for a range of shrinkage parameters. The OOS EPO chose *w* by using only past data.

(following Moskowitz and Grinblatt 1999) with notional weights given by $XSMOM_t^i$ in Equation 24, and a standard MVO using unshrunk correlations.

In all cases in Table 5, the OOS EPO portfolio outperformed 1/*N*, INDMOM, and the standard MVO portfolios—often by a substantial margin. The robustness of these results is noteworthy in light of the range of specifications. Recall that Equity 1 through Equity 3 varied the risk model from 40 days to 60 months—a broad span of risk models. Equity 4 and Equity 5 used different ways to scale the signals about expected returns. Equity 6 and Equity 7 used different implementations of the EPO method—the anchored EPO rather than the simple EPO—while considering different anchors. Finally, Equity 8 was based on a more granular set of test assets—that is, two portfolios per industry.

The OOS EPO portfolio comes close to realizing the highest in-sample Sharpe ratio among all EPO portfolios with a constant shrinkage parameter in all samples except for Equity 2, which shows the robustness of the process. Also, note that all the OOS EPO portfolios realized higher Sharpe ratios than all five Fama–French factors, despite the fact that the Fama–French factors are based on individual stocks whereas the EPO factors rely only on industry returns. In fact, the best OOS EPO factors even outperformed a portfolio that simultaneously invested in all five Fama–French factors (equal weighted) over the comparable time period.[22]

Alpha to standard factors. **Table 6** reports returns after controlling for a nonoptimized INDMOM portfolio (the anchor) as well as the Fama–French five-factor model. The alpha is positive in all cases. Furthermore, the positive alphas are statistically significant at the 5% level in all samples except for Equity 6, where the *t*-statistic of 1.80 is significant only at the 10% level. For Equity 2 through Equity 4, the *t*-statistic is greater than 6, which is highly statistically significant. The weaker risk-adjusted return of Equity 6 may

Table 6. Alpha of EPO for Equity Portfolios, 1963–2018

	Dependent Variable: OOS EPO Portfolio							
	Equity 1	Equity 2	Equity 3	Equity 4	Equity 5	Equity 6	Equity 7	Equity 8
Alpha (annualized)	3.82%	8.09%	7.65%	6.25%	2.50%	1.07%	1.31%	4.40%
	(4.41)	(6.68)	(6.29)	(6.07)	(2.38)	(1.80)	(2.49)	(5.07)
INDMOM	0.78	0.53	0.53	0.69	0.67	0.31	0.33	0.78
	(32.11)	(15.66)	(15.64)	(24.00)	(22.71)	(18.85)	(22.43)	(32.21)
Mkt – RF	0.08	–0.09	–0.07	–0.07	–0.04	0.85	0.91	0.10
	(3.12)	(–2.38)	(–1.95)	(–2.10)	(–1.23)	(45.87)	(55.57)	(3.69)
SMB	–0.06	–0.04	–0.02	–0.04	–0.07	0.16	0.09	–0.05
	(–2.27)	(–1.14)	(–0.66)	(–1.32)	(–2.14)	(9.33)	(5.88)	(–2.04)
HML	–0.01	0.10	0.10	0.04	0.01	0.03	0.05	–0.03
	(–0.44)	(2.04)	(2.12)	(0.94)	(0.23)	(1.40)	(2.31)	(–1.01)
CMA	–0.14	–0.12	–0.12	–0.05	–0.01	–0.02	0.00	–0.10
	(–4.05)	(–2.43)	(–2.35)	(–1.22)	(–0.27)	(–0.84)	(0.12)	(–2.73)
RMW	–0.04	–0.04	–0.04	–0.03	0.01	0.06	0.08	–0.03
	(–1.42)	(–1.04)	(–1.06)	(–0.95)	(0.40)	(3.33)	(5.16)	(–1.10)
Information ratio	0.63	0.96	0.90	0.87	0.34	0.26	0.36	0.73
R^2	64%	29%	28%	49%	46%	83%	87%	64%

Notes: Performance of the OOS EPO portfolios is presented after controlling for standard factors. The Equity 1 through Equity 8 portfolios are described in Table 1. Each column reports a multivariate regression of EPO on a standard industry momentum factor (INDMOM) and the Fama–French five-factor model (Mkt-RF, SMB, HML, CMA, and RMW), from 1963 to 2018. Note that the Fama–French five-factor model data series does not begin until 1963. All variables are ex post standardized to an annualized full-sample volatility of 10% to make coefficients comparable. This scaling did not affect the *t*-statistics reported in parentheses.

result from the fact that in this specification, the EPO is anchored to the long-only 1/*N* portfolio, which creates two issues: (1) a large market loading of 0.85 and (2) a trade-off (in the choice of the shrinkage parameter) between stabilizing the optimization and moving toward a long-only portfolio, which does not exploit signals about expected returns. Nevertheless, the EPO portfolios delivered strong performance across a range of settings, and this strong performance cannot be explained by standard factors.

Conclusion: A Practical Guide to Optimization

We developed a simple and transparent method to make portfolio optimization work in practice. The method is essentially as simple as standard mean–variance optimization. The simple EPO method uses a single extra input—namely, a correlation shrinkage parameter, which is chosen to maximize risk-adjusted returns in past data. EPO improves portfolio performance by accounting for noise in the investor's estimates of risk and expected return. The method encompasses several optimization procedures in the literature—notably, Black–Litterman (1992), robust optimization, and regularization methods used in machine learning—so it demystifies, unifies, and simplifies much of this literature.

To illuminate why standard MVO techniques often fail, we identified the problem portfolios, to which MVO gives large weight despite their poor performance. Our EPO method addresses this issue via correlation shrinkage, which, perhaps surprisingly, downweights the problem portfolios.

Despite the method's simplicity, EPO delivers powerful results empirically. Applying our EPO method to several realistic examples, we found surprisingly large performance improvements in optimized industry momentum and time-series momentum portfolios relative to standard benchmarks and predictors used in the literature. When applied to global assets, our EPO time-series momentum portfolio substantially outperformed the market portfolio, the $1/N$ portfolio, and even relatively sophisticated benchmarks that already perform substantially better than the $1/N$ portfolio. Indeed, the EPO method delivered significant alpha relative even to volatility-scaled long-only and standard time-series momentum portfolios. These sophisticated benchmarks already deliver high Sharpe ratios because they exploit the lowest hanging fruits of optimization by (1) using information about expected returns, (2) controlling for volatility differences across assets and over time, (3) potentially exploiting market risk premiums and risk-parity effects, and (4) potentially readjusting asset class weights. This benchmark is a tough one to beat, yet EPO beat it.

When applied to equities, our EPO industry momentum portfolio substantially outperformed the market portfolio, the $1/N$ benchmark, and a standard industry momentum portfolio. This strong outperformance of EPO cannot be explained by exposure to existing factors in the literature, such as the Fama–French factors. Furthermore, the performance enhancements are robust to a range of specifications. Although for simplicity we focused on momentum predictors, future research could use this approach to enhance other predictors.

Appendix A. Summary of Notation and Auxiliary Results

In addition to a summary of the notation, we discuss the construction of TSMOM factors and random matrix theory.

Summary of Notation

Symbol	Meaning
$\mathbf{r} = \left(r^1,\ldots,r^n\right)'$	Vector of excess returns
$\mathbf{x} = \left(x^1,\ldots,x^n\right)'$	Vector of portfolio holdings
γ	Relative risk aversion
$\mathbf{s} = \left(s^1,\ldots,s^n\right)'$	Vector of signals about expected excess returns
$\mathbf{\Sigma} = \mathrm{var}(r\|s)$	Variance–covariance matrix
$\tilde{\mathbf{\Sigma}}$	Enhanced risk estimate
$\mathbf{\Sigma}_w = (1-w)\tilde{\Sigma} + wV$	Shrunk variance–covariance matrix
w	EPO shrinkage parameter
$\boldsymbol{\sigma} = \mathrm{diag}\left(\sqrt{\Sigma^{11}},\ldots,\sqrt{\Sigma^{nn}}\right) = \mathrm{diag}(\sigma^1,\ldots,\sigma^n)$	Diagonal matrix of volatilities
$\mathbf{V} = \boldsymbol{\sigma}^2$	Diagonal matrix of variances
$\mathbf{\Omega} = \mathbf{PDP}'$	Correlation matrix
$\mathbf{P}$	Matrix whose columns are principal-component portfolio weights (eigenvectors)
$\mathbf{D}$	Diagonal matrix of variances of principal-component portfolios (eigenvalues)
$\mathbf{a}$	Anchor portfolio
$\boldsymbol{\mu}$	True, but unobserved, expected return
τ	Variation in true expected returns
$\mathbf{\Lambda}, \lambda$	Error in the estimation of expected returns

Standard Time-Series Momentum (TSMOM) Factors

Following Moskowitz et al. (2012), we used the "global assets" data to construct standard TSMOM factors. In particular, we considered the equal-notional-weighted TSMOM factor with the following notional positions:

$$x_t^{TSMOM,\ equal\text{-}notional\text{-}weighted} = \frac{1}{n_t}\operatorname{sign}\left(r_{t-12,t}^i\right). \quad \text{(A1)}$$

This factor goes long or short depending on the sign of the past 12 months' excess returns and invests equally across the n_t available assets. Notional-weighted portfolios are not common in practice when investing across asset classes with large cross-sectional dispersion of volatilities. In such cases, a notional-weighted portfolio's risk may be dominated by a few assets or asset classes with higher volatilities. Nevertheless, we include comparisons to notional-weighted portfolios because they are the standard benchmark in the academic literature, are close to the 1/*N* portfolio, and are still used by some investors in practice (e.g., many investors who are benchmarked to a 60/40 stock/bond portfolio). We also considered the equal-volatility-weighted TSMOM factor with notional positions given by

$$x_t^{\text{TSMOM, equal-volatility-weighted}} = \frac{1}{n_t}\frac{40\%}{\sigma_t^i}\operatorname{sign}\left(r_{t-12,t}^i\right). \quad \text{(A2)}$$

Baltas (2015) and Yang et al. (2019) considered equal-risk-contribution TSMOM portfolios by extending the concept of "equal risk contribution" to long–short portfolios. In contrast, our anchor portfolio simply targets equal standalone volatility in each asset, thus matching the Moskowitz et al. (2012) implementation.

Random Matrix Theory

The subsection "Problem Portfolios" in the section "Identifying the Problem with Standard Optimization" shows that errors in the estimated risk model lead to problems for MVO. Specifically, small eigenvalues of the variance–covariance matrix give rise to "problem portfolios." These problem portfolios may be accommodated by stabilizing the correlation matrix, but what is the best way to do this?

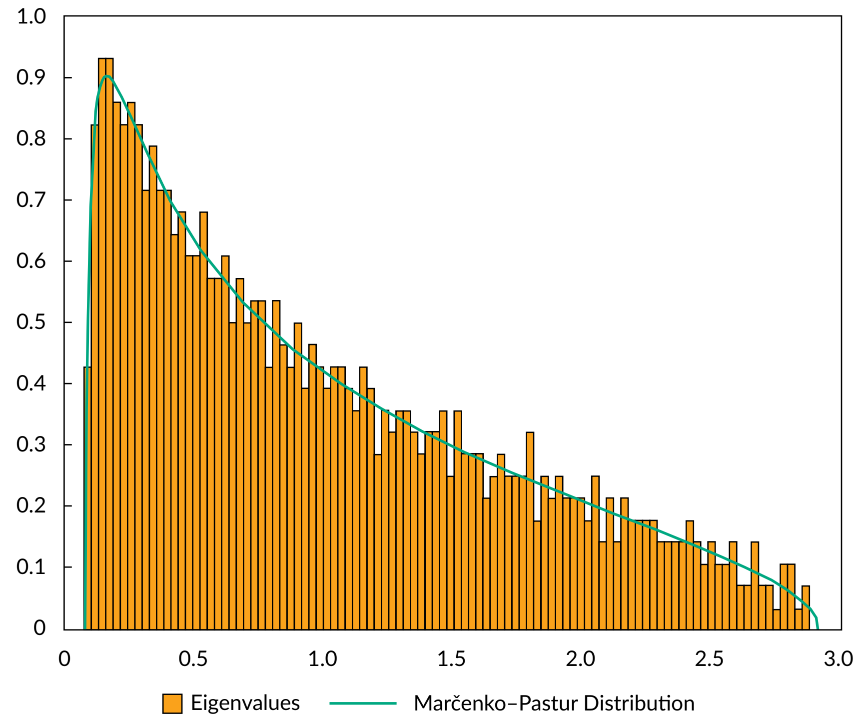
The subsection "Shrinking Correlations: The Simple EPO" discusses a simple way to stabilize risk—namely, by shrinking correlations toward zero. How do we choose the shrinkage parameter, θ? One approach is to choose a parameter that works well empirically (by looking at past data), but one can also use *random matrix theory* to derive an asymptotically optimal choice (Ledoit and Wolf 2004). Furthermore, RMT can be used to derive more general forms of stabilized correlation matrices, such as a nonlinear shrinkage of the eigenvalues (see El Karoui 2008; Ledoit and Wolf 2017; Bun et al. 2017; and the references in them).

Whereas standard statistics relies on estimates to be close to the true values when the number of time periods, *T*, is large, RMT, instead, deals with the "big data" environment of modern financial markets—that is, when we have large values of both the number of securities, *n*, and the number of time periods. Specifically, RMT considers what happens when $T \to \infty$ and $n \to \infty$ such that $n/T \to q$, where the number q is typically in (0,1). In practice, this aspect of the theory means that we can learn a lot about a variance–covariance matrix simply from knowing the ratio of the number of securities to the number of time periods used for estimation.

In line with our analysis in "Shrinking Correlations: The Simple EPO," RMT is focused on the eigenvalues of the matrix. A basic result is that if all returns are independent across securities and over time, then the asymptotic distribution of the eigenvalues is known explicitly and is given by Marčenko and Pastur (1967). As shown in the example in **Figure A1**, the Marčenko–Pastur distribution fits the distribution of the observed eigenvalues well even in a single sample. This characteristic is called the self-averaging property of random matrices.

Of course, security returns from real financial data are not independent, so the distribution of eigenvalues from real data does not closely fit Marčenko and Pastur (1967). The point is that on the basis of Marčenko and Pastur, we know what random noise in eigenvalues looks like. In particular, the "bulk" of small eigenvalues inside the Marčenko–Pastur distribution are probably just noise, whereas the larger eigenvalues outside the bulk are more likely to reflect true common return factors. Interestingly, we can talk about a specific "bulk" because the Marčenko–Pastur distribution is concentrated on a bounded interval; this distribution is very different from the

Figure A1. Distribution of Eigenvalues for Independent Securities



Notes: This figure is a histogram of the eigenvalues for the correlation matrix for 1,000 securities with returns simulated over 2,000 days, where the returns are assumed to be independently and identically distributed (i.i.d.) normal. The true eigenvalues are all 1 for a correlation matrix of independent securities, but estimation noise creates randomness (smaller and larger estimated eigenvalues), which is well captured by Marčenko and Pastur (1967).

normal distribution that we are used to seeing as a limiting distribution in standard statistics.

RMT offers various methods to "clean" the correlation matrix in the following two steps. First, we replace the estimated eigenvalues, $(D_1,\ldots,D_n)$, with cleaned eigenvalues $(\tilde{D}_1,\ldots,\tilde{D}_n)$ while, typically, leaving the eigenvectors, **P**, unchanged.[23] For this cleaning of eigenvalues, we focus here on the "IWs" method described in Bun et al. (2017), which is essentially the same as the RIE (rotationally invariant estimator) method described in Box 1 of Bun et al. (2016), with the extra steps of sorting the cleaned eigenvalues by size (to ensure that the ordering of the cleaned eigenvalues matches that of the original eigenvalues) and rescaling the cleaned eigenvalues to ensure that their sum matches that of the original eigenvalues. Then, we recover the cleaned correlation matrix as $\tilde{\Omega} = \mathbf{P}\tilde{\mathbf{D}}\mathbf{P}^{-1}$ and the cleaned variance–covariance matrix as $\tilde{\Sigma} = \sigma\tilde{\Omega}\sigma$.

To understand Marčenko and Pastur (1967) in more detail, we start with the estimated correlation matrix, $\boldsymbol{\Omega}$, for n i.i.d. random returns observed over T time periods:

$$\Omega_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{r_t^i-\bar{r}^i}{\sigma^i}\right)\left(\frac{r_t^j-\bar{r}^j}{\sigma^j}\right), \tag{A3}$$

where $\bar{r}^i$ is the average return of security i and σ^i is the standard deviation of the return. In standard "frequentist statistics," we then let the number of time periods go to infinity, concluding that the estimated correlation matrix converges to the population counterpart (and having access to the central limit theorem).

RMT, instead, considers the limit $T, n \to \infty$ such that $n/T \to q$. The remarkable result is that the empirical distribution of eigenvalues of $\boldsymbol{\Omega}$ converges to the Marčenko–Pastur distribution. When ratio q satisfies $q \in (0,1)$, then density f of the Marčenko–Pastur distribution is given by

$$f(d) = \frac{\sqrt{(q_+ - d)(d - q_-)}}{2\pi q d} \tag{A4}$$

for $d \in (q_-, q_+)$, and otherwise, $f(d) = 0$, where $q_- = \left(1-\sqrt{q}\right)^2$ and $q_+ = \left(1+\sqrt{q}\right)^2$. A slightly more complicated result holds for $q > 1$. This density is plotted in Figure A1 together with a histogram of estimated eigenvalues. This is a form of central limit theorem for RMT.

Appendix B. Proofs

Proof of Proposition 1

This model yields the following posterior mean for $\boldsymbol{\mu}$:

$$\begin{aligned} E(\mu|\mathbf{s}) &= E(\mu) + \mathrm{cov}(\mu,\mathbf{s})\,\mathrm{var}(\mathbf{s})^{-1}(\mathbf{s} - E(\mathbf{s})) \\ &= \gamma\Sigma\mathbf{a} + \tau\Sigma(\tau\Sigma+\Lambda)^{-1}(\mathbf{s}-\gamma\Sigma\mathbf{a}) \\ &= \Sigma(\tau\Sigma+\Lambda)^{-1}\tau\mathbf{s} + \gamma\Sigma\left[\mathbf{I} - (\tau\Sigma+\Lambda)^{-1}\tau\Sigma\right]\mathbf{a} \\ &= \Sigma(\tau\Sigma+\Lambda)^{-1}\tau\mathbf{s} + \gamma\Sigma(\tau\Sigma+\Lambda)^{-1}\Lambda\mathbf{a} \\ &= \Sigma(\tau\Sigma+\Lambda)^{-1}(\tau\mathbf{s}+\gamma\Lambda\mathbf{a}), \end{aligned}$$

where the first equality is due to the standard formula for conditional means of normally distributed random variables (or, equivalently, the standard ordinary least-squares, OLS, formula for regressing $\boldsymbol{\mu}$ on $\mathbf{s}$) and the fourth equality uses the Woodbury matrix identity.[24]

Proof of Proposition 2

We first solve the minimization problem inside Equation 15. For this, consider the Lagrangian:

$$L = (\mathbf{x}-\mathbf{a})'\mu + l\left[(\mu-\mathbf{s})'\Lambda^{-1}(\mu-\mathbf{s}) - c^2\right],$$

where l is the Lagrange multiplier. Differentiating with respect to $\boldsymbol{\mu}$, we get the first-order condition:

$$0 = (\mathbf{x}-\mathbf{a}) + 2l\Lambda^{-1}(\mu-\mathbf{s})$$

so $\mu = \mathbf{s} - \frac{1}{2l}\Lambda(\mathbf{x}-\mathbf{a})$. Choosing l so that the constraint specifying the uncertainty region is satisfied with equality, we see that the solution to the minimization problem is

$$\mu = \mathbf{s} - \frac{c}{\sqrt{(\mathbf{x}-\mathbf{a})'\Lambda(\mathbf{x}-\mathbf{a})}}\Lambda(\mathbf{x}-\mathbf{a}).$$

Based on this solution to the minimization problem, we can write the robust portfolio problem in the following way:

$$\max_{\mathbf{x}}\left[(\mathbf{x}-\mathbf{a})'\mathbf{s} - \frac{\gamma}{2}\mathbf{x}'\Sigma\mathbf{x} - c\sqrt{(\mathbf{x}-\mathbf{a})'\Lambda(\mathbf{x}-\mathbf{a})}\right].$$

Given that c can be chosen freely, the set of solutions (as we vary the parameter c) is the same as the set of solutions where we drop the square root (see Lemma 1 below). Further, for consistency with the other sections, we replace the parameter c by the parameter τ (which we put in the denominator) and drop constant terms:

$$\max_{x}\left[\mathbf{x}'\mathbf{s} - \frac{\gamma}{2}\mathbf{x}'\Sigma\mathbf{x} - \frac{\gamma}{2\tau}(\mathbf{x}-\mathbf{a})'\Lambda(\mathbf{x}-\mathbf{a})\right].$$

The first-order condition is

$$0 = \mathbf{s} - \gamma\Sigma\mathbf{x} - \frac{\gamma}{\tau}\Lambda(\mathbf{x}-\mathbf{a}),$$

which yields the final solution to the robust portfolio optimization problem:

$$\mathbf{x} = \frac{1}{\gamma}(\tau\Sigma+\Lambda)^{-1}(\tau\mathbf{s}+\gamma\Lambda\mathbf{a}).$$

Lemma 1

For any vector $\mathbf{a} \in \mathbf{R}^{\mathbf{n}}$ and positive definite matrices $\mathbf{B},\mathbf{C} \in \mathbf{R}^{\mathbf{n}\times\mathbf{n}}$, the set of solutions to Problem A, $\left\{\mathbf{x}^{*A}(c)\right\}_{c\geq 0}$, equals the set of solutions to Problem B, $\left\{\mathbf{x}^{*B}(d)\right\}_{d\geq 0}$, where

$$\text{Problem A: } \max_{\mathbf{x}}(\mathbf{x}'\mathbf{a} - \mathbf{x}'\mathbf{B}\mathbf{x} - c\mathbf{x}'\mathbf{C}\mathbf{x})$$

$$\text{Problem B: } \max_{\mathbf{x}}(\mathbf{x}'\mathbf{a} - \mathbf{x}'\mathbf{B}\mathbf{x} - d\sqrt{\mathbf{x}'\mathbf{C}\mathbf{x}}).$$

Proof of Lemma 1

For a given c, note that the solution $\mathbf{x}^{*A}(c)$ to Problem A satisfies the first-order condition:

$$0 = \mathbf{a} - \mathbf{B}\mathbf{x} - 2c\mathbf{C}\mathbf{x}.$$

We wish to show that $\mathbf{x}^{*A}(c)$ also satisfies the first-order condition corresponding to Problem B for an appropriate choice of d:

$$0 = \mathbf{a} - \mathbf{Bx} - \frac{d}{\sqrt{\mathbf{x}'\mathbf{Cx}}}\mathbf{Cx}.$$

We see that the result holds for

$d = 2c\sqrt{\left(\mathbf{x}^{*\mathrm{A}}(c)\right)'\mathbf{Cx}^{*\mathrm{A}}(c)}$.

Similarly, for any given d with corresponding solution $\mathbf{x}^{*\mathbf{B}}(d)$ to Problem B, we see that this vector is also a solution to Problem A when we let

$c = d/\left(2\sqrt{\left(\mathbf{x}^{*\mathbf{B}}(d)\right)'\mathbf{Cx}^{*\mathbf{B}}(d)}\right)$.

Proof of Proposition 3

Parts 1 and 2 are clear. Regarding part 3, the derivation is shown in "Anchoring Expected Returns: A Bayesian Approach." Regarding the relation to Black and Litterman (1992), we use the superscript BL to indicate their notation. With the relations that $\Pi^{BL} = \gamma\Sigma\mathbf{a}$, $\mathbf{Q}^{BL} = \mathbf{s}$, $P^{BL} = \mathbf{I}$, $\Omega^{BL} = \Lambda$, $\Sigma^{BL} = \Sigma$, and $\tau^{BL} = \tau$, their expression in point 8 of their appendix can be shown to equal our expression for the conditional mean:

$$\begin{aligned}
E(\mu|\mathbf{s}) &= \Sigma(\tau\Sigma + \Lambda)^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}) \\
&= (\tau\mathbf{I} + \Lambda\Sigma^{-1})^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}) \\
&= (\tau\Lambda^{-1} + \Sigma^{-1})^{-1}\Lambda^{-1}(\tau\mathbf{s} + \gamma\Lambda\mathbf{a}) \\
&= (\tau\Lambda^{-1} + \Sigma^{-1})^{-1}(\tau\Lambda^{-1}\mathbf{s} + \gamma\mathbf{a}) \\
&= \left[(\tau\Sigma)^{-1} + \Lambda^{-1}\right]^{-1}(\tau^{-1}\gamma\mathbf{a} + \Lambda^{-1}\mathbf{s}) \\
&= \left[(\tau^{BL}\Sigma^{BL})^{-1} + (\Omega^{BL})^{-1}\right]^{-1} \\
&\quad \cdot\left[(\tau^{BL}\Sigma^{BL})^{-1}\Pi^{BL} + (\Omega^{BL})^{-1}\mathbf{Q}^{BL}\right].
\end{aligned}$$

Regarding our part 4, the derivation of robust optimization is in "Anchoring Expected Returns: Robust Optimization," using Lemma 1, which is stated and proved in this appendix.

Regarding part 5, note that a ridge regression is a method used to mitigate noise and collinearity in a regression setting. Specifically, consider the regression $\mathbf{y} = \mathbf{z}\beta + \varepsilon$, where β is the vector of regression coefficients. The ridge regression chooses the β that minimizes the sum of squared errors plus a scalar, say λ, times the sum of squared regression coefficients, $(\mathbf{y} - \mathbf{z}\beta)'(\mathbf{y} - \mathbf{z}\beta) + \lambda\beta'\beta$. The solution is $\hat{\beta}^{\text{ridge}} = (\mathbf{z}'\mathbf{z} + \lambda\mathbf{I})^{-1}\mathbf{z}'\mathbf{y}$, so we see that the symmetric matrix $\mathbf{z}'\mathbf{z}$ is being pushed toward the identity matrix $\mathbf{I}$, ensuring invertibility. So, if expected returns (summarized by $\mathbf{s}$) are estimated in a regression, then a ridge regression can be used to stabilize the parameter estimates. This is related to, but somewhat different from, the stabilization of the optimization behind the EPO solution.

To see the direct relation to EPO, recall that we seek to solve the first-order condition for the optimal portfolio problem (Equation 3), $\mathbf{s} = \gamma\Sigma\mathbf{x}$. That is, we need to solve for the optimal portfolio $\mathbf{x}$ based on the noisy data on $\boldsymbol{\Sigma}$ and $\mathbf{s}$. We rewrite this equation as $\frac{1}{\gamma}\Sigma^{-1/2}\mathbf{s} = \Sigma^{1/2}\mathbf{x} + \varepsilon$, introducing an error term $\boldsymbol{\varepsilon}$ in order to interpret this equation as a regression (and to indicate that we are willing to accept that the equation does not hold with equality, in exchange for robustness).[25] We interpret the left-hand side as the dependent variable in a regression and the right-hand side as the independent variable multiplied by the "regression coefficient" $\mathbf{x}$. The ridge regression estimator is $\mathbf{x} = \frac{1}{\gamma}(\Sigma + \lambda\mathbf{I})^{-1}\mathbf{s}$, which is closely related to the EPO solution.

The Tikhonov regularization introduces a matrix $\boldsymbol{\Gamma}$ (instead of the multiple of the identity matrix, $\lambda\mathbf{I}$) and minimizes $(\mathbf{y} - \mathbf{z}\beta)'(\mathbf{y} - \mathbf{z}\beta) + \beta'\Gamma'\Gamma\beta$ with solution $\hat{\beta}^{\text{Tikhonov}} = (\mathbf{z}'\mathbf{z} + \Gamma'\Gamma)^{-1}\mathbf{z}'\mathbf{y}$. In our context, we can use the same regression as above with $\Gamma = \sqrt{\lambda}\sigma$, which yields $\mathbf{x} = \frac{1}{\gamma}(\Sigma + \lambda\mathbf{V})^{-1}\mathbf{s}$ using $\mathbf{V} = \sigma'\sigma$. This solution is proportional to the simple EPO—that is, it is the simple EPO solution with a different risk aversion.

Next, consider the Lavrentiev regularization (which is a generalized version of the Tikhonov regularization when $\mathbf{z}$ is symmetric and positive definite), which generally solves $\mathbf{y} = \mathbf{z}\beta + \varepsilon$ by choosing $\boldsymbol{\beta}$ in order to minimize $\|\mathbf{z}\beta - \mathbf{y}\|^2_{\mathbf{z}^{-1}} + \|\beta - \beta_0\|^2_{\mathbf{Q}}$, where the norm is defined as $\|\mathbf{x}\|^2_Q = \mathbf{x}'\mathbf{Qx}$, $\mathbf{Q}$ is a symmetric matrix, and $\boldsymbol{\beta}_0$ is a base-case parameter choice. The solution is $\hat{\boldsymbol{\beta}}^{\text{Lavrentiev}} = (\mathbf{z} + \mathbf{Q})^{-1}(\mathbf{y} + \mathbf{Q}\boldsymbol{\beta}_0)$. Next, consider this regularization for the regression $\frac{1}{\gamma}\mathbf{s} = \Sigma\mathbf{x} + \varepsilon$, where again, we are solving for $\mathbf{x}$, letting the anchor portfolio $\mathbf{a}$ play the role of β_0 and $\frac{1}{\tau}\Lambda$ play the role of $\mathbf{Q}$.

Then, we have $\mathbf{x} = \left(\Sigma + \frac{1}{\tau}\Lambda\right)^{-1}\left(\frac{1}{\gamma}\mathbf{s} + \frac{1}{\tau}\Lambda\mathbf{a}\right)$, which is exactly equal to the EPO portfolio.

Lastly, consider the regression of a vector of ones, $\mathbf{1}$, on a matrix, $\mathbf{R}$, of realized excess returns for all n assets over T time periods, $\mathbf{1} = \mathbf{Rx} + \varepsilon$. As pointed out by Britten-Jones (1999), the OLS estimate, $\mathbf{x}=\left(\frac{1}{T}\mathbf{R}'\mathbf{R}\right)^{-1}\frac{1}{T}\mathbf{R}'\mathbf{1}$, is the standard MVO when we view the average realized return, $\frac{1}{T}\mathbf{R}'\mathbf{1}$, as the signal about expected returns and the realized second moment, $\left(\frac{1}{T}\mathbf{R}'\mathbf{R}\right)^{-1}$, as the variance estimate. If we use the Tikhonov regularization with $\Gamma=\sqrt{\lambda T}\sigma$, we get $\mathbf{x}=\left(\frac{1}{T}\mathbf{R}'\mathbf{R}+\lambda\mathbf{V}\right)^{-1}\frac{1}{T}\mathbf{R}'\mathbf{1}$, which is the simple EPO under the stated assumptions.

Editor's Note

Submitted 13 August 2020

Accepted 17 November 2020 by Stephen J. Brown

Notes

1. A large literature has addressed estimation noise—for example, Ledoit and Wolf (2003, 2004) on noise in variance–covariance matrices and Black and Litterman (1992) on noise in expected returns.

2. Note that this result is not simply the same as saying that averaging portfolios improves performance (as shown by Tu and Zhou 2011). We found that EPO can work even better. For example, if we first compute the standard MVO portfolio without shrinkage, $x^{w=0}$, and the solution with full shrinkage, $x^{w=1}$, and then take the average of these, $ax^{w=0}+(1-a)x^{w=1}$, the result does *not* work as well as our EPO method for any a, especially if the MVO is particularly ill behaved. The EPO method first shrinks and then optimizes, not the other way around, which is useful because shrinking the correlations stabilizes the optimization process.

3. We unify several leading approaches to optimization, but EPO obviously does not nest all methods. Roncalli (2013) and Bruder, Gaussel, Richard, and Roncalli (2013) reviewed various methods of regularizing MVO, including a discussion of the eigendecomposition of the variance–covariance matrix similar to our problem portfolios, showing that the risk of these portfolios is low. We additionally show that the expected return of problem portfolios is too high (see Panel B of Figure 1) and that large EPO shrinkage can help address both these problems. DeMiguel et al. (2009), considering 14 methods of optimization, found that none consistently outperformed the simple 1/*N* portfolio. Some methods do show promise in outperforming the 1/*N* portfolio, however, such as methods that constrain the portfolio norm (Jagannathan and Ma 2003; DeMiguel, Garlappi, Nogales, and Uppal 2009), methods based on ambiguity aversion (Garlappi, Uppal, and Wang 2007), methods that average several approaches (Tu and Zhou 2011), and methods that apply careful MVO with good inputs (Allen, Lizieri, and Satchell 2019).

4. Although a version of EPO can be shown to be equivalent to Black and Litterman (1992), there are several differences. Indeed, Black and Litterman always shrank toward the market portfolio, whereas we consider a general anchor (or no anchor); they considered long–short "view portfolios," whereas we simply consider signals about expected returns, such as industry momentum or time-series momentum, and we allow "double shrinkage"—of both the estimated expected returns and the variance–covariance matrix. Most importantly, our contribution is to unify this approach with other optimization methods by showing the link to correlation shrinkage (which is not clear from the equations in Black and Litterman, p. 42), by presenting a simple, new, and powerful way to operationalize the method, and by documenting empirically how it works.

5. Appendix A describes a method to stabilize the risk model that is more sophisticated than shrinking correlations called "random matrix theory" (RMT). We have found empirically, however, that EPO works as well with simple correlation shrinkage as with RMT.

6. The variance of $\boldsymbol{\eta}$ is proportional to $\boldsymbol{\Sigma}$ in order to capture the idea that true fluctuations in expected returns are correlated across correlated assets (similar to the assumption made in Point 7 of the appendix of Black and Litterman 1992). Expressed in a different way, the PC portfolios have expected returns $\mathbf{P}'\boldsymbol{\sigma}^{-1}\boldsymbol{\mu}=\gamma\mathbf{P}'\boldsymbol{\sigma}^{-1}\boldsymbol{\Sigma}\mathbf{a}+\mathbf{P}'\boldsymbol{\sigma}^{-1}\boldsymbol{\eta}$, where the random fluctuation term, $\mathbf{P}'\boldsymbol{\sigma}^{-1}\boldsymbol{\eta}$, has variance $\tau\mathbf{P}'\boldsymbol{\sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\sigma}^{-1}\mathbf{P}=\tau\mathbf{D}$, implying that the expected returns of the least important principal components vary the least.

7. To understand the anchor at a deeper level, consider again the case of $\boldsymbol{\eta}=0$. In this case, the expected excess return on any asset—say, asset number 1, is $E(r_1)=\gamma(1,0,\ldots,0)\boldsymbol{\Sigma}\mathbf{a}=\gamma\,\mathrm{cov}(r_1,r_a|\mathbf{s})$. Using this relationship for anchor portfolio $\mathbf{a}$ and solving for $\gamma=E(r_a)/\mathrm{var}(r_a|\mathbf{s})$, we get $E(r_1)=\left[\mathrm{cov}(r_1,r_a|\mathbf{s})/\mathrm{var}(r_a|\mathbf{s})\right]E(r_a)=:\boldsymbol{\beta}_{1,a}E(r_a)$. If $\mathbf{a}$ is the market portfolio, this relationship is simply the conditional capital asset pricing model (CAPM). Hence, Equation 12 defining $\boldsymbol{\mu}$ means that the CAPM holds, on average, but $\boldsymbol{\eta}$ pushes the expected returns around in such a way that the CAPM does not always hold exactly, resulting in trading opportunities. More generally, Equation 12 says that the anchor is the tangency portfolio when there are no shocks ($\boldsymbol{\eta}=0$).

8. To our knowledge, the specification of Equation 15 and its solution is new, but Fabozzi et al. (2010) considered a version of Equation 15 that is simpler in two ways: First, whereas we consider a general $\boldsymbol{\Lambda}$, Fabozzi et al. assumed that $\boldsymbol{\Lambda}$ equals $\boldsymbol{\Sigma}$, which means that there is no shrinkage of the variance–covariance matrix, and second, Fabozzi et al. did not have an anchor portfolio.

9. The assumption of independence of errors in the expected returns across securities, $\boldsymbol{\Lambda}=\lambda\mathbf{V}$, implies that the error in the measurement of the expected return of the principal components has a variance given by $\mathbf{P}'\sigma^{-1}(\lambda\mathbf{V})\sigma^{-1}\mathbf{P}=\lambda\mathbf{I}$,

where **I** is the identity matrix. That is, errors of all the principal components are independent and of equal magnitude.

10. Alternatively, we can think of the anchor being $\mathbf{a} = 0$, which gives the same result as Equation 20 up to a constant that can be absorbed in the risk aversion coefficient. However, we think of the anchor as also being the EPO portfolio with full shrinkage, $w = 1$, implying that $\mathbf{a} = (1/\gamma)\mathbf{V}^{-1}\mathbf{s}$ is the more natural interpretation of Equation 20.

11. Investors can also avoid specifying γ altogether by solving an equivalent optimization that maximizes expected returns subject to a maximum volatility constraint, thus specifying a volatility target in lieu of γ.

12. Choosing γ may be done in several other, related ways, some of which work better than others. For example, although γ in Equation 22 equalizes the variance of the anchor with that of $(1/\gamma)\boldsymbol{\Sigma}_w^{-1}\mathbf{s}$, one could also replace the latter with the variance of the standard MVO solution, $(1/\gamma)\tilde{\Sigma}^{-1}\mathbf{s}$, but this is a poor choice if the standard MVO is ill behaved. Ao et al. (2019) and Raponi et al. (2020) also considered methods where γ is based on variance.

13. Specifically, the general EPO is the solution to a Lavrentiev regularization (Lavrentiev 1967), and the simple EPO is the solution to a Tikhonov regularization. The simple EPO can also be seen as a ridge regression of a vector of 1s on the matrix of realized returns when risk and expected returns are estimated by their sample counterparts.

14. If we had included non-USD currency pairs, then the variance–covariance matrix would not be of full rank because, for example, EUR–USD, EUR–JPY, and USD–JPY are linked through a triangular arbitrage.

15. Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

16. The annualized variance of instrument i was estimated as $(\sigma_t^i)^2 = 261\sum_{k=0,\ldots,\infty}(1-\delta)\delta^k(r_{t-1-k}^i - \bar{r}_t^i)^2$, where $\bar{r}_t^i$ is the exponentially weighted average return computed similarly, 261 annualizes the daily returns, and δ was chosen to achieve a center of mass of $\sum_{k=0,\ldots,\infty}(1-\delta)\delta^k k = \delta/(1-\delta) = 60$ days. The correlations were estimated by first computing covariance and volatilities in the corresponding way—using 3-day returns with 150-day center of mass—and then computing the correlations as ratios of the covariances to the product of the volatilities. We required at least 300 days of data to be available for an asset before it entered the covariance matrix.

17. In other words, the covariance of assets i and j is estimated as $[1/(K-1)]\sum_{k=1,\ldots,K}\left(r_{t-k}^i - \bar{r}_t^i\right)\left(r_{t-k}^j - \bar{r}_t^j\right)$.

18. Some studies have considered longer time horizons—for example, past five-year returns. Past long-term returns, however, predict returns negatively, if at all, perhaps because securities that have risen in price over a long time have become expensive (De Bondt and Thaler 1985). Alas, comparing optimization methods using a faulty signal of expected returns is not informative.

19. Babu et al. (2020) reported a median time-series momentum Sharpe ratio per asset of 0.34 per year (i.e., 0.10 per month) for traditional assets.

20. Indeed, this coefficient implies that the EPO portfolio with full shrinkage, $EPO^s(w = 100\%) = (1/\gamma_t)\mathbf{V}_t^{-1}\mathbf{s_t}$, has a notional exposure to asset i that matches that of Moskowitz et al. (2012) given in Appendix A. That is,
$$EPO^s(w = 100\%)^i = (1/\gamma_t)\left[s_t^i/\left(\sigma_t^i\right)^2\right] = (1/n_t)\left(40\%/\sigma_t^i\right)\operatorname{sign}\left(r_{t-12,t}^i\right).$$

21. Because the number of assets in our sample varied over time, we scaled the realized and ex ante average returns and volatilities to preserve the trace of the correlation matrix—that is, ensuring that the sum of variances would equal the largest number of assets in our sample, 55.

22. From 1963 to 2018, the five Fama–French factors realized Sharpe ratios between 0.27 and 0.49 and the equal-weighted portfolio of all five factors realized a Sharpe ratio of 0.93.

23. Estimates of the eigenvectors are kept equal to the sample eigenvectors to make the estimate of the correlation matrix rotational invariant, meaning that rotating the data by some orthogonal matrix rotates the estimator in the same way (see Ledoit and Wolf 2012; Bun et al. 2017).

24. The Woodbury matrix identity shows a way to rewrite the inverse of a sum of matrices and, using the Woodbury formula, we see that
$$\left[\mathbf{I} - \left(\tau\Sigma + \Lambda\right)^{-1}\tau\Sigma\right] = \left(\mathbf{I} + \Lambda^{-1}\tau\Sigma\right)^{-1} = \left[\Lambda^{-1}\left(\Lambda + \tau\Sigma\right)\right]^{-1} = \left(\tau\Sigma + \Lambda\right)^{-1}\Lambda.$$

25. We can also write the regression in a simpler way, $\frac{1}{\gamma}\mathbf{s} = \boldsymbol{\Sigma}\mathbf{x} + \boldsymbol{\varepsilon}$, as we do when we consider the Lavrentiev regularization. When we use the standard ridge regression on this simpler equation, we get $\mathbf{x} = \frac{1}{\gamma}\left(\boldsymbol{\Sigma}^2 + \lambda\mathbf{I}\right)^{-1}\boldsymbol{\Sigma}\mathbf{s}$, so we have written the regression differently to avoid the $\boldsymbol{\Sigma}$-squared.

References

Allen, D., C. Lizieri, and S. Satchell. 2019. "In Defense of Portfolio Optimization: What If We Can Forecast?" *Financial Analysts Journal* 75 (3): 20–38.

Ao, Mengmeng, Yingying Li, and Xinghua Zheng. 2019. "Approaching Mean–Variance Efficiency for Large Portfolios." *Review of Financial Studies* 32 (7): 2890–919.

Asness, C., T. Moskowitz, and L. H. Pedersen. 2013. “Value and Momentum Everywhere.” *Journal of Finance* 68 (3): 929–85.

Babu, Abhilash, Ari Levine, Yao Hua Ooi, Lasse Heje Pedersen, and Erik Stamelos. 2020. “Trends Everywhere.” *Journal of Investment Management* 18 (1): 52–68.

Baltas, N. 2015. “Trend-Following, Risk-Parity and the Influence of Correlations.” In *Risk-Based and Factor Investing*, edited by Emmanuel Jurczenko, 65–95. Amsterdam: Elsevier.

Baltas, N., and R. Kosowski. 2020. “Demystifying Time-Series Momentum Strategies: Volatility Estimators, Trading Rules and Pairwise Correlations.” In *Market Momentum: Theory and Practice*, edited by Stephen Satchell and Andrew Grant, 30–67. Hoboken, NJ: Wiley.

Black, Fischer, and Robert Litterman. 1992. “Global Portfolio Optimization.” *Financial Analysts Journal* 48 (5): 28–43.

Britten-Jones, Mark. 1999. “The Sampling Error in Estimates of Mean–Variance Efficient Portfolio Weights.” *Journal of Finance* 54 (2): 655–71.

Bruder, Benjamin, Nicolas Gaussel, Jean-Charles Richard, and Thierry Roncalli. 2013. “Regularization of Portfolio Allocation.” (June). Available at SSRN: https://ssrn.com/abstract=2767358 or http://dx.doi.org/10.2139/ssrn.2767358.

Bun, Joël, Jean-Philippe Bouchaud, and Marc Potters. 2016. “Cleaning Correlation Matrices.” Risk.net (29 March).

——. 2017. “Cleaning Large Correlation Matrices: Tools from Random Matrix Theory.” *Physics Reports* 666: 1–109.

Clarke, R. G., H. de Silva, and S. Thorley. 2006. “Minimum–Variance Portfolios in the U.S. Equity Market.” *Journal of Portfolio Management* 33 (1): 10–24.

De Bondt, Werner F. M., and Richard Thaler. 1985. “Does the Stock Market Overreact?” *Journal of Finance* 40 (3): 793–805.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal. 2009. “A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms.” *Management Science* 55 (5): 798–812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. “Optimal versus Naive Diversification: How Inefficient Is the 1/*N* Portfolio Strategy?” *Review of Financial Studies* 22 (5): 1915–53.

El Karoui, Noureddine. 2008. “Spectrum Estimation for Large Dimensional Covariance Matrices Using Random Matrix Theory.” *Annals of Statistics* 36 (6): 2757–90.

Elton, Edwin J., Martin J. Gruber, and Jonathan Spitzer. 2006. “Improved Estimates of Correlation Coefficients and Their Impact on Optimum Portfolios.” *European Financial Management* 12 (3): 303–18.

Fabozzi, Frank J., Dashan Huang, and Guofu Zhou. 2010. “Robust Portfolios: Contributions from Operations Research and Finance.” *Annals of Operations Research* 176 (1): 191–220.

Fama, Eugene F., and Kenneth R. French. 1993. “Common Risk Factors in the Returns on Stocks and Bonds.” *Journal of Financial Economics* 33 (1): 3–56.

——. 2015. “A Five-Factor Asset Pricing Model.” *Journal of Financial Economics* 116 (1): 1–22.

Fan, Jianqing, Yingying Fan, and Jinchi Lv. 2008. “High Dimensional Covariance Matrix Estimation Using a Factor Model.” *Journal of Econometrics* 147 (1): 186–97.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang. 2007. “Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach.” *Review of Financial Studies* 20 (1): 41–81.

Gârleanu, Nicolae, and Lasse Heje Pedersen. 2013. “Dynamic Trading with Predictable Returns and Transaction Costs.” *Journal of Finance* 68 (6): 2309–40.

——. 2016. “Dynamic Portfolio Choice with Frictions.” *Journal of Economic Theory* 165 (September): 487–516.

Jagannathan, Ravi, and Tongshu Ma. 2003. “Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps.” *Journal of Finance* 58 (4): 1651–83.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. 2020. “Shrinking the Cross-Section.” *Journal of Financial Economics* 135 (2): 271–92.

Lavrentiev, M. M. 1967. *Some Improperly Posed Problems of Mathematical Physics*. New York: Springer.

Ledoit, Olivier, and Michael Wolf. 2003. “Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection.” *Journal of Empirical Finance* 10 (5): 603–21.

——. 2004. “A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices.” *Journal of Multivariate Analysis* 88 (2): 365–411.

——. 2012. “Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices.” *Annals of Statistics* 40 (2): 1024–60.

——. 2017. “Nonlinear Shrinkage of the Covariance Matrix for Portfolio Selection: Markowitz Meets Goldilocks.” *Review of Financial Studies* 30 (12): 4349–88.

Marčenko, Vladimir A., and Leonid A. Pastur. 1967. “Distribution of Eigenvalues for Some Sets of Random Matrices.” *Mathematics of the USSR-Sbornik* 1 (4): 457–83.

Markowitz, H. 1952. “Portfolio Selection.” *Journal of Finance* 7 (1): 77–91.

Michaud, R. O. 1989. “The Markowitz Optimization Enigma: Is ‘Optimized’ Optimal?” *Financial Analysts Journal* 45 (1): 31–42.

Moskowitz, T. J., and M. Grinblatt. 1999. “Do Industries Explain Momentum?” *Journal of Finance* 54 (4): 1249–90.

Moskowitz, T., Y. H. Ooi, and L. H. Pedersen. 2012. “Time Series Momentum.” *Journal of Financial Economics* 104 (2): 228–50.

Raponi, Valentina, Raman Uppal, and Paolo Zaffaroni. 2020. “Robust Portfolio Choice.” Unpublished working paper, Imperial College Business School.

Roncalli, T. 2013. *Introduction to Risk Parity and Budgeting*. Boca Raton, FL: CRC Press.

Tu, Jun, and Guofu Zhou. 2011. “Markowitz Meets Talmud: A Combination of Sophisticated and Naive Diversification Strategies.” *Journal of Financial Economics* 99 (1): 204–15.

Yang, K., E. Qian, and B. Belton. 2019. “Protecting the Downside of Trend When It Is Not Your Friend.” *Journal of Portfolio Management* 45 (5): 99–111.