Machine Learning and Portfolio Optimization

Gah-Yi Ban, Noureddine El Karoui, Andrew E. B. Lim

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 64, No. 3, March 2018, pp. 1136–1154
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Machine Learning and Portfolio Optimization

Gah-Yi Ban,[a] **Noureddine El Karoui,**[b] **Andrew E. B. Lim**[c]

[a] Management Science and Operations, London Business School, London NW1 4SA, United Kingdom; [b] Department of Statistics, University of California, Berkeley, Berkeley, California 94720; [c] Department of Decision Sciences and Department of Finance, National University of Singapore Business School, Singapore 119245

Contact: gban@london.edu (G-YB); nkaroui@stat.berkeley.edu (NEK); andrewlim@nus.edu.sg (AEBL)

Received: June 24, 2013
Revised: October 14, 2014; February 8, 2016
Accepted: August 16, 2016
Published Online in Articles in Advance: November 21, 2016

https://doi.org/10.1287/mnsc.2016.2644

Abstract. The portfolio optimization model has limited impact in practice because of estimation issues when applied to real data. To address this, we adapt two machine learning methods, regularization and cross-validation, for portfolio optimization. First, we introduce *performance-based regularization* (PBR), where the idea is to constrain the sample variances of the estimated portfolio risk and return, which steers the solution toward one associated with less estimation error in the performance. We consider PBR for both mean-variance and mean-conditional value-at-risk (CVaR) problems. For the mean-variance problem, PBR introduces a quartic polynomial constraint, for which we make two convex approximations: one based on rank-1 approximation and another based on a convex quadratic approximation. The rank-1 approximation PBR adds a bias to the optimal allocation, and the convex quadratic approximation PBR shrinks the sample covariance matrix. For the mean-CVaR problem, the PBR model is a combinatorial optimization problem, but we prove its convex relaxation, a quadratically constrained quadratic program, is essentially tight. We show that the PBR models can be cast as robust optimization problems with novel uncertainty sets and establish asymptotic optimality of both sample average approximation (SAA) and PBR solutions and the corresponding efficient frontiers. To calibrate the right-hand sides of the PBR constraints, we develop new, performance-based k-fold cross-validation algorithms. Using these algorithms, we carry out an extensive empirical investigation of PBR against SAA, as well as L1 and L2 regularizations and the equally weighted portfolio. We find that PBR dominates all other benchmarks for two out of three Fama–French data sets.

History: Accepted by Yinyu Ye, optimization.
Funding: This research was supported in part by the Eleanor Sophia Wood Traveling Scholarship from the University of Sydney (Ban), the London Business School Research and Materials Development Fund (Ban), the National Science Foundation [Grants DMS-0847647 (CAREER) and DMS-1510172] (El Karoui), Alfred P. Sloan Research Fellowship (El Karoui), and the Singapore Ministry of Education Academic Research Fund [Tier 2 MOE2016-T2-1-086] (Lim).
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2016.2644.

Keywords: machine learning • portfolio optimization • robust optimization • regularization • cross-validation • conditional value-at-risk

1. Introduction

Regularization is a technique that is commonly used to control the stability of a wide range of problems. Its origins trace back to the 1960s, when it was introduced to deal with ill-posed linear operator problems. A linear operator problem is one of finding an $x \in X$ that satisfies $Ax = b$, where A is a linear operator from a normed space X to a normed space Y, and $b \in Y$ is a predetermined constant. The linear operator problem is ill-posed if small deviations in b, perhaps due to noise, result in large deviations in the corresponding solution. Specifically, if b changes to b_δ, $\|b_\delta - b\| < \delta$, then finding an x that minimizes the functional $R(x) = \|Ax - b_\delta\|^2$ does not guarantee a good approximation to the desired solution even if δ tends to zero. Tikhonov (1963), Ivanov (1962), and Phillips (1962) discovered that if instead of minimizing $R(x)$, the most obvious choice, one minimizes the *regularized* functional

$$R^*(x) = \|Ax - b_\delta\|_2^2 + \gamma(\delta)P(x),$$

where $P(x)$ is some functional and $\gamma(\delta)$ is an appropriately chosen constant, then one obtains a sequence of solutions that does converge to the desired one as δ tends to zero. Regularization theory thus shows that whereas the self-evident method of minimizing $R(x)$ does not work, the nonself-evident method of minimizing $R^*(x)$ does.

Regularization has particularly been made known in recent years through its adoption in classification, regression, and density estimation problems. The reader may be most familiar with its recent popularity in the high-dimensional regression literature (see, e.g., Candes and Tao 2007, Belloni and Chernozhukov 2013):

$$\min_{\beta \in \mathbb{R}^p} \{\|\mathbf{y} - \mathbf{X}\beta\|_2 + \lambda P(\beta)\}, \tag{1}$$

where $P(\beta) = \|\beta\|_1$, $\mathbf{y} = [y_1, \ldots, y_n] \in \mathbb{R}^n$ are the data on the observable, $\mathbf{X} = [X_1, \ldots, X_n] \in \mathbb{R}^{n \times p}$ is the vector of covariates, $\beta \in \mathbb{R}^p$ is the regression coefficient that best fits the linear model $y = X\beta$, and $\lambda > 0$ is a parameter that controls the sparsity of the solution. The regression model (1) with $P(\beta) = \|\beta\|_1$ is known as the least absolute shrinkage and selection operator, or lasso, model, used in high-dimensional applications where sparsity of the solution β is desirable for interpretability and recovery purposes when p is large. Another common model is the Tikhonov regularization function $P(\beta) = \|\beta\|_2$, which deals with issues that arise when the data matrix $\mathbf{X}$ is ill-conditioned or singular.

In this paper, we consider regularizing the data-driven portfolio optimization problem, not for sparsity or numerical stability as in lasso or ridge regression, but for the purpose of improving the out-of-sample performance of the solution. The portfolio optimization model we consider is

$$\begin{aligned} w_0 = \arg\min_{w\in\mathbb{R}^p} \;& \mathrm{Risk}(w^\top X) \\ \text{s.t.}\;& w^\top 1_p = 1, \\ & (w^\top \mu = R), \end{aligned} \qquad \text{(PO)}$$

where $w \in \mathbb{R}^p$ is the investor's holding on p different assets, $X \in \mathbb{R}^p$ denotes the relative return on the p assets, $\mu = \mathbb{E}X$ is the mean return vector, and Risk: $\mathbb{R} \to \mathbb{R}$ is some measure of risk. The investor's wealth is normalized to 1, so $w^\top 1_p = 1$, where 1_p denotes $p \times 1$ vector of ones, and $w^\top \mu = R$ is the target return constraint, which we may or may not consider,[1] hence it is shown in parentheses. Note that short selling, i.e., having $w < 0$, is allowed in this model. Setting $\mathrm{Risk}(w^\top X) = w^\top \Sigma w$, we recover the classical model of Markowitz (1952), and setting $\mathrm{Risk}(w^\top X) = \mathrm{CVaR}(-w^\top X; \beta)$, where $\beta \in (0.5, 1)$ and

$$\mathrm{CVaR}(-w^\top X;\beta) := \min_{\alpha}\left\{\alpha + \frac{1}{1-\beta}\mathbb{E}(-w^\top X - \alpha)^+\right\}, \qquad (2)$$

we recover the conditional value-at-risk[2] (CVaR) formulation of Rockafellar and Uryasev (2000).

In practice, one does not know the true distribution of X but has access to past data: $\mathsf{X} = [X_1, \ldots, X_n]$. Assuming that these are independent and identically distributed (iid), the standard data-driven approach to solving (PO) is to solve

$$\begin{aligned} \hat{w}_n = \arg\min_{w\in\mathbb{R}^p} \;& \widehat{\mathrm{Risk}}_n(w^\top X) \\ \text{s.t.}\;& w^\top 1_p = 1, \\ & (w^\top \hat{\mu}_n = R), \end{aligned} \qquad \text{(SAA)}$$

where $\widehat{\mathrm{Risk}}_n(w^\top X)$ is the sample average estimate of the risk function, and $\hat{\mu}_n = n^{-1}\sum_{i=1}^n X_i$ is the sample average return vector. This approach is befittingly known as the sample average approximation (SAA) method in the stochastic programming literature (see Shapiro et al. 2009 for a general overview).

As is the case with ill-posed linear operator problems (which includes regression problems), the solution to the SAA approach can be highly unstable. For the portfolio optimization problem, the fact that the SAA allocation is highly unreliable has been well documented (see Frankfurter et al. 1971; Frost and Savarino 1986, 1988; Michaud 1989; Best and Grauer 1991; Chopra and Ziemba 1993; Broadie 1993 for the Markowitz problem; and Lim et al. 2011 for the mean-CVaR problem), and it has limited the widespread adoption of the model in practice, despite the conferral of a Nobel Prize to Harry Markowitz in 1990 for his seminal work.

In this paper, we propose *performance-based regularization* (PBR) to improve upon the performance of the SAA approach to the data-driven portfolio allocation problem (SAA). The idea is to constrain the sample variances of estimated quantities in a problem; for portfolio optimization, they are the estimated portfolio risk $\widehat{\mathrm{Risk}}_n(w^\top X)$ and the estimated portfolio mean $w^\top \hat{\mu}_n$. The goal of PBR is to steer the solution toward one that is associated with less estimation error in the performance. The overall effect is to reduce the chance that a solution is chosen by misleadingly high in-sample performance. Performance-based regularization is thus philosophically different from Tikhonov regularization (whose purpose is stability of the solution) and lasso regularization (whose purpose is sparsity) but is natural to the portfolio optimization problem where the ultimate goal is the *out-of-sample* performance of the decision made.

We make four major contributions in this paper. First, we propose and analyze new portfolio optimization models by introducing performance-based regularization to the mean-variance and mean-CVaR problems. This is an important conceptual development that extends the current literature on portfolio optimization. For the mean-variance problem, the PBR model involves a quartic polynomial constraint. Determining whether such a model is convex or not is an NP-hard problem, so we consider two convex approximations, one based on a rank-1 approximation and one based on the best convex quadratic approximation. We then investigate the two approximation models and analytically characterize the effect of PBR on the solution. In the rank-1 approximation model, PBR adds a bias to the optimal allocation directly, whereas in the quadratic approximation case, PBR is equivalent to shrinking the sample covariance matrix. For the mean-CVaR problem, the PBR model is a combinatorial optimization problem, but we prove that its convex relaxation, a quadratically constrained quadratic program (QCQP), is tight; hence it can be efficiently solved.

Second, we show that the PBR portfolio models can be cast as robust optimization problems, introducing uncertainty sets that are new to the literature. The PBR constraint on the mean return uncertainty is equivalent to the a constraint where the portfolio return is required to be robust to all possible values of the mean vector falling within an ellipsoid, centered about the true mean. This is a well-known result in robust optimization (see Ben-Tal et al. 2009). However, the robust counterparts of the PBR constraint on the risk have structures that have not been considered before. The robust counterparts are somewhat related to constraining estimation error in the portfolio risk; however, the robust models do not enjoy the same intuitive interpretation of the original PBR formulations. We thus not only link PBR with novel robust models but also justify the original PBR formulation in its own right, as it is motivated by the intuitive idea of cutting off solutions associated with high in-sample estimation errors, whereas the equivalent robust constraint does not necessarily enjoy intuitive interpretation.

Third, we prove that the SAA and PBR solutions are asymptotically optimal under the very mild assumption that the true solutions be well separated (i.e., identifiable). This is an important and necessary result because data-driven decisions that are not asymptotically optimal as the number of stationary observations increases are nonsensical. We also show that the corresponding performances of the SAA and PBR solutions converge to the true optimal solutions. To the best of our knowledge, this is the first paper that analytically proves the asymptotic optimality of the solutions to estimated portfolio optimization problems for general underlying return distributions. (See Jobson and Korkie 1980 for asymptotic analysis when the returns are multivariate normal.)

Finally, we make an extensive empirical study of the PBR method against SAA as well as other benchmarks, including L1 and L2 regularizations and the equally weighted portfolio of DeMiguel et al. (2009a). We use the 5, 10, and 49 industry data sets from French (2015). To calibrate the constraint right-hand side (RHS) of PBR and standard regularization models, we also develop a new, performance-based extension of the k-fold cross-validation algorithm. The two key differences between our algorithm and standard k-fold cross-validation are that the search boundaries for the PBR constraint RHS need to be set carefully in order to avoid infeasibility and having no effect, and that we validate by computing the Sharpe ratio (the main performance metric for investment in practice) as opposed to the mean squared error. In sum, we find that for the 5- and 10-industry data sets, the PBR method improves upon SAA, in terms of the out-of-sample Sharpe ratio (annualized) with statistical significance levels at 5% and 10%, for both the mean-variance and mean-CVaR problems. Also for these data sets, PBR dominates standard L1 and L2 regularizations, as well as the equally weighted portfolio of DeMiguel et al. (2009a). The results for the 49 industry portfolio data set are inconclusive, with none of the strategies considered being statistically significantly different from the SAA result. We attribute this to the high-dimensionality effect (see Ledoit and Wolf 2004, El Karoui 2010), and leave studies of mitigating for the dimensionality to future work (Ban and Chen 2016).

1.1. Survey of Literature

As mentioned in the Section 1, Tikhonov (1963), Ivanov (1962), and Phillips (1962) first introduced the notion of regularization for ill-posed linear operator problems. For details on the historical development and use of regularization in statistical problems, Vapnik (2000) is a classic text; for more recent illustrations of the technique, we refer the reader to Hastie et al. (2009).

The more conventional regularization models have been investigated for the Markowitz problem by Chopra (1993), Frost and Savarino (1988), Jagannathan and Ma (2003), and DeMiguel et al. (2009b) and for the mean-CVaR problem by Gotoh and Takeda (2010). Specifically, Chopra (1993), Frost and Savarino (1988), and Jagannathan and Ma (2003) consider imposing a no short sale constraint on the portfolio weights (i.e., require portfolio weights to be nonnegative). DeMiguel et al. (2009b) generalize this further by considering L1, L2, and A-norm regularizations, and they show that the no short sale constraint is a special case of L1 regularized portfolio. Our PBR model for the Markowitz problem extends this literature by considering performance-motivated regularization constraints. The actual PBR model is nonconvex, so we consider two convex approximations, the first being an extra affine constraint on the portfolio weights and the second being a constraint on a particular A-norm of the vector of portfolio weights. The first corresponds to adding a bias to the SAA solution, and the second corresponds to shrinking the sample covariance matrix in a specific way. Analogously, Gotoh and Takeda (2010) consider L1 and L2 norms for the data-driven mean-CVaR problem; our work also extends this literature. In Section 5, we show that PBR outperforms the standard regularization techniques in terms of the out-of-sample Sharpe ratio.

The PBR models add a new perspective on recent developments in robust portfolio optimization that construct uncertainty sets from data (Delage and Ye 2010, Goldfarb and Iyengar 2003). While the PBR constraint on the portfolio mean is equivalent to the mean uncertainty constraint considered in Delage and Ye (2010), the PBR constraint on the portfolio variance for the mean-variance problem leads to a new uncertainty set that is different from that in Delage and Ye (2010). The main difference is that Delage and Ye (2010) considers an uncertainty set for the sample covariance matrix separately from the decision, whereas PBR considers protecting against estimation errors in the portfolio variance, thereby considering both the decision and the covariance matrix together. The difference is detailed in Online Appendix B. Goldfarb and Iyengar (2003) also take the approach of directly modelling the uncertainty set of the covariance matrix, although it is different from Delage and Ye (2010) and also from our work because it starts from a factor model of asset returns and assumes that the returns are multivariate normally distributed, whereas both Delage and Ye (2010) and our work are based on a nonparametric, distribution-free setting.

Finally, Gotoh et al. (2015) show that a large class of distributionally robust empirical optimization problems with uncertainty sets defined in terms of ϕ-divergence are asymptotically equivalent to PBR problems. We note, however, that the class of models studied in Gotoh et al. (2015) does not include CVaR.

Notations. Throughout the paper, we denote convergence in probability by $\xrightarrow{\mathrm{P}}$.

2. Motivation: Fragility of SAA in Portfolio Optimization

In this paper, we consider two risk functions: the variance of the portfolio and the conditional value-at-risk. In the former case, the problem is the classical Markowitz model of portfolio optimization, which is

$$\begin{aligned} w_{\mathrm{MV}} = \arg\min_{w\in\mathbb{R}^p} \quad & w^\top \Sigma w \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & (w^\top \mu = R), \end{aligned} \tag{MV-true}$$

where μ and Σ are, respectively, the mean and the covariance matrix of X, the relative stock return, and where the target return constraint $(w^\top \mu = R)$ may or may not be imposed.

Given data $\mathsf{X} = [X_1, X_2, \dots, X_n]$, the SAA approach to the problem is

$$\begin{aligned} \hat{w}_{n,\mathrm{MV}} = \arg\min_{w\in\mathbb{R}^p} \quad & w^\top \hat{\Sigma}_n w \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & (w^\top \hat{\mu}_n = R), \end{aligned} \tag{MV-SAA}$$

where $\hat{\mu}_n$ and $\hat{\Sigma}_n$ are the sample mean and the sample covariance matrix of X, respectively.

In the latter case, we have a mean-CVaR portfolio optimization model. Specifically, the investor wants to pick a portfolio that minimizes the CVaR of the portfolio loss at level $100(1-\beta)\%$, for some $\beta \in (0.5, 1)$, while reaching an expected return R:

$$\begin{aligned} w_{\mathrm{CV}} = \arg\min_{w\in\mathbb{R}^p} \quad & \mathrm{CVaR}(-w^\top X; \beta) \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & (w^\top \mu = R), \end{aligned} \tag{CV-true}$$

where

$$\mathrm{CVaR}(-w^\top X; \beta) := \min_{\alpha} \left\{ \alpha + \frac{1}{1-\beta} \mathbb{E}(-w^\top X - \alpha)^+ \right\},$$

as in Rockafellar and Uryasev (2000).

The SAA approach to the problem is to solve

$$\begin{aligned} \hat{w}_{n,\mathrm{CV}} = \arg\min_{w\in\mathbb{R}^p} \quad & \widehat{\mathrm{CVaR}}_n(-w^\top \mathsf{X}; \beta) \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & (w^\top \hat{\mu}_n = R), \end{aligned} \tag{CV-SAA}$$

where

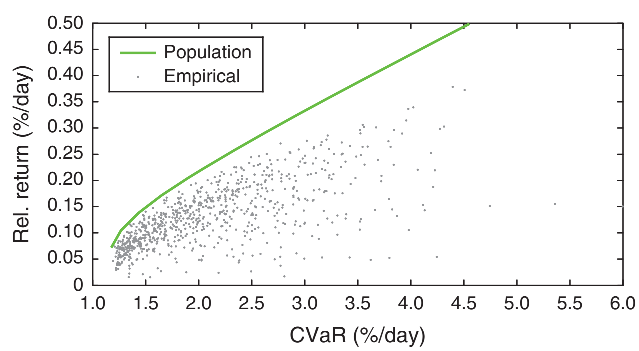
$$\widehat{\mathrm{CVaR}}_n(-w^\top \mathsf{X}; \beta) := \min_{\alpha\in\mathbb{R}} \alpha + \frac{1}{n(1-\beta)} \sum_{i=1}^{n} (-w^\top X_i - \alpha)^+$$

is the sample average estimator for $\mathrm{CVaR}(-w^\top X; \beta)$.

Asymptotically, as the number of observations n goes to infinity, we can show that the SAA solutions $\hat{w}_{n,\mathrm{MV}}$ and $\hat{w}_{n,\mathrm{CV}}$ converge in probability to w_{MV} and w_{CV}, respectively (see Section 4 for details). In practice, however, the investor has a limited number of relevant (i.e., stationary) observations (Jegadeesh and Titman 1993, Lo and MacKinlay 1990, DeMiguel et al. 2014). Solving (MV-SAA) and (CV-SAA) with a finite amount of stationary data can yield highly unreliable solutions (Lim et al. 2011). Let us illustrate this point by a simulation experiment for (CV-SAA). There are $p = 10$ stocks, with daily returns following a Gaussian distribution[3]: $X \sim \mathcal{N}(\mu_{\mathrm{sim}}, \Sigma_{\mathrm{sim}})$, and the investor has $n = 250$ iid observations of X. The experimental procedure is as follows:

- Simulate 250 historical observations from the distribution $\mathcal{N}(\mu_{\mathrm{sim}}, \Sigma_{\mathrm{sim}})$.
- Solve (CV-SAA) with $\beta = 0.95$ and some return level R to find an instance of $\hat{w}_{n,\mathrm{CV}}$.
- Plot the realized return $\hat{w}_{n,\mathrm{CV}}^\top \mu_{\mathrm{sim}}$ versus realized risk $\mathrm{CVaR}(-\hat{w}_{n,\mathrm{CV}}^\top X; \beta)$; this corresponds to one grey point in Figure 1.
- Repeat for different values of R to obtain a sample efficient frontier.

Figure 1. (Color online) Distribution of Realized Daily Return (Percent per Day) vs. Daily Risk (Percent per Day) of SAA Solutions $\hat{w}_{n,\mathrm{CV}}$ for the Target Return Range 0.107%–0.430% per Day



Note. The solid (green in the online version) line represents the population frontier, i.e., the efficient frontier corresponding to solving (CV-true).

- Repeat many times to get a distribution of the sample efficient frontier.

The result of the experiment is summarized in Figure 1. The smooth curve corresponds to the population efficient frontier. Each of the grey dots corresponds to a solution instance of (CV-SAA). There are two noteworthy observations: the solutions $\hat{w}_{n,\text{CV}}$ are suboptimal, and they are highly variable. For instance, for a daily return of 0.1%, the CVaR ranges from 1.3% to 4%.

3. Performance-Based Regularization

We now introduce PBR to improve upon (SAA). The PBR model is

$$\begin{aligned}\hat{w}_{n,\text{PBR}} = \arg\min_{w\in\mathbb{R}^p}\ & \widehat{\text{Risk}}_n(w^\top X)\\ \text{s.t.}\ & w^\top 1_p = 1,\\ & (w^\top \hat{\mu}_n = R),\\ & \text{Svar}(\widehat{\text{Risk}}_n(w^\top X)) \le U_1,\\ & (\text{Svar}(w^\top \hat{\mu}_n) \le U_2), \end{aligned} \qquad \text{(PBR)}$$

where $\text{Svar}(\cdot)$ is the sample variance operator, and U_1 and U_2 are parameters that control the degree of regularization.

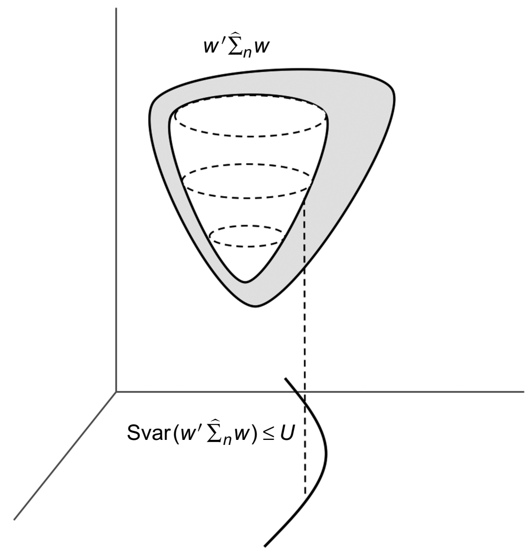
The motivation behind the model (PBR) is intuitive and straightforward: for a fixed portfolio w, the point estimate $\widehat{\text{Risk}}_n(w^\top X)$ of the objective has a confidence interval around it, which is approximately equal to the sample standard deviation of the estimator $\widehat{\text{Risk}}_n(w^\top X)$. As w varies, the error associated with the point estimate varies, as the confidence interval is a function of w. The PBR constraint $\text{Svar}(\widehat{\text{Risk}}_n(w^\top X)) \le U_1$ dictates that any solution w that is associated with a large estimation error of the objective function be removed from consideration, which is sensible since such a decision would be highly unreliable. A similar interpretation can be made for the second PBR constraint $\text{Svar}(w^\top \hat{\mu}_n) \le U_2$. A schematic of the PBR model is shown in Figure 2.

Another intuition for PBR is obtained via Chebyshev's inequality. Chebyshev's inequality tells us that, for all $\delta > 0$, for some random variable Y,

$$\mathbb{P}(|Y - \mathbb{E}Y| \ge \delta) \le \frac{\text{Var}(Y)}{\delta^2}.$$

Thus letting Y equal $\widehat{\text{Risk}}_n(w^\top X)$ or $w^\top \hat{\mu}_n$, we see that constraining their sample variances has the effect of constraining the probability that the estimated portfolio risk and return deviate from the true portfolio risk and return by more than a certain amount. In other words, the PBR constraints squeeze the SAA problem (SAA) closer to the true problem (PO) with some probability.

Figure 2. A Schematic of PBR on the Objective Only



Notes. The objective function estimated with data is associated with an error (indicated by the grey shading), which depends on the position in the solution space. The PBR constraint cuts out solutions that are associated with large estimation errors of the objective.

3.1. PBR for Mean-Variance Portfolio Optimization

The PBR model for the mean-variance problem is

$$\begin{aligned}\hat{w}_{n,\text{MV}} = \arg\min_{w\in\mathbb{R}^p}\ & w^\top \hat{\Sigma}_n w\\ \text{s.t.}\ & w^\top 1_p = 1,\\ & (w^\top \hat{\mu}_n = R),\\ & \text{Svar}(w^\top \hat{\Sigma}_n w) \le U. \end{aligned} \qquad \text{(mv-PBR)}$$

Note that we do not regularize the mean constraint, as the sample variance of $w^\top \hat{\mu}_n$ is precisely $w^\top \hat{\Sigma}_n w$, which is already captured by the objective.

The following proposition characterizes the sample variance of the sample variance of the portfolio, $\text{Svar}(w^\top \hat{\Sigma}_n w)$.

Proposition 1. *The sample variance of the sample variance of the portfolio, $\text{Svar}(w^\top \hat{\Sigma}_n w)$, is given by*

$$\text{Svar}(w^\top \hat{\Sigma}_n w) = \Sigma_{i=1}^p \Sigma_{j=1}^p \Sigma_{k=1}^p \Sigma_{l=1}^p w_i w_j w_k w_l \hat{Q}_{ijkl}, \qquad (3)$$

where

$$\hat{Q}_{ijkl} = \frac{1}{n}(\hat{\mu}_{4,ijkl} - \hat{\sigma}_{ij}^2 \hat{\sigma}_{kl}^2) + \frac{1}{n(n-1)}(\hat{\sigma}_{ik}^2 \hat{\sigma}_{jl}^2 + \hat{\sigma}_{il}^2 \hat{\sigma}_{jk}^2),$$

where $\hat{\mu}_{4,ijkl}$ is the sample average estimator for $\mu_{4,ijkl}$, the fourth central moment of the elements of X given by

$$\mu_{4,ijkl} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)(X_l - \mu_l)],$$

and $\hat{\sigma}_{ij}^2$ is the sample average estimator for σ_{ij}^2, the covariance of the elements of X given by

$$\sigma_{ij}^2 = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)].$$

Proof. See Online Appendix Section A.1.

The PBR constraint of (mv-PBR) is thus a quartic polynomial in the decision vector w. Determining whether a general quartic function is convex is an NP-hard problem (Ahmadi et al. 2013); hence it is not clear at the outset whether $\operatorname{Svar}(w^\top \hat{\Sigma}_n w)$ is a convex function in w, and thus (mv-PBR) a convex problem. We thus consider two convex approximations of (mv-PBR).

3.1.1. Rank-1 Approximation of (mv-PBR). Here, we make a rank-1 approximation of the quartic polynomial constraint:

$$(w^\top \hat{\alpha})^4 \approx \sum_{ijkl} w_i w_j w_k w_l \hat{Q}_{ijkl},$$

by matching up the diagonals; i.e., $\hat{\alpha}$ is given by

$$\hat{\alpha}_i = \sqrt[4]{\hat{Q}_{iiii}} = \sqrt[4]{\frac{1}{n}\hat{\mu}_{4,iiii} - \frac{n-3}{n(n-1)}(\hat{\sigma}_{ii}^2)^2}. \tag{4}$$

We thus obtain the following convex approximation of (mv-PBR):

$$\begin{aligned} \hat{w}_{n,\mathrm{PBR1}} = \arg\min_{w\in\mathbb{R}^p} \quad & w^\top \hat{\Sigma}_n w \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & (w^\top \hat{\mu}_n = R), \\ & w^\top \hat{\alpha} \le \sqrt[4]{U}, \end{aligned} \tag{mv-PBR-1}$$

where $\hat{\alpha}$ is given in (4).

We can state the effect of PBR constraint as in (mv-PBR-1) on the SAA solution explicitly as follows.

Proposition 2. *The solution to (mv-PBR-1) with the mean constraint $w^\top \hat{\mu}_n = R$ is given by*

$$\hat{w}_{n,\mathrm{PBR1}} = \hat{w}_{n,\mathrm{MV}} - \tfrac{1}{2}\lambda^* \hat{\Sigma}_n^{-1}(\beta_1 1_p + \beta_2 \hat{\mu}_n + \hat{\alpha}), \tag{5}$$

where $\hat{w}_{n,\mathrm{MV}}$ is the SAA solution, λ^* is the optimal Lagrange multiplier for the PBR constraint $w^\top \alpha \le \sqrt[4]{U}$, and

$$\beta_1 = \frac{\hat{\alpha}^\top \hat{\Sigma}_n^{-1}\hat{\mu}_n \cdot \hat{\mu}_n^\top \hat{\Sigma}_n^{-1} 1_p - \hat{\alpha}^\top \hat{\Sigma}_n^{-1} 1_p \cdot \hat{\mu}_n^\top \hat{\Sigma}_n^{-1}\hat{\mu}_n}{1_p^\top \hat{\Sigma}_n^{-1} 1_p \cdot \hat{\mu}_n^\top \hat{\Sigma}_n^{-1}\hat{\mu}_n - (\hat{\mu}_n^\top \hat{\Sigma}_n^{-1} 1_p)^2},$$

$$\beta_2 = \frac{\hat{\alpha}^\top \hat{\Sigma}_n^{-1}\hat{\mu}_n \cdot 1_p^\top \hat{\Sigma}_n^{-1} 1_p - \hat{\alpha}^\top \hat{\Sigma}_n^{-1} 1_p \cdot 1_p^\top \hat{\Sigma}_n^{-1}\hat{\mu}_n}{1_p^\top \hat{\Sigma}_n^{-1} 1_p \cdot \hat{\mu}_n^\top \hat{\Sigma}_n^{-1}\hat{\mu}_n - (\hat{\mu}_n^\top \hat{\Sigma}_n^{-1} 1_p)^2}.$$

The solution to (mv-PBR-1) without the mean constraint is given by

$$\hat{w}_{n,\mathrm{PBR1}} = \hat{w}_{n,\mathrm{MV}} - \tfrac{1}{2}\lambda^* \hat{\Sigma}_n^{-1}(\beta 1_p + \hat{\alpha}), \tag{6}$$

where $\hat{w}_{n,\mathrm{MV}}$ is the SAA solution, λ^* is the optimal Lagrange multiplier for the PBR constraint $w^\top \alpha \le \sqrt[4]{U}$, and

$$\beta = -\frac{1_p^\top \hat{\Sigma}_n^{-1}\hat{\alpha}}{1_p^\top \hat{\Sigma}_n^{-1} 1_p}.$$

Remark. The effect of rank-1 approximation PBR on the Markowitz problem is thus to shrink the SAA portfolio, by an amount scaled by λ^*, toward a direction that depends on the (approximated) fourth moment of the asset returns.

Proof. See Online Appendix Section A.2.

3.1.2. Best Convex Quadratic Approximation of (mv-PBR). We also consider a convex quadratic approximation of the quartic polynomial constraint:

$$(w^\top A w)^2 \approx \sum_{ijkl} w_i w_j w_k w_l \hat{Q}_{ijkl},$$

where A is a positive semidefinite (PSD) matrix. Expanding the left-hand side, we get

$$\sum_{ijkl} w_i w_j w_k w_l A_{ij} A_{kl}.$$

Let us require the elements of A to be as close as possible to the pairwise terms in Q; i.e., $A_{ij}^2 \approx \hat{Q}_{ijij}$. Then the best PSD matrix A that approximates $\hat{Q}$ in this way is given by solving the following semidefinite program (PSD):

$$A^* = \arg\min_{A\succeq 0} \|A - Q_2\|_F, \tag{Q approx}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and where Q_2 is a matrix such that its ijth element equals $\hat{Q}_{ijij}$. We thus obtain the following convex quadratic approximation of (mv-PBR):

$$\begin{aligned} \hat{w}_{n,\mathrm{PBR2}} = \arg\min_{w\in\mathbb{R}^p} \quad & w^\top \hat{\Sigma}_n w \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & (w^\top \hat{\mu}_n = R), \\ & w^\top A^* w \le \sqrt{U}. \end{aligned} \tag{mv-PBR-2}$$

We can state the effect of PBR constraint as in (mv-PBR-2) on the SAA solution explicitly as follows.

Proposition 3. *The solution to (mv-PBR-2) with the mean constraint $w^\top \hat{\mu}_n = R$ is given by*

$$\hat{w}_{n,\mathrm{PBR2}} = -\tfrac{1}{2}\tilde{\Sigma}_n(\lambda^*)^{-1}\big(\nu_1^*(\lambda^*)1_p + \nu_2^*(\lambda^*)\hat{\mu}_n\big), \tag{7}$$

where $\tilde{\Sigma}_n(\lambda^*) = \hat{\Sigma}_n + \lambda^* A^*$, λ^* is the optimal Lagrange multiplier for the PBR constraint $w^\top A^* w \le \sqrt{U}$, and

$$\nu_1^*(\lambda) = 2\frac{R\hat{\mu}_n^\top \tilde{\Sigma}^{-1} 1_p - \hat{\mu}_n^\top \tilde{\Sigma}^{-1}\hat{\mu}_n}{1_p^\top \tilde{\Sigma}^{-1} 1_p \cdot \hat{\mu}_n^\top \tilde{\Sigma}^{-1}\hat{\mu}_n - (\hat{\mu}_n^\top \tilde{\Sigma}^{-1} 1_p)^2},$$

$$\nu_2^*(\lambda) = 2\frac{-R 1_p^\top \tilde{\Sigma}^{-1} 1_p + \hat{\mu}_n^\top \tilde{\Sigma}^{-1} 1_p}{1_p^\top \tilde{\Sigma}^{-1} 1_p \cdot \hat{\mu}_n^\top \tilde{\Sigma}^{-1}\hat{\mu}_n - (\hat{\mu}_n^\top \tilde{\Sigma}^{-1} 1_p)^2}.$$

The solution to (mv-PBR-2) without the mean constraint is given by

$$\hat{w}_{n,\mathrm{PBR2}} = \frac{\tilde{\Sigma}_n(\lambda^*)^{-1} 1_p}{1_p^\top \tilde{\Sigma}_n(\lambda^*)^{-1} 1_p}, \tag{8}$$

where $\tilde{\Sigma}_n(\lambda^*) = \hat{\Sigma}_n + \lambda^* A^*$, and λ^* is the optimal Lagrange multiplier for the PBR constraint $w^\top A^* w \le \sqrt{U}$, as before.

Proof. See Online Appendix Section A.3.

For both mean-constrained and mean-unconstrained cases, notice that the solution depends on λ only through the matrix $\tilde{\Sigma}_n(\lambda^*) = \hat{\Sigma}_n + \lambda^* A^*$. We thus retrieve the unregularized SAA solution $\hat{w}_{n,\mathrm{MV}}$ when λ is set to zero. Thus the PSD approximation to (mv-PBR) is equivalent to using a different estimator for the covariance matrix than the sample covariance matrix $\hat{\Sigma}_n$. Clearly, $\tilde{\Sigma}_n(\lambda^*)$ adds a bias to the sample covariance matrix estimator. It is well known that adding some bias to a standard estimator can be beneficial, and such estimators are known as shrinkage estimators. Haff (1980) and Ledoit and Wolf (2004) have explored this idea for the sample covariance matrix by shrinking the sample covariance matrix toward the identity matrix, and they have shown superior properties of the shrunken estimator. By contrast, our PBR model shrinks the sample covariance matrix toward a direction that is approximately equal to the variance of the sample covariance matrix. Conversely, DeMiguel et al. (2009b) showed that using the shrinkage estimator for the covariance matrix as in Ledoit and Wolf (2004) is equivalent to L2 regularization; in Section 5 we compare the two methods.

3.2. PBR for Mean-CVaR Portfolio Optimization

The PBR model for the mean-CVaR problem is

$$\begin{aligned}\min_{w\in\mathbb{R}^p}\ & \widehat{\mathrm{CVaR}}_n(-w^\top \mathsf{X};\beta)\\ \text{s.t.}\ & w^\top 1_p = 1,\\ & (w^\top\hat{\mu}_n = R),\\ & \mathrm{Svar}(\widehat{\mathrm{CVaR}}_n(-w^\top\mathsf{X};\beta))\le U_1,\\ & (\mathrm{Svar}(w^\top\hat{\mu}_n)\le U_2).\end{aligned} \tag{cv-PBR}$$

The variance of $w^\top\hat{\mu}_n$ is given by

$$\mathrm{Var}(w^\top\hat{\mu}_n)=\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(w^\top X_i)=\frac{1}{n}w^\top\Sigma w;$$

hence $\mathrm{SVaR}(w^\top\hat{\mu}_n)=n^{-1}w^\top\hat{\Sigma}_n w$. The variance of

$$\widehat{\mathrm{CVaR}}_n(-w^\top\mathsf{X};\beta)$$

is given by the following proposition.

Proposition 4. *Suppose* $\mathsf{X}=[X_1,\ldots,X_n]\overset{\text{iid}}{\sim}F$, *where* F *is absolutely continuous with a twice continuously differentiable probability density function. Then*

$$\begin{aligned}&\mathrm{Var}[\widehat{\mathrm{CVaR}}_n(-w^\top\mathsf{X};\beta)]\\ &\quad=\frac{1}{n(1-\beta)^2}\mathrm{Var}\big[(-w^\top X-\alpha_\beta(w))^+\big]+O(n^{-2}),\end{aligned}$$

where

$$\alpha_\beta(w)=\inf\{\alpha\colon P(-w^\top X\ge\alpha)\le 1-\beta\},$$

the value-at-risk (VaR) *of the portfolio* w *at level* β.

Proof. See Online Appendix Section A.4.

Thus, the sample variance of $\widehat{\mathrm{CVaR}}_n(-w^\top\mathsf{X};\beta)$ is, to first order,

$$\mathrm{Svar}[\widehat{\mathrm{CVaR}}_n(-w^\top\mathsf{X};\beta)]=\frac{1}{n(1-\beta)^2}z^\top\Omega_n z,$$

where $\Omega_n=(1/(n-1))[I_n-n^{-1}1_n1_n^\top]$, with I_n being the $n\times n$ identity matrix, and $z_i=\max(0,-w^\top X_i-\alpha)$ for $i=1,\ldots,n$.

Incorporating the above formulas for the sample variances, (cv-PBR) can be written as

$$\begin{aligned}\min_{\alpha,w,z}\ & \alpha+\frac{1}{n(1-\beta)}\sum_{i=1}^n z_i\\ \text{s.t.}\ & w^\top 1_p=1,\\ & (w^\top\hat{\mu}_n=R),\\ & \frac{1}{n(1-\beta)^2}z^\top\Omega_n z\le U_1,\\ & z_i=\max(0,-w^\top X_i-\alpha),\quad i=1,\ldots,n,\\ & \frac{1}{n}w^\top\hat{\Sigma}_n w\le U_2,\end{aligned} \tag{cv-PBR′}$$

Note that (cv-PBR′) is nonconvex as a result of the cutoff variables $z_i=\max(0,-w^\top X_i-\alpha)$, $i=1,\ldots,n$. Without the regularization constraint $[n(1-\beta)^2]^{-1}z^\top\Omega_n z\le U_1$, one can solve the problem by relaxing the nonconvex constraint $z_i=\max(0,-w^\top X_i-\alpha)$ to $z_i\ge 0$ and $z_i\ge -w^\top X_i-\alpha$. However, $z^\top\Omega_n z$ is not a monotone function of z; hence it is not clear at the outset whether one can employ such a relaxation trick for the regularized problem.

We find that (cv-PBR′) is a combinatorial optimization problem because one can solve it by considering all possible combinations of $\lfloor n(1-\beta)\rfloor$ out of n observations that contribute to the worst $(1-\beta)$ of the portfolio loss (which determines the nonzero elements of z), then finding the portfolio weights that solve the problem based on these observations alone. Clearly, this is an impractical strategy; for example, there are 34,220 possible combinations to consider for a modest number of observations $n=60$ (five years of monthly data) and $\beta=0.95$.

However, it turns out that relaxing $z_i=\max(0,-w^\top X_i-\alpha)$, $i=1,\ldots,n$ does result in a tight convex relaxation. The resulting problem is a QCQP, which can be solved efficiently. Before formally stating this result, let us first introduce the convex relaxation of (cv-PBR′):

$$\begin{aligned}\min_{\alpha,w,z}\ & \alpha+\frac{1}{n(1-\beta)}\sum_{i=1}^n z_i\\ \text{s.t.}\ & w^\top 1_p=,1 && (\nu_1)\\ & (w^\top\hat{\mu}_n=R), && (\nu_2)\\ & \frac{1}{n(1-\beta)^2}z^\top\Omega_n z\le U_1, && (\lambda_1)\\ & z_i\ge 0\quad i=1,\ldots,n, && (\eta_1)\\ & z_i\ge -w^\top X_i-\alpha,\quad i=1,\ldots,n, && (\eta_2)\\ & \frac{1}{n}w^\top\hat{\Sigma}_n w\le U_2, && (\lambda_2)\end{aligned} \tag{cv-relax}$$

and its dual (where the dual variables correspond to the primal constraints as indicated above):

$$\begin{aligned}\max_{\nu_1,\nu_2,\lambda_1,\lambda_2,\eta_1,\eta_2} \quad & g(\nu_1,\nu_2,\eta_1,\eta_2,\lambda_1,\lambda_2)\\ \text{s.t.} \quad & \eta_2^\top 1_n = 1,\\ & \lambda_1 \geq 0, \quad \lambda_2 \geq 0,\\ & \eta_1 \geq 0, \quad \eta_2 \geq 0, \end{aligned} \qquad \text{(cv-relax-d)}$$

where

$$\begin{aligned} g(\nu_1,\nu_2,\lambda_1,\lambda_2,\eta_1,\eta_2) &= -\frac{n}{2\lambda_1}(\nu_1 1_p + \nu_2\hat{\mu}_n - \mathrm{X}\eta_2)^\top \hat{\Sigma}_n^{-1}(\nu_1 1_p + \nu_2\hat{\mu}_n - \mathrm{X}\eta_2)\\ &\quad - \frac{n(1-\beta)^2}{2\lambda_2}(\eta_1+\eta_2)^\top \Omega_n^\dagger(\eta_1+\eta_2)\\ &\quad + \nu_1 + R\nu_2 - U_1\lambda_1 - U_2\lambda_2, \end{aligned}$$

and $\Omega_n^\dagger$ is the Moore–Penrose pseudo inverse of the singular matrix Ω_n.

We now state the result that (cv-PBR′) is a tractable optimization problem because its convex relaxation is essentially tight.

Theorem 1. *Let $(\alpha^*, w^*, z^*, \lambda_1^*, \lambda_2^*, \eta_1^*, \eta_2^*)$ be the primal-dual optimal point of (cv-PBR′) and (cv-relax-d). If $\eta_2^* \neq 1_n/n$, then (α^*, w^*, z^*) is an optimal point of (cv-PBR′). Otherwise, if $\eta_2^* = 1_n/n$, we can find the optimal solution to (cv-PBR′) by solving (cv-relax-d) with an additional constraint $\eta_2^\top 1_n \geq \delta$, where δ is any constant $0 < \delta < 1$.*

Proof. See Online Appendix Section A.5.

Remark. Theorem 1 shows that one can solve (cv-PBR′) via at most two steps. The first step is to solve (cv-PBR′); if the dual variables corresponding to the constraints $z_i \geq -w^\top X_i - \alpha$, $i = 1, \ldots, n$ are all equal to $1/n$, then we solve (cv-relax-d) with an additional constraint $\eta_2^\top 1_n \geq \delta$, where δ is any constant $0 < \delta \ll 1$; otherwise, the relaxed solution is feasible for the original problem and hence optimal. For the record, all problem instances solved in the numerical section, Section 5, were solved in a single step.

3.3. Robust Counterparts of PBR Models

In this section, we show that the three PBR portfolio optimization models can be transformed into robust optimization (RO) problems.

Proposition 5. *The convex approximations to the Markowitz PBR problem (mv-PBR) have the following robust counterpart representation:*

$$\begin{aligned}\hat{w}_{n,\mathrm{PBR1}} = \arg\min_{w\mathbb{R}^p} \quad & w^\top \hat{\Sigma}_n w\\ \text{s.t.} \quad & w^\top 1_p = 1,\\ & (w^\top \hat{\mu}_n = R),\\ & \max_{u\in\mathcal{U}} w^\top u \leq \sqrt[4]{U}, \end{aligned} \qquad \text{(mv-PBR-RO)}$$

where $\mathcal{U}$ is the ellipsoid

$$\mathcal{U} = \{u \in \mathbb{R}^p \mid u^\top P^\dagger u \leq 1, \ (I - PP^\dagger)u = 0\},$$

with $P = \alpha\alpha^\top$ for (mv-PBR-1) and $P = A^*$ for (mv-PBR-2), and $P^\dagger$ denotes the Moore–Penrose pseudoinverse of the matrix P (which equals the inverse if P is invertible, which is the case for $P = A^*$).

Proof. See Online Appendix Section A.6.

Proposition 6. *The mean-CVaR PBR problem (cv-PBR) has the following robust counterpart representation:*

$$\begin{aligned}\min_{\alpha,w,z} \quad & \alpha + \frac{1}{n(1-\beta)}\sum_{i=1}^n z_i\\ \text{s.t.} \quad & w^\top 1_p = 1,\\ & (w^\top \hat{\mu}_n = R),\\ & \max_{u\in\mathcal{U}_1} z^\top u \leq \sqrt{U_1},\\ & z_i = \max(0, -w^\top X_i - \alpha), \quad i = 1, \ldots, n\\ & \Big(\max_{\tilde{\mu}\in\mathcal{U}_2} w^\top(\tilde{\mu} - \mu) \leq \sqrt{U_2}\Big), \end{aligned} \qquad \text{(cv-PBR-RO)}$$

where $\mathcal{U}_1$ is the ellipsoid

$$\mathcal{U}_1 = \{\mu \in \mathbb{R}^p \mid (\tilde{\mu} - \mu)^\top \hat{\Sigma}_n^{-1}(\tilde{\mu} - \mu) \leq 1\},$$

and $\mathcal{U}_2$ is the ellipsoid

$$\mathcal{U}_2 = \{u \in \mathbb{R}^n \mid u^\top \Omega_n^\dagger u \leq 1, \ 1_p^\top u = 0\},$$

where $\Omega_n^\dagger$ is the Moore–Penrose pseudoinverse of the matrix Ω_n.

Proof. One can follow steps similar to those in the proof of Proposition 5.

While the PBR constraint on the portfolio mean is equivalent to the mean uncertainty constraint considered in Delage and Ye (2010), the PBR constraint on the portfolio variance for the mean-variance problem leads to a new uncertainty set, which is different from Delage and Ye (2010). The main difference is that Delage and Ye (2010) considers an uncertainty set for the sample covariance matrix separately from the decision, whereas PBR considers protecting against estimation errors in the portfolio variance, thereby considering both the decision and the covariance matrix together. The difference is detailed in Online Appendix B.

4. Asymptotic Optimality of SAA and PBR Solutions

In this section, we show that the SAA solution $\hat{w}_n$ and the PBR solutions are asymptotically optimal under the mild condition that the true solution be well separated (i.e., identifiable). In other words, we show that the SAA solution converges in probability to the true optimal w_0 as the number of observations n tends to infinity. We then show that the performances of the estimated solutions also converge to that of the true optimal; i.e., the return-risk frontiers corresponding to $\hat{w}_n$ and $\hat{w}_{n,\mathrm{PBR}}$ converge to the efficient frontier of w_0.

For ease of exposition and analysis, we will work with the following transformation of the original problem:

$$\min_{\theta=(\alpha,v)\in\mathbb{R}\times\mathbb{R}^{p-1}} M(\theta) = \min_{\theta=(\alpha,v)\in\mathbb{R}\times\mathbb{R}^{p-1}} \mathbb{E}[m_\theta(X)], \qquad (\mathrm{PO}')$$

where we have reparameterized w to $w = w_1 + Lv$, where $L = [0_{(p-1)\times 1}, I_{(p-1)\times(p-1)}]^\top$, $v = [w_2, \ldots, w_p]^\top$, and $w_1 = [1 - v^\top 1_{(p-1)}, 0_{1\times(p-1)}]^\top$, and

$$m_\theta(x) = \big((w_1+Lv)^\top x - (w_1+Lv)^\top \mu\big)^2 - \lambda_0(w_1+Lv)^\top x \qquad (9)$$

for the mean-variance problem (MV-true) and

$$m_\theta(x) = \alpha + \frac{1}{1-\beta} z_\theta(x) - \lambda_0(w_1+Lv)^\top x \qquad (10)$$

for the mean-CVaR problem (CV-true), where $z_\theta(x) = \max(0, -(w_1 + Lv)^\top x - \alpha)$. In other words, we have transformed (PO) to a global optimization problem, where $\lambda_0 > 0$ determines the investor's utility on the return. Without loss of generality, we restrict problem (PO′) to optimizing over a compact subset Θ of $\mathbb{R}\times\mathbb{R}^{p-1}$.

We now prove asymptotic optimality of the SAA solution to the mean-variance and mean-CVaR problems.

Theorem 2 (Asymptotic Optimality of SAA Solution of Mean-Variance Problem). *Consider (PO′) with $m_\theta(\cdot)$ as in (9). Denote the solution by θ_{MV}. Suppose, for all $\epsilon > 0$,*

$$\sup_{\theta\in\Theta}\big\{M(\theta): \|\theta, \theta_{\mathrm{MV}}\|_2 \geq \epsilon\big\} < M(\theta_{\mathrm{MV}}). \qquad (*)$$

Then, as n tends to infinity,

$$\hat{\theta}_{n,\mathrm{MV}} \xrightarrow{\mathrm{P}} \theta_{\mathrm{MV}},$$

where $\hat{\theta}_{n,\mathrm{MV}}$ is the solution to the SAA *problem*

$$\min_{\theta\in\Theta} M_n(\theta) = \min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}\big((w_1+Lv)^\top X_i - (w_1+Lv)^\top \hat{\mu}_n\big)^2 - \lambda_0(w_1+Lv)^\top X_i.$$

Theorem 3 (Asymptotic Optimality of SAA Solution of Mean-CVaR Problem). *Consider (PO′) with $m_\theta(\cdot)$ as in (10). Denote the solution by θ_{CV}. Suppose, for all $\epsilon > 0$,*

$$\sup_{\theta\in\Theta}\{M(\theta): \|\theta, \theta_{\mathrm{CV}}\|_2) \geq \epsilon\} < M(\theta_{\mathrm{CV}}). \qquad (**)$$

Then, as n tends to infinity,

$$\hat{\theta}_{n,\mathrm{CV}} \xrightarrow{\mathrm{P}} \theta_{\mathrm{CV}},$$

where $\hat{\theta}_{n,\mathrm{CV}}$ is the solution to the SAA *problem*

$$\min_{\theta\in\Theta} M_n(\theta) = \min_{\theta\in\Theta}\bigg\{\alpha + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\beta} z_\theta(X_i) - \lambda_0(w_1+Lv)^\top X_i\bigg\}.$$

Sketch of the Proofs of Theorems 2 and 3. Theorems 2 and 3 are statements about the asymptotic consistency of estimated quantities $\hat{\theta}_{n,\mathrm{MV}}$ and $\hat{\theta}_{n,\mathrm{CV}}$ to their true respective quantities θ_{MV} and θ_{CV}. While proving (statistical) convergence of sample average-type estimators for independent samples is straightforward, proving convergence of *solutions* of estimated optimization problems is more involved.

In mathematical statistics, the question of whether solutions of estimated optimization problems converge arises in the context of maximum likelihood estimation, whose study goes back as far as seminal works of Fisher (1922, 1925). Huber initiated a systematic study of M-estimators (where "M" stands for maximization; i.e., estimators that arise as solutions to maximization problems) with Huber (1967), which subsequently led to asymptotic results that apply to more general settings (e.g., nondifferentiable objective functions) that rely on the theory of empirical processes. Van der Vaart (2000) gives a clean, unified treatment of the main results in the theory of M-estimation, and we align our proof to the setup laid out in this book.

In particular, van der Vaart (2000) gives general conditions under which the solution of a static optimization problem estimated from data converges to the true value as the sample size grows. In other words, the conditions correspond to the near-optimality of the estimator (for the estimated problem), the true parameter value being well defined, and the estimated objective function converging uniformly to the true objective function over the domain. The first condition is satisfied because we assume that $\hat{\theta}_{n,\mathrm{MV}}$ and $\hat{\theta}_{n,\mathrm{CV}}$ are optimal for the estimated problem. The second condition is an identifiability condition, which we assume holds for our problems via (*) and (**). This is a mild criterion that is necessary for statistical inference; e.g., it suffices that θ_{MV} (respectively, θ_{CV}) be unique, Θ compact, and $M(\cdot)$ continuous. The third and final condition is the uniform convergence of the estimated objective function $M_n(\cdot)$ to its true value $M(\cdot)$. This is not a straightforward result, especially if the objective function is not differentiable, which is the case for the mean-CVaR problem. Showing uniform convergence for such functions requires intricate arguments that involve bracketing numbers (see van der Vaart 2000, Chapter 19). The proofs of Theorems 2 and 3 can be found in Online Appendix C.

4.1. Asymptotic Optimality of PBR Solutions

Let us now consider the PBR portfolio optimization problem. With similar global transformation as above, the PBR problem becomes

$$\hat{\theta}_{n,\mathrm{PBR}} = \underset{\theta=(\alpha,v)\in\mathbb{R}\times\mathbb{R}^{p-1}}{\arg\min}\; M_n(\theta;\lambda_1,\lambda_2), \qquad (\mathrm{PBR}')$$

where

$$M_n(\theta;\lambda_1,\lambda_2) = w^\top\hat{\Sigma}_n w - \lambda_0 w^\top\hat{\mu}_n + \lambda_1 w^\top\alpha, \qquad (11)$$

where α is as in (4), for the mean-variance problem (mv-PBR-1);

$$M_n(\theta;\lambda_1,\lambda_2) = w^\top \hat{\Sigma}_n w - \lambda_0 w^\top \hat{\mu}_n + \lambda_1 w^\top A^* w, \quad (12)$$

where A^* is as in (Q approx), for the mean-variance problem (mv-PBR-2); and

$$\begin{aligned} M_n(\theta;\lambda_1,\lambda_2) &= \frac{1}{n}\sum_{i=1}^{n} m_\theta(X_i) + \frac{\lambda_1}{n} w^\top \hat{\Sigma}_n w \\ &+ \frac{\lambda_2}{n(n-1)(1-\beta)^2}\sum_{i=1}^{n}\left(z_\theta(X_i) - \frac{1}{n}\sum_{j=1}^{n} z_\theta(X_j)\right)^2, \quad (13) \end{aligned}$$

for the mean-CVaR problem (cv-PBR). Note $\lambda_1, \lambda_2 \geq 0$ are parameters that control the degree of regularization; they play the same role as U_1 and U_2 in the original problem formulation.

We now prove the asymptotic optimality of the PBR solutions.

Theorem 4. *Assume* (*) *and* (**) *hold. Then, as n tends to infinity,*

$$\hat{\theta}_{n,\mathrm{PBR}}(\lambda_1,\lambda_2) \xrightarrow{\mathrm{P}} \theta_0,$$

where $\hat{\theta}_{n,\mathrm{PBR}}(\lambda_1,\lambda_2)$ *are minimizers of* $M_n(\theta,\lambda_1,\lambda_2)$ *equal to* (11)–(13), *and* θ_0 *is the corresponding true solution.*

The following result is an immediate consequence of Theorem 4, by the continuous mapping theorem.

Corollary 1 (Convergence of Performance of PBR Solutions). *Assume the same setting as in Theorem 4. Then the performance of the* PBR *solution also converges to the true performance of the true optimal solution; i.e.,*

$$\left|\hat{w}_{n,\mathrm{PBR}}^\top \mu - w_0^\top \mu\right| \xrightarrow{\mathrm{P}} 0$$

and

$$\left|\mathrm{Risk}(\hat{w}_{n,\mathrm{PBR}}^\top X;\beta) - \mathrm{Risk}(w_0^\top X;\beta)\right| \xrightarrow{\mathrm{P}} 0$$

as n tends to infinity, where $\hat{w}_{n,\mathrm{PBR}}$ *is the portfolio allocation corresponding to* $\hat{\theta}_{n,\mathrm{PBR}}$.

5. Results on Empirical Data

In this section, we compare the PBR method against a number of key benchmarks on three data sets: the 5, 10, and 49 industry portfolios from Ken French's website, which report monthly excess returns over the 90-day nominal U.S. T-bill (French 2015). We take the most recent 20 years of data, covering the period from January 1994 to December 2013. Our computations are done on a rolling-horizon basis, with the first 10 years of observations used as training data ($N_{\mathrm{train}} = 120$) and the last 10 years of observations used as test data ($N_{\mathrm{test}} = 120$). All computations were carried out on MATLAB2013a with the solver MOSEK and CVX, a package for specifying and solving convex programs Grant and Boyd (2008, 2013) on a Dell Precision T7600 workstation with two Intel Xeon E5-2643 processors, each of which has four cores and 32.0 GB of RAM.

5.1. Portfolio Strategies Considered for the Mean-Variance Problem

We compute the out-of-sample performances of the following eight portfolio allocation strategies:

1. SAA, which solves the sample average approximation problem (MV-SAA).
2. PBR (*rank*-1), which solves the rank-1 approximation problem (mv-PBR-1). The RHS of the PBR constraint, $\sqrt[4]{U}$, is calibrated using the out-of-sample performance-based k-fold cross-validation algorithm (OOS-PBCV), which we explain in detail in Section 5.4.
3. PBR (PSD), which solves the convex quadratic approximation problem (mv-PBR-2). The RHS of the PBR constraint, $\sqrt[2]{U}$, calibrated using OOS-PBCV.
4. NS, which solves problem (MV-SAA) with the no short-selling constraint $w \geq 0$, as in Jagannathan and Ma (2003).
5. L1 regularization, which solves the SAA problem (MV-SAA) with the extra constraint $\|w\|_1 \leq U$, where U is also calibrated using OOS-PBCV.
6. L2 regularization, which solves the SAA problem (MV-SAA) with the extra constraint $\|w\|_2 \leq U$, where U is also calibrated using OOS-PBCV.
7. *Minimum variance*, which solves the above (SAA, PBR (rank-1), PBR (PSD), NS, L1, and L2) for the global minimum variance problem, which is (MV-true) without the mean return constraint. We do this because the difficulty in estimating the mean return is a well-known problem (Merton 1980), and some more recent works in the Markowitz literature have shown that removing the mean constraint altogether can yield better results (e.g., Jagannathan and Ma 2003).
8. *Equally weighted portfolio*, where DeMiguel et al. (2009a) have shown that the naive strategy of equally dividing up the total wealth (i.e., investing in a portfolio w with $w_i = 1/p$ for $i = 1,\ldots,p$) performs very well relative to a number of benchmarks for the data-driven mean-variance problem. We include this as a benchmark.

5.2. Portfolio Strategies Considered for the Mean-CVaR Problem

We compute the out-of-sample performances of the following eight portfolio allocation strategies:

1. SAA, which solves the sample average approximation problem (CV-SAA).
2. PBR *only on the objective*, which solves problem (cv-PBR) with no regularization of the mean constraint; i.e., $U_2 = \infty$. The RHS of the objective regularization constraint, U_1, is calibrated using the OOS-PBCV (see Section 5.4).
3. PBR *only on the constraint*, which solves problem (cv-PBR) with no regularization of the objective function; i.e., $U_1 = \infty$. The RHS of the mean regularization constraint, U_2, is calibrated using OOS-PBCV.

4. PBR *on both the objective and the constraint*, which solves problem (cv-PBR). Both regularization parameters U_1 and U_2 are calibrated using OOS-PBCV.

5. L1 *regularization*, which solves the sample average approximation problem (cv-PBR) with the extra constraint $\|w\|_1 \le U$, where U is also calibrated using OOS-PBCV.

6. L2 *regularization*, which solves the sample average approximation problem (cv-PBR) with the extra constraint $\|w\|_2 \le U$, where U is also calibrated using OOS-PBCV.

7. *Equally weighted portfolio*, where DeMiguel et al. (2009a) have shown that the naive strategy of equally dividing up the total wealth (i.e., investing in a portfolio w with $w_i = 1/p$ for $i = 1, \ldots, p$) performs very well relative to a number of benchmarks for the data-driven mean-variance problem. We include this as a benchmark.

8. *Global minimum* CVaR *portfolio*, which solves the sample average approximation problem (CV-SAA) without the target mean return constraint $w^\top \hat{\mu}_n = R$. We do this because the difficulty in estimating the mean return is a well-known problem (Merton 1980), and some more recent works in the Markowitz literature have shown that removing the mean constraint altogether can yield better results (Jagannathan and Ma 2003). Thus, as an analogy to the global minimum variance problem, we consider the global minimum CVaR problem.

5.3. Evaluation Methodology

We evaluate the various portfolio allocation models on a rolling-horizon basis. In other words, we evaluate the portfolio weights on the first N_{train} asset return observations (the "training data"), then compute its return on the $(N_{\text{train}}+1)$th observation. We then roll the window by one period, evaluate the portfolio weights on the second to $(N_{\text{train}}+1)$th return observations, then compute its return on the $(N_{\text{train}}+2)$th observation, and so on, until we have rolled forward N_{test} number of times. Let us generically call the optimal portfolio weights solved over N_{test} number of times $\hat{w}_1, \ldots, \hat{w}_{N_{\text{test}}} \in \mathbb{R}^p$ and the asset returns $X_1, \ldots, X_{N_{\text{test}}} \in \mathbb{R}^p$. Also define

$$\hat{\mu}_{\text{test}} := \frac{1}{N_{\text{test}}} \sum_{t}^{N_{\text{test}}} \hat{w}_t^\top X_t,$$

$$\hat{\sigma}_{\text{test}}^2 := \frac{1}{N_{\text{test}}-1} \sum_{t}^{N_{\text{test}}} (\hat{w}_t^\top X_t - \hat{\mu}_{\text{test}})^2,$$

i.e., the out-of-sample mean and variance of the portfolio returns.

We report the following performance metrics:

1. *Sharpe ratio*: We compute annualized Sharpe ratio as

$$\textit{Sharpe} = \frac{\hat{\mu}_{\text{test}}}{\hat{\sigma}_{\text{test}}}. \tag{14}$$

2. *Turnover*: The portfolio turnover, averaged over the testing period, is given by

$$\textit{Turnover} = \frac{1}{N_{\text{test}}} \sum_{t=1}^{N_{\text{test}}} \sum_{j=1}^{p} \left| \hat{w}_{t+1,j} - \hat{w}_{t,j} \right|^{+}. \tag{15}$$

For further details on these performance measures, we refer the reader to DeMiguel et al. (2009a).

5.4. Calibration Algorithm for U: Performance-Based k-Fold Cross-Validation

One important question in solving (PBR) is how to choose the right-hand side of the regularization constraints U_1 and U_2. If they are set too small, the problem is infeasible; if set too large, regularization has no effect, and we retrieve the SAA solution. Ideally, we want to choose U_1 and U_2 so that it constrains the SAA problem just enough to maximize the out-of-sample performance. Obviously, one cannot use the actual test data set to calibrate U_1 and U_2, and we need to calibrate them on the training data set via a cross-validation (CV) method.

A common CV technique used in statistics is the k-fold CV. It works by splitting the training data set into k equally sized bins, training the statistical model on every possible combination of $k-1$ bins, and then validating on the remaining bin. Any parameter that needs to be tuned is tuned via the prediction accuracy on the validation data set.

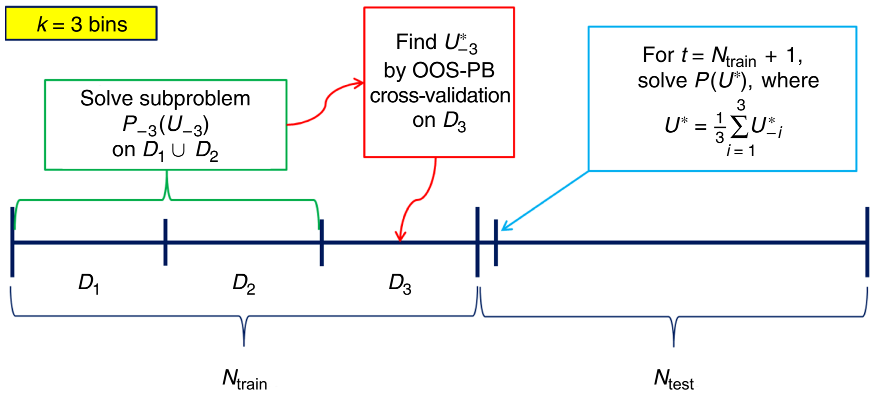
Here, we develop a performance-based k-fold CV method to find U_1 and U_2 that maximize the out-of-sample Sharpe ratio on the validation data set. The two key differences between our algorithm and the standard k-fold CV is that (i) the search boundaries for U_1 and U_2 need to be set carefully to avoid infeasibility and having no effect, and (ii) we validate by computing the Sharpe ratio (the main performance metric for investment in practice) as opposed to some measure of error.

For simplicity, we explain the algorithm for the case of having just one regularization constraint on the objective. We thus omit the subscript and refer to the RHS by U instead of U_1. Generalization to the two-dimensional case is straightforward. Figure 3 displays a schematic explaining the main parts of the algorithm, for the case $k=3$. Let $D = [X_1, \ldots, X_{N_{\text{train}}}] \in \mathbb{R}^{p \times N_{\text{train}}}$ be the training data set of stock returns. This is split into k equally sized bins, $D_1, D_2, \ldots, D_k$. Let $P_{-i}(U_{-i})$ denote the PBR problem solved on the data set $D \backslash D_i$ with RHS $U = U_{-i}$. We find the optimal U, denoted by U^*, on the whole data set D by the following steps:

1. Set a search boundary for U_{-i}, $[\underline{U}_{-i}, \bar{U}_{-i}]$.

2. Solve $P_{-i}(U_{-i})$ on $D \backslash D_i$ starting at $U_{-i} = \bar{U}_{-i}$, computing the Sharpe ratio of the solution on D_i, then repeating the process with progressively smaller U_{-i} via a descent algorithm. Find $U_{-i}^* \in [\underline{U}_{-i}, \bar{U}_{-i}]$ by a stopping criterion.

Figure 3. (Color online) A Schematic Explaining the OOS-PBCV Algorithm Used to Calibrate the Constraint HRS, U, for the Case $k = 3$



Note. The training data set is split into k bins, and the optimal U for the entire training data set is found by averaging the best U found for each subset of the training data.

3. Average over the k results to get $U^* = (1/k) \cdot \sum_{i=1}^{k} U^*_{-i}$.

We elaborate on these three parts of the CV algorithm below.

1. *Set a search boundary for* U_{-i}, $[\underline{U}_{-i}, \bar{U}_{-i}]$. As previously mentioned, setting the correct search boundary for U_{-i} is very important. We require the boundary for the ith subproblem to be contained within the allowable range for the problem on the entire data set, i.e., $[\underline{U}_{-i}, \bar{U}_{-i}] \subset [\underline{U}, \bar{U}]$. This is because if we solve the PBR problem on the whole training data set with $U > \bar{U}$, then PBR will not have any effect, and likewise, if we solve the PBR problem with $U < \underline{U}$, then the problem will be infeasible.

The upper bound on U is given by

$$\bar{U} = \widehat{\text{Risk}}_n(-\hat{w}_n^\top \mathrm{X}),$$

recalling that $\hat{w}_n$ is the SAA solution. In other words, the upper bound is set to be the value of the PBR penalty if the penalty were not imposed. To find $\underline{U}$, the minimum possible PBR parameter, we solve

$$\begin{aligned} \underline{U} = \min_{w} \quad & w^\top \alpha \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & w^\top \hat{\mu}_n = R, \end{aligned} \tag{U-min-mv1}$$

for (mv-PBR-1);

$$\begin{aligned} \underline{U} = \min_{w} \quad & w^\top A^* w \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & w^\top \hat{\mu}_n = R, \end{aligned} \tag{U-min-mv2}$$

for (mv-PBR-2); and

$$\begin{aligned} \underline{U} = \min_{w,z} \quad & z^\top \Omega_n z \\ \text{s.t.} \quad & w^\top \hat{\mu}_n = R, \\ & w^\top 1_p = 1, \\ & z_i \geq -w^\top X_i - \alpha, \quad i = 1, \ldots, n, \\ & z_i \geq 0, \quad i = 1, \ldots, n, \end{aligned} \tag{U-min-cv}$$

for (cv-PBR).

To find the upper bound on the subproblem, $\bar{U}_{-i}$, we compute (SAA) on data set $D \backslash D_i$ for $\hat{w}_{-i}$, then set

$$\bar{U}_{-i} = \min[\bar{U}, \widehat{\text{Risk}}_n(-\hat{w}_{-i}^\top \mathrm{X})].$$

To find $\underline{U}_{-i}$, we first solve

$$\begin{aligned} \underline{U}_{tmp} = \min_{w} \quad & w^\top \alpha \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & w^\top \hat{\mu}_{-i} = R, \end{aligned}$$

for (mv-PBR-1), where $\hat{\mu}_{-i}$ the sample mean computed on $D \backslash D_i$;

$$\begin{aligned} \underline{U}_{tmp} = \min_{w} \quad & w^\top A^* w \\ \text{s.t.} \quad & w^\top 1_p = 1, \\ & w^\top \hat{\mu}_{-i} = R, \end{aligned}$$

for (mv-PBR-2); and

$$\begin{aligned} \underline{U}_{tmp} = \min_{w,z} \quad & z^\top \Omega_{-i} z \\ \text{s.t.} \quad & w^\top \hat{\mu}_{-i} = R, \\ & w^\top 1_p = 1, \\ & z_i \geq -w^\top X_i - \alpha, \quad i \in C \backslash C_i, \\ & z_i \geq 0, \quad i \in C \backslash C_i, \end{aligned}$$

for (cv-PBR), where Ω_{-i} is the sample variance operator computed on $D \backslash D_i$, and C and C_i are sets of labels of the elements in D and D_i, respectively. We then set

$$\underline{U}_{-i} = \max[\underline{U}_{tmp}, \underline{U}].$$

The pseudocode for this part of the CV algorithm is shown in Algorithm 1.

Algorithm 1 (A pseudocode for the out-of-sample performance-based k-fold cross-validation algorithm (OOS-PBCV))

Initialize

Choose number of bins k

Solve PBR(U) on D_{train} to get $\hat{w}_{\text{train}}$;
  set $\bar{U} = (\hat{w}_{\text{train}})^\top \hat{\Sigma} \hat{w}_{\text{train}}$
Solve U-min-mv1(U) [U-min-mv2(U) or U-min-cv(U)]
  on D_{train} to get $\hat{w}_{U\min}$; set $\underline{U} = (\hat{w}_{U\min})^\top \hat{\Sigma} \hat{w}_{U\min}$
Divide up D_{train} randomly into k equal bins,
  D^b_{train}, $b = 1, \ldots, k$
Let D^{-b}_{train} denote the training data minus the bth bin
for $b \leftarrow 1$ **do** k **do**
  Solve PBR(U) on D^{-b}_{train} to get $\hat{w}_{-b}$;
    set $\bar{U}_{-b} = (\hat{w}_{-b})^\top \hat{\Sigma} \hat{w}_{-b}$
  if $\bar{U}_{-b} < \underline{U}$ **then** $U^*_{-b} = \underline{U}$ and terminate
    else Solve U-min-mv1(U)
      [U-min-mv2(U) or U-min-cv(U)]
      on D^{-b}_{train} to get $\hat{w}^{-b}_{U\min}$;
      set $\underline{U}_{-b} = (\hat{w}^{-b}_{U\min})^\top \hat{\Sigma} \hat{w}^{-b}_{U\min}$;
  end
  if $\underline{U}_{-b} > \bar{U}$ **then** $U^*_{-b} = \bar{U}$ and terminate
    else Compare and update boundaries:
    $\bar{U}_{-b} = \min(\bar{U}_{-b}, \bar{U})$
    $\underline{U}_{-b} = \max(\underline{U}_{-b}, \underline{U})$
    Run **(OOS-PBSD)** with boundaries $[\underline{U}_{-b}, \bar{U}_{-b}]$
      to find U^*_{-b};
  end
end
Return $U^* = \frac{1}{k}\sum_{i=1}^k U^*_{-i}$.

2. *Finding* $U^*_{-i} \in [\underline{U}_{-i}, \bar{U}_{-i}]$. To find the optimal parameter for the ith subproblem that *maximizes the out-of-sample Sharpe ratio*, we employ a backtracking line search algorithm (see Boyd and Vandenberghe 2004, Chapter 9.2), which is a simple yet effective descent algorithm. We start at the maximum $\bar{U}_{-i}$ determined in the previous step and descend by step size $t\Delta U := t(\bar{U}_{-i} - \underline{U}_{-i})/Div$, where Div a preset granularity parameter, t is a parameter that equals 1 initially and then is backtracked at rate β, a parameter chosen between 0 and 1, until the stopping criterion

$$Sharpe(U - t\Delta U) < Sharpe(U) + \alpha t \Delta U \frac{dSharpe(U)}{dU}$$

is met.

Computing $dSharpe(U)/dU$, the marginal change in the out-of-sample Sharpe ratio with a change in U, is slightly tricky, as we do not know how the out-of-sample Sharpe ratio depends on U analytically. Nevertheless, we can compute it numerically by employing the chain rule:

$$\frac{dSharpe(U)}{dU} = \nabla_{\hat{w}^*} Sharpe(\hat{w}^*(U))^\top \left[\frac{d\hat{w}^*(U)}{dU}\right],$$

where $\hat{w}^*(U)$ is the optimal PBR solution when the RHS is set to U. The first quantity, $\nabla_{\hat{w}^*} Sharpe(\hat{w}^*(U))$, can be computed explicitly, as we know the formula for the Sharpe ratio as a function of w. Suppressing the dependency of w on U, we have

$$\nabla_w Sharpe(w) = \frac{(w^\top \Sigma w)\mu - (w^\top \mu)\Sigma w}{(w^\top \Sigma w)^{3/2}}.$$

The second quantity, $dw^*(U)/dU$, is the marginal change in the optimal solution $\hat{w}^*$ as the RHS U changes. We approximate this by solving (PBR) with $(1 - bit)U$, where $0 < bit \ll 1$ is a predetermined parameter, then computing

$$\frac{d\hat{w}^*(U)}{dU} \approx \frac{\hat{w}^*(U) - \hat{w}^*\big((1-bit)U\big)}{bit \times U},$$

where $\hat{w}^*((1 - bit)U)$ is the new optimal allocation when the PBR constraint RHS is set to $(1 - bit)U$.

The pseudocode for this part of the CV algorithm is shown in Algorithm 2.

Algorithm 2 (A pseudocode for the out-of-sample performance-based steepest descent algorithm (OOS-PBSD), which is a subroutine of (OOS-PBCV))

Initialize
Choose backtracking parameters $\alpha \in (0, 0.5)$,
  $\beta \in (0, 1)$
Choose stepsize Div
Choose perturbation size $bit \in (0, 0.5)$
for $b \leftarrow 1$ **to** k **do**
  Set $U = \bar{U}_{-b}$, $\Delta U := t(\bar{U}_{-b} - \underline{U}_{-b})/Div$, $t = 1$
  Compute

$$\frac{dSharpe(U)}{dU} = \nabla_w Sharpe(\hat{w}_{-b}(U))^\top \left[\frac{d\hat{w}_{-b}(U)}{dU}\right],$$

  where

$$\nabla_w Sharpe(\hat{w}_{-b}(U)) = \frac{((\hat{w}_{-b})^\top \Sigma_{-b} \hat{w}_{-b})\mu_{-b} - (\hat{w}'_{-b}\mu)\Sigma_{-b}\hat{w}_{-b}}{((\hat{w}_{-b})^\top \Sigma_{-b} \hat{w}_{-b})^{3/2}}$$

$$\frac{d\hat{w}_{-b}(U)}{dU} = \frac{\hat{w}_{-b}(U) - \hat{w}_{-b}((1-bit)U)}{bit \times U}$$

  while

$$Sharpe(U - t\Delta U) < Sharpe(U) + \alpha t \Delta U \frac{dSharpe(U)}{dU}$$

  do
    $t = \beta t$
  end
end
Return $U^*_{-b} = U - t\Delta U$.

In our computations, we used the parameters $\alpha = 0.4$, $\beta = 0.9$, $Div = 5$, and $bit = 0.05$, and we considered $k = 2$ and $k = 3$ bins. It took on average approximately two seconds to solve one problem instance for all problem sizes and bin numbers considered in this paper.

5.5. Discussion of Results: Mean-Variance Problem

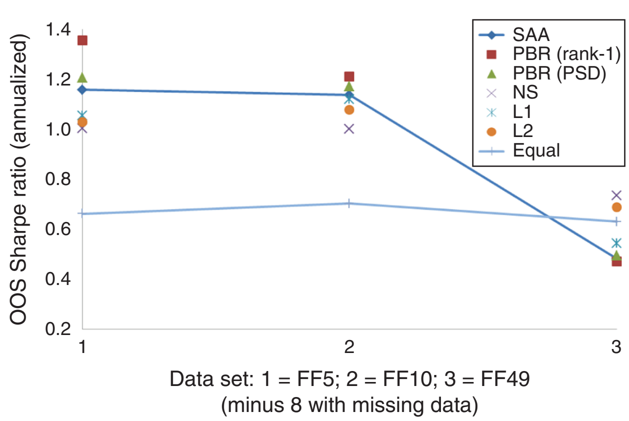
5.5.1. Out-of-Sample Sharpe Ratio. Table 1 reports the out-of-sample Sharpe ratios of the eight strategies listed in Section 5.1. For $p = 5$, the rank-1 approximation PBR performs the best, with a Sharpe ratio of 1.3551, followed by best convex quadratic approximation PBR (1.2052), then SAA (1.1573). For this data set, standard regularizations (L1, L2, and no short-selling) and the equally weighted portfolio all perform below these strategies. Similarly, for $p = 10$, the rank-1 approximation PBR performs the best, with a Sharpe ratio of 1.2112, followed by best convex quadratic approximation PBR (1.1696), then SAA (1.1357); the other strategies again relatively underperform. See Figure 4 for a graphical representation.

The $p = 41$ data set yields results that are quite different from those of $p = 5$ and $p = 10$, evidencing that dimensionality (i.e., the number of assets) is a significant factor in its own right (this has been observed in other studies; see, e.g., Jagannathan and Ma 2003; El Karoui 2010, 2013.). While we could rank the strategies by their average out-of-sample performances, they are statistically indistinguishable at the 5% level from the SAA method (all p-values are quite large, the smallest being 0.3178). Hence we cannot make any meaningful conclusions for this data set, and we leave the study of regularizing for dimensionality to future work.

From the perspective of an investor looking at the results of Table 2, the takeaway is clear: Focus on a small number of assets (the Fama–French (FF) 5 industry portfolio) and optimize using the PBR method on both the objective and mean constraints to achieve the highest Sharpe ratio.

5.5.2. Portfolio Turnover. Table 3 reports the portfolio turnovers of the eight strategies listed in Section 5.1. For obvious reasons, the equally weighted portfolio achieves the smallest portfolio turnover. For all three data sets, we find that the two PBR approximations generally have greater portfolio turnovers than SAA, whereas the standard regularization methods (L1, L2, and no short-selling) have lower turnovers than SAA. This is reflective of the fact that standard regularization is by design a *solution* stabilizer, whereas PBR is not.

Figure 4. (Color online) The Out-of-Sample Sharpe Ratios (Annualized) for the Strategies Considered for the Mean-Variance Problem, for Three Data Sets



Note. Detailed results are in Table 1.

5.6. Discussion of Results: Mean-CVaR Problem

5.6.1. Out-of-Sample Sharpe Ratio. Table 2 reports the out-of-sample Sharpe ratios of the eight strategies listed in Section 5.2. For $p = 5$ and $p = 10$ data sets, we find that PBR on both the objective and the constraint dominate the SAA solution. For example, the best Sharpe ratio for $p = 5$ for the SAA method is achieved by setting a return target of $R = 0.08$, yielding a Sharpe ratio of 1.2487, whereas the best PBR result for the same data set and target return has a Sharpe ratio of 1.2715, the difference of which is statistically significant at the 5% level (the exact p-value is 0.0453). Likewise, for $p = 10$, the best SAA Sharpe ratio of 1.0346 is dominated by the best PBR Sharpe ratio of 1.1506. This difference is statistically significant at the 10% level (the exact p-value is 0.0607). Also for the $p = 5$ and $p = 10$ data sets, the PBR method consistently dominates both L1 and L2 regularizations across all problem target returns and choice of the number of bins used for cross-validation. In addition, both the equally weighted portfolio and the global minimum CVaR portfolios underperform SAA, and hence also PBR on these data sets. See Figure 5 for a graphical representation.

The $p = 41$ data set yields results that are quite different from those of $p = 5$ and $p = 10$, signaling that dimensionality is an important parameter in its own right. First of all, the highest Sharpe ratio of all strategies across all target return levels and choice of bins is achieved by the equally weighted portfolio, with 0.6297. Second, all regularizations—PBR, L1, and L2—yield results that are statistically indistinguishable from the SAA method (all p-values are quite large, the smallest being 0.6249). Hence we cannot make any meaningful conclusions for this data set, and we leave the study of regularizing for dimensionality to future work.

Finally, let us comment on the effects of PBR on the objective and the mean estimations separately. The question that comes to mind is whether one constraint dominates the other—i.e., whether PBR on the objective only consistently dominates PBR on the mean, or vice versa. The answer is a yes, but the exact relationship depends on the data set: for $p = 5$ and $p = 10$, the Sharpe ratios of PBR on CVaR are better than that of PBR on the mean for each target return (and taking the best of the two bin results), whereas for $p = 41$, the opposite is true. This pattern seems to indicate that for

Table 1. Sharpe Ratios for Empirical Data for the Mean-Variance Problem

	FF 5 industry $p=5$		FF 10 industry $p=10$		FF 49 industry $p=41$ (−8 assets with missing data)	
Mean-variance $R=0.04$						
SAA	1.1459		1.1332		0.4744	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	1.2603	1.3254	1.1868	1.2098	0.4344	**0.4712**
	(0.0411)	(0.0286)	(0.0643)	(0.0509)	(0.5848)	(0.5386)
PBR (PSD)	1.1836	1.1831	1.1543	1.1678	0.4776	0.4825
	(0.0743)	(0.071)	(0.0891)	(0.0816)	(0.5593)	(0.5391)
NS	**1.0023**		0.9968		0.7345	
	(0.1404)		(0.1437)		(0.2977)	
L1	1.0136	1.0386	1.1185	1.1175	0.5419	0.5211
	(0.1568)	(0.1396)	(0.1008)	(0.1017)	(0.5044)	(0.5216)
L2	0.9711	1.0268	1.0579	1.0699	0.6672	0.6009
	(0.1781)	(0.1452)	(0.1482)	(0.1280)	(0.3950)	(0.4455)
Mean-variance $R=0.06$						
SAA	1.1535		**1.1357**		0.4468	
	Two bins	Three bins	Two bins	Three bins	Two bins	3 bins
PBR (rank-1)	1.2945	1.3362	1.1870	**1.2112**	0.4011	0.4515
	(0.0297)	(0.0244)	(0.0629)	(0.0503)	(0.6136)	(0.5530)
PBR (PSD)	1.1912	**1.2052**	1.1532	**1.1696**	0.4585	0.4587
	(0.0689)	(0.0638)	(0.0898)	(0.0809)	(0.5757)	(0.5598)
NS	0.9853		0.9699		0.7124	
	(0.1422)		(0.1537)		(0.3247)	
L1	0.9963	1.0198	1.0902	1.1010	0.4991	0.4941
	(0.1535)	(0.1394)	(0.1124)	(0.1101)	(0.5490)	(0.5448)
L2	0.9713	1.0265	1.0642	1.0755	0.6313	0.5701
	(0.1735)	(0.1425)	(0.1425)	(0.1238)	(0.4250)	(0.4696)
Markowitz $R=0.08$						
SAA	**1.1573**		1.1225		0.4253	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	1.3286	**1.3551**	1.1743	1.2018	0.3927	0.4253
	(0.0223)	(0.0208)	(0.0668)	(0.0510)	(0.6142)	(0.5778)
PBR (PSD)	1.1813	1.1952	1.1467	1.1575	0.4477	0.4366
	(0.0648)	(0.0614)	(0.0893)	(0.0844)	(0.5852)	(0.5804)
NS	0.9664		0.9405		0.6600	
	(0.1514)		(0.1577)		(0.3790)	
L1	0.9225	0.9965	1.0318	1.0779	0.4770	0.4930
	(0.1857)	(0.1403)	(0.1332)	(0.1181)	(0.5649)	(0.5379)
L2	0.9703	**1.0284**	1.0671	**1.0776**	0.6098	0.5522
	(0.1649)	(0.1398)	(0.1398)	(0.1209)	(0.4369)	(0.4785)
Minimum variance						
SAA	1.1454		1.1331		**0.4816**	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	1.2580	1.3269	1.1922	1.2086	0.4409	0.4683
	(0.0420)	(0.0288)	(0.0603)	(0.0505)	(0.5795)	(0.5472)
PBR (PSD)	1.1883	1.1882	1.154	1.1657	**0.4942**	0.4903
	(0.0710)	(0.0693)	(0.0892)	(0.0823)	(0.5400)	(0.5322)
NS	1.0022		**1.0012**		**0.7347**	
	(0.1405)		(0.1447)		(0.3178)	
L1	1.0321	**1.0546**	**1.1199**	1.1111	**0.5424**	0.5260
	(0.1455)	(0.1286)	(0.1000)	(0.1026)	(0.5017)	(0.5151)
L2	0.9945	1.0140	1.0543	1.0760	**0.6886**	0.6204
	(0.1632)	(0.1472)	(0.1488)	(0.1236)	(0.3761)	(0.4276)
Equal	**0.6617**		**0.7019**		**0.6297**	

Notes. This table reports the annualized out-of-sample Sharpe ratios of solutions to the mean-variance problem solved with the methods described in Section 5.1 for three different data sets for target returns, $R=0.04, 0.06, 0.08$. For each data set, the highest Sharpe ratio attained by each strategy is highlighted in boldface. To set the degree of regularization, we use the performance-based k-fold cross-validation algorithm detailed in Section 5.4, with $k=2$ and 3 bins. In parentheses we report the p-values of tests of differences from the SAA method. We also report the Sharpe ratio of the equally weighted portfolio.

Table 2. Sharpe Ratios for Empirical Data for the Mean-CVaR Problem

	FF 5 industry $p=5$		FF 10 industry $p=10$		FF 49 industry $p=41$ (−8 assets with missing data)	
Mean-CVaR $R=0.04$						
SAA	1.2137		1.0321		**0.3657**	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (CVaR only)	1.2113	1.1733	1.0506	1.1381	0.1304	0.1304
	(0.0554)	(0.0674)	(0.0638)	(0.0312)	(0.7908)	(0.7908)
PBR (mean only)	1.2089	1.1802	1.0994	1.0519	0.2732	0.3682
	(0.0746)	(0.0790)	(0.1051)	(0.1338)	(0.7518)	(0.6454)
PBR (both)	1.2439	1.2073	1.1112	1.1422	**0.3607**	0.2247
	(0.0513)	(0.0601)	(0.0691)	(0.0648)	(0.7054)	(0.7667)
L1	1.0112	**1.0754**	0.9254	0.9741	0.4048	0.4642
	(0.1497)	(0.1366)	(0.2293)	(0.1880)	(0.6874)	(0.6242)
L2	0.9650	1.0636	1.0031	0.9835	**0.3982**	0.3586
	(0.1780)	(0.1287)	(0.1512)	(0.1598)	(0.7087)	(0.6878)
Mean-CVaR $R=0.06$						
SAA	1.2179		1.0321		**0.3657**	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (CVaR only)	1.2223	1.2063	1.0518	1.1451	0.1265	0.1300
	(0.0503)	(0.0527)	(0.0633)	(0.0294)	(0.7920)	(0.7909)
PBR (mean only)	1.2205	1.1902	1.0988	1.0466	0.2704	0.3771
	(0.0699)	(0.0746)	(0.1053)	(0.1358)	(0.7531)	(0.6359)
PBR (both)	1.2450	1.2043	1.1122	**1.1506**	0.3503	0.2267
	(0.0504)	(0.0581)	(0.0686)	(0.0607)	(0.7102)	(0.7656)
L1	0.9404	1.0464	0.9276	0.9746	0.3888	0.4635
	(0.1812)	(0.1395)	(0.2282)	(0.1887)	(0.7001)	(0.6249)
L2	0.9271	1.0627	1.0146	0.9794	0.3842	0.3571
	(0.1977)	(0.1286)	(0.1432)	(0.1621)	(0.7175)	(0.6886)
Mean-CVaR $R=0.08$						
SAA	**1.2487**		**1.0346**		**0.3657**	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (CVaR only)	1.2493	1.2098	1.0551	1.1433	0.1304	0.1304
	(0.0434)	(0.0462)	(0.0579)	(0.0323)	(0.7908)	(0.7908)
PBR (mean only)	1.2480	1.2088	1.0987	1.0470	0.2675	0.3738
	(0.0591)	(0.0693)	(0.1053)	(0.1384)	(0.7541)	(0.6391)
PBR (both)	**1.2715**	1.2198	1.1122	1.1449	0.2656	0.2285
	(0.0453)	(0.0544)	(0.0664)	(0.0639)	(0.7618)	(0.7647)
L1	0.8921	0.9836	0.9416	**1.0087**	0.3855	**0.4872**
	(0.1964)	(0.1572)	(0.2122)	(0.1645)	(0.7008)	(0.6128)
L2	0.9367	**1.0801**	**1.0278**	0.9947	0.3784	0.3588
	(0.1989)	(0.1179)	(0.1323)	(0.1530)	(0.7177)	(0.6870)
Global minimum CVaR	**1.2137**		**1.0321**		**0.3657**	
Equal	**0.6617**		**0.7019**		**0.6297**	

Notes. This table reports the annualized out-of-sample Sharpe ratios of the solutions to the mean-CVaR problem solved with SAA; PBR with the regularization of the objective ("CVaR only"), the constraint ("mean only"), and both the objective and the constraint ("both"); and L1 and L2 regularization constraints for three different data sets and for target returns $R=0.04, 0.06, 0.08$. For each data set, the highest Sharpe ratio attained by each strategy is highlighted in boldface. To set the degree of regularization, we use the performance-based k-fold cross-validation algorithm detailed in Section 5.4, with $k=2$ and 3 bins. In parentheses we report the p-values of tests of differences from the SAA method. We also report the Sharpe ratio of the equally weighted portfolio and the solution to the global minimum CVaR problem (no mean constraint).

a smaller number of assets, CVaR estimation is more of an issue, whereas mean estimation is more problematic for a larger number of assets.

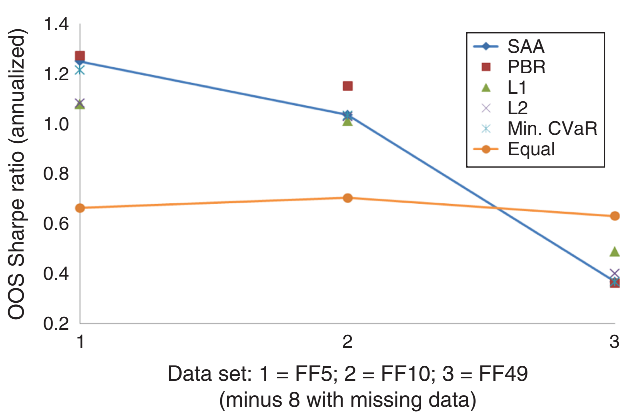
5.6.2. Portfolio Turnover. Table 4 reports the portfolio turnovers of the eight strategies listed in Section 5.2. For obvious reasons, the equally weighted portfolio achieves the smallest portfolio turnover. For the $p=5$ data set, the PBR method is consistently lower than the SAA, L1, and L2 regularization methods for each target return level and across the two bin sizes considered. The opposite is true for $p=10$ or $p=41$, however, with PBR having consistently higher turnovers than the SAA, L1, and L2 regularization methods for each target return level and across the two bin sizes

Table 3. Turnovers for Empirical Data for the Mean-Variance Problem

	FF 5 industry $p = 5$		FF 10 industry $p = 10$		FF 49 industry $p = 41$ (−8 assets with missing data)	
Mean-variance $R = 0.04$						
SAA	0.0935		0.1325		0.5188	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	0.1213	0.1292	0.1746	0.1851	0.5442	0.6611
PBR (PSD)	0.1002	0.0988	0.1415	0.1523	0.5201	0.4999
NS	0.0391		0.0544		0.0833	
L1	0.0986	0.0848	0.1158	0.1208	0.5167	0.4453
L2	0.1171	0.0901	0.1255	0.1071	0.4704	0.4079
Mean-variance $R = 0.06$						
SAA	0.1034		0.1339		0.5289	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	0.1397	0.1357	0.1741	0.1841	0.5646	0.6427
PBR (PSD)	0.1132	0.1086	0.1442	0.1513	0.5301	0.5042
NS	0.0417		0.0711		0.0859	
L1	0.1206	0.0963	0.1256	0.1205	0.4992	0.4439
L2	0.1267	0.0992	0.1379	0.1121	0.4809	0.4110
Mean-variance $R = 0.08$						
SAA	0.1288		0.1475		0.5434	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	0.1775	0.1504	0.1894	0.1959	0.5721	0.5434
PBR (PSD)	0.1147	0.1344	0.1689	0.1547	0.5414	0.5204
NS	0.0511		0.0965		0.1122	
L1	0.1476	0.1246	0.1480	0.1392	0.5064	0.4567
L2	0.1582	0.1241	0.1470	0.1229	0.5118	0.4200
Minimum variance						
SAA	0.1034		0.1325		0.5146	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (rank-1)	0.1245	0.1311	0.1756	0.1807	0.5393	0.6065
PBR (PSD)	0.1221	0.1182	0.1609	0.1682	0.5138	0.5022
NS	0.0391		0.0524		0.0835	
L1	0.0995	0.0886	0.1138	0.1219	0.4956	0.4435
L2	0.1213	0.0910	0.1255	0.1061	0.4575	0.4070
Equal	0.0427		0.0382		0.0483	

Notes. This table reports the portfolio turnovers (defined in Equation (15)) of the solutions to the mean-variance problem solved with the methods described in Section 5.1 for three different data sets and for target returns $R = 0.04, 0.06, 0.08$. We also report the turnovers of the equally weighted portfolio.

Figure 5. (Color online) The Out-of-Sample Sharpe Ratios (Annualized) for the Strategies Considered for the Mean-CVaR Problem, for Three Data Sets



Note. Detailed results are in Table 2.

considered. Global minimum variance portfolios have turnovers greater than the equally weighted portfolio but generally less than the SAA method.

6. Conclusion

We introduced performance-based regularization and performance-based cross-validation for the portfolio optimization problem and investigated them in detail. The PBR models constrain sample variances of estimated quantities in the problem—namely, the portfolio risk and return. The PBR models are shown to have equivalent robust counterparts, with new, nontrivial robust constraints for the portfolio risk. We have shown that PBR with performance-based cross-validation is highly effective at improving the finite-sample performance of the data-driven portfolio decision compared with SAA as well as other benchmarks known benchmarks in the literature. We conclude that PBR is

Table 4. Turnovers for Empirical Data for the Mean-CVaR Problem

Sharpe ratios	FF 5 industry $p=5$		FF 10 industry $p=10$		FF 49 industry $p=41$ (−8 assets with missing data)	
Mean-CVaR $R=0.04$						
SAA	0.2857		0.3534		1.6833	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (CVaR only)	0.1834	0.1985	0.4049	0.5586	1.7773	1.7773
PBR (mean only)	0.1230	0.1274	0.3104	0.2731	1.3173	1.5023
PBR (both)	0.1387	0.1388	0.3700	0.3682	1.7492	1.4158
L1	0.1992	0.1581	0.3415	0.2722	1.5158	1.3731
L2	0.1565	0.1469	0.2288	0.2270	1.2192	1.1217
Mean-CVaR $R=0.06$						
SAA	0.2909		0.3534		1.6833	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (CVaR only)	0.1918	0.2074	0.4071	0.5616	1.7922	1.7778
PBR (mean only)	0.1364	0.1414	0.3100	0.2729	1.3188	1.5147
PBR (both)	0.1519	0.1526	0.3724	0.3672	1.7610	1.4170
L1	0.2263	0.1754	0.3498	0.2723	1.5232	1.3730
L2	0.1842	0.1532	0.2407	0.2401	1.2374	1.1220
Mean-CVaR $R=0.08$						
SAA	0.2980		0.3615		1.6833	
	Two bins	Three bins	Two bins	Three bins	Two bins	Three bins
PBR (CVaR only)	0.2148	0.2242	0.4486	0.6472	1.7775	1.7775
PBR (mean only)	0.1517	0.1575	0.3066	0.2827	1.3228	1.5038
PBR (both)	0.1693	0.1681	0.4099	0.4034	1.6887	1.4190
L1	0.3100	0.2395	0.3628	0.3042	1.5370	1.3731
L2	0.2451	0.1835	0.2588	0.2633	1.1774	1.1227
Global minimum CVaR	0.2857		0.3534		1.6833	
Equal	0.0427		0.0382		0.0483	

Notes. This table reports the portfolio turnovers (defined in Equation (15)) of the solutions to the mean-CVaR problem solved with SAA; PBR with regularization of the objective ("CVaR only"), the constraint ("mean only"), and both the objective and the constraint ("both"); and L1 and L2 regularization constraints for three different data sets and for target returns $R=0.04, 0.06, 0.08$. We also report the turnovers of the equally weighted portfolio and the solution to the global minimum CVaR problem (no mean constraint).

a promising modeling paradigm for handling uncertainty, and it is worthy of further study to generalize to other decision problems.

Acknowledgments

The opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the supporting organizations.

Final note: Author Gah-Yi Ban has formerly spelled her last name as Vahn.

Endnotes

[1] There is empirical evidence that ignoring the mean return constraint yields better solutions (see Jorion 1985).

[2] Also known as expected shortfall; see Acerbi and Tasche (2002).

[3] The parameters are the sample mean and covariance matrix of data from 500 daily returns of 10 different U.S. stocks from January 2009 to January 2011.

References

Acerbi C, Tasche D (2002) Expected shortfall: A natural coherent alternative to value at risk. *Econom. Notes* 31(2):379–388.

Ahmadi AA, Olshevsky A, Parrilo PA, Tsitsiklis JN (2013) NP-hardness of deciding convexity of quartic polynomials and related problems. *Math. Programming* 137(1–2):453–476.

Ban G-Y, Chen CJ (2016) Portfolio optimization in high dimensions: Aggregate then optimize. Working paper, London Business School, London.

Belloni A, Chernozhukov V (2013) Least squares after model selection in high-dimensional sparse models. *Bernoulli* 19(2): 521–547.

Ben-Tal A, El Ghaoui L, Nemirovski A (2009) *Robust Optimization* (Princeton University Press, Princeton, NJ).

Best MJ, Grauer RR (1991) On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results. *Rev. Financial Stud.* 4(2):315–342.

Boyd SP, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Broadie M (1993) Computing efficient frontiers using estimated parameters. *Ann. Oper. Res.* 45(1):21–58.

Candes E, Tao T (2007) The dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* 35(6):2313–2351.

Chopra VK (1993) Improving optimization. *J. Investing* 2(3):51–59.

Chopra VK, Ziemba WT (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. *J. Portfolio Management* 19(2):6–11.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.

DeMiguel V, Garlappi L, Uppal R (2009a) Optimal versus naive diversification: How inefficient is the $1/n$ portfolio strategy? *Rev. Financial Stud.* 22(5):1915–1953.

DeMiguel V, Nogales FJ, Uppal R (2014) Stock return serial dependence and out-of-sample portfolio performance. *Rev. Financial Stud.* 27(4):1031–1073.

DeMiguel V, Garlappi L, Nogales FJ, Uppal R (2009b) A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Sci.* 55(5):798–812.

El Karoui N (2010) High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: Risk underestimation. *Ann. Statist.* 38(6):3487–3566.

El Karoui N (2013) On the realized risk of high-dimensional Markowitz portfolios. *SIAM J. Financial Math.* 4(1):737–783.

Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London, Ser. A* 222:309–368.

Fisher RA (1925) Theory of statistical estimation. *Math. Proc. Cambridge Philos. Soc.* 22(5):700–725.

Frankfurter GM, Phillips HE, Seagle JP (1971) Portfolio selection: The effects of uncertain means, variances and covariances. *J. Financial Quant. Anal.* 6(5):1251–1262.

French KR (2015) Data library. Accessed August 20, 2015, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Frost PA, Savarino JE (1986) An empirical bayes approach to efficient portfolio selection. *J. Financial Quant. Anal.* 21(3):293–305.

Frost PA, Savarino JE (1988) For better performance: Constrain portfolio weights. *J. Portfolio Management* 15(1):29–34.

Goldfarb D, Iyengar G (2003) Robust portfolio selection problems. *Math. Oper. Res.* 28(1):1–38.

Gotoh JY, Takeda A (2010) On the role of norm constraints in portfolio selection. *Comput. Management Sci.* 8:323–353.

Gotoh J, Kim MJ, Lim AEB (2015) Robust empirical optimization is almost the same as mean-variance optimization. Working paper, Chuo University, Tokyo.

Grant M, Boyd S (2008) Graph implementations for nonsmooth convex programs. Blondel V, Boyd S, Kimura H, eds. *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, Vol. 371 (Springer-Verlag, London), 95–110. http://stanford.edu/~boyd/graph_dcp.html.

Grant M, Boyd S (2013) CVX: Matlab software for disciplined convex programming, version 2.1. Accessed June 24, 2013, http://cvxr.com/cvx.

Haff L (1980) Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* 8(3):586–597.

Hastie T, Tibshirani R, Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. (Springer-Verlag, New York).

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.*, Vol. 1 (University of California Press, Berkeley), 221–233.

Ivanov VK (1962) On linear problems which are not well-posed. *Soviet Math. Dokl.* 145(2):981–983.

Jagannathan R, Ma T (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps. *J. Finance* 58(4): 1651–1684.

Jegadeesh N, Titman S (1993) Returns to buying winners and selling losers: Implications for stock market efficiency. *J. Finance* 48(1):65–91.

Jobson JD, Korkie B (1980) Estimation for Markowitz efficient portfolios. *J. Amer. Statist. Assoc.* 75(371):544–554.

Jorion P (1985) International portfolio diversification with estimation risk. *J. Bus.* 58(3):259–278.

Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* 88(2): 365–411.

Lim AEB, Shanthikumar JG, Vahn G-Y (2011) Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Oper. Res. Lett.* 39(3):163–171.

Lo AW, MacKinlay AC (1990) When are contrarian profits due to stock market overreaction? *Rev. Financial Stud.* 3(2):175–205.

Markowitz H (1952) Portfolio selection. *J. Finance* 7(1):77–91.

Merton RC (1980) On estimating the expected return on the market: An exploratory investigation. *J. Financial Econom.* 8(4):323–361.

Michaud R (1989) The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts J.* 45(1):31–42.

Phillips DL (1962) A technique for the numerical solution of certain integral equations of the first kind. *J. ACM* 9(1):84–97.

Rockafellar R, Uryasev S (2000) Optimization of conditional value-at-risk. *J. Risk* 2(3):21–41.

Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming: Modeling and Theory* (Society for Industrial and Applied Mathematics, Philadelphia).

Tikhonov A (1963) Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* 5:1035–1038.

van der Vaart A (2000) *Asymptotic Statistics* (Cambridge University Press, Cambridge, UK).

Vapnik V (2000) *The Nature of Statistical Learning Theory* (Springer, New York).