Robustifying Convex Risk Measures for Linear Portfolios: A Nonparametric Approach

Article *in* Operations Research · October 2014

DOI: 10.1287/opre.2014.1323

Submitted to *Operations Research*
manuscript OPRE-2012-01-014

Robustifying Convex Risk Measures for Linear Portfolios: A Non-Parametric Approach

David Wozabal
TUM School of Management, Technische Universität München, Germany
{david.wozabal@tum.de}

This paper introduces a framework for robustifying convex, law invariant risk measures. The robustified risk measures are defined as the worst case portfolio risk over neighborhoods of a reference probability measure, which represent the investors beliefs about the distribution of future asset losses. It is shown that under mild conditions, the infinite dimensional optimization problem of finding the worst-case risk can be solved analytically and closed-form expressions for the robust risk measures are obtained. Using these results, robust versions of several risk measures including the standard deviation, the Conditional Value-at-Risk, and the general class of distortion functionals are derived. The resulting robust risk measures are convex and can be easily incorporated into portfolio optimization problems and a numerical study shows that in most cases they perform significantly better out-of-sample than their non-robust variants in terms of risk, expected losses, and turnover.

Key words: Robust optimization, Kantorovich distance, norm-constrained portfolio optimization, soft robust constraints

History:

---

1. Introduction

Since Markowitz published his seminal work on portfolio optimization, the scientific community and financial industry have proposed a plethora of methods to find risk-optimal portfolio decisions in the face of uncertain future asset losses. Most of these, similar to the Markowitz model, treat uncertain losses as random variables. Although they recognize the uncertainty of the losses, these methods usually assume that the distribution of losses is known to the decision maker, so that there is no uncertainty about the *nature of the randomness.* However, in most cases, the distribution of the losses is unknown and thus typically replaced by an estimate. It was recognized already in early papers that the estimation of distributions, underlying the stochastic programs in question, introduces an additional level of model uncertainty (see Dupačová 1977). The estimation errors thus introduced at the level of the loss distributions can lead to dramatically erroneous portfolio decisions, as is well documented for the classical Markowitz portfolio selection problem (see for example Chopra and Ziemba 1993).

In accordance with recent literature, we use the term *ambiguity* to refer to this type of (epistemic) uncertainty, to distinguish it from the *normal* (aleatoric) uncertainty about the outcomes of random variables. Possible ways to deal with ambiguity in portfolio optimization can be categorized roughly into three classes: robust estimation, norm-constrained portfolio optimization, and robust optimization.

Robust estimation aims to dampen estimation errors that might have an adverse effect on the resulting stochastic optimization problem. For portfolio optimization, examples of this approach include various modifications of the Markowitz portfolio selection problem, such as the application of Bayesian shrinkage type estimators proposed by Jorion (1986) and more recent approaches by DeMiguel and Nogales (2009).

Norm-constrained portfolio optimization follows a slightly different approach: Instead of robustifying the estimation, this method changes the corresponding risk minimization problems in order to mitigate the effects of estimation error on the results of the optimization problem by *artificially* restricting optimal portfolio weights. This line of research was triggered by Jagannathan and Ma (2003), who argue that restricting portfolio weights is equivalent to using shrinkage type estimators to estimate the covariance matrix in a Markowitz model. Similar approaches can be found in DeMiguel et al. (2009a) and Gotoh and Takeda (2011).

The third approach uses robust optimization to immunize stochastic optimization problems against estimation error. In contrast to the models discussed before, an ambiguity set, i.e. a set of distributions assumed to contain the true distribution, is explicitly specified and the objective function is changed to the worst-case outcome for distributions in the ambiguity set. Hence, decisions are optimal in a minimax sense as they result in *best worst-case* outcomes. Initial research in this direction includes papers by Dupačová (see for example Dupačová 1977), followed by more recent contributions by Shapiro and Kleywegt (2002), El Ghaoui et al. (2003), Goldfarb and Iyengar (2003), Maenhout (2004) and Shapiro and Ahmed (2004). While most approaches make strong assumptions about the nature of the ambiguity, there is also some research that uses non-parametric methods (see Calafiore 2007, Pflug and Wozabal 2007, Delage and Ye 2010, Zymler et al. 2011, Wozabal 2012, Li and Kwon 2013).

Various notions of ambiguity call for a number of different solution techniques that in turn lead to different robust solutions. As an example, take robustification of the Conditional Value-at-Risk (CVaR), which recently attracted some attention in the literature. Lim et al. (2011) show that the CVaR is sensitive to misspecification of the underlying loss distribution. In Zhu and Fukushima (2009), the authors treat the problem by considering uncertainty sets that comprise either of discrete distributions with a fixed set of atoms or of combinations of a fixed set of given densities. Natarajan et al. (2009) study similar problems but use partial information on the moments of the loss distributions to define ambiguity sets. Gotoh and Takeda (2011) and Gotoh et al. (2013) study robustified Value-at-Risk and CVaR optimization where the discrete support of the random returns is ambiguous and the ambiguity set is described by a norm ball. Using the dual definition of the norm, the authors arrive at robustified version of the CVaR that resembles the robustified version of the CVaR obtained in this paper.

In this paper, we adopt a robust optimization approach with the ambiguous parameter being the joint distribution of the asset losses. We assume the existence of a distributional model $\hat{P}$ that represents a *best guess* for the true distribution of the losses, which we refer to as the *reference distribution*. As ambiguity sets we use neighborhoods of this reference distribution which are consistent with the notion of weak convergence. Hence, the method is non-parametric in the sense that it does not restrict the ambiguity sets to subsets of a parametric family of distributions. The ambiguity set is used to robustify a portfolio optimization problem involving a convex, law invariant risk measure. Although the notion of ambiguity is rather general, we attain closed-form expressions of the robustified risk measures, which can be used in place of the original risk measures to solve the robustified problem. Our approach works for various risk measures, including the standard deviation, general distortion functionals such as the CVaR, the Wang functional and the Gini functional. The results in this paper are based on theoretical findings in Pflug et al. (2012) obtained to study certain qualitative features of naive diversification heuristics in portfolio optimization.

One of the advantages of the proposed robust measures is that they derive from a very general notion of ambiguity, which requires only weak conditions regarding the real distribution of asset losses. Furthermore, the obtained analytical expressions for the robustified risk measures lead to convex, computationally tractable robustified stochastic programming problems, which are often in the same problem class as the unrobustified problems. The computational simplicity of the proposed robust risk measures also makes them applicable in a multitude of contexts as we show by demonstrating that soft robustification of risk constraints (see Ben-Tal et al. 2010) leads to computationally tractable problems that can be solved as a single convex programming problem. These favorable computational properties arise because the obtained robust risk measures are essentially regularized versions of the original measures and therefore have a close connection to the norm-constrained portfolios in the extant literature. In fact, we show that using the robustified standard deviation is equivalent to some of the models proposed in DeMiguel et al. (2009a). Similarly, using our results, we can interpret the norm constrained CVaR optimization problem in Gotoh and Takeda (2011) and in Gotoh et al. (2013) as robustified CVaR optimization. This paper thus yields an alternative interpretation of norm constraints in portfolio optimization.

A remarkably stable pattern reappearing in many different settings is that robustification of optimization problems is, in some sense, equivalent to regularization of the original problem, as is also the case here. Regularization is a widely used tool for stabilization and robustification of estimators. Apart from classical statistical regularizations, such as for example in ridge regression and shrinkage estimators, authors in the field of machine learning as well as the robust optimization recently discovered links between robustness and regularization. In Bousquet and Elisseeff (2002), the authors show that adding a regularizing term to a learning algorithm yields algorithms that are *uniformly stable*, i.e., robust with respect to changes in the test data set. Similar results were obtained by Caramanis et al. (2009) who show that regularized support vector machines are equivalent to robustified support vector machines. El Ghaoui and Lebret (1997) investigate a linear regression problem with ambiguous data, where the ambiguity set is defined by the Froebnius norm and use conic programming techniques to show that the problem is equivalent to Tikhonov regularization. Bertsimas et al. (2004) show that robust linear programming problems where the ambiguity sets are balls in a normed space can be reformulated to convex optimization problems involving the dual norm, which can also be interpreted as regularized versions of the original problems. Similar results are obtained in Gotoh and Takeda (2011), Gotoh et al. (2013).

The remainder of this paper is structured as follows: Section 2 outlines the non-parametric notion of ambiguity which leads to the specification of ambiguity sets and robustified risk measures. Section 3 is dedicated to robustifying convex measures of risk and deriving closed-form expressions for the robustified risk measures of many commonly used convex risk measures. In this section, we also establish a connection between robust risk measures and norm-constrained portfolio optimization and demonstrate how robustified risk measures can be used to define soft robust constraint in a computationally efficient way. In Section 4, we provide a comparison of the out-of-sample performance of several robustified risk measures with their respective non-robustified counterparts. We also discuss how to choose the size of the ambiguity set for robustified risk measures. Section 5 concludes and suggests some avenues for further research.

2. Setting

Let $(\Omega, \mathcal{F}, \mu)$ be an arbitrary uncountable probability space that admits a uniform random variable, and let $X^P : (\Omega, \mathcal{F}, \mu) \to \mathbb{R}^N$ be the random losses of N assets comprising the asset universe, i.e., the set of assets from which the decision maker may choose. The notation X^P indicates that the image measure of X^P is the measure P on $\mathbb{R}^N$, or $\mu(X^P \in A) = P(A)$ for all Borel sets $A \subseteq \mathbb{R}^N$. Our assumptions about the probability space ensure that for every Borel measure P on $\mathbb{R}^N$, there exists a random variable X^P (see Pflug et al. 2012). Let $L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}^n)$ be the Lebesgue space with exponent p containing random variables $X : (\Omega, \mathcal{F}, \mu) \to \mathbb{R}^n$ and $L^p(\Omega, \mathcal{F}, \mu)$ the space

$L^p(\Omega,\mathcal{F},\mu;\mathbb{R})$. Throughout our discussion, we choose q to be the conjugate of p, i.e., such that $1/p+1/q=1$. We denote the norm in $L^p(\Omega,\mathcal{F},\mu;\mathbb{R}^n)$ by $||\cdot||_{L^p}$ to distinguish it from the p-norm in $\mathbb{R}^n$, which we denote by $||\cdot||_p$.

We are interested in robustifying convex measures of risk, defined as follows.

Definition 1 (Convex Risk Measure). Let $1 \le p < \infty$ and X, $Y \in L^p(\Omega,\mathcal{F},\mu)$. A functional $\mathcal{R}: L^p(\Omega,\mathcal{F},\mu) \to \mathbb{R}$, which is

1. convex, $\mathcal{R}(\lambda X + (1-\lambda)Y) \le \lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(Y)$ for all $\lambda \in [0,1]$;
2. monotone, $\mathcal{R}(X) \ge \mathcal{R}(Y)$ if $X \ge Y$ a.s.; and
3. translation equivariant, $\mathcal{R}(X+c) = \mathcal{R}(X) + c$ for all $c \in \mathbb{R}$,

is called a convex risk measure.

Convex risk measures were first introduced in Föllmer and Schied (2002) and are generalizations of coherent risk measures (see Artzner et al. 1999). Because of their convexity there exists a rich duality theory for these functionals. In particular, one can show that every convex risk measure $\mathcal{R}$ has a dual representation of the form

$$\mathcal{R}(X^P) = \sup\left\{\mathbb{E}(X^P Z) - \rho(Z) : Z \in L^q(\Omega,\mathcal{F},\mu)\right\} \tag{1}$$

where ρ is a penalty function defined on $L^q(\Omega,\mathcal{F},\mu)$ and $\mathbb{E}$ is the expectation operator (for details see Föllmer and Schied 2002, Pflug and Römisch 2007). It is worth noting that, using the above representation, a convex risk measure can be interpreted as a robustified expectation operator, where the robustification takes place via a change of measure induced by multiplying with Z inside the expectation. In this paper, we want to go one step further and robustify convex risk measures with respect to the measure P.

We denote a generic risk measure by $\mathcal{R}$ and assume that $\mathcal{R}$ is law invariant (see Kusuoka 2007), and therefore is a statistical functional that only depends on the distribution of the random variables. More specifically, we assume that $\mathcal{R}(Y) = \mathcal{R}(Y')$ for all random variables Y and Y' with the same image measure on $\mathbb{R}$. This assumption is innocuous and correspondingly, it is fulfilled by all meaningful risk measures.

We start by analyzing the following generic portfolio optimization problem:

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} & \mathcal{R}(\langle X^P, w\rangle) \\ \text{s.t.} & w\in\mathcal{W} \end{array} \tag{2}$$

where $\langle\cdot,\cdot\rangle:\mathbb{R}^N\times\mathbb{R}^N\to\mathbb{R}$ is the inner product, and $\mathcal{W}$ is the feasible set of the problem. The set $\mathcal{W}$ may represent arbitrary constraints on the portfolio weights such as budget constraints, upper and lower bounds on asset holdings of single assets, cardinality constraints, or minimum holding constraints for certain assets. The only restriction we impose on $\mathcal{W}$ is that it must not depend on the probability measure P, which rules out feasible sets defined using probability functionals as well as optimization problems with probabilistic constraints.

If the distribution P of the asset losses is known, then (2) is a stochastic optimization problem that can be solved by techniques that depend on $\mathcal{R}$, $\mathcal{W}$, and P. However, if P is ambiguous, then the solution of problem (2), with P replaced by an estimate $\hat{P}$, is subject to model uncertainty, and the resulting decisions are in general not optimal for the true measure P. Although statistical methods, analysis of fundamentals, and expert opinions may suggest beliefs about the measure P, the true distribution remains ambiguous in most cases.

Therefore it is reasonable to assume that the decision maker takes the available information into account but also accounts for model uncertainty when making decisions. We model this uncertainty by specifying a set of possible loss distributions, given the prior information represented by a distribution $\hat{P}$. This set of distributions is referred to as the ambiguity set, and $\hat{P}$ is called the reference probability measure. We define the ambiguity set as the set of measures whose distance to the reference measure does not exceed a certain threshold. To this end, we use $\mathcal{P}^p(\mathbb{R}^N)$ to denote the space of all Borel probability measures on $\mathbb{R}^N$ with finite p-th moment, and

$$d(\cdot,\cdot):\mathcal{P}^p(\mathbb{R}^N)\times\mathcal{P}^p(\mathbb{R}^N)\to[0,\infty) \tag{3}$$

to represent a metric on this space. The ambiguity set for a risk measure $\mathcal{R}:L^p(\Omega,\mathcal{F},\mu)\to\mathbb{R}$ is defined as $\mathcal{B}^p_\kappa(\hat{P})=\left\{Q\in\mathcal{P}^p(\mathbb{R}^N):d(\hat{P},Q)\le\kappa\right\}$, i.e., the *ball* of radius κ around the reference measure $\hat{P}$ in the space of measures $\mathcal{P}^p(\mathbb{R}^N)$.

More specifically, we use the Kantorovich metric to construct ambiguity sets. For $1 \le p < \infty$, the Kantorovich metric $d_p(\cdot,\cdot)$ is defined as

$$d_p(P,Q) = \inf \left\{ \left(\int_{\mathbb{R}^N \times \mathbb{R}^N} ||x-y||_p^p d\pi(x,y) \right)^{\frac{1}{p}} : \mathrm{proj}_1(\pi) = P,\ \mathrm{proj}_2(\pi) = Q \right\} \tag{4}$$

where the infimum runs over all joint distributions π on $\mathbb{R}^N \times \mathbb{R}^N$, such that the marginal distributions $\mathrm{proj}_1(\pi)$ and $\mathrm{proj}_2(\pi)$ of the first and last N components are P and Q respectively. Each feasible measure π can be regarded as a transportation plan, transporting the probability mass from the measure P to the measure Q (or the other way around) in the following way: For arbitrary sets A, $B \subseteq \mathbb{R}^N$, $\pi(A,B)$ indicates how much of the probability mass $P(A)$ is transported to the set B. In this way, the problem (4) can be interpreted as a problem of optimal transport in $\mathbb{R}^N$ for which the cost of transport is measured by $||\cdot||_p^p$. It can be shown that the infimum in (4) is always attained. For an excellent introduction to the subject of optimal transport see Villani (2003).

The Kantorovich metric d_p metricizes weak convergence on sets of probability measures on $\mathbb{R}^N$, for which $x \mapsto \|x\|_p^p$ is uniformly integrable (see Villani 2003). In particular, the empirical measure $\hat{P}_n$, based on n observations, approximates P in the sense that $d_p(P,\hat{P}_n) \to 0$ as $n \to \infty$, if the p-th moment of P exists. This property justifies the use of d_p to construct ambiguity sets as a stronger metric would not necessarily reduce the degree of ambiguity by collecting more data.

Using the preceding definition of the ambiguity set, we arrive at the robust counterpart of (2):

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} \sup_{Q\in\mathcal{B}_\kappa^p(P)} & \mathcal{R}(\langle X^Q, w\rangle) \\ \text{s.t.} & w \in \mathcal{W}. \end{array} \tag{5}$$

We then define the solution of the inner problem as the robustified version $\mathcal{R}^\kappa$ of $\mathcal{R}$, such that for any given risk measure $\mathcal{R}$ and $\kappa > 0$

$$(P,w) \mapsto \mathcal{R}^\kappa(P,w) := \sup_{Q\in\mathcal{B}_\kappa^p(P)} \mathcal{R}(\langle X^Q, w\rangle). \tag{6}$$

Note that the robustified risk measure takes two inputs: a measure P and portfolio weights w. For a given reference measure $\hat{P}$, the mapping $w \mapsto \mathcal{R}^\kappa(\hat{P},w)$ is convex in w, so problem (5) has a convex objective.

3. Robust Risk Measures

In this section, we derive explicit expressions for the robustified versions of convex, law-invariant risk measures defined above. We consider risk measures $\mathcal{R}$ with a subdifferential representation of the form (1)

$$\mathcal{R}(X) = \sup\{\mathbb{E}(XZ) - \rho(Z) : Z \in L^q(\Omega, \mathcal{F}, \mu)\} \tag{7}$$

for some convex function $\rho : L^q(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$. If $\mathcal{R}$ is lower semi-continuous, then it admits a representation of the form (7), with $\rho = \mathcal{R}^*$ where $\mathcal{R}^*$ is the convex conjugate of $\mathcal{R}$. If $\rho = \mathcal{R}^*$ and X is in the interior of the domain $\{X \in L^p(\Omega, \mathcal{F}, \mu) : \mathcal{R}(X) < \infty\}$, then $\operatorname{argmax}_Z \{\mathbb{E}(XZ) - \rho(Z)\} = \partial\mathcal{R}(X)$ where $\partial\mathcal{R}(X)$ is the set of subgradients of $\mathcal{R}$ at X. Consequently, we denote the set of maximizers of (7) at X by $\partial\mathcal{R}(X)$.

In the following, we give some examples of convex risk measures. A more detailed exposition and derivations of the subdifferential representation can be found in Ruszczyński and Shapiro (2006) as well as in Pflug and Römisch (2007). We start with the simplest risk measure: the expectation operator.

Example 1 (Expectation). The expectation $\mathbb{E}(X) : L^1(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ is used by risk-neutral agents but since it fulfills all the conditions of Definition 1, it is a convex measure of risk. The subdifferential representation is trivial with $\partial\mathbb{E}(X) = \{1\}$ for all $X \in L^1(\Omega, \mathcal{F}, \mu)$.

The next risk measure relates closely to the classical Markowitz functional, with the only difference being that the variance is replaced by the standard deviation.

Example 2 (Expectation-corrected standard deviation). The expectation-corrected standard deviation $S_\gamma : L^2(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ is defined as $S_\gamma(X) = \gamma \operatorname{Std}(X) + \mathbb{E}(X)$. The subdifferential representation of S_γ is given by

$$S_\gamma(X) = \sup\left\{\mathbb{E}(XZ) : \mathbb{E}(Z) = 1, \ ||Z||_{L^2} = \sqrt{1+\gamma^2}\right\}. \tag{8}$$

We also address the CVaR, the prototypical example of a coherent risk measure in the sense of Artzner et al. (1999).

Example 3 (Conditional Value-at-Risk). The Conditional Value-at-Risk (see Rockafellar and Uryasev (2000)) is defined as

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(t)dt, \tag{9}$$

where F_X is the cumulative distribution function of the random variable X, and F_X^{-1} denotes its inverse distribution function. Because CVaR is defined as a risk measure, we are concerned with the upper tail of the loss distribution, such that α is typically chosen close to 1. The dual representation of CVaR is given by

$$\mathrm{CVaR}_\alpha(X) = \sup\left\{\mathbb{E}(XZ) : \mathbb{E}(Z) = 1, 0 \le Z \le \frac{1}{1-\alpha}\right\} \tag{10}$$

for $0 \le \alpha < 1$.

Next we discuss a class of examples, called distortion functionals that are predominantly used in the insurance and pricing literature.

Example 4 (Distortion functionals). The definition of a general distortion functional is based on the following representation of the mean of the random losses

$$\mathbb{E}(X) = \int_0^1 F_X^{-1}(p)dp = \int_0^1 F_X^{-1}(p)d\,\mathrm{id}(p) \tag{11}$$

with $\mathrm{id}(p) = p$ the identity. In a distortion functional the mean is distorted in the sense that instead of id some monotonically increasing, convex function $H : [0,1] \to \mathbb{R}$ is used in the above integral. A general distortion functional, dependent on a function H, is therefore given as

$$\mathcal{R}_H(X) = \int_0^1 F_X^{-1}(p)dH(p). \tag{12}$$

Note that H introduces an increased weight on the higher losses and thereby reflects the risk aversion of the decision maker.

It can be shown that if $H(p) = \int_0^p h(t)dt$, then

$$\mathcal{R}_H(X) = \sup\{\mathbb{E}(XZ) : Z = h(U),\ U \text{ uniform on } [0,1]\} \tag{13}$$

is the subdifferential representation of $\mathcal{R}_H$.

Note that the CVaR is a distortion functional with $H(p) = \max\left(\frac{p-(1-\alpha)}{\alpha}, 0\right)$. The next two examples directly make use of the idea of distorting the expectation by assigning higher weights to higher losses and have their origin in the insurance pricing literature.

Example 5 (Wang transform). Let Φ be the cumulative distribution of the standard normal distribution. The Wang transform $W_\lambda : L^2(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ defined as

$$W_\lambda(X) = \int_0^\infty \Phi\left(\Phi^{-1}(1 - F_X(t)) + \lambda\right) dt \tag{14}$$

for $\lambda > 0$, was originally introduced by Wang (2000) for positive random variables X. It can be shown that

$$W_\lambda(X) = \int_0^1 F_X^{-1}(p) dH_\lambda(p) \tag{15}$$

with $H_\lambda(p) = -\Phi\left[\Phi^{-1}(1-p) + \lambda\right]$. Note that (15) is also meaningful for general random variables, i.e., the restriction to positive random variables can be relaxed.

Example 6 (Proportional hazards transform or power distortion). The proportional hazards transform or power distortion $P_r : L^1(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ for $0 < r \leq 1$ is defined as

$$P_r(X) = \int_0^\infty (1 - F_X(t))^r dt, \tag{16}$$

and was introduced by Wang (1995) for positive random variables. Similar to the case of the Wang transform, it can be shown that

$$P_r(X) = \int_0^1 F_X^{-1}(p) dH_r(p), \tag{17}$$

with $H_r(p) = -(1-p)^r$.

The next two risk measures also turn out to be distortion risk measures although their original definition is not based on the idea of a distorted expectation.

Example 7 (Gini measure). The Gini measure was first studied by Yitzhaki (1982) and proposed as an alternative to classical mean-variance type optimization because of its close links to second order stochastic dominance. The expectation-corrected Gini measure $\mathrm{Gini}_r : L^1(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ is defined as

$$\mathrm{Gini}_r(X) = \mathbb{E}(X) + r\mathbb{E}(|X - X'|) \tag{18}$$

where X' is an independent copy of X. It can be shown that

$$\mathrm{Gini}_r(X) = \mathcal{R}_H(X) = \int_0^1 F_X^{-1}(p)dH(p) \tag{19}$$

with $H(p) = (1-r)p + rp^2$.

Example 8 (Deviation from the median). The deviation from the median $\mathrm{DM}_a : L^1(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$, introduced by Denneberg (1990), can be regarded as a L^1-variant of the standard deviation and is defined as

$$\begin{aligned}\mathrm{DM}_a(X) &= \mathbb{E}(X) + a\mathbb{E}\left(|X - F_X^{-1}(1/2)|\right) &(20)\\ &= \int_0^1 F_X^{-1}(p)dp + a\int_0^1 |F_X^{-1}(p) - F_X^{-1}(1/2)|dp &(21)\\ &= \int_0^{1/2} F_X^{-1}(p)(1-a)dp + \int_{1/2}^1 F_X^{-1}(p)(1+a)dp = \int_0^1 F_X^{-1}(p)dH(p) &(22)\end{aligned}$$

with

$$H(p) = \begin{cases} p(1-a), & p < 1/2 \\ \frac{1}{2}(1-a) + \frac{p-1}{2}(1+a), & p \geq 1/2. \end{cases} \tag{23}$$

We proceed by investigating the robust portfolio selection problem. For this purpose, let the portfolio weights w and the measure $\hat{P}$ be given. The idea behind calculating robustified risk measures is to define a measure Q such that

$$\langle X^Q, w\rangle = \langle X^{\hat{P}}, w\rangle + c|Z|^{q/p}\,\mathrm{sign}(Z), \tag{24}$$

for $Z \in \partial\mathcal{R}(\langle X^{\hat{P}}, w\rangle)$ and $c \in [0, \infty)$. In (24), the portfolio losses under $\hat{P}$, $\langle X^{\hat{P}}, w\rangle$, are shifted in the *worst direction* with respect to $\mathcal{R}$, such that the parameter c determines the distance of Q to P. If $Z \in \partial\mathcal{R}(\langle X^Q, w\rangle)$, i.e. continues to be the direction of steepest ascent of $\mathcal{R}$ at the point $\langle X^Q, w\rangle$, then Q is the worst case measure, in the sense that $\mathcal{R}^\kappa(\hat{P}, w) = \mathcal{R}(\langle X^Q, w\rangle)$ with $\kappa = d(\hat{P}, Q)$. The next proposition formalizes this intuition. The key assumption is that the norm of the subgradients of $\mathcal{R}$ stays constant, which ensures that $Z \in \partial\mathcal{R}(\langle X^Q, w\rangle)$, i.e., that the risk measure can be maximized in the loss distribution by *going up along* the subgradient Z starting from $\langle X^{\hat{P}}, w\rangle$.

PROPOSITION 1. *Let $\mathcal{R}: L^p(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$ be a convex, law-invariant risk measure and $1 \le p < \infty$ and q be defined by $\frac{1}{p} + \frac{1}{q} = 1$. Let further $\hat{P}$ be the reference probability measure on $\mathbb{R}^N$. If $\kappa > 0$ and either*

1. *$p > 1$ and*

$$||Z||_{L^q} = C \text{ for all } Z \in \bigcup_{X \in L^p} \partial\mathcal{R}(X) \text{ with } \rho(Z) < \infty, \text{ or} \tag{25}$$

2. *$p = 1$ and*

$$||Z||_{L^\infty} = C \text{ and } |Z| = C \text{ or } |Z| = 0, \text{ a.s.,} \tag{26}$$

then the solution to the inner problem (6) *is*

$$\mathcal{R}^\kappa(\hat{P}, w) = \mathcal{R}(\langle X^{\hat{P}}, w\rangle) + \kappa C||w||_q. \tag{27}$$

Proof. This follows directly from Lemma 1 and Propositions 1 and 2 in Pflug et al. (2012). □

Note that the risk measures treated in this paper share the property of having subgradients with constant norm with the L^p-norm itself: The dual representation of the p-norm is given by $||x||_p = \sup\{\langle x, y\rangle : ||y||_q = 1\}$. The conditions on the subgradients are not fulfilled for all functionals on $L^p(\Omega, \mathcal{F}, \mu)$, as is demonstrated for the variance in Section 3.1. However, all examples discussed above fulfill condition (25) or (26) for some $p' \ge p$ and therefore we can derive robust versions of the discussed risk measures based on Proposition 1.

PROPOSITION 2. 1. *The robustified expectation operator $\mathbb{E}^\kappa: L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by $\mathbb{E}^\kappa(P, w) = \mathbb{E}(\langle X^P, w\rangle) + \kappa||w||_\infty$.*

2. *The robustified expectation-corrected standard deviation $S^\kappa_\gamma: L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by $S^\kappa_\gamma(P, w) = S_\gamma(\langle X^P, w\rangle) + \kappa\sqrt{1+\gamma^2}||w||_2$.*

3. *The robustified Conditional Value-at-Risk $\mathrm{CVaR}^\kappa_\alpha: L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by*

$$\mathrm{CVaR}^\kappa_\alpha(P, w) = \mathrm{CVaR}_\alpha(\langle X^P, w\rangle) + \frac{\kappa}{1-\alpha}||w||_\infty. \tag{28}$$

4. *For $1 < p < \infty$ and a general distortion measure $\mathcal{R}_H : L^p(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$, the robustified version $\mathcal{R}^\kappa_H : L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by*

$$\mathcal{R}^\kappa_H(P, w) = \mathcal{R}_H(\langle X^P, w\rangle) + \kappa ||h(U)||_{L^q} ||w||_q, \tag{29}$$

with $H(p) = \int_0^p h(t)dt$ and U representing a uniform random variable on $[0,1]$.

5. *The robustified Wang transform $W^\kappa_\lambda : L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by $W^\kappa_\lambda(P, w) = W_\lambda(\langle X^P, w\rangle) + \kappa e^{\lambda^2/2} ||w||_2$.*

6. *For $1/2 < r \leq 1$, the robustified power distortion $P^\kappa_r : L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by $P^\kappa_r(P, w) = P_r(\langle X^P, w\rangle) + \frac{\kappa r}{\sqrt{2r-1}} ||w||_2$.*

7. *The robustified Gini measure $\mathrm{Gini}^\kappa_r : L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$ is given by $\mathrm{Gini}^\kappa_r(P, w) = \mathrm{Gini}_r(\langle X^P, w\rangle) + \kappa \sqrt{\frac{3+r^2}{3}} ||w||_2$.*

8. *The robustified deviation from the median, $\mathrm{DM}^\kappa_a : L^2(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R}^N \to \mathbb{R}$, is given by $\mathrm{DM}^\kappa_a(P, w) = \mathrm{DM}_a(\langle X^P, w\rangle) + \kappa \sqrt{1 + a^2} ||w||_2$.*

Proof. 1 and 2 follow directly from Proposition 1 and the corresponding subdifferential representations. To show 3, we note that if we choose a set $A \subseteq \Omega$ such that $\mu(A) = 1 - \alpha$ and $X(\omega) \geq F_X^{-1}(\alpha)$ for all $\omega \in A$, it is easy to see that

$$Z(\omega) = \begin{cases} \frac{1}{1-\alpha}, & \omega \in A \\ 0, & \text{otherwise} \end{cases} \in \partial\, \mathrm{CVaR}_\alpha(X). \tag{30}$$

Hence, condition (26) is fulfilled, and $||Z||_\infty = \frac{1}{1-\alpha}$. Since all subgradients have to be pointwise smaller than $(1-\alpha)^{-1}$, any Z' that doesn't put the whole mass on the worst $(1-\alpha)$ percent of the outcomes cannot represent a direction of steepest descent, i.e., all the subgradients are a.s. equivalent to Z, which proves 3.

From the subdifferential representation of distortion measures, it follows that subgradients are of the form $h(U)$, with U uniform on $[0,1]$. In particular, all subgradients have the same distribution and therefore the same q-norm. Hence, 4 follows.

To calculate the robust version of the Wang functional, note that

$$h_\lambda(p) = \frac{dH_\lambda(p)}{dp} = \exp\left(\frac{-2\lambda\Phi^{-1}(1-p) - \lambda^2}{2}\right). \tag{31}$$

We then compute

$$||h_\lambda||_2^2 = \int_0^1 \exp\left(-2\lambda\Phi^{-1}(1-p) - \lambda^2\right) dp = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-2\lambda x - \lambda^2 - \frac{x^2}{2}\right) dx \tag{32}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x+2\lambda)^2 - 2\lambda^2}{2}\right) dx = e^{\lambda^2}. \tag{33}$$

Therefore, 5 follows. The points 6, 7 and 8 can be proven analogous to 5. □

Note that the dual norm together with the constant C represents the Lipschitz constant of the respective risk measure with respect to the Kantorovich distance. Hence, the constant C can be interpreted as the sensitivity of the risk measure with respect to the change in the image measure of the asset losses. Clearly, some risk measures are more sensitive to ambiguity than others. For example, the expectation is relatively robust (with a coefficient of 1) as compared to the CVaR (with a coefficient of $(1-\alpha)^{-1}$). This is the case, since the CVaR only depends on a small part of the distribution and hence, by changing this small part, the CVaR can be dramatically influenced while the overall distribution does not change much. In contrast, for a unit change in expectation a larger change in the overall distribution is necessary.

Note that the first four risk measures in the above proposition, are defined on their respective maximal domains. Clearly, a smaller p can not be chosen, since the corresponding Kantorovich ball would contain measures for which the risk is not defined. If instead a higher exponent p is chosen for the domain $L^p(\Omega, \mathcal{F}, \mu)$, then the corresponding ambiguity set gets smaller and the robust risk measure gets *less robust.*

For the power distortion, the Wang transform, the Gini measure and the expectation-corrected deviation from the median the domains of the robustified measures are smaller than the domain of the non-robustified risk measures – in particular, these measures are no longer defined on $L^1(\Omega, \mathcal{F}, \mu)$. The reason lies in the more restrictive, pointwise conditions on the subgradients for the case $p = 1$ in Proposition 2, which is not fulfilled by the subgradients of the respective measures. We therefore define the robust measures on $L^2(\Omega, \mathcal{F}, \mu)$, though any other p, with $1 < p < \infty$, is a possible choice as well.

3.1. The case of variance and standard deviation

Although not formally a risk measure in the sense of Definition 1, the variance and the standard deviation are often used as measures of risk. In this section, we show that the standard deviation can be robustified using Proposition 1, whereas the variance cannot be treated within the outlined framework. We also investigate the relation of norm-constrained portfolio optimization to the robustifications we propose.

We start with a subdifferential representation of the standard deviation. The standard deviation, $\mathrm{Std}(X): L^2(\Omega, \mathcal{F}, \mu) \to \mathbb{R}$, is defined as

$$\mathrm{Std}(X) = ||X - \mathbb{E}(X)||_2 = \sup\left\{\mathbb{E}\left[(X - \mathbb{E}(X))Z\right] : ||Z||_{L^2} = 1\right\}. \tag{34}$$

Clearly, the maximizer Z^* in (34) is equal to

$$Z^* = \frac{X - \mathbb{E}(X)}{||X - \mathbb{E}(X)||_2}. \tag{35}$$

Because $\mathbb{E}(Z^*) = 0$, we can rewrite (34) as

$$\mathrm{Std}(X) = \sup\left\{\mathbb{E}\left[X(Z - \mathbb{E}(Z))\right] : ||Z||_{L^2} = 1,\, \mathbb{E}(Z) = 0\right\} \tag{36}$$

$$= \sup\left\{\mathbb{E}\left[XZ\right] : ||Z||_{L^2} = 1,\, \mathbb{E}(Z) = 0\right\}. \tag{37}$$

This representation is of the form (7), and because translation equi-variance is not used in the proof of Proposition 1, we can write the robustified standard deviation as

$$\mathrm{Std}^{\kappa}(P, w) = \mathrm{Std}(\langle X^P, w\rangle) + \kappa||w||_2. \tag{38}$$

The situation differs for the variance: The applicability of Proposition 1 requires that all elements in $\partial\mathcal{R}$ have the same q-norm. While this requirement is met for most common risk measures, it is not true for the variance, because

$$\mathrm{Var}(X) = ||X - \mathbb{E}(X)||_2^2 = \sup\left\{\mathbb{E}(XZ) - \frac{1}{4}\mathrm{Var}(Z) : \mathbb{E}(Z) = 0\right\} \tag{39}$$

and $Z^* = 2(X - \mathbb{E}(X))$, with $||2(X - \mathbb{E}(X))||_2 = 2\,\mathrm{Std}(X)$, i.e., the subgradients do not have constant norms. The variance therefore does not fit into the framework proposed in this paper.

However, minimizing the variance is equivalent to minimizing the standard deviation, which can be robustified as demonstrated above.

Inspecting the robustified risk measures, we note that the robustification is achieved by *penalizing* with the corresponding dual norm of the portfolio, multiplied by a constant. Therefore, we can relate the robustified risk measures to a problem of norm-constrained portfolio optimization (see DeMiguel et al. 2009a, Gotoh and Takeda 2011). In particular, minimizing (38) is equivalent to solving a 2-norm-constrained Markowitz problem: If we denote by Σ the covariance matrix of the N assets under measure $\hat{P}$, we can define the 2-norm-constrained Markowitz problem proposed by DeMiguel et al. (2009a) as

$$\begin{array}{ll} \min_w & w^\top \Sigma w \\ \text{s.t.} & \langle w, \mathbb{1} \rangle = 1,\ ||w||_2 \leq c, \end{array} \tag{40}$$

where $\mathbb{1} = (1, \ldots, 1)^\top \in \mathbb{R}^N$. Comparing the KKT condition of (40) to the KKT conditions of

$$\begin{array}{ll} \min_w & w^\top \Sigma w + \kappa ||w||_2 \\ \text{s.t.} & \langle w, \mathbb{1} \rangle = 1, \end{array} \tag{41}$$

it is easy to show that for every $c \geq 1/\sqrt{N}$, there exists a $\kappa \geq 0$, such that (40) is equivalent to (41). Conversely, for every $\kappa \geq 0$, there exists a c such that (41) is equivalent to (40). This principle can be extended to other measures of risk, which demonstrates the equivalence of the robust optimization approach with models that penalize high norms of the portfolio weights.

3.2. Soft robust constraints

The motivating problem (5) centered on robustifying the objective function of a stochastic optimization problem. In this section, we show that the robustified risk measures found in Proposition 2 can also be used in the constraints of stochastic optimization problems. For a fixed reference measure $\hat{P}$ and a convex risk measure $\mathcal{R}$, we consider a robustified problem of the form

$$\begin{array}{ll} \inf_{w \in \mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w \rangle) \\ \text{s.t.} & \mathcal{R}(\langle X^Q, w \rangle) \leq \beta,\ \forall Q \in \mathcal{B}^p_\kappa(\hat{P}) \\ & w \in \mathcal{W}. \end{array} \tag{42}$$

Using the robust risk measure $\mathcal{R}^k$, we can rewrite (42) as the following convex problem:

$$\begin{array}{ll} \inf_{w \in \mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w \rangle) \\ \text{s.t.} & \mathcal{R}^\kappa(\hat{P}, w) \leq \beta \\ & w \in \mathcal{W}. \end{array} \tag{43}$$

For $\kappa = 0$, problem (43) reduces to the nominal instance, i.e., a classical mean risk problem. However, for a given $\kappa > 0$, the constraint has to be fulfilled for all distributions $Q \in \mathcal{B}_\kappa^p(\hat{P})$, regardless of their distance from $\hat{P}$. This approach leaves little flexibility to trade-off the robustness of the constraint against performance: Although decreasing κ decreases robustness and typically increases performance, this necessarily implies that measures whose distance from $\hat{P}$ is greater than κ will not be taken into account.

A possible remedy for this dilemma has been proposed by Ben-Tal et al. (2010), who define what they call a soft robust approach by considering the problem

$$\begin{array}{ll} \inf_{w \in \mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w \rangle) \\ \text{s.t.} & \mathcal{R}(\langle X^Q, w \rangle) \le f(\kappa),\ \forall Q \in \mathcal{B}_\kappa^p(\hat{P}),\ \forall \kappa \in [0, \delta] \\ & w \in \mathcal{W}, \end{array} \tag{44}$$

where $f: \mathbb{R} \to \mathbb{R}$ is a convex function. The authors discuss several choices for f and the ambiguity set. For their numerical results they choose $f(\kappa) = \kappa$ and the entropy distance as a notion of distance between probability measures to define the ambiguity set. The resulting problems are solved in an iterative fashion requiring the solution of one *standard robust* problem per iteration.

For a decision w to fulfill the soft robust constraint for a risk measure $\mathcal{R}$ with Lipschitz constant C, we require that

$$\max_{\kappa \in [0,\delta]} \mathcal{R}^\kappa(\hat{P}, w) \le f(\kappa), \tag{45}$$

or equivalently $\mathcal{R}(\langle X^{\hat{P}}, w \rangle) + \max_{\kappa \in [0,\delta]} \{\kappa C ||w||_q - f(\kappa)\} \le 0$. Because f is convex, it turns out that we can find one κ^*, such that the infinitely many constraints in (44) can be replaced by a single one. We either have

$$\max_{\kappa \in [0,\delta]} \{\kappa C ||w||_q - f(\kappa)\} = \begin{cases} \delta C ||w||_q - f(\delta), & \text{for } \kappa^* = \delta \\ -f(0), & \text{for } \kappa^* = 0, \end{cases} \tag{46}$$

i.e., a boundary solution, or the maximum is given by the first-order condition

$$C ||w||_q - \frac{\partial f}{\partial \kappa} = 0. \tag{47}$$

We could for example choose $\delta = \infty$ and $f(\kappa) = d\kappa^2 + \beta$, which leads to $\kappa^* = \frac{C||w||_q}{2d}$. Consequently, (44) becomes

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w\rangle) \\ \text{s.t.} & \mathcal{R}(\langle X^{\hat{P}}, w\rangle) + \frac{C^2||w||_q^2}{4d} \le \beta, \\ & w \in \mathcal{W}. \end{array} \tag{48}$$

In general, problem (48) is a convex problem with finitely many constraints, which can be solved efficiently for the risk measures discussed herein. We note that f also could be chosen as a linear function or as an arbitrary convex polynomial. The chosen form gives the modeler the freedom to model the trade-off between performance and robustness: The parameter β represents the risk bound for the nominal model and d offers the possibility of weakening the risk constraints for the other measures. Measures that are far away from the reference measure have to fulfill looser risk limits than measures that are closer to the reference measure. Thus, the robustification is not restricted to measures in a prespecified neighborhood of $\hat{P}$ but rather takes into account all measures according to their distance from $\hat{P}$.

4. Numerical Study

In this section, we numerically test a selected set of risk measures against their robust counterparts. As is common in the literature, we use a rolling horizon analysis to evaluate the out-of-sample performance of different portfolio selection criteria. This *as if* analysis permits us to assess what would have happened, had we applied a specific portfolio selection criterion in the past. The selection of data sets is motivated by a similar out-of-sample analysis performed by DeMiguel et al. (2009a). All computations are carried out in MATLAB 2012a using GUROBI 5.0 as a solver and YALMIP (see Löfberg (2004)) to formulate the optimization problems.

We test the portfolio selection rules S_γ, CVaR, standard deviation, and deviation from the median against their respective robust counterparts. The selection of the first three measures is motivated by their importance in finance literature; the mean absolute deviation from the median is interesting, because it is a L^1-equivalent of S_γ.

As a benchmark, we use the 1/N investment strategy, investing uniformly in all available assets, which has received significant attention in recent literature on portfolio selection (e.g. DeMiguel et al. 2009b). Pflug et al. (2012) show that the 1/N rule eventually becomes optimal if ambiguity about the true distribution of the asset returns increases. The uniform portfolio allocation and the nominal problem thus can be seen as two extremes with respect to ambiguity in the loss distribution: The former assumes no information at all about the distribution, whereas the latter assumes complete information. Optimally, a robustified portfolio selection rule outperforms both extremes by incorporating the available information $\hat{P}$ while also insuring against misspecification of the model.

Additionally, we use the short-sale constrained versions of the non-robustified problems as benchmarks. Preventing short-sales makes portfolios with negative and therefore also with extremely large positive positions impossible and therefore the corresponding solutions are less prone to estimation error as was also noted for example in DeMiguel et al. (2009a). In fact, Jagannathan and Ma (2003) showed that for Markowitz optimization no-short-sale constraints are equivalent to shrinking the covariance matrix – a procedure which is often advocated to mitigate the negative effects of estimation error (see for example Jorion 1986, DeMiguel and Nogales 2009).

This section comprises four subsections: First we outline the setup of the rolling horizon study followed by a discussion of the data sets used to conduct the study and the parameter choice for the different portfolio selection rules. The third section describes how to choose the parameter κ for the robustified risk measures. In the final section, we analyze the numerical results.

4.1. Out-of-sample evaluation

We use historical loss data $x_t \in \mathbb{R}^N$ over T periods and choose an estimation window of length L, with $L < T$. Starting at period $L+1$, we use the data on the first L historical losses $(x_1, \ldots, x_L)$ as an estimate of the future loss distribution to compute the portfolio position w_{L+1} for period $L+1$. Specifically, we choose $\hat{P}$ to be the uniform distribution on the scenarios $(x_1, \ldots, x_L)$, such that $\hat{P}(x_i) = 1/L$ for all $1 \leq i \leq L$. In the next step, we evaluate the portfolio against the actual historical losses in period $L+1$ to arrive at the portfolio loss $l_{L+1} = \langle w_{L+1}, x_{L+1} \rangle$. Subsequently, we adopt a *rolling* estimation window for the data by removing the first return and adding x_{L+1} to our data for estimation. Continuing in this manner, we cover the whole data set and obtain a sequence of portfolio decisions $(w_{L+1},\dots,w_T)$ and a sequence of realized losses $(l_{L+1},\dots,l_T)$, which we use to assess the quality of the portfolio selection mechanism.

For the rolling horizon analysis, we solve the problem

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} & \mathcal{R}^\kappa(\hat{P},w) \\ \text{s.t.} & \langle w,\mathbb{1}\rangle = 1 \end{array} \tag{49}$$

for the risk and deviation measures mentioned previously. We compare the results for $\kappa = 0$, i.e., the nominal case with the robustified results for $\kappa > 0$. For the short-selling constrained problems, we additionally add the constraint $w \geq 0$.

In practice, a portfolio manager would impose many more restrictions on feasible portfolio weights than it is done here. However, because we want to analyze the impact of robustification on the performance of $\mathcal{R}$ as a portfolio selection criteria, we refrain from diluting the results by imposing further constraints.

We use three performance criteria to assess the quality of a portfolio selection rule: the risk and the returns associated with the empirical distribution of $(l_{L+1},...,l_T)$, and the average turnover resulting from the respective decisions. We use the arithmetic mean to calculate the average returns from the realized losses in the rolling horizon study. The turnover is defined as follows: Let $w_t^+\in\mathbb{R}^N$ be the relative portfolio weights after the losses l_t have been realized but before the rebalancing decision in period $t+1$,

$$w_t^+ = \frac{w_t\odot(1-l_t)}{\langle w_t,(1-l_t)\rangle}, \tag{50}$$

where $\odot$ is the component-wise or Hadamard product. Then the turnover is defined as

$$\text{turnover} = \frac{1}{T-L-1}\sum_{t=L+1}^{T-1}\langle |w_t^+ - w_{t+1}|,\mathbb{1}\rangle. \tag{51}$$

The turnover is a measure of stability of the portfolio over time. Portfolio strategies that yield a high turnover are undesirable because of the induced transaction costs and, in extreme cases, the practical infeasibility of the resulting decisions.

Note that neither the expected return nor the turnover appear explicitly in the optimization problems and it could be argued that if these are of relevance to the portfolio manager, then this

Abbr.	Description	Range	Freq.	T	L
10Ind	10 US industry portfolios	07.1963–12.2010	Monthly	570	240
48Ind	48 US industry portfolios	07.1963–12.2010	Monthly	570	240
6SBM	6 portfolios formed on size and book-to-market	07.1963–12.2010	Monthly	570	240
25SBM	25 portfolios formed on size and book-to-market	07.1963–12.2010	Monthly	570	240
100SP	100 S&P assets	04.1983–12.2010	Weekly	1445	500

Table 1 Overview of used historical data sets. The first four data sets were obtained from the homepage of Kenneth French, `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The last data set was obtained from Yahoo Finance.

should be addressed directly in the problem formulation. However, we still report these figures, since they reveal interesting insights into the characteristics of the solutions and the differences between the approaches.

4.2. Input data & parameters

We use the data sets described in Table 1 for our numerical studies. The data for the first four portfolios are available on Kenneth French's webpage. The portfolios 6SBM and 25SBM are discussed in Fama and French (1992). The data for 10Ind, 48Ind, 6SBM, and 25SBM each consist of $T = 570$ monthly returns from July 1963 until December 2010. The number of assets ranges from 6 to 45, thus representing small- to medium-scale asset universes. The largest data set 100SP consists of weekly returns for 100 randomly selected S&P assets from April 1983 to December 2010, i.e., $T = 1445$ data points.

Because short windows for estimation often lead to unrealistic estimates of the loss distributions (cf. Kritzman et al. 2010), we choose $L = 240$ for the data sets consisting of monthly losses and $L = 500$ for 100SP, such that the forecast window covers a time span of 20 and approximately 10 years in the past, respectively. Consequently, we obtain 330 and 945 portfolio decisions and realized losses on which we base our analysis. We note that, in line with the findings in the literature, shorter time horizons are expected to lead to worse outcomes both for the robustified and the unrobustified risk minimization approaches, i.e., make the uniform portfolio strategy seem more attractive.

We choose the functional S_1, CVaR with parameter $\alpha = 0.95$, the deviation from the median with parameter $a = 2$, and upper semi-standard deviation as risk measures for our numerical tests.

4.3. Choice of κ

The choice of the parameter κ is crucial when using the robustified risk measures. Portfolio optimization problems with differently sized asset universes, different degree of stability of the stochastic process over time, and different additional constraints call for tailored choices of the robustness parameter κ.

Increasing κ can mitigate the effects of estimation error resulting from using a wrong distribution of losses $\hat{P}$ in the optimization problem. However, choosing $\kappa > 0$ introduces a bias in the form of the penalization term. Thus, when choosing κ the modeler must weigh contradictory goals of minimizing estimation error and minimizing bias – a situation reminiscent of many statistical procedures.

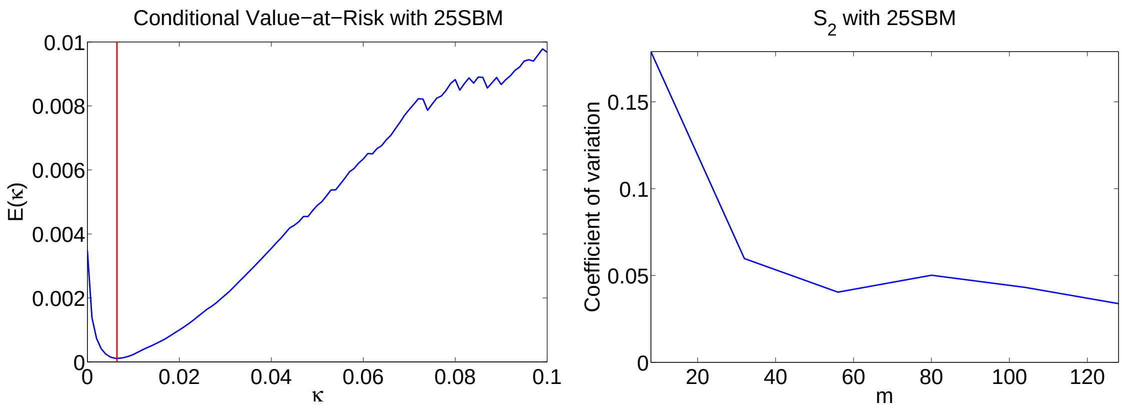
In this paper, we choose κ by a cross validation type procedure based on S in-sample losses $X_1, \ldots, X_S$: For a set $\mathcal{S} \subseteq \{1, \ldots, S\}$, denote by $X^{\mathcal{S}}$ the random variable with $\mu(X^{\mathcal{S}} = X_j) = |\mathcal{S}|^{-1}$ for all $j \in \mathcal{S}$. Now consider randomly drawn partitions $\mathcal{S}_1^i$, $\mathcal{S}_2^i$ of $\{1, \ldots, S\}$, i.e. $\{1, \ldots, S\} = \mathcal{S}_1^i \cup \mathcal{S}_2^i$ and $\mathcal{S}_1^i \cap \mathcal{S}_2^i = \emptyset$ for $i = 1, \ldots, m$. For $\kappa > 0$, define the following measure of estimation error

$$E(\kappa) = \sum_{i=1}^{m} \Big(\mathcal{R}^{\kappa}(\langle X^{\mathcal{S}_1^i}, w^{\kappa}_{\mathcal{S}_1^i} \rangle) - \mathcal{R}(\langle X^{\mathcal{S}_2^i}, w^{\kappa}_{\mathcal{S}_1^i} \rangle) \Big)^2 \tag{52}$$

where $w^{\kappa}_{\mathcal{S}_1^i}$ are the optimal portfolio weights found using $X^{\mathcal{S}_1^i}$ and κ in the robustified portfolio optimization problem. The optimal value κ^* is chosen such that

$$E(\kappa^*) = \min_{\kappa} E(\kappa). \tag{53}$$

As is common with cross-validation, the above optimization problem of finding κ^* has no special structure that can be exploited by optimization routines. In fact, in general it can not even be guaranteed that the problem is convex. However, since the optimization involves only one variable and the violations of convexity are typically *mild*, standard line search routines do not have a



(a) (b)

Figure 1 Panel (a): The cross validation function E for the first CVaR problem of the rolling horizon study for the data set 25SBM. The red line indicates the optimal value found by the line search routine *fminbnd*; Panel (b): Dependence of the coefficient of variation of 30 runs of (53) on the number of partitions m for the first problem of the rolling horizon study for the data set 25SBM and S_2.

problem finding the minimum. For the numerical examples, we use the MATLAB function *fminbnd* to solve (53). A typical example of an objective function, including mild non-convexities, is depicted in Figure 1, Panel (a). As can be seen from the figure the *fminbnd* routine finds the minimal κ.

The above method *simulates* an in-sample/out-of-sample situation using only in-sample data. κ is chosen in such a way so as to strike a balance between variance and bias: If κ is too large the robustified in-sample risk $\mathcal{R}^\kappa(\langle X^{\mathcal{S}_1^i}, w^\kappa_{\mathcal{S}_1^i}\rangle)$ will be much larger than the out-of-sample risks $\mathcal{R}(\langle X^{\mathcal{S}_2^i}, w^\kappa_{\mathcal{S}_1^i}\rangle)$ for many i. If κ is too small, there will be many partitions i such that the estimate of the optimal robust in-sample risk is too small and therefore $\mathcal{R}^\kappa(\langle X^{\mathcal{S}_1^i}, w^\kappa_{\mathcal{S}_1^i}\rangle)$ is smaller than $\mathcal{R}(\langle X^{\mathcal{S}_2^i}, w^\kappa_{\mathcal{S}_1^i}\rangle)$. Both situations are penalized by E.

Clearly, since the partitions used in the calculation of E are randomly sampled, the solution to (53) is random as well. By increasing the number of partitions m it is possible to consider more different configurations of the data points and therefore also get a more complete picture of the out-of-sample error. One would therefore expect that the outcome of the optimization (53) will be less dependent on the chosen random seed if m is large. In other words, if we solve (53) several times, we would expect a small variance of the obtained optimal values κ^*.

We choose $m = 100$, since this seems to consistently keep the coefficients of variations of κ^* below 5% for all the considered risk measures and data sets. See Figure 1, Panel (b) for a depiction of the typical dependence of the coefficient of variation of κ^* on the parameter m. For the numerical results presented in the next section, we use the whole rolling horizon window for finding κ and the partitions are chosen to have equally sized parts, i.e., $|\mathcal{S}_1^i| = |\mathcal{S}_2^i|$ for all $i = 1, \ldots, m$ and $S = L$.

4.4. Results

Table 2 shows the out-of-sample risk for the different asset universes in Table 1. The reported figures are calculated for the set of out-of-sample losses generated by the rolling horizon study; for the robustified measures $\mathcal{R}^\kappa$, we report the value of the unrobustified measure $\mathcal{R}^0$ calculated for losses generated by the robustified problems. The corresponding risks for the 1/N strategy and the short-sale constrained problems, which serve as benchmarks, can also be found in the table. We use a standard MATLAB implementation of the two-sided bootstrapping test, based on 5000 samples, to test whether the risk of the respective robustified versions of the risk measures differ significantly from the non-robustified versions.

We note that both the robustified risk measures and the non-robustified risk measures outperform the 1/N rule in most cases and are never significantly outperformed by the 1/N rule. This finding is interesting, especially for the non-robustified risk measures, because it implies that the chosen measure $\hat{P}$ is close enough to the real data generating process to result in sensible decisions. Thus, it confirms our choice of the window size.

Turning to the comparison between the robustified and non-robustified measures, we note that in most cases, the former yield a lower out-of-sample risk than latter. There are some exceptions, but the general picture indicates that larger data sets yield larger (more significant) differences between the two risk measures, and the robustified measures are unambiguously better for large data sets. For data sets with fewer assets, the situation is less clear though. For 6SBM, the non-robustified risk measures even fare better than the robustified measures. These results indicate that for this rather small data set, the information encoded in $\hat{P}$ is accurate to a degree that the gains from

		10Ind	48Ind	6SBM	25SBM	SP100
Nominal	S_2	0.0645	0.0719	**0.0650**	0.0575	0.0363
	CVaR$_{.95}$	0.0887	0.1187	**0.0850**	0.1104	0.0587
	DM$_2$	0.0467	0.0555	**0.0434**	0.0420	0.0272
	Std	0.0375	0.0383	**0.0406**	0.0370	0.0189
No short-sales	S_2	0.0648	0.0643	0.0783	0.0757	0.0340
	CVaR$_{.95}$	0.0876	**0.0888**	0.1037	0.1020	**0.0437**
	DM$_2$	0.0465	0.0467	0.0539	0.0527	0.0237
	Std	0.0375	0.0372	0.0444	0.0432	0.0183
1/N	S_2	0.0757	0.0856	0.0861	0.0888	0.0449
	CVaR$_{.95}$	0.0990	0.1121	0.1142	0.1167	0.0562
	DM$_2$	0.0537	0.0599	0.0608	0.0631	0.0305
	Std	0.0428	0.0478	0.0481	0.0497	0.0239
Robustified	S_2	**0.0633**	**0.0622*****	0.0665**	**0.0568**	**0.0333*****
	CVaR$_{.95}$	**0.0836**	0.0897**	0.0969**	**0.0817****	0.0497*
	DM$_2$	**0.0458*****	**0.0443*****	0.0442***	**0.0385*****	**0.0231*****
	Std	**0.0368****	**0.0355*****	0.0412**	**0.0369**	**0.0176*****

Table 2 Risks for the non-robustified and robustified measures, for the models with no short-sale restrictions, as well as the 1/N rule. The best results for a combination of risk measure and data sets are printed in bold font. Stars indicate the significance of the difference of the robustified and unrobustified solutions (***: 1%, **: 5%, *: 10%).

robustification are smaller than the losses that result from the distortion of the objective function in the robustified problem using the κ found by cross-validation. For larger sets of assets, this effect reverses, as is evident from the results for 100SP and 48Ind.

Looking at the results for the short-sale constrained problems, we see that those yield smaller risk than the robustified risk measures only in two instances. A comparison between the unrobustified risk measures and the short-sale constrained problems yields mixed results: In some instances the results of the short-sale constrained problems are better than the results for the unrobustified problems (especially for the data sets with more assets), while in some instances the restriction to positive asset weights leads to worse results in terms of risk. In summary, it can be said that for larger data sets the short-sale constraint mitigates some of the problems connected with estimation error but it generally tends performs worse than the robustified risk measures in this respect.

Although maximizing returns was not the goal of this experiment, we report the expected out-of-sample average returns in Table 3. Note that the quantities reported in Table 3 are returns, i.e., the negative of the losses used to calculate the risk. Hence, larger values are better. Surprisingly, the

		10Ind	48Ind	6SBM	25SBM	SP100
Nominal	S_2	**0.0110**	0.0081	**0.0163**	**0.0191**	0.0016
	CVaR$_{.95}$	**0.0121**	0.0084	**0.0170**	**0.0241**	0.0012
	DM$_2$	**0.0112**	0.0076	**0.0160**	**0.0196**	0.0016
	Std	**0.0111**	0.0073	**0.0151**	**0.0162**	0.0018
No short-sales	S_2	0.0099	0.0098	0.0104	0.0105	**0.0022**
	CVaR$_{.95}$	0.0104	0.0101	0.0097	0.0095	**0.0023**
	DM$_2$	0.0100	0.0097	0.0108	0.0105	**0.0022**
	Std	0.0100	0.0099	0.0102	0.0105	**0.0023**
1/N		0.0099	0.0099	0.0101	0.0105	0.0030
Robustified	S_2	0.0104**	**0.0101****	0.0153***	0.0167***	0.0019
	CVaR$_{.95}$	0.0089***	**0.0116**	0.0141***	0.0179***	0.0020
	DM$_2$	0.0102*	**0.0101****	0.0152**	0.0166***	0.0020
	Std	0.0105**	**0.0099*****	0.0142***	0.0148***	0.0020

Table 3 Returns for the non-robustified and robustified measures, for the models with no short-sale restrictions, as well as the 1/N rule. For explanations regarding the formatting see Table 2.

results are quite similar to the results on the out-of-sample risks, although less pronounced: The robustified measures outperform the unrobustified measures on the two largest data sets. Comparing the returns to those from the 1/N strategy, we find that both the unrobsutified and the robust risk measures outperform the 1/N strategy except on 100SP in which the 1/N policy consistently outperforms all other strategies. With the exception of the data set SP100, the short-sale constrained problems yield lower expected returns than both the robustified and the unrobustified problems and the 1/N strategy is narrowly outperformed by various short-sale constrained problems in about half of the cases.

The situation is clearer for the turnovers, presented in Table 4: Except for two results, turnover for the robustified portfolio selection rules are consistently and significantly better than the respective non-robustified counterparts. Therefore the portfolio compositions that arise from the robustified portfolio selection rules are much more stable and require less rebalancing, hence, incurring less transaction costs than portfolios found using the original measures. Unsurprisingly, the 1/N policy significantly outperforms both the robustified and the unrobustified problems in terms of turnover. The short-sale restricted portfolios exhibit significantly less turnover than the nominal and the robustified risk measures, which is due to the inability to leverage which restricts the absolute size of the portfolio.

		10Ind	48Ind	6SBM	25SBM	SP100
Nominal	S_2	0.1086	0.4222	0.1827	0.5349	0.1961
	CVaR$_{.95}$	0.2300	1.1079	0.3064	1.2437	0.5717
	DM$_2$	0.1878	0.9811	0.2686	1.1463	0.5438
	Std	0.1050	0.3932	0.1507	0.4445	0.1859
No short-sales	S_2	**0.0358**	**0.0481**	**0.0316**	**0.0458**	**0.0470**
	CVaR$_{.95}$	**0.0435**	**0.0598**	**0.0415**	**0.0473**	**0.0634**
	DM$_2$	**0.0768**	**0.1037**	**0.0520**	**0.0784**	0.1208
	Std	**0.0373**	**0.0481**	**0.0333**	**0.0437**	**0.0487**
1/N		0.0498	0.0498	0.0407	0.0426	0.0327
Robustified	S_2	0.0733***	0.1629***	0.2015***	0.3002***	0.0674***
	CVaR$_{.95}$	0.2390	0.6086***	0.2007***	0.8173***	0.1634***
	DM$_2$	0.0940***	0.1902***	0.2673***	0.3427***	**0.0790*****
	Std	0.0755***	0.1734***	0.1675***	0.2611***	0.0751***

Table 4 Turnovers for the non-robustified and robustified measures, for the models with no short-sale restrictions, as well as the 1/N rule. For explanations regarding the formatting see Table 2.

	Mean				Standard Deviation			
	S_2	CVaR$_{.95}$	DM$_2$	Std	S_2	CVaR$_{.95}$	DM$_2$	Std
10Ind	0.0036	0.0028	0.0038	0.0036	0.0006	0.0026	0.0007	0.0005
48Ind	0.0155	0.0311	0.0152	0.0150	0.0018	0.0091	0.0020	0.0015
6SBM	0.0014	0.0028	0.0013	0.0016	0.0004	0.0015	0.0003	0.0005
25SBM	0.0040	0.0069	0.0040	0.0040	0.0004	0.0026	0.0003	0.0003
100SP	0.0174	0.0612	0.0156	0.0170	0.0022	0.0266	0.0013	0.0022

Table 5 Means and standard deviations for κ^* found in (53) for the rolling horizons study.

In summary, the robustified risk measures perform very well for larger data sets. For data sets with only a few assets, the results are mixed, because the approximation of the data generating process by $\hat{P}$ seems more accurate.

Table 5 shows the average values of κ^* that are chosen by the cross-validation type procedure discussed in Section 4.3 for the different data sets and risk measures. What can be seen is that κ^* is generally higher for data sets with more assets, which can be attributed to the fact that the L^p norms in the corresponding spaces are larger. Furthermore, we note that the κ^* for the risk measures S$_2$, DM$_2$, and Std are quite similar and that, except for 10Ind, the κ^* for CVaR$_{.95}$ are relatively larger. Also the variability of the κ^* over time, expressed in terms of the standard deviation, is larger for the CVaR$_{.95}$ than for the other measures, which can be attributed to the sensitivity of CVaR with respect to small changes in the (upper tail of) distribution.

5. Conclusion and further work

We propose a framework for solving portfolio optimization problems under ambiguous loss distributions. The problems are solved as worst case over a set of distributions called the ambiguity set, constructed as a Kantorovich ball around a reference measure $\hat{P}$. In contrast to most other approaches, the ambiguity sets are constructed without any assumptions regarding the membership of the true distribution in any parametric family, such that the ambiguity sets are fully non-parametric.

Despite the generality of the approach, we obtain closed-form expressions for a large class of robustified risk measures. Furthermore, these closed-form expressions are typically numerically tractable, in the sense that they can be incorporated as convex objective functions or constraints in portfolio optimization problems. We also provided a numerical study showing that the robustified portfolio selection problems usually yield better results than the nominal problems, unless the data sets have very few assets.

The robustified optimization problems bear a close resemblance to the norm-constrained problems proposed by DeMiguel et al. (2009a). The results in this paper thus yield an alternative interpretation to norm-constrained portfolio selection rules and thereby of Bayesian shrinkage type estimators in portfolio selection – an aspect that deserves further attention and may be an interesting topic for future research. More generally, the connection between robustification with respect to the Kantorovich metric and norm-regularization seems to be an interesting topic to explore, for example in regression, where it can be shown, using results of this paper, that classical ridge regression is equivalent to a robustified OLS estimator.

Acknowledgements

The author is grateful to the editorial team and two anonymous referees for their helpful suggestions and comments which greatly improved the paper. Furthermore, the author would also like to thank Johan Löfberg for his excellent support of the YALMIP package.

Biography

David Wozabal is an assistant professor at the Technische Universität München, Munich, Germany. His research interests are in robust and stochastic optimization, risk management, and energy markets modeling.

References

P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3): 203–228, 1999.

A. Ben-Tal, D. Bertsimas, and D.B. Brown. A soft robust model for optimization under ambiguity. *Operations Research*, 58:1220–1234, 2010.

D. Bertsimas, D. Pachamanova, and M. Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510 – 516, 2004.

O. Bousquet and A. Elisseeff. Stability and generalization. *Journal of Machine Learning Research*, 2:499–526, 2002.

G.C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18 (3):853–877, 2007.

C. Caramanis, H. Xu, and S. Mannor. Robustness and regularization of support vector machines. *Journal of Machine Learning Research*, 10:1485–1510, 2009.

V. K. Chopra and W. T. Ziemba. The effect of errors in mean, variance and covariance estimates on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11, 1993.

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

V. DeMiguel and F.J. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3): 560–577, 2009.

V. DeMiguel, L. Garlappi, F.J. Nogales, and R. Uppal. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, 55(5):798–812, 2009a.

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009b.

D. Denneberg. Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, 20:181–190, 1990.

J. Dupačová. The minimax problem of stochastic linear programming and the moment problem. *Československá Akademie Věd. Ekonomicko-Matematický Obzor*, 13(3):279–307, 1977.

L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. 18(4):1035–1064, 1997.

L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operations Research*, 51(4):543–556, 2003.

E.F. Fama and K.R. French. The cross-section of expected stock returns. *The Journal of Finance*, 47(2): 427–465, 1992.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4): 429–447, 2002.

D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28 (1):1–38, 2003.

J. Gotoh and A. Takeda. On the role of norm constraints in portfolio selection. *Computational Management Science*, 8:323–353, 2011.

J. Gotoh, K. Shinozaki, and A. Takeda. Robust portfolio techniques for mitigating the fragility of CVaR minimization and generalization to coherent risk measures. *Quantitative Finance*, 13:1621–1635, 2013.

R. Jagannathan and T. Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 58:1651–1684, 2003.

P. Jorion. Bayes-stein estimation for portfolio analysis. *Journal of Financial and Quantitative Analysis*, 21 (03):279–292, September 1986.

M. Kritzman, S. Page, and D. Turkington. In Defense of Optimization: The Fallacy of 1/N. *Financial Analysts Journal*, 66(2):31–39, 2010.

S. Kusuoka. A remark on law invariant convex risk measures. volume 10 of *Advances in mathematical economics*, pages 91–100. Springer, 2007.

J.Y. Li and R.H. Kwon. Portfolio selection under model uncertainty: a penalized moment-based optimization approach. *Journal of Global Optimization*, 56(1):131–164, 2013.

A.E.B. Lim, J.G. Shanthikumar, and G.Y. Vahn. Conditional value-at-risk in portfolio optimization: Coherent but fragile. *Operations Research Letters*, 39(3):163 – 171, 2011.

J. Löfberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. URL `http://users.isy.liu.se/johanl/yalmip`.

P.J. Maenhout. Robust portfolio rules and asset pricing. *Review of Financial Studies*, 17(4):951–983, 2004.

K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141, 2009.

G. Ch. Pflug and W. Römisch. *Modeling, Measuring and Managing Risk.* World Scientific, Singapore, 2007.

G.Ch. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.

G.Ch. Pflug, A. Pichler, and D. Wozabal. The 1/n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.

R. T. Rockafellar and S. Uryasev. Optimization of Conditional Value-at-Risk. *The Journal of Risk*, 2(3): 21–41, 2000.

A. Ruszczyński and A. Shapiro. Optimization of risk measures. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 119–157. Springer, 2006.

A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14 (4):1237–1249, 2004.

A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods & Software*, 17(3):523–542, 2002.

C. Villani. *Topics in optimal transportation.* American Mathematical Society, Providence, RI, 2003.

S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, 17(1):43–54, August 1995.

S. Wang. A class of distortion operators for pricing financial and insurance risks. *The Journal of Risk and Insurance*, 67(1):15–36, 2000.

D. Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47, 2012.

S. Yitzhaki. Stochastic dominance, mean variance, and gini's mean difference. *The American Economic Review*, 72(1):178–185, 1982.

S. Zhu and M. Fukushima. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, 57:1155–1168, September 2009.

S. Zymler, B. Rustem, and D. Kuhn. Robust portfolio optimization with derivative insurance guarantees. *European Journal of Operational Research*, 210(2):410–424, 2011.