Robustifying Convex Risk Measures: A Non-Parametric Approach

David Wozabal

January 13, 2012

Abstract

This paper introduces a framework for robustifying convex, law invariant risk measures, to deal with ambiguity of the distribution of random asset losses in portfolio selection problems. The robustified risk measures are defined as the worst-case portfolio risk over the ambiguity set of loss distributions, where an ambiguity set is defined as a neighborhood around a reference probability measure representing the investors beliefs about the distribution of asset losses. 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 robustified versions of several risk measures, including the standard deviation, the Conditional Value-at-Risk, and the general class of distortion functionals. The resulting robust policies are of similar computational complexity as their non-robust counterparts. Finally, a numerical study shows that in most instances the robustified risk measures perform significantly better out-of-sample than their non-robust variants in terms of risk, expected losses, and turnover.

Keywords: Robust optimization; Kantorovich distance; Norm-constrained portfolio optimization; Soft robust constraints

1 Introduction

Since Markowitz published his seminal work on portfolio optimization, scientific communities and financial industry have proposed a plethora of policies to find risk-optimal portfolio decisions in the face of uncertain future asset losses. Most proposed policies, 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 there is no uncertainty about the *nature of the randomness.* However, in most cases, the distribution of the losses is actually unknown to the decision maker and thus typically replaced by an estimate. It was recognized already in early papers that the estimation of the distributions, underlying the stochastic programs in question, introduce an additional level of model uncertainty into the problem (see Dupačová, 1977, 1980). The estimation errors thus introduced at the level of the loss distributions can lead to dramatically erroneous portfolio decisions, as has been well documented for the classical Markowitz portfolio selection problem (see Michaud; Broadie, 1993; 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 the random variables. Possible ways to deal with such ambiguity in portfolio optimization can be categorized roughly into three classes: robust estimation, norm-constrained portfolio optimization, and robust optimization.

Robust estimation tries 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 Welsch and Zhou (2007) and 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 ideas to immunize stochastic optimization problems with respect to estimation error. In contrast to models with restricted portfolio weights, 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 the ambiguity set. Hence, decisions are optimal in a minimax sense as they have best worst-case outcome. Initial research in this line 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); Shapiro and Ahmed (2004); Calafiore (2007); Pflug and Wozabal (2007); Zhu and Fukushima (2009). These authors each define the ambiguity sets differently and accordingly apply various methods to solve the resulting optimization problems. While most approaches make strong assumptions about the nature of the ambiguity to deal with the robustified problems, there is also some research that uses non-parametric methods (see Calafiore, 2007; Pflug and Wozabal, 2007; Wozabal, 2010; Zymler et al., 2011).

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 of the true distribution of the losses, which we refer to as the reference distribution. As a ambiguity set, we use a neighborhood of this reference distribution which is consistent with the notion of weak convergence. This 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 then 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 Conditional Value-at-Risk, the Wang functional and the Gini functional. The results in this paper are based on theoretical findings in Pflug et al. (2011) 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. In contrast, most other approaches require the real distribution to be in a specific family of distributions or differ from $\hat{P}$ only in a certain way (e.g., different covariance structure).

Furthermore, the obtained analytical expressions for the robustified risk measures lead to robustified stochastic programming problems with the similar computational complexity as the nominal, non-robustified problems. In contrast, in most other robust optimization approaches, the robustified problem tends to be harder to solve than the nominal problem instance. 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 (Ben-Tal et al. (2010)), is possible and leads to computationally tractable problems that can be solved as a single convex programming problem of the same complexity as the original, non-robustified problem. These favorable computational properties of the robustified strategies arise because the obtained robust risk measures have a close connection to the norm-constrained portfolios proposed in previous literature. In fact, we show that using the robustified standard deviation is equivalent to some of the models proposed in DeMiguel et al. (2009a). This paper thus yields a compelling alternative interpretation of norm constraints in portfolio optimization.

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 most commonly used convex risk measures. We also establish a connection between robust risk measures and norm-constrained portfolio optimization. We also demonstrate how robustified risk measures can be used to define soft robust constraints, and the resulting problems can be solved efficiently. In Section 4, a numerical experiment provides a comparison of the out-of-sample performance of several robustified risk measures with respective non-robustified counterparts. In this section, 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

$$\mu(X^P \in A) = P(A) \tag{1}$$

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., 2011). Because the investment policies that we consider only depend on the image measure P, we use P and X^P interchangeably. Let $L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}^n)$ be the Lebesgue space of exponent p containing random variables $X : (\Omega, \mathcal{F}, \mu) \to \mathbb{R}^n$. We denote by $L^p(\Omega, \mathcal{F}, \mu)$ the space $L^p(\Omega, \mathcal{F}, \mu; \mathbb{R})$. Throughout our discussion, we choose q to be the conjugate of p, i.e. choose q such that $1/p + 1/q = 1$. We denote the norm in this space by $|| \cdot ||_{L^p}$ to distinguish it from the p-norm in $\mathbb{R}^n$, which we denote by $|| \cdot ||_p$. With a little abuse of notation, we will sometimes write $P \in L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}^N)$ instead of $X^P \in L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}^N)$.

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

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

1. *convex, $\mathcal{R}(\lambda X + (1 - \lambda)Y) \leq \lambda \mathcal{R}(X) + (1 - \lambda)\mathcal{R}(Y)$ for all $\lambda \in [0, 1]$;*
2. *monotone, $\mathcal{R}(X) \geq \mathcal{R}(Y)$ if $X \geq Y$ 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.

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

$$\mathcal{R}(Y) = \mathcal{R}(Y') \tag{2}$$

for all random variables Y and Y' with the same image measure on $\mathbb{R}$. This assumption is rather innocuous, because it is fulfilled by all meaningful risk measures.

We therefore start by analyzing the following generic portfolio optimization problem:

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} & \mathcal{R}(\langle X^P, w\rangle) \\ s.t. & w \in \mathcal{W}. \end{array} \tag{3}$$

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 vector X^P of random losses is assumed to be in $L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}^N)$. The set $\mathcal{W}$ may represent arbitrary, possibly non-convex conditions 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, for example. 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 (3) 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 (3), 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, analyses of fundamentals, and expert opinions may suggest beliefs about the measure P, the true distribution remains ambiguous in most cases.

It is therefore reasonable to assume that the decision maker takes the available information into account but also accounts for model uncertainty in 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

$$d(\cdot,\cdot) : \mathcal{P}^p(\mathbb{R}^N) \times \mathcal{P}^p(\mathbb{R}^N) \to \mathbb{R}^+ \cup \{0\} \tag{4}$$

to represent a metric on this space (for an introduction to probability metrics, see Gibbs and Su, 2002) . The ambiguity set for a risk measure $\mathcal{R} : L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}) \to \mathbb{R}$ is then defined as

$$\mathcal{B}^p_\kappa(P) = \left\{ Q \in \mathcal{P}^p(\mathbb{R}^N) : d(\hat{P}, Q) \leq \kappa \right\}, \tag{5}$$

i.e. the *ball* of radius κ around the reference measure $\hat{P}$ in the space of measures $\mathcal{P}^p(\mathbb{R}^N)$.

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

$$d_p(P,Q) = \inf \left\{ \left(\int_{\mathbb{R}^N\times\mathbb{R}^N} ||x-y||_p^p d\pi(x,y) \right)^{\frac{1}{p}} : \text{proj}_1(\pi) = P,\ \text{proj}_2(\pi) = Q \right\} \tag{6}$$

where the infimum runs over all transportation plans, viz. joint distributions π on $\mathbb{R}^N \times \mathbb{R}^N$. Accordingly, $\text{proj}_1(\pi)$ and $\text{proj}_2(\pi)$ are the marginal distributions of the first and last N components respectively. The infimum in this definition is always attained (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

$$d_p(P, \hat{P}_n) \overset{n\to\infty}{\longrightarrow} 0 \tag{7}$$

if the p-th moment of P exists. This property justifies the use of d_p to construct ambiguity sets; a stronger metric would not necessarily reduce the degree of ambiguity by collecting more data. Furthermore, the Kantorovich metric plays an important role in stability results in stochastic programming (e.g. Mirkov and Pflug, 2007; Heitsch and Römisch, 2009).

With the preceding definition of the ambiguity set and $\kappa > 0$, we arrive at the robustified problem, the robust counterpart of (3):

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} \sup_{Q\in\mathcal{B}_\kappa^p(P)} & \mathcal{R}(\langle X^Q, w\rangle) \\ s.t. & w \in \mathcal{W}. \end{array} \tag{8}$$

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 $\kappa > 0$

$$(P, w) \mapsto \mathcal{R}^\kappa(P, w) := \sup_{Q\in\mathcal{B}_\kappa^p(P)} \mathcal{R}(\langle X^Q, w\rangle). \tag{9}$$

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

$$w \mapsto \mathcal{R}^\kappa(\hat{P}, w) \tag{10}$$

is convex in w, so problem (8) has a convex objective.

3 Robust Risk Measures

In this section, we derive explicit expressions for the worst-case equivalents of convex, law-invariant risk measures. We consider risk measures $\mathcal{R}$ with a subdifferential representation of the form

$$\mathcal{R}(X) = \sup\{\mathbb{E}(XZ) - R(Z) : Z \in L^q(\Omega, \mathcal{F}, \mu; \mathbb{R})\} \tag{11}$$

for some convex function $R : L^q(\Omega, \mathcal{F}, \mu; \mathbb{R}) \to \mathbb{R}$. If $\mathcal{R}$ is lower semi-continuous, then it admits a representation of the form (11), with $R = \mathcal{R}^*$ where $\mathcal{R}^*$ is the convex conjugate of $\mathcal{R}$. If $R = \mathcal{R}^*$ and X is in the interior of the domain $\{X \in L^p(\Omega, \sigma, \mu; \mathbb{R}) : \mathcal{R}(X) < \infty\}$, then

$$\text{argmax}_Z \{\mathbb{E}(XZ) - R(Z)\} = \partial\mathcal{R}(X)$$

where $\partial\mathcal{R}(X)$ is the set of subgradients of $\mathcal{R}$ at X. Consequently, we denote the set of maximizers of (11) 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 Ruszczynski and Shapiro (2006) as well as in Pflug and Römisch (2007). We start with the simplest risk measure: the expectation operator.

Example 1 (Expectation). *As a linear functional,* $\mathbb{E}(X) : L^1(\Omega, \sigma, \mu) \to \mathbb{R}$ *is not a classical risk measure. The subdifferential representation is trivial with* $\mathcal{Z} = \{1\}$.

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* $S_\gamma : L^2(\Omega, \sigma, \mu) \to \mathbb{R}$ *is defined as*

$$S_\gamma(X) = \gamma\,\mathrm{Std}(X) + \mathbb{E}(X). \tag{12}$$

The subdifferential representation of S_γ *is given by*

$$S_\gamma(X) = \sup\left\{\mathbb{E}(XZ) : \mathbb{E}(Z) = 1,\ ||Z||_{L^2} = \sqrt{1+\gamma^2}\right\}. \tag{13}$$

We also address the Conditional Value-at-Risk (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 (also called the Average Value-at-Risk)*

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(t)dt, \tag{14}$$

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 values in the upper tail of the loss distribution, such that* α *typically is chosen close to* 1. *The dual representation of CVaR is given by*

$$\mathrm{CVaR}_\alpha(X) = \sup\left\{\mathbb{E}(XZ) : \mathbb{E}(Z) = 1, 0 \le Z \le \frac{1}{1-\alpha}\right\} \tag{15}$$

for $0 < \alpha \le 1$.

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

Example 4 (Distortion Functionals). *Let $H : [0,1] \to \mathbb{R}$ be a convex function, then*

$$\mathcal{R}_H(X) = \int_0^1 F_X^{-1}(p) dH(p) \tag{16}$$

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

$$\mathcal{R}_H(X) = \sup\{\mathbb{E}(XZ) : Z = h(U),\ U \text{ uniform on } [0,1]\} \tag{17}$$

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

Note that the CVaR is a distortion functional with $H(p) = \max\left(\frac{p-(1-\alpha)}{\alpha}, 0\right)$. Two other prominent examples of distortion functionals appear next.

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

$$W_\lambda(X) = \int_0^\infty \Phi\left(\Phi^{-1}(1 - F_X(t)) + \lambda\right) dt \tag{18}$$

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

$$W_\lambda(X) = \int_0^1 F_X^{-1}(p) dH_\lambda(p) \tag{19}$$

with $H_\lambda(p) = -\Phi\left[\Phi^{-1}(1-p) + \lambda\right]$. Note that (19) 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 hazard transform or power distortion $P_r : L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}) \to \mathbb{R}$ for $0 < r \leq 1$ is defined as*

$$P_r(X) = \int_0^\infty (1 - F_X(t))^r dt, \tag{20}$$

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

$$P_r(X) = \int_0^1 F_X^{-1}(p) dH_r(p), \tag{21}$$

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

Finally, two less known risk measures also fall in the category of distortion measures as introduced by Denneberg (1990).

Example 7 (Gini measure). *The expectation-corrected Gini measure* $\mathrm{Gini}_r : L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}) \to \mathbb{R}$ *is defined by*

$$\mathrm{Gini}_r(X) = \mathbb{E}(X) + r\mathbb{E}(|X - X'|) \tag{22}$$

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

$$\text{Gini}_r(X) = \mathcal{R}_H(X) = \int_0^1 F_X^{-1}(t)dH(t) \tag{23}$$

with $H(t) = (1-r)t + rt^2$.

Example 8 (Deviation from the median)**.** *The deviation from the median* $\text{DM}_a : L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}) \to \mathbb{R}$ *is defined by*

$$\begin{aligned} \text{DM}_a(X) &= \mathbb{E}(X) + a\mathbb{E}\left(|X - F_X^{-1}(0.5)|\right) &(24)\\ &= \int_0^1 F_X^{-1}(t)dt + a\int_0^1 |F_X^{-1}(t) - F_X^{-1}(0.5)|dt &(25)\\ &= \int_0^{1/2} F_X^{-1}(t)(1-a)dt + \int_{1/2}^1 F_X^{-1}(t)(1+a)dt = \int_0^1 F_X^{-1}(t), dH(t) &(26)\end{aligned}$$

with

$$H(p) = \begin{cases} p(1-a), & p < 0.5 \\ \frac{1}{2}(1-a) + \frac{p-1}{2}(1+a), & p \geq 0.5. \end{cases} \tag{27}$$

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 in calculating robustified risk measures is to define a measure Q such that

$$\langle X^Q, w\rangle = \langle X^{\hat{P}}, w\rangle + c|Z|^{q/p}\,\text{sign}(Z), \tag{28}$$

for $Z \in \partial\mathcal{R}(\langle X^{\hat{P}}, w\rangle)$. In (28) 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 ascend 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 $Z \in \mathcal{R}(\langle X^Q, w\rangle)$.

Proposition 1. *Let $\mathcal{R} : L^p(\Omega, \sigma, \mu; \mathbb{R}) \to \mathbb{R}$ be a convex, law-invariant risk measure and $1 \leq p < \infty$ 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*

$$||Z||_{L^q} = C \text{ for all } Z \in \bigcup_{X \in L^p} \partial\mathcal{R}(X) \text{ with } R(Z) < \infty, \text{or} \tag{29}$$

2. *$p = 1$ and*

$$||Z||_{L^\infty} = C \text{ and } |Z| = C \text{ or } |Z| = 0, \tag{30}$$

then the solution to the inner problem (9) *is*

$$\mathcal{R}^\kappa(\hat{P}, w) = \mathcal{R}(\langle X^{\hat{P}}, w\rangle) + \kappa C||w||_q. \tag{31}$$

Proof. This follows directly from Lemma 1 and Propositions 1 and 2 in Pflug et al. (2011). □

Note that all discussed examples fulfill condition (29) or (30) and therefore we can derive robust version of the discussed risk measures based. 1.

Proposition 2. 1. *The robustified expectation operator* $\mathbb{E}^\kappa : L^1(\Omega, \mathcal{F}, \mu; \mathbb{R}^N) \times \mathbb{R} \to \mathbb{R}$ *is given by*

$$\mathbb{E}^\kappa(P, w) = \mathbb{E}(\langle X^P, w\rangle) + \kappa||w||_\infty. \tag{32}$$

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*

$$S^\kappa_\gamma(P, w) = S_\gamma(\langle X^P, w\rangle) + \kappa\sqrt{1+\gamma^2}||w||_2. \tag{33}$$

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*

$$\mathrm{CVaR}^\kappa_\alpha(P, w) = \mathrm{CVaR}_\alpha(\langle X^P, w\rangle) + \frac{\kappa}{1-\alpha}||w||_\infty. \tag{34}$$

4. *For* $1 < p < \infty$ *and a general distortion measure* $\mathcal{R}_H : L^p(\Omega, \mathcal{F}, \mu; \mathbb{R}) \to \mathbb{R}$, *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*

$$\mathcal{R}^\kappa_H(P, w) = \mathcal{R}_H(\langle X^P, w\rangle) + \kappa||h(U)||_{L^q}||w||_q, \tag{35}$$

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*

$$W^\kappa_\lambda(P, w) = W_\lambda(\langle X^P, w\rangle) + \kappa e^{\lambda^2/2}||w||_2. \tag{36}$$

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*

$$P^\kappa_r(P, w) = P_r(\langle X^P, w\rangle) + \frac{\kappa r}{\sqrt{2r-1}}||w||_2. \tag{37}$$

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*

$$\mathrm{Gini}^\kappa_r(P, w) = \mathrm{Gini}_r(\langle X^P, w\rangle) + \kappa\sqrt{\frac{3+r^2}{3}}||w||_2. \tag{38}$$

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*

$$\mathrm{DM}^\kappa_a(P, w) = \mathrm{DM}_a(\langle X^P, w\rangle) + \kappa\sqrt{1+a^2}||w||_2. \tag{39}$$

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

$$Z(\omega) = \begin{cases} \frac{1}{1-\alpha}, & \omega \in A \\ 0, & \text{otherwise} \end{cases} \in \partial\, \mathrm{CVaR}_\alpha(X). \tag{40}$$

Hence, condition (30) is fulfilled, and $||Z||_\infty = \frac{1}{1-\alpha}$, 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

$$h_\lambda(p) = \frac{dH_\lambda(p)}{dp} = \exp\left(\frac{-2\lambda\Phi^{-1}(1-p) - \lambda^2}{2}\right). \tag{41}$$

We then compute

$$||h_\lambda||_2^2 = \int_0^1 \exp\left(-2\lambda\Phi^{-1}(1-p) - \lambda^2\right) dp = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-2\lambda x - \lambda^2 - \frac{x^2}{2}\right) dx \tag{42}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x+2\lambda)^2 - 2\lambda^2}{2}\right) dx = e^{\lambda^2} \tag{43}$$

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

The domain of the robustified power distortion, the robustified Gini measure, and the robustified expectation corrected deviation from the median is smaller than the domain of the non-robustified risk measures – in particular, they are no longer defined on $L^1(\Omega, \mathcal{F}, \mu; \mathbb{R})$. The reason lies in the more restrictive, pointwise conditions on the subgradients for the case $p = 1$ in Proposition 1, which is not fulfilled for the subgradients of the respective measures. We therefore define the robust measures on $L^2(\Omega, \mathcal{F}, \mu; \mathbb{R})$, 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 norm-constrained portfolio optimization and the robustifications we propose. In particular, we show that a norm-constrained problem proposed in DeMiguel et al. (2009a) is equivalent to the problem of minimizing the robustified standard deviation defined herein.

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

$$\mathrm{Std}(X) = ||X - \mathbb{E}(X)||_2 = \sup \left\{ \mathbb{E}\left[(X - \mathbb{E}(X))Z \right] : ||Z||_{L^2} = 1 \right\}. \tag{44}$$

Clearly, the maximizer Z^* in (44) is equal to

$$Z^* = \frac{X - \mathbb{E}(X)}{||X - \mathbb{E}(X)||_2}. \tag{45}$$

Becasue $\mathbb{E}(Z^*) = 0$, we can rewrite (44) as

$$\begin{aligned} \mathrm{Std}(X) &= \sup \left\{ \mathbb{E}\left[X(Z - \mathbb{E}(Z)) \right] : ||Z||_{L^2} = 1,\ \mathbb{E}(Z) = 0 \right\} && (46) \\ &= \sup \left\{ \mathbb{E}\left[XZ \right] : ||Z||_{L^2} = 1,\ \mathbb{E}(Z) = 0 \right\}. && (47) \end{aligned}$$

This representation is of the form (11), and because translation equi-variance is not used in the proof of Proposition 1, we can write the robustified standard deviation as

$$\mathrm{Std}^{\kappa}(P, w) = \mathrm{Std}(\langle X^P, w \rangle) + \kappa ||w||_2. \tag{48}$$

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

$$\mathrm{Var}(X) = ||X - \mathbb{E}(X)||_2^2 = \sup \left\{ \mathbb{E}(XZ) - \frac{1}{4}\mathrm{Var}(Z) : \mathbb{E}(Z) = 0 \right\} \tag{49}$$

and $Z^* = 2(X - \mathbb{E}(X))$, with $||2(X - \mathbb{E}(X))||_2 = 2\,\mathrm{Std}(X)$. That is 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 we demonstrated previously.

Inspecting the robustified risk measures, we note that the robustification is achieved by *penalizing* by the corresponding dual norm of the portfolio, multiplied by a constant (the Lipschitz constant of the risk measure with respect to the Kantorovich distance). 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, we show that minimizing (48) is equivalent to solving a 2-norm-constrained Markowitz problem. Specifically, 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

$$\begin{array}{ll} \min_w & w^\top \Sigma w \\ s.t. & \langle w, \mathbb{1} \rangle = 1 \\ & ||w||_2 \leq c, \end{array} \tag{50}$$

and show the following

Proposition 3. *For every $c \geq 1/\sqrt{N}$, there exists a κ, such that* (50) *is equivalent to*

$$\begin{array}{ll} \min_w & w^\top \Sigma w + \kappa ||w||_2 \\ s.t. & \langle w, \mathbb{1} \rangle = 1. \end{array} \tag{51}$$

Conversely, for every κ, there exists a c such that (51) *is equivalent to* (50).

Proof. Because $\min \{||w||_2 : \langle w, \mathbb{1} \rangle = 1\} = 1/\sqrt{N}$, problem (50) is feasible for $c \geq 1/\sqrt{N}$. A portfolio w is optimal for (50), iff the following KKT conditions hold for some ν and η

$$\begin{aligned} 2\Sigma w + \eta \mathbb{1} + 2\nu w &= 0 \\ \langle w, \mathbb{1} \rangle &= 1 \\ ||w||_2 &\leq c \\ \nu &\geq 0. \end{aligned} \tag{52}$$

Similarly optimality for (51) is equivalent to the existence of a η, such that

$$\begin{aligned} 2\Sigma w + \eta \mathbb{1} + 2\kappa w &= 0 \\ \langle w, \mathbb{1} \rangle &= 1 \end{aligned} \tag{53}$$

Clearly, conditions (52) imply (53) with $\nu = \kappa$, and (53) implies (52) with $c = ||w||_2$. □

The underlying principle can be extended to other measures of risk, which demonstrates the equivalence of the robust optimization approach we have pursued with models that penalize high norms of the portfolio.

3.2 Soft robust constraints

The motivating problem (8) 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 to formulate constraints for arbitrary stochastic optimization problems. For a fixed reference measure $\hat{P}$, we consider a robustified problem of the form

$$\begin{array}{ll} \inf_{w \in \mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w \rangle) \\ s.t. & \mathcal{R}(\langle X^Q, w \rangle) \leq \beta, \quad \forall Q \in \mathcal{B}^p_\kappa(\hat{P}) \\ & w \in \mathcal{W}. \end{array} \tag{54}$$

for a convex risk measure that fulfills the necessary subgradient conditions. Using notation from the previous sections, we can rewrite (54) as the following convex problem:

$$\begin{array}{ll} \inf_{w \in \mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w \rangle) \\ s.t. & \mathcal{R}^\kappa(\hat{P}, w) \leq \beta \\ & w \in \mathcal{W}. \end{array} \tag{55}$$

For $\kappa = 0$, problem (55) reduces to a nominal instance of a classical mean risk problem. However, for a given $\kappa > 0$ the constraint must be fulfilled for all distributions $Q \in \mathcal{B}^p_\kappa(\hat{P})$, regardless of their distance from $\hat{P}$. Although this is standard for robustifying constraints, we might argue that it 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 some measures whose distance 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

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w\rangle) \\ s.t. & \mathcal{R}(\langle X^{Q}, w\rangle) \leq f(\kappa), \quad \forall Q \in \mathcal{B}^p_\kappa(\hat{P}),\ \forall \kappa \in [0,\delta] \\ & w \in \mathcal{W}, \end{array} \tag{56}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a convex function. These authors choose $f(\kappa) = \kappa$ and solve the resulting problem by iteratively solving *standard robust* problems with the entropy distance as as a notion of distance between probability measures.

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

$$\max_{\kappa\in[0,\delta]} \mathcal{R}^\kappa(\hat{P}, w) \leq f(\kappa), \tag{57}$$

or equivalently,

$$\mathcal{R}(\langle X^{\hat{P}}, w\rangle) + \max_{\kappa\in[0,\delta]} \{\kappa C||w||_q - f(\kappa)\} \leq 0. \tag{58}$$

Because f is convex, it turns out that we can find one κ^*, such that the infinitely many constraints in (56) can be replaced by a single one. We have either

$$\max_{\kappa\in[0,\delta]} \{\kappa C||w||_q - f(\kappa)\} = \delta C||w||_q - f(\delta), \tag{59}$$

i.e. the boundary solution $\kappa^* = \delta$, or the maximum is given by the first-order condition

$$C||w||_q - \frac{\partial f}{\partial \kappa} = 0. \tag{60}$$

We choose $\delta = \infty$ and $f(\kappa) = d\kappa^2 + \beta$, which leads to $\kappa^* = \frac{C||w||_q}{2d}$. Consequently, (56) becomes

$$\begin{array}{ll} \inf_{w\in\mathbb{R}^N} & \mathbb{E}(\langle X^{\hat{P}}, w\rangle) \\ s.t. & \mathcal{R}(\langle X^{\hat{P}}, w\rangle) + \frac{C^2||w||_q^2}{4d} \leq \beta, \\ & w \in \mathcal{W}. \end{array} \tag{61}$$

In general, problem (61) 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 an arbitrary convex polynomial, for example. The chosen quadratic 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 all measures into account 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 prior 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 notation and selection of data sets both are motivated by a similar out-of-sample analysis performed by DeMiguel et al. (2009a).

We test the portfolio selection rule 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 also 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. (2011) 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.

Accordingly, this section comprises four subsections: the setup of the rolling horizon study, followed by the data sets used to conduct the study, as well as the parameter choice for the different portfolio selection rules. The third section briefly touches on how to choose the parameter κ for the robustified policies. Finally, we offer a discussion of 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 $(x_1, \dots, x_L)$ 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 $(x_1, \dots, x_L)$, 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-base 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

$$\begin{array}{ll} \inf_{w \in \mathbb{R}^N} & \mathcal{R}^{\kappa}(\hat{P}, w) \\ s.t. & \langle w, \mathbb{1} \rangle = 1 \end{array} \tag{62}$$

for the risk and deviation measures mentioned previously. We compare the results for $\kappa = 0$, which is the nominal case, with the results for $\kappa > 0$, i.e. the robustified case. See Section 4.3 for a discussion of the choice of κ.

In practice, a portfolio manager would impose many more restrictions on feasible portfolio weights than (62). 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, such as short-selling constraints or constraints on the maximum size of single positions.

We use three performance criteria to assess the quality of a portfolio selection rule: the risk $(l_{L+1}, \dots, l_T)$, the expected losses, and the average turnover. 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 $t+1$,

$$w_t^+ = \frac{w_t \odot (1 - l_t)}{\langle w_t, (1 - l_t) \rangle}, \tag{63}$$

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

$$\text{turnover} = \frac{1}{T - L - 1} \sum_{t=L+1}^{T-1} \langle |w_t^+ - w_{t+1}|, \mathbb{1} \rangle. \tag{64}$$

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.

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.

4.2 Input data

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 consist of monthly returns, and all data sets start in July 1963 and end in December 2010, such that each data set consists of $T = 570$ data points. The number of assets ranges from 6 to 45, so that they represent 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 or approximately 10 full years in the past, respectively. Consequently, we obtain 450 and 945 portfolio decisions and realized losses on which to base our analysis.

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 policies 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.

The numerical value of κ has no immediate interpretation. Let $\kappa^* > 0$ be such that the optimal portfolio is equal to the 1/N strategy (cf. Pflug et al., 2011). Reasonable choices for κ range from 0 to κ^* and increasing κ can mitigate the effects of estimation error that results 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 small estimation error and small bias – a situation reminiscent of many statistical procedures.

In choosing κ, it is theoretically possible to work directly with a ratio κ/κ^* between 0 and 1. Alternatively, we consider three different ways to choose κ.

1. The Kantorovich distance has a close connection to the weak convergence of probability measures. Using the empirical measure $\hat{P}_n$ as a reference measure, we can employ existing finite sample versions of the Glivenko-Cantelli theorem, formulated in terms of the Kantorovich distance. Examples of such bounds can be found for example in Bolley et al. (2007). This approach makes it possible to interpret the ambiguity set around the empirical measures as a *confidence ball* in which the real measure lies with a certain probability. However, though some bounds extant literature are of the exponential type, they are generally too loose to be useful in a practical context, especially for problems with many assets.

2. Similar DeMiguel et al. (2009a), we could choose the parameter κ to coincide with the parameter that would have worked best in the past few periods.

3. The notion of calmness of the stochastic process of asset losses can be quantified by measuring the Kantorovich distance $d_p(\hat{P}_1, \hat{P}_2)$ between two subsamples $\hat{P}_1$ and $\hat{P}_2$ of points that constitute the empirical measure $\hat{P}$. For example, we might use the and the second halves of the sample as $\hat{P}_1$ and $\hat{P}_2$, respectively. This approach would offer a measure of the representativeness of data from past observations for future realizations. The parameter κ then can be chosen as a fraction of $d_p(\hat{P}_1, \hat{P}_2)$.

In our examples, we use the third method and choose $\kappa = 0.01 \times d_p(\hat{P}_1, \hat{P}_2)$; this relatively low factor of 0.01 proved beneficial in numerical tests.

4.4 Results

We start our discussion with the results of the 1/N strategy, which serve, together with the non-robustified measures, as a benchmark for the robustified risk measures. If the description of uncertainty is sufficiently accurate, the 1/N strategy should not outperform the portfolios obtained by the respective non-robustified optimization approaches. Table 2 lists the average losses, the turnover, and the risks of the 1/N strategy, measured with the four risk measures.

Table 3 shows the out-of-sample risk, calculated from the rolling horizon study for the different asset universes in Table 1. The reported figures are calculated for the set of out-of-sample

	Return	Turnover	S_1	$\text{CVaR}_{0.95}$	DM_2	Std
10Ind	0.0100	0.0415	0.0329	0.0990	0.0537	0.0428
48Ind	0.0099	0.0498	0.0379	0.1121	0.0599	0.0478
6FF	0.0101	0.0407	0.0380	0.1142	0.0608	0.0481
25FF	0.0105	0.0426	0.0391	0.1167	0.0631	0.0497
SP100	0.0030	0.0327	0.0209	0.0562	0.0305	0.0239

Table 2: Risks, returns, and turnover for the 1/N strategy.

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 policies. Furthermore, 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 consider two quantities significantly different if the p value of the bootstrapping test is less than 0.1.

Comparing the values in Table 3 with the risks reported in Table 2, 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 10Ind, the robustified risk measures are better, but the differences are rather small and in some cases not significant. For 6SBM, the non-robustified risk measures fare slightly better than the robustified measures, and in three of four cases, the difference is significant. These results indicate that for a small set of assets, the information encoded in $\hat{P}$ is accurate to a degree that the gains from robustification are smaller than the losses that result from the distortion of the objective function in the robustified problem. For larger sets of assets, this effect is reverses – most prominently for 100SP.

Although maximizing returns was not the goal of this experiment, we report the expected out-of-sample returns in Table 4. The comparison between the robustified and non-robustified measures yields ambiguous results. In some cases, the returns for the non-robustified measures

	Nominal				Robustified			
	S_1	$\text{CVaR}_{0.95}$	DM_2	Std	S_1^{κ}	$\text{CVaR}_{0.95}^{\kappa}$	DM_2^{κ}	Std^{κ}
10Ind	0.0275	0.0887	0.0463	0.0375	**0.0271**	**0.0812**	**0.0457**	**0.0372**
p-value					0.0500	0.0700	0.0000	0.0000
48Ind	0.0378	0.1187	0.0551	0.0383	**0.0318**	**0.0866**	**0.0427**	**0.0367**
p-value					0.0000	0.0000	0.0000	0.0000
6FF	**0.0239**	**0.0850**	**0.0482**	**0.0406**	0.0244	0.0970	0.0492	0.0409
p-value					0.1100	0.0000	0.0000	0.0600
25FF	0.0222	0.1104	0.0485	0.0370	**0.0196**	**0.0844**	**0.0466**	**0.0366**
p-Value					0.0000	0.0700	0.0000	0.0900
SP100	0.0189	0.0587	0.0273	0.0189	**0.0168**	**0.0449**	**0.0241**	**0.0180**
p-value					0.0000	0.0000	0.0000	0.0000

Table 3: Risks for the non-robustified and robustified measures and p-values for the significance of the difference between the two values. To facilitate the direct comparison between the robustified and non-robustified measures is, this table presents the better values in bold font.

are slightly higher, while in others, it is the other way around. For portfolios with more assets, the results either do not differ significantly, as for SP100, or are better for the robustified portfolio selection criteria. For the smaller asset universes, this effect reverses. In many cases though, the results are not significant. Comparing the returns to those from the 1/N strategy, we find that the robust policies outperform the 1/N strategy more often then do the non-robustified policies. The only data sets for which the 1/N rule yields higher returns than the robustified policies are 48Ind and 100SP, and then only for some risk measures. Notably, in the later data set, the 1/N policy consistently outperforms all other strategies in terms of out-of-sample returns.

The situation is clearer for the turnovers, presented in Table 5: 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, such that they incur more transaction costs than portfolios found using the original measures. Unsurprisingly, the 1/N policy significantly outperforms all other policies in terms of turnover.

In summary, the robustified portfolio policies perform very well for data sets with more than 20 assets. For data sets with fewer assets, the results are mixed, because the approximation of the data generating process by $\hat{P}$ seems more accurate.

	Nominal				Robustified			
	S_1	$\text{CVaR}_{0.95}$	DM_2	Std	S_1^κ	$\text{CVaR}_{0.95}^\kappa$	DM_2^κ	Std^κ
10Ind	**0.0109**	**0.0121**	**0.0112**	**0.0111**	0.0107	0.0102	0.0108	0.0109
p-value					0.0500	0.0900	0.2600	0.0400
48Ind	0.0091	0.0084	0.0064	0.0073	**0.0097**	**0.0109**	**0.0106**	**0.0086**
p-value					0.4200	0.3200	0.0000	0.0000
6FF	**0.0176**	**0.0170**	**0.0135**	**0.0151**	0.0170	0.0124	0.0131	0.0148
p-value					0.0000	0.0000	0.1000	0.0000
25FF	**0.0228**	**0.0241**	0.0120	**0.0162**	0.0206	0.0149	**0.0122**	0.0156
p-Value					0.0000	0.0000	0.8400	0.0000
SP100	0.0013	0.0012	0.0026	0.0018	**0.0014**	**0.0019**	**0.0026**	**0.0019**
p-value					0.5000	0.2700	0.9200	0.6300

Table 4: Return for the unrobustified and the robustified measure as well as p-values for the significance of the difference between the two values. The direct comparison between the robustified and unrobustified measures is facilitate by the boldface print of the respective better values.

5 Conclusion and further work

We offer 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 with most other approaches dealing with model uncertainty, the ambiguity sets are constructed without any assumptions about the membership of the true distribution in any parametric family, such that the ambiguity sets are fully non-parametric. Despite this constriction and 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 used as objective functions or constraints in portfolio optimization problems just as easily as their non-robust counterparts can. We also have provided a numerical study showing that the robustified portfolio selection problems usually seem to yield better results than the non-robustified policies, 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.

	Nominal				Robustified			
	S_1	$\text{CVaR}_{0.95}$	DM_2	Std	S_1^κ	$\text{CVaR}_{0.95}^\kappa$	DM_2^κ	Std^κ
10Ind	0.1257	0.2300	0.2004	0.1050	**0.1054**	**0.1619**	**0.1317**	**0.0939**
p-value					0.0000	0.0000	0.0000	0.0000
48Ind	0.5751	1.1078	0.7472	0.3932	**0.3408**	**0.2418**	**0.1627**	**0.2893**
p-value					0.0000	0.0000	0.0000	0.0000
6FF	0.2377	0.3064	0.2118	0.1508	**0.1950**	**0.1764**	**0.1385**	**0.1283**
p-value					0.0000	0.0000	0.0000	0.0000
25FF	0.7881	1.2438	0.7528	0.4445	**0.5403**	**0.2929**	**0.2091**	**0.3466**
p-Value					0.0000	0.0000	0.0000	0.0000
SP100	0.2492	0.5719	0.3951	0.1859	**0.1385**	**0.1024**	**0.0469**	**0.1288**
p-value					0.0000	0.0000	0.0000	0.0000

Table 5: Turnover for non-robustified and robustified measures, as well as p-values for the significance of the difference between the two values. To facilitate a direct comparison between the robustified and non-robustified measures, this table presents the better values in bold font.

Another interesting topic for further research is the choice of the robustness parameter κ. Altough the method applied in the numerical examples seems to work quite well, a more structured approach would be desirable.

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.

F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137:541–593, 2007.

M. Broadie. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45:21–58, 1993.

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

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.

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.

J. Dupačová. On minimax decision rule in stochastic linear programming. In *Studies on mathematical programming (Papers, Third Conf. Math. Programming, Mátrafüred, 1975)*, volume 1 of *Math. Methods Oper. Res.*, pages 47–60. Akad. Kiadó, Budapest, 1980.

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.

A.L. Gibbs and F.E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 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*, In Press, Corrected Proof:–, 2011.

H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118(2, Ser. A):371–406, 2009.

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.

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

R.O. Michaud. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1).

R. Mirkov and G. Ch. Pflug. Tree approximations of dynamic stochastic programs. *SIAM Journal on Optimization*, 18(3):1082–1105, 2007.

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*, In Press, Corrected Proof:–, 2011.

A. Ruszczynski 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*, volume 58 of *Graduate Studies in Mathematics.* 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.

R.E. Welsch and X. Zhou. Application of robust statistics to asset allocation models. *REVSTAT – Statistical Journal*, 5(1):97–114, 2007.

D. Wozabal. A framework for optimization under ambiguity. *Annals of Operations Research*, page In Print, 2010.

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.