Enhanced Portfolio Optimization*

Lasse Heje Pedersen, Abhilash Babu, and Ari Levine

Abstract

Portfolio optimization should provide large benefits to investors, but standard mean-variance optimization (MVO) works so poorly in practice that optimization is often abandoned. Several approaches have been developed to address this important issue, but they are often surrounded by mystique regarding how, why, and whether they really work. We seek to demystify, simplify, and enhance optimization: we identify the portfolios that cause problems in standard MVO and develop a simple enhanced portfolio optimization (EPO) method that addresses the problems. Applying EPO across equities and global asset classes, we find that EPO significantly enhances the performance of industry momentum and time series momentum factors, adding significant alpha beyond the market, the 1/N portfolio, risk parity, and standard asset pricing factors.

Keywords: portfolio choice, optimization, robustness, Black-Litterman, machine learning

JEL: C58, C61, G11, G14

* Lasse H. Pedersen is at AQR Capital Management, Copenhagen Business School, and CEPR. Abhilash Babu and Ari Levine are at AQR Capital Management. AQR Capital Management is a global investment management firm, which may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the author and not necessarily those of AQR. Pedersen gratefully acknowledges support from FRIC Center for Financial Frictions (grant no. DNRF102).

Investors seek to construct portfolios that optimally trade off risk and expected return. A standard tool to achieve this goal is mean-variance optimization (MVO), but MVO often produces large and unintuitive bets that perform poorly in practice. Indeed, it has proven surprisingly difficult to find optimization methods that beat the simple 1/N portfolio that allocates capital (or risk) equally across securities (DeMiguel, Garlappi, and Uppal 2007). Perhaps as a result, many investors skip optimization altogether. Likewise, standard academic factors that bet on such characteristics as value (HML), size (SMB), and momentum (UMD) are constructed without the use of optimization or, in fact, any use of 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 is performance improved?

This paper seeks 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 methods in the literature, and (4) that the fix works surprisingly well. More specifically: (1) it is well-known that the problems with standard MVO arise due to 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 down-weight these problem portfolios. We provide a simple closed-form solution that makes EPO as simple to implement as standard MVO. (3) The method unifies and extends a broad range of existing methods, showing how these methods work by shrinking the problem portfolios. (4) Finally, we find empirically that 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 notional-weighted or risk-weighted time-series momentum portfolios. Likewise in equities, we find large performance improvements relative to standard factors when applying 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 stronger academic factors.

[1] There is a large literature on estimation noise, see e.g. Ledoit and Wolf (2003, 2004) on noise in variance-covariance matrices and Black and Litterman (1992) on noise in expected returns.

To understand the poor performance of standard MVO, consider how optimization works in practice. An investor (or researcher) first identifies the securities that she likes and dislikes, or, said differently, estimates securities' expected returns. Then she estimates securities' risks (volatilities and correlations). All these estimates naturally have measurement errors, which cause problems for MVO. Indeed, partly due to measurement errors, the risk model may imply that certain (long-short) portfolios have little risk at a given time, and nothing ensures that the estimated expected return of these portfolios is correspondingly small. Hence, these portfolios appear to have sizeable Sharpe ratios (SRs) driven by noise, which is the source of trouble for MVO. Indeed, MVO seeks to take sizeable risk in such portfolios with sizeable estimated SRs, and, to achieve this, the optimizer takes large notional positions (possibly applying significant leverage). These large notional positions are problematic because the true SRs of these noise-driven portfolios are usually close to zero.

What are these problem portfolios? 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 ranked by their importance, namely their variance. Working with principal components greatly simplifies the diagnosis of the problems with standard MVO, since principal components are by definition uncorrelated, which in turn means 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 the standard MVO. Indeed, these portfolios have the lowest estimated risk and, as a result, their risks tend to be slightly underestimated as seen in Figure 1.A. (Figure 1 is explained in detail in Section III.B.) Further, while 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 seen in Figure 1.B. As a result, from the perspective of standard MVO, these problem portfolios have large estimated Sharpe ratios as seen in Figure 1.C. Therefore, MVO takes large risks in these portfolios as seen in Figure 1.D. These large risk exposures are problematic because these bets perform poorly in practice as seen on their low realized Sharpe ratios in Figure 1.C.

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 Figure 1.C. Reducing estimated Sharpe ratios of the least important principal components can be achieved by increasing their estimated volatilities. Further, 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 new insights are: (i) that correlation shrinkage can fix both errors in risk and expected return, and (ii) this can be achieved by choosing the shrinkage parameter to maximize the portfolio's Sharpe ratio (out of sample), in contrast to the existing literature that chooses correlation shrinkage to maximize the fit of the correlation (or variance-covariance) matrix. Tuning to maximize Sharpe ratio yields a much larger shrinkage parameter, which yields a large performance improvement empirically, and is motivated by the theory that we develop. Indeed, recall that shrinking correlations of the original assets corresponds to shrinking the ex ante volatilities of the problem portfolios, which further corresponds to shrinking their Sharpe ratios, so this shrinkage addresses *both* errors in correlations and expected returns.

This new insight – the power of tuning correlation shrinkage to maximize risk-adjusted returns, not just risk – has deeper theoretical foundations based on Bayesian estimation and robust optimization. Indeed, we solve a new form of robust optimization, showing that uncertainty about expected returns leads endogenously to shrinkage of correlations, even when correlations are known without error. Further, we show that the solution to this robust optimization equals the solution to the seminal model of Black and Litterman (1992). In addition to unifying these approaches, a key contribution is to explain *why* these methods work, namely because they shrink correlations, which fixes the problem portfolios (which is not transparent from the equations of Black and Litterman (1992), who focus on shrinkage toward the market portfolio, a separate effect discussed below, which appears less important in our applications).

To see how the simple EPO works in practice, consider a shrinkage parameter $w \in [0,1]$. First, the off-diagonal correlation Ω_{ij} between any pair of assets i and j is replaced by $(1-w)\Omega_{ij}$, and then we perform MVO using this modified variance-covariance matrix. That's it! Note that this is very easy to do.

When the EPO parameter is $w = 0$, there is no shrinkage, so this 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 standard Fama-French factors and even more similar to 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.

Note that this result is not simply saying that "averaging portfolios works".[2] 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, $ax^{w=0} + (1-a)x^{w=1}$, then this 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. This order is useful because shrinking the correlations stabilizes the optimization process.

How much shrinkage is needed? The simple answer is that this is an empirical question. We empirically choose w out-of-sample as follows: each time period, we estimate what choice of w would have produced the highest SR in the time period up until today, and then use this estimate in the next time period. In several applications, $w = 50\%$ works well. Our theory provides some intuition for this finding. First of all, 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% to 10%. So why do we need a much larger shrinkage of around 50%? As explained above, 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 find strong optimization improvements when we use a surprisingly large amount of shrinkage (surprising from the perspective of what is needed to fix the correlation matrix from a pure risk perspective).

We also develop a more general form of EPO, which 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, hence using the benchmark as an anchor. Or, an investor may have a heuristic way to construct his portfolio – say splitting his money equally among good stocks (1/N) – and may wish that the optimized portfolio stays close to that anchor.

Empirically, we apply our EPO method to optimize momentum portfolios using several realistic data sets, showing that EPO produces significant performance gains relative to standard benchmarks in the literature. When applied to a universe of global equity indices, bonds, currencies, and commodities, the EPO time series momentum portfolio substantially outperforms several benchmarks that are known to be

[2] Averaging portfolios can produce a portfolio that outperforms each input as shown by Tu and Zhou (2011), but we find that EPO works even better in our samples.

difficult to beat. Indeed, EPO outperforms 1/N portfolios, notional-weighted time series momentum factors, risk-weighted time series momentum, standard MVO, and MVO with enhanced risk models.

Furthermore, in the context of equity industry portfolios, the EPO industry momentum portfolio significantly outperforms 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 has significant alpha relative to the Fama-French Five-Factor model augmented with a standard industry-momentum factor.

Related literature. Our paper is related to several literatures and, indeed, one of our theoretical contributions is to unify and demystify these seemingly different frameworks.[3] First, some papers focus on improving the variance-covariance estimate using shrinkage (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, Karoui 2008, and Bun, Bouchaud, and Potters 2017). We find that the EPO solution significantly outperforms such approaches since EPO uses a much larger shrinkage to account for noise in estimates of *both* risk and expected returns (as discussed above).

Second, Black and Litterman (1992) pioneered the focus on noise in expected returns. Despite the fame of this paper, it remains mysterious to many readers who find it difficult to apply and difficult to understand where the result is coming from, including what is being assumed and what the parameters mean. While seemingly different, we show that the EPO solution is in fact equivalent to Black and Litterman (1992), 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] Further, we demystify the whole approach by proposing an easy and transparent method (the simple EPO) and by illustrating how it fixes the "problem portfolios".

[3] While we unify several leading approaches to optimization, EPO obviously does not nest all methods. Regarding other methods, see DeMiguel, Garlappi, and Uppal (2007) who consider 14 methods of optimization, finding that none consistently outperform 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 2006) and methods that average several approaches (Tu and Zhou 2011).

[4] While a version of EPO can be shown to be equivalent to Black and Litterman (1992), there are several differences. Indeed, Black and Litterman (1992) always shrink toward the market portfolio while we consider a general anchor (including no anchor), they consider long-short "view portfolios" while we simply consider signals about expected returns, and we allow "double shrinkage" – both of the estimated expected returns and variance-covariance matrix. More importantly, our contribution is to unify this approach with other optimization methods, by showing the link

Third, we link our approach to the literature on robust optimization (see the survey by Fabozzi, Huang, and Zhou 2010 and references therein) 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 (2006) provide an axiomatic foundation for robustness based on ambiguity aversion and uncover a connection between their approach and shrinkage estimators. Raponi, Uppal, and Zaffaroni (2020) is a recent study with strong results for robust portfolio optimization.

Fourth, Britten-Jones (1999) shows that standard MVO can be seen as the regression coefficient when regressing a constant on realized returns. Machine learning has many ways to regularize regressions, and Ao, Li, and Zheng (2019) find that a so-called LASSO regression significantly improves performance. These papers assume that assets have constant expected returns, while we allow signals to vary over time. Further, 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. In a similar spirit, Kozak, Nagel, and Santosh (2020) use a regression with an elastic net penalty for factor selection.

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). Future research can use this approach to enhance other investment approaches or other factors (e.g., Fama and French 1993, 2015).

In summary, we contribute to the literature by (i) identifying the problem portfolios that plague standard MVO, (ii) proposing a very simple method to address the issue (the simple EPO), showing that it works surprisingly well, (iii) demystifying, simplifying, and unifying Black and Litterman (1992), robust optimization, and methods from machine learning, showing both their equivalence and how they work through shrinking the problem portfolios, and (iv) presenting several interesting new empirical applications, showing that EPO leads to a surprisingly large performance improvement for industry momentum and time series momentum factors.

to correlation shrinkage (which is not clear from the equations on their page 42), by presenting a simple, new, and powerful way to operationalize the method, and by documenting empirically how it works.

I. Identifying the problem with standard optimization

We first lay out the standard portfolio choice framework and then show how to identify problem portfolios. The appendix contains a summary of our notation.

A. Standard mean-variance optimization

We 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 collected in the vector of random variables denoted by $r = (r^1, .., r^n)'$.

The investor receives a signal s about the assets and, using this signal, computes the vector of the risky assets' conditional expected excess returns, $\alpha = \mathrm{E}(r|s)$. For now, we assume that the investor ignores potential noise in the signal. Further, rather than considering an abstract signal, we assume for simplicity that the signal is already scaled to be the conditional expected excess return, that is, $\alpha = s$.

Similarly, the investor computes the conditional variance-covariance matrix of excess returns, $\Sigma = \mathrm{var}(r|s)$. The investor starts with a wealth of W_0 and chooses a portfolio, $x = (x^1, .., x^n)'$. Specifically, x^i is the fraction of capital invested in security i, or, said differently, the investor buys $x^i W_0$ dollars worth of security i. Given this portfolio choice, the investor's future wealth is

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

The investor seeks to maximize her mean-variance utility over final wealth with absolute risk aversion parameter $\bar{\gamma}$, which can be written as follows:

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

$$= W_0\left(1 + r^f + x's - \frac{\gamma}{2} x'\Sigma x\right)$$

where $\gamma = \bar{\gamma} W_0$ is the relative risk aversion. Hence, we see that the optimization problem boils down to the following:

$$\max_x \left(x's - \frac{\gamma}{2} x'\Sigma x\right) \tag{3}$$

Based on the first-order condition, $0 = s - \gamma\Sigma x$, we get the standard mean-variance optimal portfolio

$$x^{MVO} = \frac{1}{\gamma}\Sigma^{-1}s \tag{4}$$

This portfolio has the highest possible Sharpe ratio (SR) among all portfolios if the expected excess return and variance are measured correctly, but the MVO portfolio is sensitive to measurement errors. The riskiness of the portfolio naturally depends on the risk aversion γ. A lower risk aversion γ corresponds to a higher leverage, but varying γ does not affect the SR or the relative portfolio weights of the risky securities. If we vary the risk tolerance $1/\gamma$ from zero to infinity, the resulting MVO portfolios trace out the efficient frontier.

B. Identifying "problem portfolios"

We first show how the "problem portfolios" can be identified using principal components of the correlation matrix. To see this, note that the variance-covariance matrix Σ can be decomposed into the correlation matrix Ω and the diagonal matrix of asset volatilities, $\sigma = \mathrm{diag}(\sqrt{\Sigma^{11}}, \dots, \sqrt{\Sigma^{nn}})$, that is,

$$\Sigma = \sigma \Omega \sigma \tag{5}$$

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

By way of background on principal components, we note that the first principal component maximizes the function $h'\Omega h$ subject to $h'h = 1$. In other words, it maximizes the variance $h'\Omega h$ of any portfolio h (in the space of assets that have been scaled to unit volatility, given that we are working with the correlation 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 $h'\Omega h$ 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 Section III.B.

To identify the principal components, we then consider the eigen-decomposition of the correlation matrix,

$$\Omega = PDP^{-1} \tag{6}$$

where P is a matrix whose columns are the principal components (also called eigenvectors) and 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 one, that is, $PP' = I$, so that $P^{-1} = P'$. We consider the portfolio returns given by the principal component portfolios:

$$p = P'\sigma^{-1}r \tag{7}$$

This vector of portfolios has an expected excess return of $s^p = P'\sigma^{-1}s$ and a variance-covariance matrix of D. Since D is diagonal, these principal components portfolios are uncorrelated (by construction). The portfolio optimization problem can be written as

$$x's - \frac{\gamma}{2}x'\Sigma x = (P'\sigma x)'s^p - \frac{\gamma}{2}(P'\sigma x)'D(P'\sigma x) = z's^p - \frac{\gamma}{2}z'Dz \tag{8}$$

where $z = P'\sigma x$ is the vector of portfolio weights for the principal components. We see that the optimal portfolio weight, z, for the principal components is:

$$z^{MVO} = \frac{1}{\gamma}D^{-1}s^p \tag{9}$$

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

$$\underbrace{z_i^{MVO}}_{\substack{\text{notional}\\ \text{position in}\\ \text{portfolio } i}} = \underbrace{\frac{1}{\gamma}\;\underbrace{\frac{s_i^P}{\sqrt{D_i}}}_{\substack{\text{Sharpe}\\ \text{ratio of}\\ \text{portfolio } i}}}_{\substack{\text{desired}\\ \text{volatility for}\\ \text{portfolio } i}}\;\underbrace{\frac{1}{\sqrt{D_i}}}_{\substack{\text{leverage}\\ \text{needed to}\\ \text{achieve a}\\ \text{volatility of 1}\\ \text{for portfolio } i}} \tag{10}$$

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

II. Addressing the problem: Enhanced portfolio optimization

We first discuss how to address the noise in the estimate of risk (section A), then address the noise in the estimate of expected returns (sections B and C), and put the pieces together in a simple way that we use in the empirical section (section D). Finally, Section E summarizes how our method unifies many forms of optimization.

Our theory is like a treasure hunt where we search the world, but, in the end, we find the treasure close to where we started, so readers who are mostly interested in the empirical results can go directly to Section III. More specifically, we work through many advanced techniques, but, in the end, we find the solution close to the MVO where we started, just with shrunk correlations. Further, we propose a very simple, yet powerful, way to choose the shrinkage, namely to maximize the portfolio's out-of-sample performance, but this is all you need to know to read the empirical Section III. Readers who are interested in *why* this simple EPO approach works well and how different optimization techniques are connected should continue right here.

A. Stabilizing correlations: Shrinkage

As discussed above, principal components 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 their estimated variances toward their average. The average variance of these principal component portfolios is 1 (because they are the principal components of the correlation matrix, which has ones along the diagonal). Hence, we can use the modified risks of the principal components:

$$\widetilde{D} = (1-\theta)D + \theta I \tag{11}$$

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:

$$\widetilde{\Omega} = P\widetilde{D}P' = P\big((1-\theta)D + \theta I\big)P' = (1-\theta)\Omega + \theta I \tag{12}$$

Hence, we see that the adjusted correlation matrix is simply the original matrix Ω shrunk toward the identity. 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 problem portfolios while lowering the volatility of the important PC portfolios is the same as multiplying all the correlations of the original assets by* $1-\theta$.

Finally, the corresponding adjusted variance-covariance matrix is

$$\tilde{\Sigma} = \sigma\tilde{\Omega}\sigma \tag{13}$$

In summary, the first insight is that problem portfolios are the least important principal components, and, second, we can mitigate the under-estimation of their risk simply by shrinking correlations.

In the appendix, we explain how modern statistics offers a more sophisticated way to stabilize correlations using what is called “random matrix theory” (RMT). However, we find empirically that EPO with simple correlation shrinkage works as well as EPO based on a RMT risk model.[5]

B. Anchoring expected returns: A Bayesian approach

We next explicitly acknowledge that the investor’s signal s is observed with noise. This section considers a Bayesian approach following Black and Litterman (1992), although with a different way of expressing the solution (and different notation). Section C considers robust optimization, which, surprisingly, delivers the same solution, and Section D shows a simple way to combine enhancements of risk and return estimates.

We first describe the assumptions and then provide some intuition. The investor observes a vector of signals s, which is the true expected return vector μ plus noise:

$$s = \mu + \epsilon \tag{14}$$

[5] RMT is a beautiful theory ideally suited to fix large correlation matrices and the performance of standard MVO with a RMT risk model does outperform MVO with a naïve risk estimate. However, when we use the optimal EPO shrinkage, there is little difference whether we start with the naïve correlation matrix or the sophisticated RMT matrix. Said differently, we can think of the EPO shrinkage of w=50% as being 5% shrinkage to fix the risk model and 45% shrinkage to adjust for errors in expected returns. If we start with the RMT risk model, we do the first 5% in a much more clever way, but, unfortunately, this does not appear to matter given the large total amount of overall EPO shrinkage. The good news is that, either way, the simple EPO method works so well empirically.

where the noise term ϵ is normally distributed with a mean of zero and a covariance of Λ. Standard MVO is based on the idea that the best estimate of the unobserved true expected return μ is simply the signal s. However, if we incorporate a set of "prior beliefs" about μ, we can make a better guess at the true expected return μ after observing the signal s. Thus, we assert the following prior beliefs about the assets' true (unobserved) expected returns vector μ:

$$\mu = \gamma\Sigma a + \eta \tag{15}$$

Here, η represents a random fluctuation in true expected returns. Specifically, η is normally distributed with mean zero and a covariance of $\tau\Sigma$ for some constant τ, which captures the idea that true fluctuations in expected returns are correlated across correlated assets (similar to the assumption of Black and Litterman, 1992, Appendix, point 7).[6] The first term in (15), $\gamma\Sigma a$, is the unconditional average return, which is written (without loss of generality) as a product of the risk aversion γ (defined in section I.A), the variance-covariance matrix of returns Σ (also defined in I.A), and an "anchor portfolio" a. Writing the average return in this way means that the anchor is the investor's "typical portfolio" as explained below.

Let us provide some intuition before we solve this model. The investor is aware that her signal is estimated with error and has a framework (14)-(15) for the nature of this error. This framework involves some standard parameters (the risk, Σ, the signal about expected returns, s, and risk aversion, γ) and some mysterious parameters (Λ, τ, and the anchor portfolio, a). To explain the mysterious parameters, 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 Λ is the amount of measurement error. However, we need not worry too much about these parameters since we show in Section II.D how the "simple EPO" makes all these mysterious parameters disappear! Hence, the simple EPO is easy to apply in practice with no need to consider these parameters. We also consider an "anchored EPO", which makes all the mysterious parameters disappear, except the anchor – since having an anchor can be useful in practice, e.g., to control how much an optimized portfolio deviates from a benchmark.

To understand the anchor, let us consider what happens when there is no shock to expected returns, that is, $\eta = 0$. In this case, the optimal portfolio is $x = \frac{1}{\gamma}\Sigma^{-1}\mu = a$. Hence, the anchor portfolio is the

[6] Expressed in a different way, the true expected returns of principal component portfolios is $P'\sigma^{-1}\mu = \gamma P'\sigma^{-1}\Sigma a + P'\sigma^{-1}\eta$, where the noise term $P'\sigma^{-1}\eta$ has variance $\tau P'\sigma^{-1}\Sigma\sigma^{-1}P = \tau D$, implying that the expected returns of the least important principal components vary the least.

optimal portfolio when no fluctuations in expected returns have created unusual investment opportunities. Hence, we can think of the anchor as the investor's benchmark, strategic asset allocation, or her typical investment strategy. The investor deviates from this anchor when there are special investment opportunities captured by fluctuations in expected returns, $\eta \neq 0$.

To understand the anchor at a deeper level, consider again the case of $\eta = 0$. In this case, the expected excess return on any asset, say asset number 1, is:

$$E(r_1) = \gamma\ (1,0,\dots,0)\Sigma a = \gamma\ \mathrm{cov}(r_1, r_a|s)$$

Using this relation for the anchor portfolio a and solving for $\gamma = E(r_1)/\mathrm{var}(r_a|s)$, we get

$$E(r_1) = \frac{\mathrm{cov}(r_1,r_a|s)}{\mathrm{var}(r_a|s)} E(r_a) =: \beta_{1,a} E(r_a) \tag{16}$$

If the a is the market portfolio, this relation is simply the conditional capital asset pricing model (CAPM). Hence, the equation defining μ means that the CAPM holds on average, but η pushes the expected returns around such that the CAPM does not exactly hold in each state of nature, thus creating trading opportunities. More generally, (16) says that the anchor is the tangency portfolio when there are no shocks ($\eta = 0$).

To solve the model, we first compute the investor's view on expected returns based on her signal and prior given in (15), namely $E(\mu|s)$. Given that the investor maximizes her mean-variance utility (as defined in Section I), the solution to the enhanced portfolio optimization problem is then $\frac{1}{\gamma}\Sigma^{-1}E(\mu|s)$. The following proposition summarizes the result, and all proofs can be found in the appendix.

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

$$E(\mu|s) = \Sigma(\tau\Sigma + \Lambda)^{-1}(\tau s + \gamma\Lambda a) \tag{17}$$

and the solution to the enhanced portfolio optimization problem is

$$x = \frac{1}{\gamma}(\tau\Sigma + \Lambda)^{-1}(\tau s + \gamma\Lambda a) \tag{18}$$

Interestingly, the optimal portfolio (18) looks like the solution to an MVO where *both* the mean and variance have been modified even though, here, we have only assumed that there are errors in the means. That is, errors in expected returns alone lead to the shrinkage of correlations, even when correlations are assumed to be known without error.

C. Anchoring expected returns: Robust optimization

An alternative approach to address noise in the expected return 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 are interested in maximizing quadratic utility, but we want to be robust to potential errors in the signal about expected returns. One way to do this is to say that we want to choose the portfolio that gives the highest utility even if the expected return is the worst possible, within some uncertainty region:

$$\max_x \min_\mu \left((x-a)'\mu - \frac{\gamma}{2}x'\Sigma x\right) \text{ s.t. } \mu \in \{\bar{\mu} \mid (\bar{\mu}-s)'\Lambda^{-1}(\bar{\mu}-s) \leq c^2\} \tag{19}$$

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 μ can deviate from s, and we wish to ensure a good performance even for the worst possible μ. The expected return μ lies somewhere in the "uncertainty region" (as it is called in robust optimization) around the observed signal s. The uncertainty region is ellipsoidal and the parameters Λ and c control its shape and size. Said differently, Λ and c measure the amount of measurement error, but worry not, these mysterious parameters disappear in the simple EPO in Section II.D. Lastly, we have introduced the "anchor portfolio" a in the portfolio problem. We interpret a as a benchmark portfolio that we wish to outperform (or are afraid of underperforming), e.g., the market portfolio.[7] The solution is given in the following proposition.

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

$$x = \frac{1}{\gamma}(\tau\Sigma + \Lambda)^{-1}(\tau s + \gamma\Lambda a) \tag{20}$$

where τ depends on c (and the set of solutions for $c \in \mathbb{R}_{++}$ equals the set of solution for $\tau \in \mathbb{R}_{++}$).

Surprisingly, the optimal portfolio (20) is exactly the same as the solution in Section II.C. This main result of this section is useful for two reasons: First, it shows how robust optimization can be done via shrinkage of the mean and variance-covariance matrices. Second, it provides a new link between robust optimization and Bayesian optimization. What is the intuition behind this link? Both methods capture the ideas that the signal s contains imperfect information about the conditional expected returns, that the

[7] To our knowledge the specification (19) and its solution is new, but Fabozzi, Huang, and Zhou (2010) consider a version of (19) that is simpler in two ways: first, whereas we consider a general Λ, Fabozzi et al. assume that Λ equals Σ, which means that there is no shrinkage of the variance-covariance matrix, and, second, Fabozzi et al. does not have an anchor portfolio.

amount of noise in the signal is related to Λ, and that there exists an anchor portfolio a that one might not want to deviate too much from.

D. Putting optimization to work: EPO, the simple EPO, and the anchored EPO

We have discussed above that estimation errors occur in both the variance-covariance matrix and in expected returns. Hence, our solution is to first fix the problem with the variance-covariance matrix as discussed in Section II.A, giving rise to the enhanced risk estimate $\tilde{\Sigma}$. Second, we use the methods for anchoring the expected returns in Sections II.B-C, leading to the EPO solution:

$$EPO = \frac{1}{\gamma}\left(\tau\tilde{\Sigma} + \Lambda\right)^{-1}(\tau s + \gamma\Lambda a) \tag{21}$$

We next discuss some simple ways to apply EPO in practice, which will also be useful in our empirical tests. The general EPO solution in (21) depends on several parameters, some of which are straightforward to estimate, while others are more tricky, so it is useful to provide some guidance on the tricky ones. 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 Section II.A (or using random matrix theory as discussed in appendix). The signal about expected returns, s, is your favorite predictor of returns. (To be clear, predicting returns is never easy, but you would probably not be interested in portfolio optimization if you didn't have some predictors to optimize.). The more tricky 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, $\Lambda = \lambda V$, where λ is a constant and V is the diagonal matrix of variances. Recall that we use the notation σ for the diagonal matrix of volatilities so we have $V = \sigma^2$. 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.[8] Under this assumption, the EPO solution can be written as

$$EPO(w) = {\Sigma_w}^{-1}\left([1-w]\frac{1}{\gamma}s + wVa\right) \tag{22}$$

where Σ_w is a shrunk variance-covariance matrix

[8] The assumption of independence of errors in the expected returns across securities, $\Lambda = \lambda V$, implies that the error in the measurement of the expected return of the principal components has a variance given by $P'v^{-1}(\lambda V)v^{-1}P = \lambda I$, that is, errors of all principal components are independent and of equal magnitude.

$$\Sigma_w = [1-w]\tilde{\Sigma} + wV = \sigma\{[1-w]\tilde{\Omega} + wI\}\sigma \tag{23}$$

and we denote $w = \lambda/(\tau+\lambda) \in [0,1]$ as the "EPO shrinkage parameter". Let us provide some intuition for this result. First, we see from (23) that shrinking the variance-covariance matrix $\tilde{\Sigma}$ towards the diagonal matrix of variances, V, means that the correlation matrix, $\tilde{\Omega}$, is shrunk toward the identity, that is, all correlations are shrunk toward zero.

Second, the EPO solution (22) looks as if the variance-covariance matrix has been shrunk and expected returns have been modified to

$$[1-w]s + w\gamma Va \tag{24}$$

This "pseudo expected return" is a weighted average of the signal s and γVa, where we say "pseudo" because the actual posterior expected returns are given by $E(\mu|s)$ in equation (17) above. In other words, we see that the EPO portfolio has components of the MVO portfolio $\frac{1}{\gamma}\tilde{\Sigma}^{-1}s$ and the anchor portfolio $a = V^{-1}(Va)$, but, rather than blending these portfolios directly (as in $[1-w]\frac{1}{\gamma}\Sigma^{-1}s + wa$), the EPO solution instead separately "blends" the variance-covariance matrices and the expected returns. This separate blending is useful, because the standard MVO can be a wildly ill-behaved portfolio, so blending at the portfolio level may not work.

The shrinkage parameter w plays a key role in our empirical implementation. We see from (22) that w is the weight of the anchor a relative to that of the signal s. For example, a shrinkage of $w = 0$ gives the standard MVO solution, while $w = 100\%$ yields the anchor portfolio. More generally, the shrinkage parameter w summarizes the information in the magnitudes of the two sources of uncertainty, namely the error in the estimation of expected returns, λ, and true variation in expected returns, τ. A benefit of this expression is that two of the tricky parameters (λ and τ) disappear, and we only need to keep track of their relative magnitude via the shrinkage w. In the empirical implementation, we will choose the shrinkage in a pragmatic way, namely the value that yields the best risk-adjusted returns, showing how to do this both in-sample and out-of-sample. Since w becomes an empirical choice variable, we write the EPO solution as a function of w, that is, $EPO(w)$.

The remaining tricky parameters are the anchor and the risk aversion. A particularly simple expression arises if we choose the anchor portfolio[9] as $a = \frac{1}{\gamma}V^{-1}s$. In this case, we have

$$EPO^s(w) = \frac{1}{\gamma}\Sigma_w^{-1}s \qquad (25)$$

where Σ_w is the shrunk variance-covariance matrix from (23), and EPO^s stands for "simple EPO". Remarkably, this is the same expression as the as the solution to a standard MVO, except that the correlations (or variance-covariance matrix) have been shrunk. So, surprisingly, both errors in the estimation of mean and variance make it helpful to shrink the correlations. I.e., 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, and these errors may be much larger than the errors in the correlation matrix itself.

Further, the simple EPO solution given in equation (25) is linear in the risk tolerance, so performance statistics such as the Sharpe ratio do not depend on the risk aversion, γ. Therefore, this expression is straightforward to implement, e.g. by setting $\gamma = 1$ or any other number that corresponds to a desirable level of risk.

The more general EPO solution (22) is useful in our empirical implementation if we don't want to tie the anchor to the signal. For example, we may have a signal s about the assets' expected returns based on their momentum, and an anchor a based on the 1/N portfolio. In this case, we know all the inputs in (22), except w, which we choose empirically, and the 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.[10] However, using equation (22) with a γ chosen based on risk aversion, say $\gamma = 3$, requires that the signal s is measured in the right "units". Specifically, the signal must not just predict returns, it should be scaled so that 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 don't really know the scale. For example, if an asset's past momentum predicts that it will outperform in the future, but we don't know by how much. Or, as another example, if the signal is a relative ranking of securities based on their

[9] Alternatively, we can think of the anchor being $a = 0$, which gives the same as (25) 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 $a = \frac{1}{\gamma}V^{-1}s$ is the more natural interpretation of (25).

[10] 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.

valuations. In this case, 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 (22) is essentially a mixture of the anchor portfolio a and the portfolio $\Sigma_w^{-1}\frac{1}{\gamma}s$, so we can pick γ to equalize the variance of these portfolios:[11]

$$\gamma = \frac{\sqrt{s'\Sigma_w^{-1}\tilde{\Sigma}\Sigma_w^{-1}s}}{\sqrt{a'\tilde{\Sigma}a}} \tag{26}$$

which yields:

$$EPO^a(w) = \Sigma_w{}^{-1}\left([1-w]\frac{\sqrt{a'\tilde{\Sigma}a}}{\sqrt{s'\Sigma_w^{-1}\tilde{\Sigma}\Sigma_w^{-1}s}}s + wVa\right) \tag{27}$$

Here, $EPO^a(w)$ indicates our "anchored EPO" solution for the anchor a based on the shrinkage parameter w, where risk aversion is chosen endogenously. We next turn to our empirical implementations, which is based on the simple EPO from equation (25) and the anchored EPO from equation (27).

E. A unified approach to optimization

In summary, we have derived a general enhanced portfolio optimization method (21) and two straightforward ways to implement this method, the simple EPO (25) and the anchored EPO (27). We have already seen that the EPO method is related to several other approaches to portfolio optimization and, as seen in the following proposition, the method in fact has even broader links to the literature.

Proposition 3. *The EPO solution (21) is equal to*

- *a)* ***standard MVO*** *when the estimate of variance has no noise so $\tilde{\Sigma} = \Sigma$ and the signal of expected returns has no noise so $\Lambda = 0$.*
- *b) the anchor when $\tau = 0$ as in* ***reverse MVO****.*
- *c) the* ***Bayesian estimator*** *from Section II.B, which is equivalent to* ***Black-Litterman*** *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.*

---

[11] There are several other related ways to choose γ, some of which work better than others. For example, while γ in (32) equalizes the variance of the anchor with that of $\frac{1}{\gamma}\Sigma_w^{-1}$s, one could also replace the latter with the variance of the standard MVO solution, $\frac{1}{\gamma}\tilde{\Sigma}^{-1}$s, but this is a poor choice if the standard MVO is ill-behaved. Ao, Li, and Zheng (2019) and Raponi, Uppal, and Zaffaroni (2020) also consider methods where γ is based on variance.

d) *the solution to* ***robust optimization*** *with ellipsoidal uncertainty set as defined in Section II.C.*

e) *a* ***generalized ridge regression*** *(a form of regularization used in machine learning) of expected returns on the variance-covariance matrix. Specifically, the general EPO is the solution to a* ***Lavrentyev regularization*** *and the simple EPO to a* ***Tikhonov regularization****. The simple EPO can also be seen as a ridge regression of a vector of ones on the matrix of realized returns when risk and expected returns are estimated by their sample counterparts.*

The proposition shows how the EPO method helps unify seemingly unrelated strands of literature. It is interesting to further discuss these methods. Regarding parts a) and b), EPO obviously contains as special cases the standard MVO and the anchor, which is trivial in itself, but the point is that, by nesting these approaches, you get an enhanced version of things you already know. Further, when using expected returns that imply that the anchor is the optimal portfolio (part b), this is called reverse-MVO in the context of an optimization that includes a set of constraints. This is because the optimal portfolio is taken as given, and optimization is performed with the "implied expected returns," $E(\mu|s) = \tilde{\Sigma}a$, which is the expected return that makes the anchor portfolio optimal in the absence of constraints.

Regarding point c), we see that the Bayesian estimator from Section II.B is connected to the Black and 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 and Litterman (1992). Black and Litterman (1992) always take the anchor portfolio to be the market portfolio, they consider certain "view portfolios" rather our direct focus on a signal about expected returns, they consider a relatively small set of assets, and they ignore noise in the variance-covariance matrix. In contrast, we focus on different anchors including the simple EPO where the anchor essentially disappears, we use the shrinkage parameter as the key tuning variable, we consider noise in both estimates of risk and expected return, and we consider a number of data sets with many more assets.

Regarding d) and e), it is interesting, and far from obvious, that the Bayesian estimator corresponds to both to robust optimization (as derived in Section II.C) and regularization methods used in other strands of statistics and machine learning (not discussed previously in the paper, but the proof in the appendix describes ridge regressions and Lavrentyev/Tikhonov regularizations). Further, the machine learning literature has methods that may be useful for future research on the optimal choice of the EPO shrinkage parameter (e.g., cross-validation, methods to select the Tikhonov factor, and method for generalized ridge regressions such as Golub, Heath, and Wahba, 1979).

III. EPO in Practice: Empirical results

A. Data and methodology

Our empirical implementation constructs optimized industry momentum and time series momentum portfolios using 11 different samples that differ in terms of their test assets and methodology as summarized in Table 1. The first three samples “Global 1” to “Global 3” consist of equity indices, bond futures, commodities, and currencies, while “Equity 1” through “Equity 8” consist of equity portfolios, as we describe in detail next. The samples consist of different data sets and methodologies in order to examine the robustness of the EPO method.

Test assets and data. Our data for Global 1-3 consists of 55 liquid futures and forwards described in Moskowitz, Ooi, and Pedersen (2012). Specifically, we include every equity, commodity, and bond futures contract used in Moskowitz, Ooi, and Pedersen (2012), as well as the nine currency pairs that involve the US dollar (USD). We exclude non-USD cross-currency pairs to ensure that the variance-covariance matrix is full rank.[12] For each instrument, we construct a return series by computing the daily excess return of the currently most liquid contract, and then compound daily returns to a cumulative return index from which we can compute returns at any horizon. The data starts in 1970 and we extend this data through 2018. Following Moskowitz, Ooi, and Pedersen (2012), we start the backtest in 1985, at which time we have data for a broad set of instruments. Further, having the earlier data allows us to choose an initial out-of-sample EPO shrinkage parameter without shortening the time series relative to the earlier study.

The samples for Equity 1-7 are the 49 value-weighted US equity industry portfolios from Ken French’s website. Equity 8 splits each industry portfolio into two components, for a total of $2 \times 49 = 98$ test assets. Specifically, using the CRSP data on the underlying stocks, we compute a “high-momentum” and “low-momentum” portfolio within each of these 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 similarly for the high-momentum portfolio. To calculate excess returns of all equity portfolios, we subtract the one-month US Treasury bill rate, also sourced from Ken French’s website. The equity portfolio data begins in 1927 and ends in 2018. To ensure that there is enough data to select an initial out-of-sample EPO shrinkage parameter using only past information, we evaluate EPO performance using a

[12] If we include non-USD currency pairs, then the variance-covariance matrix is not of full rank since, for example, EUR-USD, EUR-JPY, and USD-JPY are linked through a triangular arbitrage.

sample period beginning 15 years after data is first available (as we do for Global Assets 1-3), meaning that all equity backtests run from 1942 to 2018.

Benchmark factors. We evaluate the returns of optimized equity industry momentum portfolios relative to the Fama-French Five-Factor Model, using monthly returns from Ken French's website. We also evaluate optimized global asset time series momentum portfolios relative to time series momentum benchmarks constructed using the aforementioned "global assets" data following Moskowitz, Ooi, and Pedersen (2012). In particular, we consider the equal-notional-weighted TSMOM factor with the following notional positions:

$$x_t^{\text{TSMOM, notional-weighted}} = \frac{1}{n_t}\, sign(r_{t-12,t}^i) \tag{34}$$

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. Likewise, we compare to the equal-risk weighted TSMOM factor with notional positions given by

$$x_t^{\text{TSMOM, risk-weighted}} = \frac{1}{n_t}\frac{40\%}{\sigma_t^i}\, sign(r_{t-12,t}^i) \tag{35}$$

Optimization methods. Table 1 also shows the optimization methods that we consider. To show the robustness of our results, we consider different optimization methods, different signals about expected returns, and different ways to estimate the risk. Specifically, we use the simple EPO method from equation (25) in the sample of Global Assets as well as in Equity 1-5, and Equity 8, while we use the anchored EPO method from equation (27) in Equity 6-7. 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 gives equal risk weight to each industry portfolio. Specifically, this portfolio has a notional weight on industry i given by $(\sigma_t^i)^{-1}/\sum_j(\sigma_t^j)^{-1}$, where σ_t^i is the estimated volatility of industry i at time t.

Risk models. Table 1 further shows how risk is estimated, again considering different methods for robustness. For Global 1, we use a method similar to that of commercial risk models. The volatility of each instrument is estimated using exponentially-weighted daily returns with a 60-day center-of-mass. The correlations, $\widetilde{\Omega}^{\text{Global 1}}$, are estimated using exponentially-weighted 3-day overlapping returns with a 150-day center-of-mass.[13] We use 3-day returns, $r_{i,t}^{3d} = \sum_{k=0}^{2} r_{t-k}^i$, to mitigate the effects of asynchronous

[13] The annualized variance of instrument i is estimated as $(\sigma_t^i)^2 = 261\sum_{k=0,\dots,\infty}(1-\delta)\delta^k(r_{t-1-k}^i - \bar{r}_t^i)^2$, where $\bar{r}_t^i$ is the is the exponentially-weighted average return computed similarly, 261 annualizes the daily returns, and δ is

trading among global assets, which affects correlations, but not volatilities. For Global 2, we use the same risk model as Global 1, except that all non-diagonal correlations are shrunk 5% toward zero, $\widetilde{\Omega}^{\text{Global 2}} = .95\, \widetilde{\Omega}^{\text{Global 1}} + .05\, I$. For Global 3, we start with the risk model of Global 1, and then enhance the model using random matrix theory, $\widetilde{\Omega}^{\text{Global 3}} = RIE(\widetilde{\Omega}^{\text{Global 1}})$, as described in the appendix, with n set to the number of securities available at each given point in time, and $T = 300$ is set to twice the center-of-mass of 150 days. We then combine each of these correlation matrices with the diagonal matrix σ of volatility estimates to arrive at the variance-covariance matrix $\tilde{\Sigma}$, as defined in equation (13).

For the equity samples, we start with the standard equal-weighted estimates of variances and covariances using, respectively, 60-months, 40-days, and 120-days of data.[14] We then shrink all off-diagonal correlations (or, equivalently, covariances) 5% toward zero.

Signals about expected returns. Lastly, we need a signal about expected returns in each sample. To have a simple signal that we know correlates with future returns, we consider past 12-month returns.[15] While we use past 12-month returns as our signal for simplicity throughout this analysis, note that our EPO method is general and can be used to optimize any predictor of future returns (or combination thereof), not just those predictors based on past returns. For Global 1-3, we consider time series momentum (TSMOM), meaning that the signal of expected return for each instrument is related to its past 12-month excess return. Specifically, we use the following signal about the expected return of instrument i in month t:

$$s_t^i = 0.1 \times \sigma_t^i \times sign(r_{t-12,t}^i) \tag{36}$$

This assumption means that each instrument has a positive expected excess return when the sign of the past 12-month excess return is positive, and otherwise a negative expected excess return, and that the monthly Sharpe ratio for each asset is constant and equal to 0.1. The constant Sharpe ratio assumption is consistent with the implicit assumption of Moskowitz et. al (2012) since they use a constant volatility

---

chosen to achieve a center of mass of $\sum_{k=0,\dots}(1-\delta)\delta^k k = \delta/(1-\delta) = 60$ days. The correlations are 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 ratio of the covariance to the product of volatilities. We require that at least 300 days of data are available for an asset before it enters the covariance matrix.

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

[15] Some studies consider longer time horizons, e.g. past 5-year returns. However, past long-term returns 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 based on a faulty signal of expected returns is not informative.

target for each asset. The scaling of 0.1 is consistent with the average realized Sharpe ratios reported by Moskowitz et. al (2012) and, more recently, by Babu et al. (2019), but this is inconsequential for the Sharpe ratio of the final EPO portfolio.[16] To ensure that the anchor portfolio exactly matches the TSMOM strategy of Moskowitz et. al (2012), we use a risk aversion coefficient of $\gamma_t = \frac{n_t}{4}$, where n_t is the number of instruments at time t. Indeed, this coefficient implies that the anchor portfolio, $a_t = \frac{1}{\gamma_t} V_t^{-1} s_t$, has a notional exposure to asset i that matches that of Moskowitz et. al (2012) given in equation (35), that is, $a_t^i = \frac{1}{\gamma_t} \frac{s_t^i}{(\sigma_t^i)^2} = \frac{1}{n_t} \frac{40\%}{\sigma_t^i}\ sign(r_{t-12,t}^i)$.

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

$$s_t^i = XSMOM_t^i := c_t (r_{t-12,t}^i - \frac{1}{n} \sum_{j=1,\ldots,n} r_{t-12,t}^j) \quad (37)$$

where the scaling factor c_t is chosen so that the positive and negative signals sum to one, that is, $\sum_i s_t^i 1_{\{s_t^i > 0\}} = \sum_i |s_t^i| 1_{\{s_t^i < 0\}} = 1$.

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

$$s_t^i = \sigma_t^i \times XSMOM_t^i \quad (38)$$

This choice of multiplying by volatility is similar in spirit to the scaling of TSMOM in (34). To see again that this is a natural choice, consider the implications for the anchor portfolio. This anchor has a notional weight of each industry i given by

$$a_t^i = \frac{1}{\gamma} \frac{s_t^i}{(\sigma_t^i)^2} = \frac{XSMOM_t^i}{\gamma\, \sigma_t^i} \quad (39)$$

which is proportional to the Sharpe ratio of its outperformance – an intuitive scaling. Further, the anchor's risk weight in industry i is $\sigma_t^i a_t^i = XSMOM_t^i / \gamma$, implying that, when we sum the absolute value of these risk weights over all instruments, this total risk weight is constant over time (due to the definition of c_t). So this is also an expression of an intuitive scaling if we think that the investment opportunity set is not

[16] Babu et al. (2019) report a median time-series momentum Sharpe ratio per asset of 0.34 per year for traditional assets, i.e., 0.10 per month.

varying much over time. Lastly, for Equity 5, we let $s_t^i = (\sigma_t^i)^2 \times XSMOM_t^i$, which implies that the anchor portfolio $a_t^i = XSMOM_t^i/\gamma$ is proportional to each industry's outperformance.

B. Results for global asset classes: Beating time series momentum

Performance of EPO vs. Benchmark Portfolios. We are ready to present 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 delivers a Sharpe ratio of 0.44, arising from the equity risk premium and similar risk premia in other asset classes. The 1/Sigma portfolio invests an equal amount of risk in each asset; for example, the notional exposure in asset i is $(\sigma_t^i)^{-1}/\sum_j(\sigma_t^j)^{-1}$. We see that this portfolio delivers a higher Sharpe ratio of 0.76. Turning to the standard time series momentum factors, we see that the equal-notional TSMOM factor has a Sharpe ratio of 0.74 and the risk-weighted TSMOM factor has a Sharpe ratio of 1.09. The risk-weighted TSMOM factor already has a very high SR since it already does several of the things that an optimizer can hope 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. So this is a tough benchmark to beat! The only information that it does not use is correlation. Said differently, while the 1/N portfolio is normally difficult to beat, we consider benchmarks that already beat 1/N hands down – so these benchmarks set a high bar.

Nevertheless, the out-of-sample EPO significantly outperforms TSMOM by 14%, delivering a SR of 1.24 in Global 1-2 and 1.23 in Global 3. Recall that these samples differ in their estimation of the risk model, where Global 2 shrinks correlations and Global 3 uses random matrix theory. This performance of the EPO TSMOM portfolio is remarkably strong.

The EPO portfolio relies on a single parameter, the EPO shrinkage parameter w. The out-of-sample EPO chooses this parameter in an expanding fashion, only using data available before each month to decide on the parameter to use next month. It is also informative to consider the performance of EPO when using a constant w, which we call in-sample EPO. The unshrunk EPO with $w = 0$ corresponds to standard MVO, and we see that MVO performs worse than the equal risk TSMOM. In other words, standard MVO does not work here. The fully shrunk EPO with $w = 1$ means that we invest in the anchor portfolio, which is the TSMOM factor by construction. With shrinkage factors in between, we see that performance improves, peaking at an even higher level than the out-of-sample EPO (OOS EPO). Over time, the shrinkage parameter used by the OOS EPO method approaches the optimal in-sample value, but initially OOS EPO uses a lower shrinkage and it takes some time for the out-of-sample process to settle on the optimal shrinkage parameter, explaining why the performance of OOS EPO is a bit below the in-sample maximum SR.

Figure 2 also shows how the realized Sharpe ratios of optimized portfolios vary with the choice of EPO shrinkage parameter. In each of the three samples, Global 1-3, we see that the EPO performance is strong for a wide range of shrinkage parameters, reflecting the robustness of the process. Further, we see that the enhancements of the correlation matrix in Global 2-3 improve the performance relative to Global 1 in the case of $w = 0$ (the left side of the graph) that 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 seen that standard MVO performs poorly while EPO performs strongly, so it is interesting to understand the source of this difference. For simplicity, we illustrate problem portfolios for the sample in Global 1.

Following the ideas in Section I.B, we uncover the problem portfolios as follows. Each month t, we first estimate the volatilities and correlation matrix of global assets Ω_t as described in III.A. We then compute the eigendecomposition of the correlation matrix, $\Omega_t = P_t D_t {P_t}^{-1}$, where $P_t = (P_t^1, \ldots, P_t^{n_t})$ is the matrix, where each column is a principal component (PC) portfolio. We then study the expected returns, ex ante volatilities, and ex ante Sharpe ratios of these monthly-rebalanced PC portfolios, comparing these numbers to the realized counterparts. To compute these statistics, we consider the assets rescaled to have unit volatility, ${\sigma_t}^{-1} r_{t+1}$, which have an ex ante variance-covariance matrix equal to the correlation matrix (where we recall that σ_t is the diagonal matrix of volatilities). Similarly, PC portfolio i has a return $(P_t^i)' {\sigma_t}^{-1} r_{t+1}$ so, based on this time series, we can compute the realized average excess return, volatility, and Sharpe ratio. The ex ante expected return is $(P_t^i)' {\sigma_t}^{-1} s_t$, where the signal s_t about the expected return is given in (36). 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.[17]

[17] Since the number of assets in our sample varies with time, we scale 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 equal the largest number of assets in our sample, 55.

Figure 1 plots the results. Let us first consider Panel A, showing the volatilities of the principal component (PC) portfolios. By construction, PC#1 has the highest ex ante volatility and PC#55 has 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, reflecting that the risk model works reasonably well. However, we do see systematic errors: the least important PCs (those with highest numbers) have higher realized volatilities than their average ex ante volatility. This is due to the fact 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.

We next consider principal component returns, plotted in Figure 1, Panel B. Naturally, realized returns are noisy while 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 less important principal components. Further, we see that realized returns approach zero faster than the expected returns. Said differently, the expected returns appear too high for the least important PCs, adding to the problem identified in Panel A. Any noise in the expected returns of the actual assets leads to non-zero expected return of the unimportant PCs and, since the optimizer can always choose the sign of the portfolio to make a non-zero expected return into a positive expected return, the optimizer wants to take a large position in these noise PCs.

Finally, we see how the problems with risk and expected return interact by looking at the corresponding Sharpe ratios in Figure 1, Panel C. We see a dramatic difference between ex ante and realized Sharpe ratios: realized Sharpe ratios *decrease* with the PC number, while 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 while 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, Figure 1, Panel D shows the relative importance of each principal component for the MVO and EPO portfolios, respectively. Specifically, we plot the realized risk for each principal component of the standard MVO portfolio and the out-of-sample EPO portfolio (where both portfolios are 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 ex-post turns out to be largely betting on noise in past data. Further, the notional weights on the unimportant PC portfolios are even larger, because these portfolios need to be leveraged due to their low risk-per-notional (not shown in the figure). 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 in turn leads to much smaller amounts of realized risk in the unimportant PC portfolios as seen in Figure 1.D.

The Alpha of EPO. Having studied the underlying cause of EPO's economically significant performance improvements, we now calculate the alphas of EPO over passive market exposures and other known factors. Table 3 shows the alphas of out-of-sample EPO for TSMOM (using the Global 2 sample) over several benchmarks (with all variables ex-post standardized to 10% volatility, for comparability of coefficients). Column 1 simply controls for the volatility-adjusted TSMOM factor. We see that the 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-squared. Column 2 further adjusts for the volatility-adjusted long-only portfolio (called 1/Sigma), which also has a good performance, to see if EPO simply benefits from being more long passive market exposures. We see that the alpha remains statistically significant. Column 3 then controls for volatility-adjusted TSMOM strategies in each of the four asset classes to see if EPO statically exploits a different asset allocation strategy. This is a stringent test since we are now controlling for 5 high-performance volatility-adjusted strategies that already implicitly do part of the job that we hope that 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, finding an insignificant alpha, consistent with the dominant performance of EPO.

Leverage and Turnover. Finally, to show that EPO produces realistic and implementable portfolios, we consider the turnover and gross leverage profiles of EPO portfolios. Table 4 shows leverage and turnover statistics for the benchmark portfolios, the out-of-sample EPO portfolio, and the in-sample EPO as a function of the EPO shrinkage parameter. 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. We see that lower EPO shrinkage parameters exhibit larger turnover and more gross leverage. For example, the standard MVO portfolio arising from an EPO shrinkage parameter of 0% has substantially more turnover and leverage than the anchor portfolio arising an EPO shrinkage parameter of 100%. Nevertheless, an EPO shrinkage parameter of 90% yields turnover and leverage similar to the anchor, with a substantial improvement in performance as shown in Table 2. The out-of-sample 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, and substantial performance improvements over the standard TSMOM factors in the literature.

C. 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 range of choices on optimization, risk estimation, and signals about expected returns.

EPO Performance vs. Benchmarks. Table 5 reports the Sharpe ratios of the out-of-sample EPO portfolio, a range of in-sample EPO portfolios depending on the shrinkage parameters, and three benchmarks portfolios. The benchmark portfolios are the 1/N portfolio, a more standard industry-momentum (INDMOM) portfolio (following Moskowitz and Grinblatt, 1999) with notional weights given by $XSMOM_t^i$ in (37), and a standard MVO using unshrunk correlations.

In all cases, we see that the out-of-sample EPO portfolio outperforms 1/N, INDMOM, and the standard MVO portfolio, often by a substantial margin. This robustness of the results is noteworthy given the range of specifications. Recall that Equity 1-3 vary the risk model from 40 days to 60 months, a broad span of risk models. Equity 4 and 5 consider different ways to scale of the signals about expected returns. Equity 6 and 7 consider different implementations of the EPO method, using the anchored EPO rather than the simple EPO, while considering different anchors. Finally, Equity 8 is based on a more granular set of test assets, looking at two portfolios per industry.

The out-of-sample EPO portfolio comes close to realizing the highest Sharpe ratio among all in-sample EPO portfolios in all samples, except for Equity 2, showing the robustness of process. Also, we note that all the out-of-sample EPO portfolios realize higher Sharpe ratios than all the five Fama-French factors despite the fact that the Fama-French factors are based on individual stocks while the EPO factors only rely on industry returns. In fact, the best out-of-sample EPO factors even outperform a portfolio that simultaneously invests in all five Fama-French factors (equal-weighted) over the comparable time period.[18]

Alpha to Standard Factors. Table 6 considers the return after controlling for non-optimized INDMOM portfolio (the anchor) as well as the Fama-French 5-factor model. We see that the alpha is positive in all cases. Further, 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 only significant at a 10% level. For Equity 2-4, the *t*-statistic is above 6, which is highly statistically significant. We note that the weaker risk-adjusted return of Equity 6 may arise due to 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 tradeoff (in the choice of the shrinkage parameter) between stabilizing the optimization and moving toward a long-only portfolio, rather than an INDMOM portfolio, which does not exploit signals about expected returns. Nevertheless, EPO portfolios deliver strong performance across a range of settings, and this strong performance cannot be explained by standard factors.

IV. Conclusion: A Practical Guide to Optimization

We develop 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, robust optimization, and regularization methods used in machine learning – so it demystifies, unifies, and simplifies much of this literature.

To illuminate why standard mean-variance (MVO) optimization techniques often fail, we identify the "problem portfolios" that MVO gives large weight despite their poor performance. Our EPO method addresses this issue via correlation shrinkage, which, perhaps surprisingly, down-weights the problem portfolios.

[18] From 1963 to 2018, the five Fama-French factors realize Sharpe ratios between 0.27 and 0.49, and the equal-weighted portfolio of all five factors realizes a Sharpe ratio of 0.93.

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

When applied to equities, our EPO industry momentum portfolio substantially outperforms the market portfolio, 1/N benchmark, and standard industry momentum. This strong outperformance of EPO cannot be explained by exposures to existing factors in the literature such as the Fama-French factors. Further, the performance enhancements are robust to range of different specifications. While we focus on momentum predictors for simplicity, future research can use this approach to enhance other predictors.

In summary, we provide a simple, yet powerful, method for portfolio construction. We unify several forms of optimization, showing in a transparent way how they work by controlling problem portfolios. Further, we present empirical findings that shed new light on the problems with standard optimization and illustrate the potential gains from enhanced portfolio optimization.

References

Ao, Mengmeng, Li Yingying, and Xinghua Zheng. "Approaching mean-variance efficiency for large portfolios." The Review of Financial Studies 32.7 (2019): 2890-2919.

Asness, C., T. Moskowitz and L.H. Pedersen. “Value and Momentum Everywhere.” The Journal of Finance, Vol. 68, No. 3 (2013), pp. 929–985.

Babu, Abhilash, Ari Levine, Yao Hua Ooi, Lasse Heje Pedersen, and Erik Stamelos, “Trends Everywhere” Journal of Investment Management, forthcoming, 2019.

Black, Fischer, and Robert Litterman. "Global portfolio optimization." Financial Analysts Journal 48, no. 5 (1992): 28-43.

Britten-Jones, Mark. "The sampling error in estimates of mean-variance efficient portfolio weights." The Journal of Finance 54.2 (1999): 655-671.

Bun, Joël, Jean-Philippe Bouchaud, and Marc Potters. "Cleaning correlation matrices." Risk Magazine (2016).

Bun, Joël, Jean-Philippe Bouchaud, and Marc Potters. "Cleaning large correlation matrices: tools from random matrix theory." Physics Reports 666 (2017): 1-109.

De Bondt, Werner FM, and Richard Thaler. "Does the stock market overreact?." The Journal of finance 40.3 (1985): 793-805.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?." The Review of Financial Studies 22, no. 5 (2007): 1915-1953.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal. "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms." Management Science 55, no. 5 (2009): 798-812.

Elton, Edwin J., Martin J. Gruber, and Jonathan Spitzer. "Improved estimates of correlation coefficients and their impact on optimum portfolios." European Financial Management 12.3 (2006): 303-318.

Fabozzi, Frank J., Dashan Huang, and Guofu Zhou. "Robust portfolios: contributions from operations research and finance." Annals of Operations Research 176, no. 1 (2010): 191-220.

Fan, Jianqing, Yingying Fan, and Jinchi Lv. "High dimensional covariance matrix estimation using a factor model." Journal of Econometrics 147.1 (2008): 186-197.

Fama, Eugene F., and Kenneth R. French. "Common risk factors in the returns on stocks and bonds." Journal of Financial Economics 33.1 (1993): 3-56.

Fama, Eugene F., and Kenneth R. French. "A five-factor asset pricing model." Journal of Financial Economics 116.1 (2015): 1-22.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang. "Portfolio selection with parameter and model uncertainty: A multi-prior approach." The Review of Financial Studies 20, no. 1 (2006): 41-81.

Gârleanu, Nicolae, and Lasse Heje Pedersen. "Dynamic trading with predictable returns and transaction costs." The Journal of Finance 68, no. 6 (2013): 2309-2340.

Gârleanu, Nicolae, and Lasse Heje Pedersen. "Dynamic portfolio choice with frictions." Journal of Economic Theory 165 (2016): 487-516.

Golub, G., M. Heath, and G. Wahba. "Generalized cross-validation as a method for choosing a good ridge parameter." Technometrics. 21 no. 2 (1979): 215–223.

Hastie, T., R. Tibshirani, and J. Friedman. “The elements of statistical learning: data mining, inference, and prediction.” (2009). Springer Science & Business Media.

Jagannathan, Ravi, and Tongshu Ma. "Risk reduction in large portfolios: Why imposing the wrong constraints helps." The Journal of Finance 58.4 (2003): 1651-1683.

Karoui, Noureddine El. "Spectrum estimation for large dimensional covariance matrices using random matrix theory." The Annals of Statistics (2008): 2757-2790.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh. "Shrinking the cross-section." Journal of Financial Economics 135.2 (2020): 271-292.

Ledoit, Olivier, and Michael Wolf. "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection." Journal of Empirical Finance 10.5 (2003): 603-621.

Ledoit, Olivier, and Michael Wolf. "A well-conditioned estimator for large-dimensional covariance matrices." Journal of Multivariate Analysis 88.2 (2004): 365-411.

Ledoit, Olivier, and Michael Wolf. "Nonlinear shrinkage estimation of large-dimensional covariance matrices." The Annals of Statistics 40.2 (2012): 1024-1060.

Ledoit, Olivier, and Michael Wolf. "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks." The Review of Financial Studies 30.12 (2017): 4349-4388.

Marčenko, Vladimir A., and Leonid A. Pastur. "Distribution of eigenvalues for some sets of random matrices." Mathematics of the USSR-Sbornik 1.4 (1967): 457-483.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? The Journal of Finance 54, 1249–1290.

Moskowitz, T., Y.H. Ooi and L.H. Pedersen. “Time Series Momentum.” Journal of Financial Economics, Vol. 104, No. 2 (2012), pp. 228–250.

Robust Portfolio Choice——————————————

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

Stein, C. “Inadmissibility of the usual estimator for the mean of a multivariate distribution.” Proc. Third Berkeley Symp. Math. Statist. Prob., 1 (1956): 197–206,

Tu, Jun, and Guofu Zhou. "Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies." Journal of Financial Economics 99, no. 1 (2011): 204-215.

Appendix: Summary of notation, auxiliary results, and proofs

Summary of Notation.

Symbol	Meaning
$r = (r^1, .., r^n)'$	Vector of excess returns
$x = (x^1, .., x^n)'$	Vector of portfolio holdings
γ	Relative risk aversion
$s = (s^1, .., s^n)'$	Vector of signals about expected excess returns
$\Sigma = \text{var}(r\|s)$	Variance-covariance matrix
$\tilde{\Sigma}$	Enhanced risk estimate, e.g. based on RMT
$\Sigma_w = [1 - w]\tilde{\Sigma} + wV$	Shrunk variance-covariance matrix
w	EPO shrinkage parameter
$\sigma = \text{diag}(\sqrt{\Sigma^{11}}, \dots, \sqrt{\Sigma^{nn}}) = \text{diag}(\sigma^1, \dots, \sigma^n)$	Diagonal matrix of volatilities
$V = \sigma^2$	Diagonal matrix of variances
$\Omega = PDP'$	Correlation matrix
P	Matrix whose columns are principal component portfolio weights (eigenvectors)
D	Diagonal matrix of variances of principal component portfolios (eigenvalues)
a	Anchor portfolio
μ	True, but unobserved, expected return
τ	Variation in true expected returns
Λ, λ	Error in the estimation of expected returns

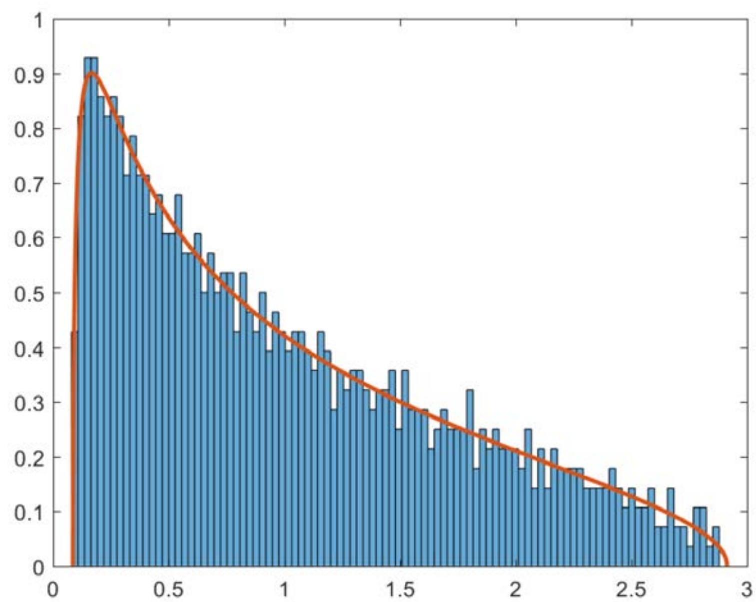
Random matrix theory. Section I.B shows that errors in the estimated risk model leads to problems for MVO. Specifically, small eigenvalues of the variance-covariance matrix give rise to "problem portfolios." These problem portfolios can be accommodated by stabilizing the correlation matrix, but what is the best way to do this?

Section II.A discusses a simple way to stabilize risk, namely shrinking correlations toward zero. How do we choose the shrinkage parameter θ? One approach is to choose a parameter that works well empirically (looking at past data), but one can also use random matrix theory to derive an asymptotically optimal choice (Ledoit and Wolf, 2004). Further, random matrix theory can be used to derive more general forms of stabilized correlation matrices, such as a non-linear shrinkage of the eigenvalues (see Karoui 2008, Ledoit and Wolf 2017, Bun, Bouchaud, and Potters 2017, and references therein).

While standard statistics relies on estimates to be close to the true values when the number of time periods T is large, random matrix theory instead deals with the "big data" environment of modern financial markets, namely when we have large values of both the number of securities n and the number of time periods. Specifically, random matrix theory 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 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 Section II.A, random matrix theory is focused on the eigenvalues of the matrix. One basic result is that, if all returns are independent across securities and time, then the asymptotic distribution of the eigenvalues is known explicitly, and given by Marčenko and Pastur (1967). As seen in the example in Figure A.1, the Marčenko-Pastur distribution fits the distribution of the observed eigenvalues well even in a single sample. This is called the self-averaging property of random matrices.

Figure A.1. Distribution of eigenvalues for independent securities. This figure shows a histogram (blue bars) for the eigenvalues for the correlation matrix for 1000 securities with returns simulated over 2000 days, where the returns are assumed to be iid. Normal. The red line is the Marčenko-Pastur distribution given in (A.1). The true eigenvalues are all one for a correlation matrix of independent securities, but estimation noise creates randomness (smaller and larger estimated eigenvalues), which is well captured by Marčenko-Pastur.



Of course, security returns from real financial data are not independent so the distribution of eigenvalues from real data does not closely fit Marčenko-Pastur. The point is that, based on Marčenko-Pastur, we know what random noise in eigenvalues looks like. In particular, the “bulk” of small eigenvalues inside the Marčenko-Pastur distribution are likely just noise, while 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 normal distribution that we are used to seeing as a limiting distribution in standard statistics.

Random matrix theory offers various methods to "clean" the correlation matrix in the following two steps. First, we replace the estimated eigenvalues $(D_1, \dots, D_n)$ with cleaned eigenvalues $(\widetilde{D}_1, \dots, \widetilde{D}_n)$, while typically leaving the eigenvectors P unchanged.[19] For this cleaning of eigenvalues, we focus here on the "IWs" method described in Bun, Bouchaud, and Potters (2017), which is essentially the same as the RIE method described in Box 1 of Bun, Bouchaud, and Potters (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. We finally recover the cleaned correlation matrix as $\widetilde{\Omega} = P\widetilde{D}P^{-1}$ and the cleaned variance-covariance matrix as $\widetilde{\Sigma} = \sigma\widetilde{\Omega}\sigma$.

To understand Marčenko and Pastur (1967) in more detail, we start with the estimated correlation matrix Ω for n iid. random returns observed over T time periods:

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

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

Random matrix theory 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 Ω converges to the Marčenko-Pastur distribution. When the ratio q satisfies $q \in (0,1]$, then the density f of the Marčenko-Pastur distribution is given by:

[19] Estimates of the eigenvectors are kept equal to the sample eigenvectors in order 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 and Bun, Bouchaud, and Potters, 2017).

$$f(d) = \frac{\sqrt{(q_+ - d)(d - q_-)}}{2\pi q d} \quad \text{(A.1)}$$

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

Proofs.

Proof of Proposition 1. This model yields the following posterior mean for μ:

$$\begin{aligned} E(\mu|s) &= E(\mu) + Cov(\mu, s)\, Var(s)^{-1}(s - E(s)) \\ &= \gamma\Sigma a + \tau\Sigma(\tau\Sigma + \Lambda)^{-1}(s - \gamma\Sigma a) \\ &= \Sigma(\tau\Sigma + \Lambda)^{-1}\tau s + \gamma\Sigma[I - (\tau\Sigma + \Lambda)^{-1}\tau\Sigma]a \\ &= \Sigma(\tau\Sigma + \Lambda)^{-1}\tau s + \gamma\Sigma(\tau\Sigma + \Lambda)^{-1}\Lambda a \\ &= \Sigma(\tau\Sigma + \Lambda)^{-1}(\tau s + \gamma\Lambda a) \end{aligned}$$

where the first equality is due to the standard formula for conditional means of normally distributed random variables (or equivalently, the standard OLS formula for regressing μ on s) and the fourth equality uses the Woodbury matrix identity.[20] ■

Proof of Proposition 2. We first solve the minimization problem inside (19). For this, consider the Lagrangian:

$$L = (x - a)'\mu + l((\mu - s)'\Lambda^{-1}(\mu - s) - c^2)$$

where l is the Lagrange multiplier. Differentiating with respect to μ, we get the first-order condition

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

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

[20] Using Woodbury, we see that $[I - (\tau\Sigma + \Lambda)^{-1}\tau\Sigma] = (I + \Lambda^{-1}\tau\Sigma)^{-1} = (\Lambda^{-1}(\Lambda + \tau\Sigma))^{-1} = (\tau\Sigma + \Lambda)^{-1}\Lambda$.

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

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

$$\max_x \left((x-a)'s - \frac{\gamma}{2}x'\Sigma x - c\sqrt{(x-a)'\Lambda(x-a)}\right)$$

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:

$$\max_x \left(x's - \frac{\gamma}{2}x'\Sigma x - \frac{\gamma}{2\tau}(x-a)'\Lambda(x-a)\right)$$

The first-order condition is

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

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

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

Lemma 1. *For any vector $a \in \mathbb{R}^n$ and positive definite matrices $B, C \in \mathbb{R}^{n\times n}$, the set of solutions to Problem A, $\{x^{*\mathrm{A}}(c)\}_{c\geq 0}$, equals the set of solutions to Problem B, $\{x^{*\mathrm{B}}(d)\}_{d\geq 0}$, where*

Problem A: $\max_x (x'a - x'Bx - c\, x'Cx)$

Problem B: $\max_x \left(x'a - x'Bx - d\sqrt{x'Cx}\right)$

Proof of Lemma 1. For a given c, note that the solution $x^{*\mathrm{A}}(c)$ to Problem A satisfies the first-order condition:

$$0 = a - Bx - 2c\, Cx$$

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

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

We see that the result holds for $d = 2c\sqrt{(x^{*\mathrm{A}}(c))'Cx^{*\mathrm{A}}(c)}$. Similarly, for any given d with corresponding solution $x^{*\mathrm{B}}(d)$ to Problem B, we see that this vector is also a solution to Problem A when we let $c = d/(2\sqrt{(x^{*\mathrm{B}}(d))'Cx^{*\mathrm{B}}(d)})$.■

Proof of Proposition 3. Part a), b) and f) are clear. Regarding part c), the derivation is shown in Section II.B. Regarding the relation to Black and Litterman (1992), we use the superscript BL to indicate their notation. With the relations that $\Pi^{\mathrm{BL}} = \gamma\Sigma a$, $Q^{\mathrm{BL}} = s$, $P^{\mathrm{BL}} = I$, $\Omega^{\mathrm{BL}} = \Lambda$, $\Sigma^{\mathrm{BL}} = \Sigma$, and $\tau^{\mathrm{BL}} = \tau$, their expression in point 8 of their appendix can be shown to equal our expression for the conditional mean:

$$\begin{aligned}
E(\mu|s) &= \Sigma(\tau\Sigma + \Lambda)^{-1}(\tau s + \gamma\Lambda a) \\
&= (\tau I + \Lambda\Sigma^{-1})^{-1}(\tau s + \gamma\Lambda a) \\
&= (\tau\Lambda^{-1} + \Sigma^{-1})^{-1}\Lambda^{-1}(\tau s + \gamma\Lambda a) \\
&= (\tau\Lambda^{-1} + \Sigma^{-1})^{-1}(\tau\Lambda^{-1}s + \gamma a) \\
&= ((\tau\Sigma)^{-1} + \Lambda^{-1})^{-1}(\tau^{-1}\gamma a + \Lambda^{-1}s) \\
&= ((\tau^{\mathrm{BL}}\Sigma^{\mathrm{BL}})^{-1} + (\Omega^{\mathrm{BL}})^{-1})^{-1}((\tau^{\mathrm{BL}}\Sigma^{\mathrm{BL}})^{-1}\Pi^{\mathrm{BL}} + (\Omega^{\mathrm{BL}})^{-1}Q^{\mathrm{BL}})
\end{aligned}$$

Regarding d), the derivation of robust optimization is in Section II.C, using Lemma 1, which is stated and proved in this appendix.

Regarding e), a ridge regression is a method used to mitigate noise and collinearity in a regression setting. Specifically, consider the regression $y = z\beta + \varepsilon$, 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, $(y - z\beta)'(y - z\beta) + \lambda\beta'\beta$. The solution is $\hat{\beta}^{\mathrm{ridge}} = (z'z + \lambda I)^{-1}z'y$, so we see that the symmetric matrix $z'z$ is being pushed toward the identity matrix I, ensuring invertibility. So if expected returns (summarized by 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 (3), $s = \gamma\Sigma x$. That is, we need to solve for the optimal portfolio x based on the noisy data on Σ and s. We rewrite this equation as $\frac{1}{\gamma}\Sigma^{-1/2}s = \Sigma^{1/2}x + \varepsilon$, introducing an error term ε 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).[21] 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" x. The ridge regression estimator is $x = \frac{1}{\gamma}(\Sigma + \lambda I)^{-1}s$, which is closely related to the EPO solution.

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

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

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

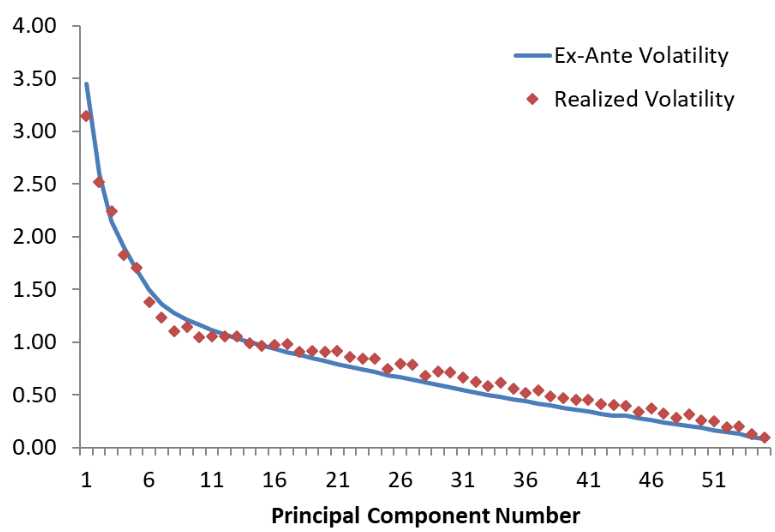
---

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

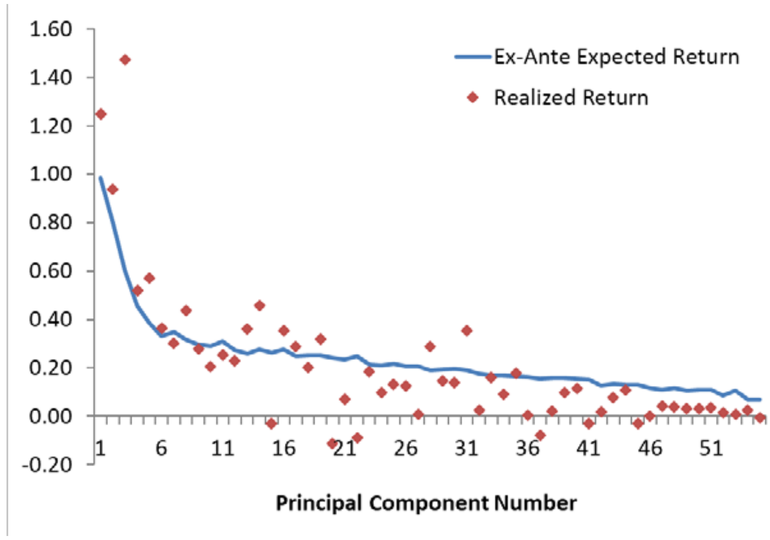
Figures and Tables

Figure 1. Understanding Problem Portfolios. This figure shows a comparison of realized and ex ante performance of principal component portfolios (PCs). The figure shows realized and ex ante volatilities (Panel A), returns (Panel B), and Sharpe ratios (Panel C). Additionally, Panel D shows how much risk that standard MVO takes in each principal component and, similarly, how much risk the out-of-sample EPO takes (both standardized to 10% full-sample volatility, for comparability). The sample consists of monthly data for 55 global equities, bonds, commodities, and currencies, 1985-2018, as specified in Table 1, Global 1. Table 1 also summarizes the method for estimation of risk and expected return. The least important PCs are those with highest numbers, and these are the "problem portfolios" because the ex ante risk model underestimates their realized risk, the ex ante expected return overestimates the realized returns, and the ex ante Sharpe ratios are higher than those of the more intuitive factors (low-numbered PCs) while the reverse is true for realized Sharpe ratios. Therefore, standard MVO invests too heavily in problem portfolios while EPO does not, as illustrated in Panel D.

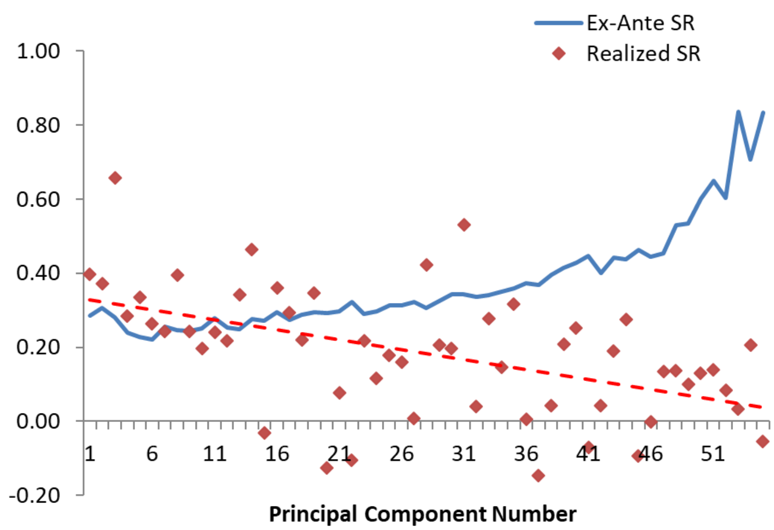
Panel A: Average Ex Ante Volatility and Realized Volatility by Principal Component



Panel B. Average Ex Ante Expected Return and Realized Return by Principal Component



Panel C. Average Annualized Ex Ante Sharpe Ratio and Annualized Realized Sharpe Ratio of TSMOM by Principal Component



Panel D. Realized Risk Allocation of EPO and Standard MVO by Principal Component

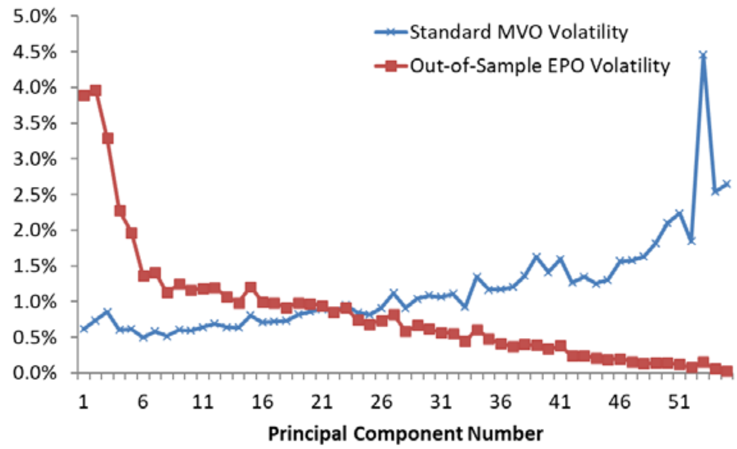


Figure 2. Performance of Optimized TSMOM Portfolios.

This figure shows how the Sharpe ratio of optimized TSMOM portfolios varies with the in-sample EPO shrinkage parameter. We consider the Global 1-3 samples described in Table 1, which consists of three different correlation matrices: the standard sample correlation matrix ("Standard"), a correlation matrix with 5% correlation shrinkage applied ("Shrunk"), and a cleaned correlation matrix based on random matrix theory ("RMT"). For comparison, we also show the performance of selecting the optimal out-of-sample EPO shrinkage parameter using only past data ("EPO (OOS)"). A shrinkage of 0% is standard mean-variance optimization, 100% shrinkage is the anchor portfolio, and in between is EPO.

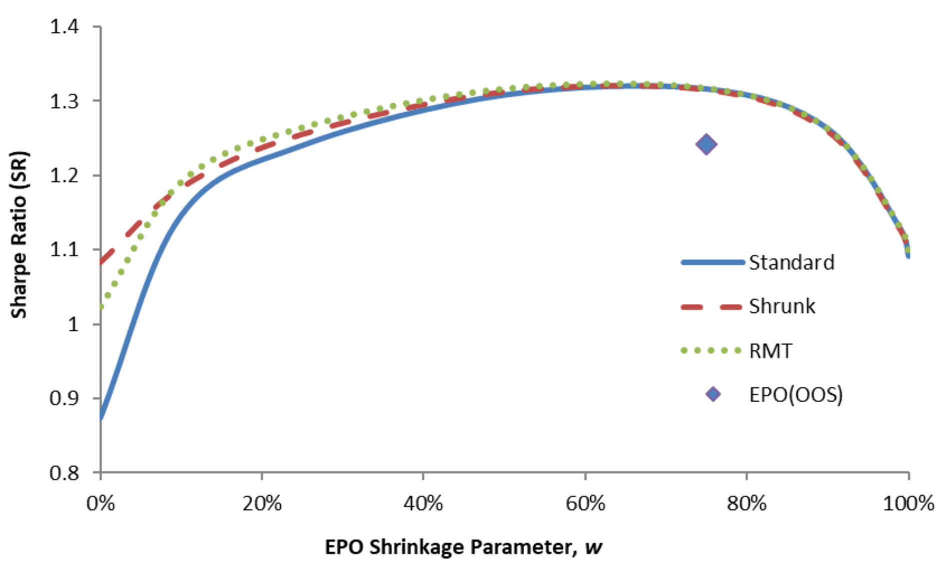


Table 1. Samples and Summary Statistics.

This table illustrates the samples used in our empirical analysis. The table lists the data used, the number of test assets, the methods used, the start date of the data, and the start date of our backtests. All backtests begin 15 years after the earliest initial data is available so we always have at least 15 years of data to select an out-of-sample EPO shrinkage parameter.

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

Table 2. Performance of Optimized TSMOM Portfolios.

This table shows the realized gross Sharpe ratios of different portfolios based on the Global 1-3 samples described in Table 1. The portfolios are the long-only 1/N portfolio that invests with equal notional exposure across all assets, the 1/Sigma portfolio that invests with equal risk in each asset, the TSMOM strategy that invests with equal notional exposure in each asset, the TSMOM strategy that invests with equal risk in each asset, and a range of optimized portfolios. The optimized portfolios are the simple out-of-sample EPO and a range of in-sample EPO portfolios that differ based on their EPO shrinkage parameter w. The out-of-sample EPO chooses w using only past data. With Global 1, the correlation matrix is estimated in the standard way, Global 2 uses a 5% shrunk correlation matrix, and Global 3 uses a cleaned correlation matrix based on random matrix theory (RMT). In all columns, we see that the out-of-sample EPO has a large outperformance relative to standard mean-variance optimization (MVO) and all the benchmark portfolios.

Portfolio	Global 1	Global 2 (Shrunk)	Global 3 (RMT)
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 Risk	1.09	1.09	1.09
EPO^s : Out-Of-Sample	1.24	1.24	1.23
$EPO^s(w)$: In Sample with Shrinkage w			
0% (Naïve 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

Table 3. Alpha of Out-of-Sample EPO for TSMOM.

This table shows the alphas of out-of-sample EPO for TSMOM over various benchmarks using the Global 2 sample described in Table 1. The table reports the alphas of out-of-sample EPO 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 in asset class. We also report alphas of the volatility-scaled TSMOM portfolios over out-of-sample EPO. All variables are ex-post standardized to an annualized full-sample volatility of 10% to make the alphas comparable. This scaling does not affect the *t*-statistics reported in parentheses.

	Dependent Variable				
	EPO			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/Sigma)		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-Squared	83%	84%	85%	83%	83%

Table 4. Leverage and Turnover of Optimized TSMOM Portfolios.

This table reports the gross leverage and annualized turnover of optimized TSMOM portfolios, and various benchmarks using the Global 2 sample from Table 1. For comparability, all statistics are reported for portfolios that are 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.

Portfolio	Gross Leverage per 10% Volatility	Annualized Turnover as % of Avg. Gross NAV
Long Only: 1/N	135%	26%
Long Only: 1/Sigma	267%	43%
TSMOM: Equal Notional Weight	167%	153%
TSMOM: Equal Risk	358%	163%
EPO^s : Out-Of-Sample	457%	254%
$EPO^s(w)$: In Sample with Shrinkage w		
0% (Naïve 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%

Table 5. Performance of Optimized Equity Portfolios.

This table shows the realized gross Sharpe ratios of equity portfolios based on the Equity 1-8 samples described in Table 1. As benchmark portfolios, the table includes the long-only 1/N portfolio that invests with equal notional exposure across all industries, a standard industry momentum (INDMOM) portfolio that is long industries that outperformed over the past 12 months and short industries that underperformed, and a portfolio based on standard mean-variance optimization (MVO) without correlation shrinkage. The optimized portfolios are the out-of-sample and in-sample EPO portfolios. We consider in-sample EPO portfolios for a range of EPO-shrinkage parameters, w. The out-of-sample EPO chooses w using only past data. In all columns, we see that the out-of-sample EPO has a large outperformance relative to the benchmark portfolios.

Portfolio	Equity 1	Equity 2	Equity 3	Equity 4	Equity 5	Equity 6	Equity 7	Equity 8
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 w/ 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

Table 6. Alpha of EPO for Equity Portfolios.

This table shows the performance of out-of-sample EPO portfolios controlling for standard factors using the Equity 1-8 portfolios 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 (since the Fama-French Five Five-Factor model data begins in 1963). All variables are ex-post standardized to an annualized full-sample volatility of 10% to make coefficients comparable. This scaling does not affect the *t*-statistics reported in parentheses.

	Dependent Variable: Out-of-Sample EPO Portfolios							
	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-Squared	64%	29%	28%	49%	46%	83%	87%	64%