The effect of regularization in portfolio selection problems

Bernardo K. Pagnoncelli[1] · Felipe del Canto[2] · Arturo Cifuentes[3]

Abstract

Portfolio selection problems have been thoroughly studied under the risk-and-return paradigm introduced by Markowitz. However, the usefulness of this approach has been hindered by some practical considerations that have resulted in poorly diversified portfolios, or, solutions that are extremely sensitive to parameter estimation errors. In this work, we use sampling methods to cope with this issue and compare the merits of two approaches: a sample average approximation approach and a performance-based regularization (PBR) method that appeared recently in the literature. We extend PBR by incorporating three different risk metrics—integrated chance-constraints, quantile deviation, and absolute semi-deviation—and deriving the corresponding regularization formulas. Additionally, a numerical comparison using index-based portfolios is presented using historic data that includes the subprime crisis.

Keywords Portfolio optimization · Regularization · Cross-validation · Risk measures · Sample average approximation · Markowitz

Mathematics Subject Classification 90C15 · 46N10 · 91G10

1 Introduction

The seminal work of Markowitz (1952) changed the landscape of asset allocation problems, which up to that point were usually tackled in an ad-hoc fashion, see Kolm et al. (2014). By casting portfolio selection as a well-defined optimization problem, he established the risk-return paradigm which is still the fundamental

✉ Bernardo K. Pagnoncelli
bernardo.pagnoncelli@uai.cl

1 Escuela de Negocios, Adolfo Ibáñez University, Santiago, Chile

2 Pontificia Universidad Católica de Chile, Santiago, Chile

3 Clapes-UC, Santiago, Chile

reference framework used by most investment professionals. In essence, he created the conceptual structure that gave birth to modern portfolio theory.

While Markowitz's contribution cannot be overstated, the direct application of the method proposed in the original paper proved problematic in practice for a variety of reasons. First, the sensitivity of the solution to the inputs—expected returns and the covariance matrix—is considerable, and given that estimation errors are always present the resulting portfolio was often deemed unreliable (see Kolm et al. 2014 and references therein). Second, it has been observed that in many cases the solution was not fully diversified. It is widely accepted that diversification is highly desirable in practice since it is an efficient way of protecting investors against unexpected events (e.g. situations of extreme volatility or geopolitical turmoil). However, forcing diversification in an ad-hoc fashion can lead to poor quality solutions due to an overly constrained feasible set. In Green and Hollifield (1992) the authors derived theoretical conditions under which diversification occurs for the mean-variance problem.

On the other hand, using the variance as a risk measure brings both practical and theoretical limitations. Being symmetrical, the variance is sensitive to extreme values on both ends of the distribution. In essence, it mixes both, "positive" and "negative" events offering a distorted assessment of the true risk the investor is exposed to. In the last 60 years there has been a significant amount of research into different ways of capturing risk (see the excellent surveys Krokhmal et al. 2011; Mansini et al. 2014), and we make use of some of those risk measures in the present work.

Markowitz optimization model assumes the parameters are known exactly, which, of course, in practice is not the case. Several different alternatives have been proposed to overcome this issue: popular choices include the assumption that the parameters follow a known distribution function, or that they belong to a specific uncertainty set. In the first case a widely-used methodology is the sample average approximation (SAA) approach described in Shapiro (2003). Given a distribution function for the returns, a deterministic approximation of the original stochastic problem can be constructed drawing samples from the given distribution. Under certain conditions, it can be shown that, as the sample size increases, the optimal solution and optimal value of the SAA problem converge to their exact counterparts in the original stochastic problem. More traditional methods in stochastic programming often make assumptions regarding the underlying distribution of the problem. SAA is flexible, and it has been shown that even for problems with an astronomical number of scenarios good candidate solutions can be obtained by sampling a few thousand of those scenarios (e.g. Linderoth et al. 2006).

Another popular approach is robust optimization (RO), and its variations. First proposed by Soyster (1973) and later developed in Bertsimas and Sim (2004), RO pursues a worst-case approach—it attempts to protect the decision maker against all possible realizations of the random parameters by considering an uncertainty set. Thus, the decision maker needs to solve a deterministic problem to obtain a solution that offers protection against all possible realizations within this set. RO is also helpful to avoid estimation errors, generating portfolio less sensitive to parameter changes. Examples of RO applications in finance include DeMiguel and Nogales (2009), Fernandes et al. (2016), Goldfarb and Iyengar (2003), Kawas and Thiele (2011), Wang et al. (2016) and Quaranta and Zaffaroni (2008).

Building on these methods, more recent approaches consider risk-averse SAA formulations that combine machine learning and regularization schemes to obtain diversified portfolios with good out-of-sample performance. Machine learning tools, which are widely used in other areas such as regression analysis (see Bishop 2006) and data mining (see Witten et al. 2011), are recently making their way into portfolio management. Algorithms developed by this discipline, like cross-validation and classification algorithms, can be of great importance for estimation procedures and as decision supporting tools.

Regularization techniques, commonly used in high-dimensional regression problems (e.g. Belloni and Chernozhukov 2013; Candes and Tao 2007), have also made an entrance in portfolio optimization due to their ability to cope with numerical stability issues (e.g. Tikhonov 1963) or sparsity of the solution (e.g. Lasso, Tibshirani 1996). In Brodie et al. (2009) the authors reformulated the mean-variance problem into a constrained least-squares regression, and added a ℓ-1 penalty to the objective function that encourages sparse portfolios. The ℓ-1 regularization is the most common approach to stabilize solution of portfolio problems, see Corsaro and De Simone (2019) and Dai and Wen (2018) for additional examples. In Fastrich et al. (2015) the authors extend ℓ-1 regularization by considering Lasso, and propose new non-convex regularization terms that exhibit good performance when applied to large data sets. The authors use cross-validation to estimate the parameter. By using tools from statistical learning theory (Still and Kondor 2010) use the ℓ-2 norm of the weight vector to induce diversification and achieve stability on out-of-sample experiments.

Recently, Ban et al. (2016) proposed a method called performance-based regularization (PBR) which focuses on both sides of the optimization problem: it acknowledges parameter estimation difficulties, and aims at generating solutions that are more diversified, which is key in practice. The central idea is to exclude solutions that are in-sample optimal, but have potentially high out-of-sample variability. The authors considered the problem of minimizing two risk measures, the variance and the Conditional Value-at-Risk (CVaR), subject to having returns higher than some threshold, in addition to regularization constraints.

In this paper, we formulate an optimization problem based on a maximization-of-return approach (instead of a minimization-of-risk approach), since we believe is more intuitive from a practical standpoint. With that as background our objective is twofold.

First, we consider three different risk measures to cast the portfolio optimization problem constraints: (i) integrated chance constraints, proposed by Haneveld (1986); ii) quantile deviation (see Cotton and Ntaimo 2015); and, absolute semi-deviation, proposed in Ogryczak and Ruszczyski (2002). We prove certain properties of these measures that are key to construct the corresponding PBR version of the relevant constraints.

Second, an extensive numerical comparison between a pure SAA approach and a PBR-based approach is presented using the three risk measures just described, in addition to the CVaR. In light of the mounting evidence in favor or passive strategies vis-à-vis active portfolio selection strategies, we have chosen to cast the optimization problem based on indices rather than individual stocks or bonds (see Arnott et al. 2000;

Bogle 1995; Elton and Blake 1996; Malkiel 1995, 1996), and we rebalance the portfolio only once a year. The experiments cover the 10-year period between January 2003 and December 2012, and thus, they include the subprime crisis, the most challenging market environment of the last 50 years. The strategy is based on a rolling horizon scheme, in which past data are used to estimate the portfolio optimization parameters, and the performance of the optimal solution is tested with samples obtained from a parameterized model of returns. By using different samples and risk measures, the results show that the formulation with PBR constraints results in higher levels of diversification compared to solutions obtained with the SAA approach. Nevertheless, in periods of relative market stability, SAA outperforms PBR because the solutions with more variability end up being the ones with higher expected returns.

The rest of the paper is organized as follows. Section 2 describes common regularized mean-risk formulations, including PBR, as well as our modification based on a max-return framework. In Sect. 3 we derived the relevant expressions to use the three risk measures mentioned within the context of PBR. Sect. 4 shows the numerical simulation results, and Sect. 5 presents our conclusions.

2 Regularized mean-risk formulations

Consider the following mean-risk formulation:

$$
\begin{aligned}
\min_{w \in \mathbb{R}^p} \quad & \text{Risk}(w^T X), \\
s.t. \quad & w^T \mathbf{1}_p = 1, \\
& w^T \mu = R, \\
& (w \geq 0),
\end{aligned}
\tag{1}
$$

where $w \in \mathbb{R}^p$ is the investor's portfolio, $\mathbf{1}_p \in \mathbb{R}^p$ is a vector of ones, $X \in \mathbb{R}^p$ is a random vector representing the return of p assets, $\mu = \mathbb{E}(X)$ is a vector of averages of each component of X and Risk $: \mathcal{X} \to \mathbb{R}$ is a risk measure defined on some space of random variables, e.g. the L^1 space. The parenthesis in the last constraint indicates it is optional. For most risk measures—and most distribution functions—it is not possible to explicitly solve problem (1). The SAA formulation associated to problem (1) is given by

$$
\begin{aligned}
\min_{w \in \mathbb{R}^p} \quad & \widehat{\text{Risk}}(w^T \mathsf{X}), \\
s.t. \quad & w^T \mathbf{1}_p = 1, \\
& w^T \hat{\mu} = R, \\
& (w \geq 0),
\end{aligned}
\tag{2}
$$

where $\mathsf{X} = (X_1, \dots, X_n)$ is a random vector representing observed returns, $\widehat{\text{Risk}}(w^T \mathsf{X})$ is the sample estimator of $\text{Risk}(w^T X)$ and $\hat{\mu}$ is the vector of sample averages based on the n observations of the random vector X, $\hat{\mu} = (1/n) \sum_{i=1}^{n} X_i$.

It is well-known that the solution of problem (2) can be highly unreliable due to estimation errors. This happens when $\text{Risk}(w^T X) = w^T \Sigma w$ (the mean-variance problem) as shown in Best and Grauer (1991), Broadie (1993), Chopra and Ziemba (1993), Frankfurter et al. (1971), Frost and Savarino (1986), Frost and Savarino (1988) and Michaud (1989), and also when the risk measure is the CVaR (Lim et al. 2011). Recall that the CVaR at a confidence level α is defined as

$$\text{CVaR}_\alpha(w^T X) := \min_{\eta \in \mathbb{R}} \eta + \frac{1}{1-\alpha} \mathbb{E}\left[(w^T X - \eta)^+ \right],$$

where $(a)^+$ denotes the maximum between $a \in \mathbb{R}$ and 0. In order to overcome this problem, the work Ban et al. (2016) proposes a performance-based regularization (PBR) to control the instability of the SAA solution in terms of its out-of-sample behavior. The main idea behind their method is to constrain the variances of the estimators of $\widehat{\text{Risk}}(w^T \mathsf{X})$ and $\omega^T \hat{\mu}$ in order to move the SAA solution away from portfolios with high variability, which tend to be less diversified and yield poor out-of-sample performance in many cases. This regularization method imposes two additional constraints to problem (2), as follows:

$$\text{SV}\left[\widehat{\text{Risk}}(w^T \mathsf{X}) \right] \leq U_1, \tag{3}$$

$$\text{SV}\left[w^T \hat{\mu} \right] \leq U_2, \tag{4}$$

where $\text{SV}(\cdot)$ is the standard unbiased sample variance estimator of the variance, denoted by $\text{Var}(\cdot)$, and U_1 and U_2 are real numbers that are obtained using cross-validation on past data. The values are selected according to the solution that gives the highest Sharpe Ratio. Most importantly, the cross-validation procedure defines automatically the values of U_1 and U_2, so they are not inputs that need to be defined by the investor. Note that constraint (4) does not depend on the choice of the risk measure and can be easily converted to a quadratic constraint by noting that

$$\text{Var}\left[w^T \hat{\mu} \right] = w^T \Sigma w,$$

where Σ is the covariance matrix of X, which gives

$$\text{SV}\left[w^T \hat{\mu} \right] = w^T \Sigma_n w,$$

where Σ_n is the sample estimator of Σ.

Expression (3) is more involved and when the risk measure is the variance, or the CVaR, Ban et al. (2016) derived explicit expressions for constraint (3). Recall that one of the purposes of this paper is to derive closed-form expressions for constraint (3) when using the three risk measures mentioned before. Our first step is to consider a mean-risk formulation different from problem (2), as follows:

$$\max_{w \in \mathbb{R}^p} \quad w^T \hat{\mu},$$
$$s.t. \quad w^T \mathbf{1}_p = 1,$$
$$\widehat{\text{Risk}}(w^T \mathsf{X}) \le k,$$
$$\text{SV}\left[\widehat{\text{Risk}}(w^T \mathsf{X})\right] \le U_1, \tag{5}$$
$$w^T \Sigma_n w \le U_2,$$
$$(w \ge 0),$$

where k is a risk-tolerance parameter selected by the investor. We believe that formulating the optimization problem in reference to a return-maximization framework is more intuitive from a practical viewpoint. Additionally, we think it is easier to select a priori an appropriate risk parameter than a desired return. (Returns, unlike risk tolerance levels, can be, at least in theory, unbounded). In the next section we derive explicit convex formulations of problem (5) for the three different risk measures.

3 Extension to different risk measures

We start defining the three risk measures we will work with in this paper. In all cases let $X \in \mathbb{R}^p$ be a random vector and $w \in \mathbb{R}^p$ a vector that represents decisions. All proofs of propositions and lemmas are relegated to the appendix.

– Integrated chance constraints (ICC): In Haneveld (1986) the integrated chance constraints (ICC) are defined as

$$\mathbb{E}_\omega\left[\left(h(\omega) - w^T X(\omega)\right)^+\right],$$

where $h(\omega) \in \mathbb{R}$ is a random benchmark, e.g. an index, $(x)^+ = \max\{x, 0\}$. ICC imposes that the average violation from a target must be bounded by a given threshold, which will be given by the risk-tolerance parameter k. Let $\mathsf{Y} = (Y_1, \dots, Y_n)$ be a random vector with $Y_i = (X_{i,1}, \dots, X_{i,p}, h_i)$, $X_{i,j}$ is the return of asset j in scenario i and h_i is the value of the stochastic benchmark $h(\omega)$ in scenario i. We have $X_i = (X_{i,1}, \dots, X_{i,p})$ as the vector representing the i-th sample of asset returns and $\mathsf{X} = (X_1, \dots, X_n)$ is the random vector of returns, as in formulation (2). The sample estimator of ICC is

$$\widehat{\text{ICC}}(w; \mathsf{Y}) = \widehat{\text{ICC}}(w, h; \mathsf{X}) = \frac{1}{n} \sum_{i=1}^{n} (h_i - w^T X_i)^+. \tag{6}$$

– Absolute semideviation (ASD): In Ogryczak and Ruszczynski (2002) the absolute semideviation (ASD) is defined as

$$\text{ASD}(w) = \mathbb{E}\left[\left(w^T X - w^T \mu\right)^+\right],$$

where $\mu = \mathbb{E}[X]$ is the vector of mean returns. ASD measures the average one-sided excesses with respect to the mean. The sample estimator of ASD is

$$\widehat{\text{ASD}}(w;\mathsf{X}) = \frac{1}{n} \sum_{i=1}^{n} \left(w^T X_i - w^T \hat{\mu}\right)^+, \tag{7}$$

where $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

– Quantile deviation (QDEV): Let $f : \mathbb{R}^p \times \Omega \to \mathbb{R}$ be a function such that $\mathbb{E}\big[|f(w,X)|\big] < \infty$ for every $X \in \Omega$. Let $\alpha \in (0,1)$. In Ogryczak and Ruszczyski (2002), quantile deviation (QDEV) is defined as

$$\text{QDEV}_\alpha(w) = \mathbb{E}\Big[(1-\alpha)\big((\kappa_\alpha - 1)f(w,X)\big)^+ + \alpha\big((1-\kappa_\alpha)f(w,X)\big)^+\Big],$$

where k_α is the α-quantile of the distribution of $f(w, x)$. Similar to the CVaR, in Ruszczyski and Shapiro (2006) $\text{QDEV}_\alpha[w]$ is shown to be equivalent to the result of the following minimization problem:

$$\text{QDEV}_\alpha(w) = \min_{\eta \in \mathbb{R}} \mathbb{E}\Big[\epsilon_1\big(\eta - w^T X\big)^+ + \epsilon_2\big(w^T X - \eta\big)^+\Big], \tag{8}$$

with $f(x, W) = w^T X$, $\alpha = \frac{\epsilon_2}{\epsilon_1 + \epsilon_2}$, and $\epsilon_1, \epsilon_2 > 0$. The sample estimator of QDEV based on expression (8) is

$$\widehat{\text{QDEV}}_\alpha(w;\mathsf{X}) = \min_{\eta \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} \epsilon_1\big(\eta - w^T X_i\big)^+ + \epsilon_2\big(w^T X_i - \eta\big)^+.$$

Before presenting the expressions for SV of the risk measures defined above, we introduce a useful definition and a lemma.

Definition 1 Let $n \in \mathbb{N}$, $n \geq 2$. We define Ω_n as the matrix

$$\Omega_n := \frac{1}{n-1}\big[I_n - n^{-1} 1_n 1_n^T\big],$$

where I_n is the $n \times n$ identity matrix and $1_n = (1, 1, \dots, 1) \in \mathbb{R}^{1 \times n}$.

Lemma 1 *Let* $\mathsf{z} = (z_1, \dots, z_n)$ *be a sample from a given distribution* $F(\cdot)$. *Then the sample variance is*

$$\text{SV}[\mathsf{z}] = \mathsf{z}^T \Omega_n \mathsf{z}.$$

3.1 Regularized ICC constraint

For the ICC we have the following result:

Proposition 1 *Let* $Y_1, \dots, Y_n \overset{\text{i.i.d}}{\sim} F$, *where the cumulative distribution F has finite second moment, and let* $\widehat{\text{ICC}}_n(w;\mathsf{Y})$ *be defined as in* (6). *Then*

$$\text{Var}\left[\widehat{\text{ICC}}(w;Y)\right] = \frac{1}{n}\text{Var}\left[(h - w^T X)^+\right].$$

Corollary 1 *Under the assumptions of Proposition* 1 *we have*

$$\text{SV}\left[\widehat{\text{ICC}}(w;Y)\right] = \frac{1}{n}z^T \Omega_n z,$$

where $z = (z_1, \ldots, z_n)$ and $z_i = (h_i - w^T X_i)^+$.

Proof Direct application of Lemma 1 for $z = (z_1, \ldots, z_n)$.   □

3.2 Regularized ASD constraint

Similarly, for the ASD we have the following proposition.

Proposition 2 *Let $X_1, \ldots X_n \overset{\text{i.i.d}}{\sim} F$, where the cumulative distribution F has finite second moment, and let $\widehat{\text{ASD}}[w;X]$ be defined as in (7). Then*

$$\text{SV}\left[\widehat{\text{ASD}}(w;X)\right] = \frac{1}{n}z^T \Omega_n z,$$

where $z = (z_1, \ldots, z_n)$, $z_i = (w^T X_i - w^T \hat{\mu})^+$ and $\hat{\mu} = (1/n)\sum_{i=1}^n X_i$ is the vector of sample averages of X.

3.3 Regularized QDEV constraint

Finally, for QDEV we have

Proposition 3 *Let $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} F$ be a random vector with finite second moments, and let F denote its distribution function. Let $\widehat{\text{QDEV}}_\alpha(w;X)$ be as above and $\eta^* \in \mathbb{R}$ such that*

$$\widehat{\text{QDEV}}_\alpha(w;X) = \frac{1}{n}\sum_{i=1}^n \epsilon_1 \left(\eta^* - w^T X_i\right)^+ + \epsilon_2 \left(w^T X_i - \eta^*\right)^+.$$

Then

$$\text{Var}\left[\widehat{\text{QDEV}}_\alpha(w;X)\right] = \frac{1}{n}\left\{ \text{Var}\left[\epsilon_1 \left(\eta^* - w^T X_i\right)^+ + \epsilon_2 \left(w^T X_i - \eta^*\right)^+\right]\right\}.$$

Note that

$$1 - \alpha = 1 - \frac{\epsilon_2}{\epsilon_1 + \epsilon_2} = \frac{\epsilon_1}{\epsilon_1 + \epsilon_2}. \tag{9}$$

For simplicity, write $Z_i = w^T X_i$. From expression (9), and using $\epsilon_1, \epsilon_2 > 0$ we have

$$\widehat{\text{QDEV}}_\alpha(w;\mathsf{X}) = (\epsilon_1 + \epsilon_2)\min_{\eta\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^{n}(1-\alpha)\big(\eta - Z_i\big)^+ + \alpha\big(Z_i - \eta\big)^+. \quad (10)$$

The next lemma is essential to prove Proposition 3.

Lemma 2 *Let* $p = \lceil n\alpha \rceil - n\alpha$. *Following the notation above, if* $p > 0$, *then* $\eta^* = Z_{(\lceil n\alpha \rceil - 1)}$ *is the unique minimizer of problem* (10). *Otherwise, if* $p = 0$, *then* $\eta^* = Z_{(\lceil n\alpha \rceil)}$ *is one of the minimizers of problem* (10).

Corollary 2 *Under the assumptions of Proposition 3 we have*

$$\text{SV}\Big[\widehat{\text{QDEV}}(w;\mathsf{X})\Big] = \frac{1}{n}z^T\Omega_n z,$$

where $z = (z_1, \ldots, z_n)$ *and* $z_i = \epsilon_1(\eta(p) - w^T X_i)^+ + \epsilon_2(w^T X_i - \eta(p))^+$.

Proof Direct application of Lemma 1 for $\mathsf{z} = (z_1, \ldots, z_n)$. □

Note that for all three cases the resulting PBR constraints are quadratic, which makes the corresponding optimization problems amenable to be tackled with off-the-shelf convex commercial solvers. In the next section we use the expressions presented in Propositions 1, 2 and 3 to test the performance of PBR in practice.

4 Numerical Results

Due to the recent boom in passive investments, in addition to mounting evidence that passive (index-based) approaches tend to outperform active investment strategies, we have designed our experiment using indices instead of individual assets. An additional advantage of choosing an investment strategy based on indices is that it offers a high degree of diversification while keeping the size of the optimization problem more manageable. In our study, following Walden (2015), we have selected thirteen indices that offer a wide exposure to the most popular asset classes, namely, stocks, bonds, real estate and commodities. The indices are described in Table 1. The period considered goes from January 2000 until December 2012; thus, it includes the subprime crisis, a time period of significant market turmoil, which we judge essential to assess the virtues of any investment strategy.

4.1 Design of experiments

We use monthly returns from a 3-year period to cast the optimization problem directly, with no parametric estimation, and then we test the performance on year 4. In order to perform extensive computations for year 4 we need some parametric assumption on returns. To this end, we assume that the vector of yearly returns r follows a multivariate normal distribution $N(\mu, \sigma)$, and estimate the corresponding parameters using past data from the 3-year window. Other possibilities could

Table 1 List of indices

Index name (Vanguard Ticker Symbol)
Stocks
Total Stock Market Index (VTSMX)
Total International Stock Index (VGTSX)
Small Cap Index (NAESX)
Emerging Markets Stock Index (VEIEX)
European Stock Index (VEURX)
Pacific Stock Index (VPACX)
Developed Markets Index Admiral Fund (VTMGX)
Bonds
Total Bond Market Index (VBMFX)
High-Yield Corporate Fund (VWEHX)
Short Term Investment Grade Fund (VFSTX)
Alternatives
Inflation Protected Securities (VIPSX)
Real Estate Investment Trust (VGSIX)
Precious Metals and Mining (VGPMX)

have been selected as long as samples can be easily obtained by the parametric model. For each year starting in 2003 and for each risk measure, we sample 100 yearly returns and evaluate the optimal solutions obtained by the SAA and PBR formulations, using the same sample in both cases. Such parametric approach allows us to have a more exhaustive and robust assessment of each method.

Thus, we initially start with [2000, 2001, 2002] and test our results with actual returns in 2003. We end with the window [2009, 2010, 2011], testing on 2012, for a total of 10 comparison years in a rolling horizon fashion. In each case (each 3-year window) we solve the data-driven optimization problem (no parametric model is needed in this step) using the four metrics (the three presented in Sect. 3, plus the CVaR, based on the results derived in Ban et al. (2016)), for both the SAA and the PBR-based approach.

For the CVaR and QDEV we use $\alpha = 0.9$, which means we want to control average tail losses (beyond the $1 - 0.9 = 0.1$ percentile) given that a loss occurred. For ICC we use $h(\omega) = 5.5\%$, which is a benchmark of yearly returns. Finally, the values of k for the experiments are the smallest numbers such that problem (5) is feasible for each risk measure, which explains why they vary from experiment to experiment.

The code was written in Python 2.7.13 and the problems were solved using Gurobi (version 7.0.2) on a MacBook Pro with a 2 GHz Intel Core i5 and 8 GB of RAM. In Table 2 we report for SAA and PBR the time (in s) to simulate and solve the 100 problems, and to compute the statistics of interest for the 10 years in study. The difference in computational times across methods is remarkable. This is to be expected, considering that the cross-validation procedure that selects U_1 and U_2 involves solving several auxiliary optimization problems. Since both methods are

Table 2 Computational times for each experiment

Risk measure	Time (s)	
	SAA	PBR
Pre-crisis period		
5.5%-ICC, $k = 5$	23.52	1036.42
0.9-QDEV, $k = 7$	31.11	1762.72
0.9-CVaR, $k = 9$	23.44	941.50
ASD, $k = 8$	24.63	1066.14
The crisis of 2008		
5.5%-ICC, $k = 8$	22.96	989.20
0.9-QDEV, $k = 7$	31.11	1762.72
0.9-CVaR, $k = 10$	23.07	942.66
ASD, $k = 9$	23.81	1074.16
Post-crisis period		
5.5%-ICC, $k = 8$	22.96	989.20
0.9-QDEV, $k = 9$	30.55	1756.21
0.9-CVaR, $k = 6$	23.19	942.96
ASD, $k = 19$	23.86	1059.13

being implemented as passive investment strategies the computational times do not prevent the implementation of PBR in practice.

4.2 Diversification

From a practical viewpoint, a fundamental aspect of the resulting portfolio is its degree of diversification. Following Woerheide and Persson (1992), we use a normalized version of the complement of Herfindal's diversification index (DI) to measure this property:

$$\mathrm{DI} = \frac{1 - \sum_{i=1}^{n} w_i^2}{1 - 1/n},$$

where w_i is the weight of index i and $n > 1$ is the number of assets available for investment. According to this measure, a concentrated portfolio (only one position among the n assets) has a DI of zero; and a portfolio of equally weighted assets (the so-called $1/n$ portfolio), would have a DI of one, which corresponds to maximum diversification.

Figure 1 shows the average DI values, for the SAA and PBR portfolios, for each year between 2003 and 2012. The results indicate that the PBR portfolios are more diversified, and often the difference is significant. Moreover, in several cases the SAA portfolios have a DI = 0 for all samples, whereas PBR portfolios always have a positive average DI (in all years and for all four risk measures). Figure 1b (0.9-QDEV) reveals an extreme case—in 7 out of 10 years SAA portfolios have a DI = 0. The reason is that in cases where regularization is not present,
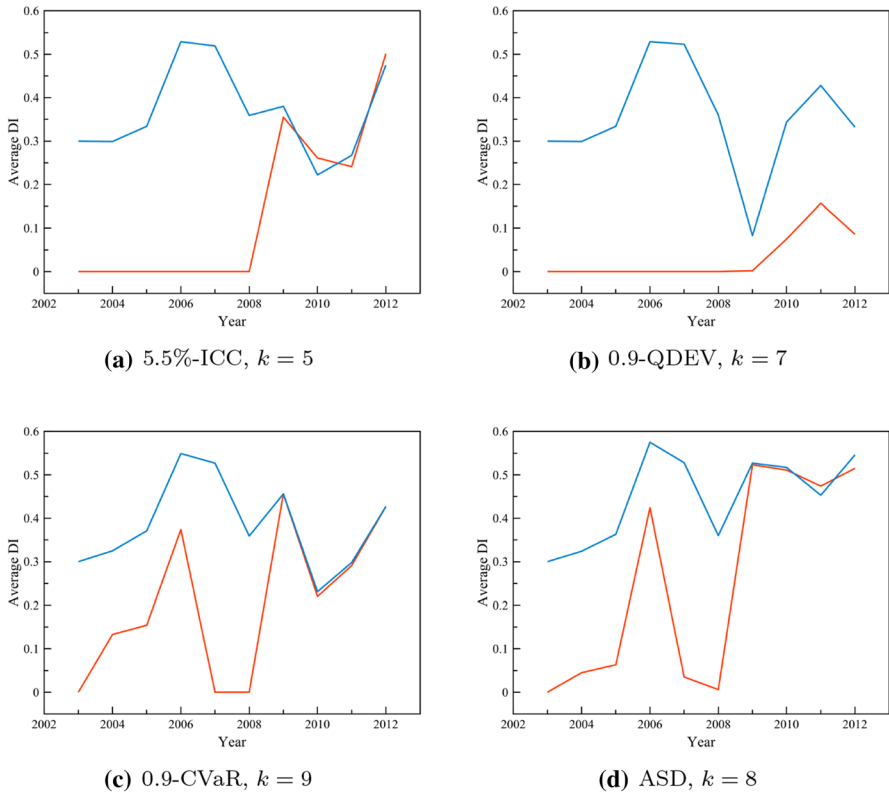
Fig. 1 Average DI values for SAA and PBR simulated portfolios, using different risk measures and values of k. The red line corresponds to SAA and the blue line to PBR

there was one asset with a high return, and being fully invested in this asset did not violate the risk constraint. PBR constraints are designed to avoid that: solutions that have high returns are often infeasible because their variability is too high.

It is also interesting to note that diversification is persistent over time, that is, the average DI for the PBR portfolios is not only higher than those of the SAA portfolios, but is also more stable over the 10 year-period under study. Figure 3 shows the area below the SAA and PBR diversification trajectories displayed in Fig. 1. Total diversification—the $1/n$ portfolio—would have an area of 10. Thus, the entries in Table 3 can be thought of as the number of years in which the strategy corresponds to complete diversification. We observe that the PBR values are roughly between 2 and 3 times those of the SAA portfolios, indicating, again, a much higher degree of diversification.

Table 3 Area below diversification trajectories

Risk measure	Area	
	SAA	PBR
5.5%-ICC, $k = 5$	1.1075	3.2965
0.9-QDEV, $k = 7$	0.277	3.216
0.9-CVaR, $k = 9$	1.8415	3.4795
ASD, $k = 8$	2.3385	4.07

4.3 Risk and returns

We now turn our attention to the performance of both methods with respect to returns and variability. The comparison will take place in three time windows: pre-crisis (2003–2007), crisis (2008) and post-crisis period (2011–2012). Results in the recovery years (2009–2010) were similar between SAA and PBR and are available from the authors upon request.

4.3.1 Pre-crisis period

Table 4 shows summary statistics for realized returns during the pre-crisis years (2003–2007). All in all, SAA portfolios outperform the PBR portfolios, with differences as high as 8% on a given year. This is not surprising since the stability provided by the PBR algorithm comes at the expense of ruling out high variability solutions, which, in turn, are the ones that yield the highest out-of-sample returns. It is also noteworthy that the differences in performance between the SAA and PBR portfolios are fairly consistent across all risk measures, which validates the robustness of both methods.

4.3.2 The crisis of 2008

During the subprime crisis, the PBR portfolios have smaller losses, not only in average terms, but also when the maximum and minimum returns are considered, as shown in Table 5. During the crisis PBR portfolios outperform SAA portfolios in every aspect and for all four risk measures. It should be noted that realized returns are significantly low with both approaches because the parameter estimation was based on data that was unable to anticipate the crisis. It is certainly beyond the scope of this study to identify or propose indicators that could predict crises. But it suffices to say that not using regularization techniques can magnify the losses in those scenarios.

4.3.3 Post-crisis period

Let us now compare the two methods in 2011 and 2012. As shown in Tables 6 and 7, the difference in performance is extraordinary, with the SAA portfolios exhibiting

Table 4 Average realized returns of optimal portfolios determined by SAA and PBR, and using four different risk measures during years pre crisis (2003–2007)

Risk measure	Year	Average realized return (%)	
		SAA	PBR
5.5%-ICC, $k = 5$	2003	50.41	42.54
		(50.41; 50.41)	(42.14; 42.88)
	2004	1.89	2.06
		(1.89; 1.89)	(2.05; 2.07)
	2005	39.00	31.51
		(39.00; 39.00)	(29.42; 33.22)
	2006	26.76	23.43
		(20.90; 27.27)	(16.77; 26.66)
	2007	31.48	22.66
		(18.68; 36.46)	(16.41; 31.97)
0.9-QDEV, $k = 7$	2003	50.41	42.55
		(50.41; 50.41)	(42.15; 42.89)
	2004	1.89	2.07
		(1.89; 1.89)	(2.06: 2.07)
	2005	39.00	31.51
		(39.00; 39.00)	(29.42; 33.23)
	2006	26.76	23.44
		(20.90; 27.27)	(16.77; 26.66)
	2007	31.48	22.64
		(18.68; 36.46)	(16.41; 31.97)
0.9-CVaR, $k = 9$	2003	50.41	42.55
		(50.41; 50.41)	(42.15; 42.89)
	2004	1.97	2.09
		(1.89; 2.25)	(2.06; 2.28)
	2005	35.538	30.38
		(22.75; 39.00)	(22.85; 33.23)
	2006	23.80	22.51
		(12.10; 27.27)	(11.44; 26.95)
	2007	31.48	22.46
		(18.68; 36.46)	(16.41; 31.97)
ASD, $k = 8$	2003	50.41	42.55
		(50.41; 50.41)	(42.15; 42.889)
	2004	1.92	2.09
		(1.89; 2.06)	(2.06; 2.20)
	2005	37.85	30.61
		(32.17; 39.00)	(26.2; 33.23)
	2006	24.02	21.33
		(20.07; 27.27)	(16.08; 26.18)
	2007	31.83	23.08
		(18.68; 36.46)	(16.41; 33.44)

The numbers in parentheses correspond to the minimum and maximum realized returns, respectively

Table 5 Average realized returns of optimal portfolios determined by SAA and PBR, and using four different risk measures in 2008. The numbers in parentheses correspond to the minimum and maximum realized returns, respectively

Risk Measure	Year	Average realized return (%)	
		SAA	PBR
5.5%-ICC, $k = 8$	2008	– 55.68	– 54.28
		(– 64.55; – 55.01)	(– 62.95; – 52.75)
0.9-QDEV, $k = 7$	2008	– 55.68	– 54.28
		(– 64.55; – 55.01)	(– 62.27; – 52.75)
0.9-CVaR, $k = 10$	2008	– 55.68	– 54.28
		(– 64.55; – 55.01)	(– 63.03; – 52.75)
ASD, $k = 9$	2008	– 55.68	– 54.3
		(– 64.55; – 55.01)	(– 63.58; – 52.75)

Table 6 Average realized returns of optimal portfolios determined by SAA and PBR, and using four different risk measures in 2011

Risk Measure	Year	Average realized return (%)	
		SAA	PBR
5.5%-ICC, $k = 8$	2011	– 2.66	– 0.25
		(– 13.57; 8.54)	(– 8.59; 7.70)
0.9-QDEV, $k = 9$	2011	5.23	5.24
		(– 1.30; 8.52)	(– 1.30; 8.52)
0.9-CVaR, $k = 6$	2011	4.623	4.64
		(.847; 7.488)	(0; 7.488)
ASD, $k = 9$	2011	– 4.45	– 1.82
		(– 10.76; 8.54)	(– 7.89; 7.70)

The numbers in parentheses correspond to the minimum and maximum realized returns, respectively

Table 7 Average realized returns of optimal portfolios determined by SAA and PBR, and using four different risk measures in 2012

Risk Measure	Year	Average realized return (%)	
		SAA	PBR
5.5%-ICC, $k = 8$	2012	13.13	11.20
		(– 4.90; 16.14)	(7.19; 16.07)
0.9-QDEV, $k = 9$	2012	14.38	13.75
		(– 17.74; 16.18)	(– 6.73; 16.16)
0.9-CVaR, $k = 6$	2012	2.97	2.97
		(1.82; 5.10)	(0.00;5.10)
ASD, $k = 19$	2012	14.408	13.774
		(– 17.741; 16.176)	(– 4.534; 16.158)

The numbers in parentheses correspond to the minimum and maximum realized returns, respectively

significantly more dispersion. Moreover, the PBR portfolios offer more protection against extreme losses, without suffering a noticeable reduction in terms of either average or maximum returns.

4.4 Discussion

For an investor dealing with an asset allocation problem it is by no means clear, unless additional information is provided, which risk measure will best fit his/her interests. The results suggest that the ASD produces more diversified portfolios, albeit with lower but more stable returns. On the other hand, the ICC results in the least diversified portfolios (lowest DI values), combined with the highest realized returns, which, not surprisingly, come with the highest variability. The CVaR portfolios are somewhere in between. Their returns are higher than those obtained by the ASD portfolios, but not as high as the ICC portfolios. Diversification for CVaR is on average higher but it is the only risk measure which generates completely concentrated portfolios (DI = 0) for some—but never for all—of the sampled returns, in some years.

Lastly, QDEV offers a much more complex behavior, and its effects depend greatly on the value of α. Our experiments suggest that when this parameter gets closer to 0 or 1, the portfolios tend to be more concentrated and riskier, but also, more rewarding. Interestingly, when α is closer to 0.5, those solutions behave similarly to those rendered by the ASD.

We close this section with a comment regarding the effects of the right-hand side constants, U_1 and U_2, in the resulting portfolio. The constraint controlled by U_2 [variance of returns, constraint (4)] is the one responsible for diversification, while U_1 [variance of risk, constraint (3)] induces minor changes in the portfolio allocations. The former is binding more often, whereas low values of U_2 are a common cause for infeasibilities. It is therefore possible to infer that controlling the variance of returns excludes unreliable solutions, while controlling the variance of the risk measure improves the out-of-sample performance of the portfolio.

An important distinction must be made between the variability constraints defined by U_1 and U_2, and the risk constraint defined by k. If infeasibilities are caused by a value of k which is too small, then, no combination of the indices will yield a portfolio with the acceptable level of risk—the only course of action is to simply increase the value of k until a solution is found. Infeasibilities induced by the U_2 constraint are completely different. First, since this parameter is set via a machine learning procedure as described in Section 5.4 of Ban et al. (2016), we have no room to maneuver. Second, and more important, the lack of feasible solutions should be taken as a warning, since it means that the observed returns being used present high variability, which should serve as an alert to reframe the optimization problem using more data.

5 Conclusions

Since the publication of Markowitz's seminal paper, the trade-off between risk and return within an optimization framework has attracted the attention of academics and investors alike. Nevertheless, the practical implementation of solutions obtained from such model has been marred by estimation errors and have often resulted in poorly diversified portfolios. Several tools have been developed to overcome this shortcoming, and in this work we studied a performance-based regularization (PBR) scheme, a novel regularization tool that incorporates machine learning to find the parameters that produce better out-of-sample performance. Building on Ban et al. (2016), we developed explicit convex expressions to test the PBR formulation in combination with three risk measures: integrated chance-constraints, absolute semi-deviation and quantile deviation.

We show in our numerical results that PBR is capable of delivering more diversified portfolios than those of SAA, and also more stable, over time. Additionally, the experiments show that PBR can effectively protect the investors from portfolios with low out-of-sample performance. In particular, during times of crises, PBR's performance was superior in maximum, average and minimum observed returns for the simulated portfolios. On more stable years, SAA outperforms PBR since the elimination of solutions with high variability—precisely the ones which perform better in those years—can damage returns when market conditions are favorable. Finally, the right-hand sides of the regularized constraints are defined via cross-validation, freeing the investor of the problematic task of having to specify those parameters.

Our findings show that pure SAA techniques are not suitable to deal with practical portfolio selection problems. It is critical to impose some regularization to the problem, and our work shows that PBR is a viable and tractable choice, especially for mid–to long–term investments. Future work should focus on comparing PBR-based methods with robust or distributionally robust optimization, in combination with machine learning techniques. Another avenue of research is to explore regularization schemes in the context of stochastic dynamic—multistage—portfolio problems. It would be interesting to compare different multistage frameworks that have been studied lately, such as Expected Conditional Risk measures (Homem-de-Mello and Pagnoncelli 2016), nested risk measures (Kozmík and Morton 2015) and Expected Conditional Stochastic Dominance (Escudero et al. 2018), and understand the effect of including PBR in each case.

A Proofs

Proof (Lemma 1) We have

$$\text{SV}[z] = \frac{1}{n-1} \sum_{i=1}^{n} \left(z_i - \frac{1}{n} \sum_{j=1}^{n} z_j \right)^2$$

$$= \frac{1}{n-1} \left\{ \sum_{i=1}^{n} z_i^2 - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j + n \cdot \frac{1}{n^2} \left(\sum_{j=1}^{n} z_j \right)^2 \right\}$$

$$= \frac{1}{n-1} \left\{ z^T I_n z - \frac{2}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} z_i z_j \right\}$$

$$= \frac{1}{n-1} \left\{ z^T I_n z - \frac{1}{n} z^T 1_n 1_n^T z \right\}$$

$$= z^T \left\{ \frac{1}{n-1} [I_n - n^{-1} 1_n 1_n^T] \right\} z$$

$$= z^T \Omega_n z.$$

$\square$

Proof (Proposition 1) From Definition 6, given that the Y_i are independent and identically distributed we have

$$\text{Var}\left[\widehat{\text{ICC}}(w;Y) \right] = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}\left[(h_i - w^T X_i)^+ \right]$$

$$= \frac{1}{n^2} \cdot n \text{Var}\left[(h - w^T X)^+ \right]$$

$$= \frac{1}{n} \text{Var}\left[(h - w^T X)^+ \right].$$

$\square$

Proof (Proposition 2) Let $Y = (Y_1, \ldots, Y_n)$ with $Y_i := -X_i$. Then the sample average of Y verifies $\hat{\mu}_Y = -\hat{\mu}$ and we have that Y_i are i.i.d. and they have finite second moment. Rewriting the expression for $\widehat{\text{ASD}}_n[w;X]$ in terms of Y we have

$$\widehat{\text{ASD}}(w;X) = \frac{1}{n} \sum_{i=1}^{n} \left(w^T X_i - w^T \hat{\mu} \right)^+$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(w^T \widehat{\mu}_Y - w^T Y_i \right)^+$$

$$= \widehat{\text{ICC}}\left(w, w^T \widehat{\mu}_Y; Y \right).$$

Thus, by Corollary 1

$$\text{SV}\left[\widehat{\text{ASD}}_n(w;X) \right] = \frac{1}{n} \tilde{z}^T \Omega_n \tilde{z},$$

where Ω_n is as above, $\tilde{z} = (\tilde{z}_1, \ldots, \tilde{z}_n)$ and $\tilde{z}_i = \left(w^T \widehat{\mu_Y} - w^T Y_i\right)^+$. Rewriting the result back in terms of X gives the desired result. $\qquad\square$

Proof (Lemma 2) The expression to be minimized in Problem (10), which we refer to as $F(\eta)$, is piecewise linear with breaking points at $Z_1, \ldots, Z_n$. For $m \in \{-\lceil n\alpha \rceil + 1, \ldots, \lceil n\alpha \rceil + n - 1\}$ we define

$$\Delta(m) := F(Z_{(\lceil n\alpha \rceil + m + 1)}) - F(Z_{(\lceil n\alpha \rceil + m)}),$$

where $Y_{(i)}$ is the ith order statistic. Note that

$$\Delta(m) = \frac{1}{n} \left\{ \sum_{i=1}^n (1 - \alpha) \left[(Z_{(\lceil n\alpha \rceil + m + 1)} - Z_i)^+ - (Z_{(\lceil n\alpha \rceil + m)} - Z_i)^+ \right] \right.$$

$$+ \alpha \left[(Z_i - Z_{(\lceil n\alpha \rceil + m + 1)})^+ - (Z_i - Z_{(\lceil n\alpha \rceil + m)})^+ \right] \right\}$$

$$= \frac{1}{n} \left\{ (1 - \alpha) \left(\lceil n\alpha \rceil + m \right) \left(Z_{(\lceil n\alpha \rceil + m + 1)} - Z_{(\lceil n\alpha \rceil + m)} \right) \right.$$

$$+ \alpha (n - \lceil n\alpha \rceil - m) \left(Z_{(\lceil n\alpha \rceil + m)} - Z_{(\lceil n\alpha \rceil + m + 1)} \right) \right\}$$

$$= \frac{1}{n} \left(Z_{(\lceil n\alpha \rceil + m + 1)} - Z_{(\lceil n\alpha \rceil + m)} \right) \left[\lceil n\alpha \rceil - n\alpha + m \right]$$

$$= \frac{1}{n} \left(Z_{(\lceil n\alpha \rceil + m + 1)} - Z_{(\lceil n\alpha \rceil + m)} \right) [p + m].$$

From the definition, $0 \le p < 1$. Thus $\Delta(m) < 0$ for $m \le -1$ and $\Delta(m) > 0$ for $m > 0$. If $p > 0$, then $\Delta(0) > 0$ and thus $\eta^* = Z_{(\lceil n\alpha \rceil - 1)}$ is unique. If $p = 0$, then $\Delta(0) = 0$, i.e.,

$$F(Z_{(\lceil n\alpha \rceil + 1)}) = F(Z_{(\lceil n\alpha \rceil)}).$$

Since $F(\cdot)$ is piecewise linear, then its minimum value is $F(Z_{\lceil n\alpha \rceil})$ and $\eta = Z_{(\rceil n\alpha \rceil)}$ is one of its minimizers, which concludes the proof. $\qquad\square$

Proof (Proposition 3) Using the notation above, Lemma 2 implies that

$$\widehat{\mathrm{QDEV}}_\alpha(w; X) = (\epsilon_1 + \epsilon_2) \frac{1}{n} \sum_{i=1}^n (1 - \alpha) \left(\eta(p) - Z_i \right)^+ + \alpha \left(Z_i - \eta(p) \right)^+$$

$$= \frac{1}{n} \sum_{i=1}^n \epsilon_1 \left(\eta(p) - Z_i \right)^+ + \epsilon_2 \left(Z_i - \eta(p) \right)^+,$$

where $\eta(p)$ can be $Z_{(\lceil n\alpha \rceil)}$ or $Z_{(\lceil n\alpha \rceil - 1)}$ depending on the value of p. The result follows from the assumption that the Z_i observations are i.i.d. $\qquad\square$

References

Arnott R, Berkin A, Ye J (2000) How well have investors been served in the 1980s and 1990s? J Portf Manag 26(4):84–91

Ban GY, Karoui NE, Lim A (2016) Machine learning and portfolio optimization. Manag Sci (2016)

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

Bertsimas D, Sim M (2004) The price of robustness. Oper Res 52(1):35–53

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

Bishop C (2006) Pattern recognition and machine learning (information science and statistics). Springer, New York

Bogle J (1995) The triumph of indexing. The Vanguard Group, Pennsylvania

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

Brodie J, Daubechies I, De Mol C, Giannone D, Loris I (2009) Sparse and stable markowitz portfolios. Proc Nat Acad Sci 106(30):12267–12272

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

Chopra V, Ziemba W (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice. J Portof Manag 19(2):6–11

Corsaro S, De Simone V (2019) Adaptive ℓ_1-regularization for short-selling control in portfolio selection. Comput Optim Appl 72(2):457–478

Cotton T, Ntaimo L (2015) Computational study of decomposition algorithms for mean-risk stochastic linear programs. Math Program Comput 7(4):471–499

Dai Z, Wen F (2018) Some improved sparse and stable portfolio optimization problems. Financ Res Lett 27:46–52

DeMiguel V, Nogales FJ (2009) Portfolio selection with robust estimation. Oper Res 57(3):560–577

Elton MGE, Blake C (1996) The persistence of risk-adjusted mutual fund performance. J Bus 69(2):133–157

Escudero LF, Monge JF, Morales DR (2018) On the time-consistent stochastic dominance risk averse measure for tactical supply chain planning under uncertainty. Comput Oper Res 100:270–286

Fastrich B, Paterlini S, Winker P (2015) Constructing optimal sparse portfolios using regularization methods. CMS 12(3):417–434

Fernandes B, Street A, Valladão D, Fernandes C (2016) An adaptive robust portfolio optimization model with loss constraints based on data-driven polyhedral uncertainty sets. Eur J Oper Res 255(3):961–970

Frankfurter G, Phillips H, Seagle J (1971) Portfolio selection: the effects of uncertain means, variances, and covariances. J Financ Quant Anal 6(5):1251–1262

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

Frost P, Savarino J (1988) For better performance: constraint portfolio weights. J Portf Manag 15(1):29–34

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

Green R, Hollifield B (1992) When will mean-variance efficient portfolios be well diversified? J Financ 47(5):1785–1809

Haneveld WK (1986) Duality in stochastic linear and dynamic programming. Springer, Berlin

Homem-de-Mello T, Pagnoncelli BK (2016) Risk aversion in multistage stochastic programming: a modeling and algorithmic perspective. Eur J Oper Res 249(1):188–199

Kawas B, Thiele A (2011) A log-robust optimization approach to portfolio management. OR Spectrum 33(1):207–233

Kolm P, Tütüncü R, Fabozzi FJ (2014) 60 years of portfolio optimization: practical challenges and current trends. Eur J Oper Res 234(2):356–371

Kozmík V, Morton DP (2015) Evaluating policies in risk-averse multi-stage stochastic programming. Math Program 152(1–2):275–300

Krokhmal P, Zabarankin M, Uryasev S (2011) Modeling and optimization of risk. Surv Oper Res Manag Sci 16(2):49–66

Lim A, Shanthikumar JG, Vahn GY (2011) Conditional value-at-risk in portfolio optimization: coherent but fragile. Oper Res Lett 39(3):163–171

Linderoth J, Shapiro A, Wright S (2006) The empirical behavior of sampling methods for stochastic programming. Ann Oper Res 142(1):215–241

Malkiel B (1995) Returns from investing in equity mutual funds 1971 to 1991. J Financ 50(2):549–572

Malkiel B (1996) Not so random. Barron, New York, p 55

Mansini R, Ogryczak W, Speranza MG (2014) Twenty years of linear programming based portfolio optimization. Eur J Oper Res 234(2):518–535

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

Michaud R (1989) The Markowitz optimization enigma: is 'optimized' optimal? Financ Anal J 45(1):31–42

Ogryczak W, Ruszczyski A (2002) Dual stochastic dominance and quantile risk measures. Int Trans Oper Res 9(5):661–680

Quaranta AG, Zaffaroni A (2008) Robust optimization of conditional value at risk and portfolio selection. J Bank Financ 32(10):2046–2056

Ruszczyski A, Shapiro A (2006) Optimization of convex risk functions. Math Oper Res 31(3):433–452

Shapiro A (2003) Monte Carlo sampling methods. Stochastic programming, handbooks in operations research and management science, vol 10. Elsevier, New York, pp 353–425

Soyster A (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. Oper Res 21(5):1154–1157

Still S, Kondor I (2010) Regularizing portfolio optimization. New J Phys 12(7):075034

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (Methodol) 58(1):267–288

Tikhonov A (1963) Solution of incorrectly formulated problems and the regularization method. Soviet Math Dokl 4(4):1035–1038

Walden M (2015) Active versus passive investment management of state pension plans: implications for personal finance. J Financ Counsel Plan 26(2):160–171

Wang Z, Glynn PW, Ye Y (2016) Likelihood robust optimization for data-driven problems. CMS 13(2):241–261

Witten IH, Frank E, Hall MA (2011) Data mining: practical machine learning tools and techniques. Morgan Kaufmann Publishers Inc., San Francisco

Woerheide W, Persson D (1992) An index of portfolio diversification. Financ Serv Rev 2(2):73–85