Tractable Counterparts of Distributionally Robust Constraints on Risk Measures

Postek, K.S.; den Hertog, Dick; Melenberg, Bertrand

Document version:
Early version, also known as pre-print

Publication date:
2014

[Link to publication](#)

Citation for published version (APA):
Postek, K. S., den Hertog, D., & Melenberg, B. (2014). Tractable Counterparts of Distributionally Robust Constraints on Risk Measures. (CentER Discussion Paper; Vol. 2014-031). Tilburg: Operations research.

TRACTABLE COUNTERPARTS OF DISTRIBUTIONALLY ROBUST CONSTRAINTS ON RISK MEASURES

By

Krzysztof Postek, Dick den Hertog,
Bertrand Melenberg

TILBURG ◆ UNIVERSITY

Tractable counterparts of distributionally robust constraints on risk measures

Krzysztof Postek[*][†]   Dick den Hertog[*]   Bertrand Melenberg[*]

May 9, 2014

Abstract

In this paper we study distributionally robust constraints on risk measures (such as standard deviation less the mean, Conditional Value-at-Risk, Entropic Value-at-Risk) of decision-dependent random variables. The uncertainty sets for the discrete probability distributions are defined using statistical goodness-of-fit tests and probability metrics such as Pearson, likelihood ratio, Anderson-Darling tests, or Wasserstein distance. This type of constraints arises in problems in portfolio optimization, economics, machine learning, and engineering. We show that the derivation of a tractable robust counterpart can be split into two parts: one corresponding to the risk measure and the other to the uncertainty set. We also show how the counterpart can be constructed for risk measures that are nonlinear in the probabilities (for example, variance or the Conditional Value-at-Risk). We provide the computational tractability status for each of the uncertainty set-risk measure pairs that we could solve. Numerical examples including portfolio optimization and a multi-item newsvendor problem illustrate the proposed approach.

Keywords: risk measure, robust counterpart, nonlinear inequality, robust optimization, support functions

JEL codes: C61

1   Introduction

Robust Optimization (RO, see Ben-Tal et al. (2009)) has become one of the main approaches to optimization under uncertainty. A particular application field is keeping risk measures of decision-dependent random variables below pre-specified limits, for instance, in finance, engineering, and economics. Often, the computation of the value of a risk measure requires knowledge of the underlying probability distribution, which is usually approximated by an estimate. Such an estimate is typically

---

[*]CentER and Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[†]Correspondence to: k.postek@tilburguniversity.edu

based on a number of past observations. Due to sampling error, this estimate approximates the true distribution only with a limited accuracy. The confidence set around the estimate gives rise to a natural uncertainty set of admissible probability distributions at a given confidence level. Robustness against this type of distributional uncertainty is the topic of this paper. We derive computationally tractable robust counterparts of constraints on a number of risk measures for various types of statistically-based uncertainty sets for discrete probability distributions.

The contribution of our paper is threefold. First, using Fenchel duality and results of Ben-Tal et al. (2012) we show that the derivation of components corresponding to the risk measure and the uncertainty set can be separated. Therefore, we derive two types of building blocks: one for the risk measures and another for the uncertainty sets. The resulting blocks may be combined arbitrarily according to the problem at hand. This allows us to cover many more risk measure-uncertainty set pairs than is captured up to now in the literature. The first building block includes negative mean return, Optimized Certainty Equivalent (with Conditional Value-at-Risk as a special case), Certainty Equivalent, Shortfall Risk, lower partial moments, mean absolute deviation from the median, standard deviation/variance less mean, Sharpe ratio, and the Entropic Value-at-Risk. The second building block encompasses uncertainty sets built using the ϕ-divergences (with the Pearson (χ^2) and likelihood ratio (G) tests as special cases), Kolmogorov-Smirnov test, Wasserstein (Kantorovich) distance, Anderson-Darling, Cramer-von Mises, Watson, and Kuiper tests.

The second contribution is dealing with the nonlinearity of several risk measures in the underlying probability distribution, including the variance, the standard deviation, the Optimized Certainty Equivalent, and the mean absolute deviation from the median. To make the use of RO methodology possible, we provide equivalent formulations of such risk measures as infimums over relevant function sets, whose elements are linear in the probabilities. A minmax result from convex analysis ensures that this operation results in an exact reformulation. For the Conditional Value-at-Risk such an approach has been applied in [33], with uncertainty sets different from the ones we consider.

As a third contribution we provide the complexity status (linear, convex quadratic, second-order conic, convex) of the robust counterparts. This is summarized in Table 1, together with a summary of the results captured in the literature up to now. As illustrated, our methodology allows for obtaining a tractable robust counterpart for most of the risk measure-uncertainty set combinations, extending the results in the field significantly.

For several types of risk measures, including the Value-at-Risk, the mean absolute deviation from the mean, and general-form distortion, coherent and spectral risk measures, we could not derive a tractable robust counterpart using our methodology. This can be seen as an indication to be careful when using them, since inability to take into account distributional uncertainty - a natural phenomenon in real-life applications - would make these risk measures less trustworthy. Otherwise, one would need to argue that such a measure is itself robust to uncertainty in the

probability distribution or use the (approximate) results obtained by other authors.

The combination of uncertain discrete probabilities and risk measures has already been investigated by several authors. Calafiore (2007) uses a cutting plane algorithm to find the optimal mean-variance and mean-absolute deviation from the mean portfolios under uncertainty specified with the Kullback-Leibler divergence. Huang et al. (2010) find the optimal worst-case Conditional Value-at-Risk under a multiple-expert uncertainty set for the probability distribution. Zhu and Fukushima (2009) provide robust constraints on the Conditional Value-at-Risk for the box and ellipsoidal uncertainty sets. Fertis et al. (2013) show how a constraint on the Conditional Value-at-Risk can be reformulated to a tractable form under generic norm uncertainty about the underlying probability measure. Pichler (2013) finds the worst-case probability measures for the negative mean of the return, the Conditional Value-at-Risk and distortion risk measures with the uncertainty set defined using the Wasserstein distance.

Wozabal (2012) combines a so-called subdifferential representation of risk measures with a Wasserstein-based uncertainty set for (discrete or continuous) probability measures, corresponding to the subdifferential representation, to derive closed-form worst-case values of risk measures. Ben-Tal et al. (2012) give results allowing to obtain constraints for the variance with ϕ-divergence- and Anderson-Darling-defined uncertainty sets. Hu et al. (2013a) develop a convex programming framework for the worst-case Value-at-Risk with uncertainty sets defined by ϕ-divergence functions. However, they do not obtain closed forms of the robust constraints. Jiang and Guan (2013) develop an efficient reformulation of ambiguous chance constraints with uncertainty defined using the Kullback-Leibler divergence. It reduces the chance-constrained problem to a problem under the nominal probability measure. Hu et al. (2013b) provide closed-form distributionally robust counterparts of constraints with a Kullback-Leibler defined uncertainty set for the probability distributions, both discrete and continuous. Wang et al. (2013) derive tractable counterparts of constraints involving linear functions of the probability vector, with uncertainty defined by the likelihood ratio test. Klabjan (2013) solves a lot-sizing problem with uncertainty defined with the χ^2 test statistic.

Natarajan et al. (2009) study the correspondence between risk measures and uncertainty sets for probability distributions, showing how risk measures can be constructed from uncertainty sets for distributions. Bazovkin and Mosler (2012) construct a geometrically-based method for solving robust linear programs with a single distortion risk measure under polytopial uncertainty sets. It is not known yet whether their results can be extended to the statistically-based uncertainty sets for probabilities because of the representation of polytopes. Bertsimas et al. (2013) construct uncertainty sets defined by statistical tests such as Kolmogorov-Smirnov, χ^2, Anderson-Darling, Watson and likelihood ratio, to obtain tight bounds on the Value-at-Risk. They utilize a cutting plane algorithm with an efficient method of evaluating the worst-case values of the decision-dependent random variables. A separate work giving tractable robust counterparts of uncertain inequalities with ϕ-divergence uncertainty, not focusing on risk measures, is Ben-Tal et al. (2013).

Table 1: Results on complexity of a tractable counterpart for risk measures and uncertainty sets. The symbol ● means that a tractable robust counterpart has been formulated in the literature and the symbol ○ means that only a partial solution was found in the literature, e.g., an efficient method of evaluating the worst-case values. The complexity symbols are: LP - linear constraints, QP - convex quadratic, SOCP - second-order conic, CP - convex. The symbol ∗ means that the right-hand side in a constraint (β in constraint (1)) must be a fixed number for the counterpart to be a system of convex constraints. The results are constructed assuming that the decision-dependent random variable $X(w)$ is linear in the decision vector w (see Section 2).

Risk measure / Uncertainty set type	ϕ-divergences	Pearson	Likelihood ratio	Kolmogorov-Smirnov	Wasserstein (Kantorovich)	Anderson-Darling	Cramer-von Mises	Watson	Kuiper
Negative mean return	● CP [6],[31]	● SOCP [6],[21],[31]	● CP [6],[18],[30],[31]	LP	● LP [24],[32]	● CP [5]	SOCP	SOCP	LP
Optimized Certainty Equivalent	CP	CP	CP	CP	CP	CP	CP	CP	CP
Conditional Value-at-Risk	○ CP [31]	● SOCP,[7],[31]	● CP,[7],[18],[31]	○ LP	● LP [24],[32]	○ CP	○ SOCP	○ SOCP	LP
Certainty Equivalent	CP∗	CP∗	CP∗	CP∗	CP∗	CP∗	CP∗	CP∗	CP∗
Shortfall risk	CP	CP	CP	CP	CP	CP	CP	CP	CP
Lower partial moment $\alpha = 1$	○ CP [31]	○ SOCP [31]	○ CP [31]	LP	LP	CP	SOCP	SOCP	LP
Lower partial moment $\alpha = 2$	CP	SOCP	CP	QP	QP	CP	SOCP	SOCP	QP
Mean absolute deviation from the median	○ CP [31]	○ SOCP [31]	○ CP [31]	LP	● LP [32]	CP	SOCP	SOCP	LP
Standard deviation less the mean	CP	SOCP	CP	SOCP	● LP [32]	CP	SOCP	SOCP	SOCP
Standard deviation	CP	SOCP	CP	SOCP	● LP [32]	CP	SOCP	SOCP	SOCP
Variance less the mean	○ CP [5]	○ SOCP [5]	○ CP [5],[8]	QP	QP	○ CP [5]	SOCP	SOCP	QP
Variance	○ CP [5]	○ SOCP [5]	○ CP [5],[8]	QP	QP	○ CP [5]	SOCP	SOCP	QP
Sharpe ratio	CP∗	SOCP∗	CP∗	SOCP∗	SOCP∗	CP∗	SOCP∗	SOCP∗	SOCP∗
Entropic Value-at-Risk	CP	CP	CP	CP	CP	CP	CP	CP	CP
Value-at-Risk	○ [10],[17]	○ [7],[10],[17]	○ [7],[10],[17],[18],[20]	○ [7]		○ [7]	○ [7]	○ [7]	○ [7]
Mean absolute deviation from the mean			○ [8]						
Distortion risk measures					● [24], [32]				
Coherent risk measures									
Spectral risk measures									

We also summarize the results with another type of uncertainty - namely in terms of the moments of the underlying random variables. This approach is frequently used since in finance, responsible for a large number of papers, it is common to specify the uncertainty in terms of the moments of asset returns. This is consistent with the type of portfolio optimization that seeks the best tradeoff between the expected return on a portfolio and its riskiness defined by the variance. Problems of such type are analyzed in, for example, [13], [15], and [29]. Worst-case bounds on the (Conditional) Value-at-Risk of random variables whose mean and variance reside within a given uncertainty set are studied, for example, in [9], [10], and [34]. A mixed type of uncertainty is studied in Wiesemann et al. (2013), who optimize the worst-case expectations of piecewise linear functions of random variables under uncertainty about the moments and probability masses corresponding to conic sets of events.

The composition of the remainder of the paper is as follows. Section 2 introduces the definitions and the main tool for deriving the computationally tractable robust counterparts. Section 3 lists the risk measures and uncertainty sets for the probability distribution that we investigate. Sections 4 and 5 include the key contributions of the paper - the results on the building blocks of the robust counterparts of the constraints on the risk measures. In Section 6, numerical examples of connecting the blocks are given. Section 7 concludes and lists the potential directions for future research.

2 Preliminaries

We study constraints on risk measures of decision-dependent random variables, where $w \in \mathbb{R}^M$ is the decision vector. The decision-dependent random variable $X(w)$, whose risk is measured, takes a value $X_n(w)$ with probability p_n for each $n \in \mathcal{N} = \{1, \ldots, N\}$. We assume that $X_n(w) = V(Y^n, w)$, where $Y^n \in \mathbb{R}^{M_Y}$ is the n-th possible outcome of the underlying random vector and $V : \mathbb{R}^{M_Y} \times \mathbb{R}^M \to \mathbb{R}$ is a function defining the dependency of X on w and Y. The uncertain parameter is the discrete probability vector $p = [p_1, \ldots, p_N]^T \in \mathbb{R}_+^N$. The reference probability vector, around which the uncertainty set for p may be specified, is denoted by $q \in \mathbb{R}_+^N$.

The constraint we shall reformulate to a tractable form is:

$$F(p, w) = \overline{F}(p, X(w)) \le \beta, \quad \forall p \in \mathcal{P}, \tag{1}$$

where $F : \mathbb{R}_+^N \times \mathbb{R}^M \to \mathbb{R}$ is a function determined by the risk measure and $\mathcal{P}$ is the uncertainty set for the probabilities defined as:

$$\mathcal{P} = \{p : \quad p = Ap', \quad p' \in \mathcal{U}\}, \tag{2}$$

where the set $\mathcal{U} \in \mathbb{R}^L$ is compact and convex and $A \in \mathbb{R}^{N \times L}$ such that $\mathcal{P} \in \mathbb{R}_+^N$. Formulation of the set $\mathcal{P}$ using the matrix A is general and encompasses cases where the set $\mathcal{U}$ has a dimension different from N.

Example 1. If the risk measure of the random variable $X(w)$ is the variance and the uncertainty set is defined with a ϕ-divergence function around the reference probability vector q (see Table 3), then the constraint is:

$$F(p, w) = \overline{F}(p, X(w)) = \sqrt{\sum_{n \in \mathcal{N}}^{n} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2} \leq \beta, \quad \forall p \in \mathcal{P},$$

with $A = I$ and

$$\mathcal{P} = \mathcal{U} = \left\{ p \geq 0 : \quad \sum_{n \in \mathcal{N}} p_n = 1, \quad \sum_{n \in \mathcal{N}} q_n \phi \left(\frac{p_n}{q_n} \right) \leq \rho \right\}. \qquad \square$$

To introduce the key theorem used in this paper we give first the definitions of the concave conjugate and the support function. The concave conjugate $f_*(.)$ of a function $f : \mathbb{R}_+^N \to \mathbb{R}$ is defined as:

$$f_*(v) = \inf_{p \geq 0} \left\{ v^T p - f(p) \right\}. \tag{3}$$

The support function $\delta^*(.|\mathcal{U}) : \mathbb{R}^L \to \mathbb{R}$ of a set $\mathcal{U}$ is defined as:

$$\delta^*(v|\mathcal{U}) = \sup_{p' \in \mathcal{U}} v^T p'. \tag{4}$$

The following theorem, adapted from [5], is the main tool for deriving the tractable robust counterparts in this paper.

Theorem 1. Let $f : \mathbb{R}_+^N \times \mathbb{R}^M \to \mathbb{R}$ be a function such that $f(., w)$ is closed and concave for each $w \in \mathbb{R}^M$. Consider a constraint of the form:

$$f(p, w) \leq \beta, \quad \forall p \in \mathcal{P}, \tag{5}$$

where $\mathcal{P}$ is defined by (2) and where it holds that:

$$\mathrm{ri}(\mathcal{P}) \in \mathbb{R}_{++}^N \tag{6}$$

Then (5) holds for a given w if and only if:

$$\exists v \in \mathbb{R}^N : \quad \delta^* \left(A^T v \middle| \mathcal{U} \right) - f_*(v, w) \leq \beta, \tag{7}$$

where $\delta^*(.|\mathcal{U})$ is the support function of the set $\mathcal{U}$ and $f_*(., w)$ is the concave conjugate of $f(., w)$ with respect to its first argument. $\square$

For a proof we refer to [5], with an extra definition that $\mathrm{dom} f(., w) = \mathbb{R}_+^N$ for all w. Theorem 1 allows for a separation of the derivation of two components: (1) the support function of the set $\mathcal{U}$ at the point $A^T v$, corresponding to the uncertainty set, and (2) the concave conjugate of $f(., w)$, corresponding to the risk measure.

If the function $F(., w)$ in (1) satisfies the concavity assumption with respect to p and we can obtain its conjugate directly from (3), then we take $f(., w) = F(., w)$. If the concavity assumption is not satisfied or the standard form of $F(., w)$ is too difficult to obtain a tractable conjugate, then we choose another function $f(.)$ such that (1) and (5) are equivalent, and Theorem 1 can be used.

6

The next section gives the potential choices for the risk measures and the uncertainty set $\mathcal{P}$.

Notation

We distinguish the vectors by using the superscripts and the components of a vector using subscripts. For example, v_k^i denotes the k-th component of the vector v^i. Also, by the symbol $v_{s:t}$ we denote the subvector of v consisting of its components with indices s through t. Throughout the paper, 1 denotes a vector of ones, consistent in dimensionality with the equation at hand, 1^k is a vector with ones on its first k positions and zeros elsewhere , 1^{-k} is defined as the vector $1 - 1^k$, and e^k denotes a vector of zeros except a single 1 as the k-th component.

3 Risk measures and uncertainty sets

3.1 Risk measures

The risk measures we analyze are given in Table 2. Some of them measure dispersion of the random variable $X(w)$ around a given level, such as the standard deviation or the mean absolute deviation from the median. Other measures, like the Certainty Equivalent, measure the overall riskiness of an uncertain position $X(w)$. In their formulations, we follow the convention that 'the smaller the risk measure, the better':

$$(\forall n \in \mathcal{N}: \quad X_n(w_1) \geq X_n(w_2)) \quad \Rightarrow \quad F(p, w_1) \leq F(p, w_2).$$

As an example, the first risk measure is the negative mean return instead of its positive counterpart. This corresponds to a situation where $X(w)$ represents gains, not losses.

The collection of risk measures analyzed in this paper exhausts a large part of practical applications. Risk measures such as the negative mean return, Conditional Value-at-Risk (the negative of the average of the worst $\alpha\%$ outcomes of a random variable, here we use the inf-formulation from [26]), lower partial moments, variance/standard deviation less the mean, Sharpe ratio (proportion of the mean to the standard deviation), and Value-at-Risk (the $\alpha\%$-quantile of the distribution of a given random variable), are examples of risk measures usually linked to portfolio optimization. Another class of risk measures is related to economics and analysis of consumer behavior. Following [3], the Certainty Equivalent denotes the negative of the 'sure amount for which a decision maker remains indifferent to the outcome of random variable $X(w)$', and the shortfall risk is the minimum amount of additional resources needed to make the expected utility of a decision maker from his portfolio nonnegative. Risk measures are used also in engineering, where a standard deviation of some quantity cannot be greater than a given value, and in statistical learning, where one minimizes the so-called empirical risk in support vector machines.

A comment is needed for the Entropic Value-at-Risk. Its definition in Table 2 does not involve p, being instead a supremum over probability vectors $\tilde{p}$ in $\mathcal{P}_q$, constructed

around a vector q. In this case the vector q shall be subject to uncertainty within a set $\mathcal{Q}$ - see the 'combined uncertainty set' in Table 3. We have chosen this formulation to make the notation of the corresponding function $f(p, w)$ (derived in Section 4) consistent with the terminology of Theorem 1. The EVaR is an upper bound on the Value-at-Risk and the Conditional Value-at-Risk with the same α (for $p = q$ in their formulations in Table 2).

Some of the measures in Table 2 are specific cases of the other ones: for instance, the Conditional Value-at-Risk is both a coherent risk measure and an example of an Optimized Certainty Equivalent. Nevertheless, a distinction has been made because of the popularity of the use of some specific cases. Also, some results can be obtained only for specific cases and it is important to state why this is so, and what the consequences are for practical applications.

3.2 Uncertainty sets for the probabilities

Table 3 presents the uncertainty sets for the discrete probabilities analyzed here. For each case we give the constraints on the vector p that define the set. Using discrete probabilities allows the use of Theorem 1 and, if needed, continuous distribution information can be transformed into discrete distribution information using techniques given in [6]. We follow the view, motivated in [27], that the formulation of an uncertainty set for a probability distribution should be supported by results in statistics. An overview of statistical goodness-of-fit tests, being the source of such statistically-based uncertainty sets, can be found in [28].

Most of the sets in Table 3, including the Pearson, likelihood ratio, Kolmogorov-Smirnov, Anderson-Darling, Cramer-von Mises, or Kuiper sets, are constructed using goodness-of-fit test statistics with the corresponding names. The Pearson and likelihood ratio sets are specific cases of the ϕ-divergence set (obtained by choosing the Kullback-Leibler or the modified χ^2 divergences, respectively), but have been distinguished here for their popularity. Examples of functions $\phi(.)$ are given in Appendix B.1.

The Wasserstein set, defined using the Wasserstein (Kantorovich) distance between distribution vectors p and q, deserves a separate explanation. The distance between p and q, defined with the use of the inf term in Table 3, can be interpreted as a minimum transport cost of the probability mass from vector p (supply) to vector q (demand), where the unit cost between the i-th cell of p and the j-th cell of q is equal to $\|Y^i - Y^j\|^d$. This type of uncertainty is studied extensively in a robust setting in [32] and the statistical advantages of its use are motivated in [27].

A separate explanation is also needed for the 'combined uncertainty set'. Its definition in Table 3 says that $\mathcal{P}^C$ has a two-stage structure. First, the vector p belongs to a set $\mathcal{P}_q$ centered around a vector q. Then, the vector q is uncertain itself and belongs to a set $\mathcal{Q}$ defined using Q convex inequalities. This class of uncertainty sets has been introduced here to derive the tractable robust counterpart of constraint on the Entropic Value-at-Risk. In this paper we shall assume that $\mathcal{P}_q$ is defined as a ϕ-divergence set around q, as in the first row of Table 3.

Table 2: Risk measures analyzed in the paper. The term $\mathbb{E}^p$ denotes expectation with respect to the probability measure induced by the vector p and $G_{X(w)}$ denotes the distribution function of the random variable $X(w)$. We define the α-quantile of a distribution of $X(w)$ as $G_{X(w)}^{-1}(\alpha) = \inf\{\kappa \in \mathbb{R} : \mathbb{P}(X(w) \geq \kappa) \geq \alpha\}$. The utility functions $u(.)$ are assumed to be defined on the entire real line.

Risk measure	Formulation $F(p, w)$
Negative mean return	$-\mathbb{E}^p(X(w))$
Optimized Certainty Equivalent (OCE)	$\inf\limits_{\kappa \in \mathbb{R}} -\kappa - \mathbb{E}^p(u(X(w) - \kappa))$, $u(.)$ concave, nondecreasing
Conditional Value-at-Risk (CVaR)	$\inf\limits_{\kappa \in \mathbb{R}} -\kappa - \mathbb{E}^p\left(\frac{1}{\alpha}\min\{X(w) - \kappa, 0\}\right), \quad 0 < \alpha < 1$
Certainty Equivalent (CE)	$-u^{-1}\left(\mathbb{E}^p u(X(w))\right)$ $u(.)$ concave, invertible, with $-\frac{u'(t)}{u''(t)}$ concave
Shortfall risk	$\inf\{\kappa \in \mathbb{R} : \mathbb{E}^p(u(X(w)) + \kappa) \geq 0\}$ $u(.)$ concave
Lower partial moment	$\mathbb{E}^p(\max\{0, \bar{\kappa} - X(w)\}^\alpha)$ $\alpha = 1, 2, \quad \bar{\kappa}$ - any value
Mean absolute deviation from the median	$\mathbb{E}^p\left|X(w) - G_{X(w)}^{-1}(0.5)\right|$
Standard deviation less the mean	$\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2} - \alpha\mathbb{E}^p(X(w)), \quad \alpha \in \mathbb{R}$
Standard deviation	$\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2}$
Variance less the mean	$\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2 - \alpha\mathbb{E}^p(X(w)), \quad \alpha \in \mathbb{R}$
Variance	$\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2$
Sharpe ratio	$\frac{-\mathbb{E}^p(X(w))}{\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2}}$
Entropic Value-at-Risk (EVaR)	$\sup\limits_{\tilde{p} \in \mathcal{P}_q} \mathbb{E}^{\tilde{p}}(-X(w)), \quad 0 < \alpha < 1$ $\mathcal{P}_q = \left\{\tilde{p} : \tilde{p} \geq 0, 1^T\tilde{p} = 1, \sum\limits_{n \in \mathcal{N}} \tilde{p}_n \log\left(\frac{\tilde{p}_n}{q_n}\right) \leq -\log\alpha\right\}$
Value-at-Risk (VaR)	$-G_{X(w)}^{-1}(\alpha), \quad 0 < \alpha < 1$
Mean deviation from the mean	$\mathbb{E}^p|X(w) - \mathbb{E}^p(X(w))|$
Distortion risk measures	$\int_0^{+\infty} g\left(1 - G_{X(w)}(t)\right)dt, \; X(w) \text{ nonnegative}, \; g: [0,1] \to [0,1]$
Coherent risk measures	$\sup\limits_{\tilde{p} \in \mathcal{C}} \mathbb{E}^{\tilde{p}}(-X(w)), \quad \mathcal{C}$ - set of probability vectors
Spectral risk measures	$-\int_0^1 G_{X(w)}^{-1}(t)\psi(t)dt,$ $\psi(.)$ nonnegative, non-increasing, right-continuous, integrable

Table 3: Uncertainty set formulations for the probabilities vector p. In each case we assume that $p \geq 0, 1^T p = 1$ hold.

Set type	Formulation	Symbol
ϕ-divergence	$\sum_{n \in \mathcal{N}} q_n \phi \left(\frac{p_n}{q_n} \right) \leq \rho$	$\mathcal{P}_q^\phi$
Pearson (χ^2)	$\sum_{n \in \mathcal{N}} \frac{(p_n - q_n)^2}{q_n} \leq \rho$	$\mathcal{P}_q^{\mathrm{P}}$
Likelihood ratio (G)	$\sum_{n \in \mathcal{N}} q_n \log \left(\frac{p_n}{q_n} \right) \leq \rho$	$\mathcal{P}_q^{\mathrm{LR}}$
Kolmogorov-Smirnov	$\max_{n \in \mathcal{N}} \left| p^T 1^n - q^T 1^n \right| \leq \rho$	$\mathcal{P}_q^{\mathrm{KS}}$
Wasserstein (Kantorovich)	$\inf_{\substack{K: K_{ij} \geq 0, \forall i,j \\ K1 = q, K^T 1 = p}} \left(\sum_{i,j \in \mathcal{N}} K_{ij} \|Y^i - Y^j\|^d \right) \leq \rho, \quad d \geq 1$	$\mathcal{P}_q^{\mathrm{W}}$
Combined set	$p \in \mathcal{P}_q, \quad q \in \mathcal{Q} = \{q: \quad h_i(q) \leq 0, \quad i = 1, ..., Q\}$	$\mathcal{P}^{\mathrm{C}}$
Anderson-Darling	$-N - \sum_{n \in \mathcal{N}} \frac{2n-1}{N} \left(\log \left(p^T 1^n \right) + \log \left(p^T 1^{-n} \right) \right) \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{AD}}$
Cramer-von Mises	$\frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} - p^T 1^n \right)^2 \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{CvM}}$
Watson	$\frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} - p^T 1^n \right)^2 - N \left(\frac{1}{N} \sum_{n \in \mathcal{N}} p^T 1^n - \frac{1}{2} \right)^2 \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{Wa}}$
Kuiper	$\max_{n \in \mathcal{N}} \left(\frac{n}{N} - p^T 1^n \right) + \max_{n \in \mathcal{N}} \left(p^T 1^{n-1} - \frac{n-1}{N} \right) \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{K}}$

Some of the formulations in Table 3 include both the vectors p and q and the others only the vector p. The first corresponds to the situation when the uncertainty set for p is defined with reference to a nominal distribution q that in principle can be chosen arbitrarily. A typical choice for q will be the empirical distribution. The other case corresponds to the goodness-of-fit tests constructed for a one-dimensional random sample $Y^1 \leq Y^2 \leq \ldots \leq Y^N$. Then, the nominal measure q is implicitly defined by the empirical distribution of the sample at hand and cannot be chosen arbitrarily. This does not mean that one can use such an uncertainty set only for the case when Y is one-dimensional. For example, such a set can easily be generalized if the marginal distributions of Y are assumed to be independent.

4 Conjugates of the risk measures

In this section we give the results on concave conjugates $f_*(v, w)$ of functions $f(p, w)$ corresponding to the risk measures from Table 2. As mentioned earlier, for some cases we take $f(p, w) = F(p, w)$. For others, such as the Optimized Certainty Equivalent or the variance, $F(.)$ is reformulated if it is possible to find an $f(.)$ linear in p:

$$f(p, w) = Z_0 + \sum_{n \in \mathcal{N}} p_n Z_n(w),$$

with Z_0 and $Z_n(w)$ to be specified. Linearity in p is a desired property since then the conjugate $f_*(v, w)$ follows directly from (3):

$$f_*(v, w) = \begin{cases} -Z_0 & \text{if } Z_n(w) \leq v_n, \quad \forall n \in \mathcal{N} \\ -\infty & \text{otherwise.} \end{cases} \tag{8}$$

Derivations for the cases where $f(.)$ is nonlinear in p are given in Appendix A. The remainder of this section distinguishes three cases, depending on the type of the functions $F(.)$ and $f(.)$: (1) when both $F(.)$ and $f(.)$ are linear in p, (2) when $F(.)$ is nonlinear in p but $f(.)$ is linear in p, and (3) when both $F(.)$ and $f(.)$ are nonlinear in p. For each conjugate function we give the complexity of the system of inequalities involved in the formulation when $V(.)$ is a linear function of w.

Case 1: $F(p, w)$ linear in p

In this subsection we analyze the risk measures for which both $F(.)$ and $f(.)$ are linear in p.

Negative mean return. For the negative mean return the function is:

$$f(p, w) = F(p, w) = \sum_{n \in \mathcal{N}} p_n \left(-X_n(w) \right).$$

Its concave conjugate is given by formula (8) with $Z_0 = 0$ and $Z_n(w) = -X_n(w)$. If $V(.)$ is linear in w, the inequalities in this formulation are linear in w.

Shortfall risk. In case of the Shortfall risk the constraint itself is imposed on the variable κ. The constraint to be reformulated is $\mathbb{E}^p u(X(w) + \kappa) \geq 0$ or, equivalently:

$$-\mathbb{E}^p u(X(w) + \kappa) \leq 0, \quad \forall p \in \mathcal{P}.$$

The function $f(.)$ we take is:

$$f(p, w) = -\sum_{n \in \mathcal{N}} p_n u(X_n(w) + \kappa).$$

Its conjugate is given by (8) with $Z_0 = 0$ and $Z_n(w) = -u(X_n(w) + \kappa)$. If $V(.)$ is linear in w then, due to the concavity of $u(.)$, the inequalities included in this formulation are convex in the decision variables.

Lower partial moment. In this case the function is:

$$f(p, w) = F(p, w) = \sum_{n \in \mathcal{N}} p_n \max \left\{ 0, \bar{\kappa} - X_n(w) \right\}^\alpha.$$

Its conjugate is given by (8) with $Z_0 = 0$ and $Z_n(w) = \max \left\{ 0, \bar{\kappa} - X_n(w) \right\}^\alpha$. If $V(.)$ is linear in w, then for $\alpha = 1$ the inequalities involved are linear, and for $\alpha = 2$ they are convex quadratic in the decision variables.

Case 2: $F(p, w)$ nonlinear in p and $f(p, w)$ linear in p

In this subsection we analyze the risk measures for which $F(.)$ is nonlinear in p but $f(.)$ is linear in p.

Optimized Certainty Equivalent. For a constraint on the OCE, the constraint is:

$$F(p, w) = \inf_{\kappa \in \mathbb{R}} \left\{ -\kappa - \sum_{n \in \mathcal{N}} p_n(u(X_n(w) - \kappa)) \right\} \leq \beta, \quad \forall p \in \mathcal{P}. \tag{9}$$

Due to Lemma 2 (see Appendix A.1), for continuous and finite-valued functions $u(.)$ and compact sets $\mathcal{P}$ (being the uncertainty set for probabilities in our case) it holds that

$$\sup_{p \in \mathcal{P}} \inf_{\kappa \in \mathbb{R}} \left\{ -\kappa - \sum_{n \in \mathcal{N}} p_n(u(X_n(w) - \kappa)) \right\} = \inf_{\kappa \in \mathbb{R}} \sup_{p \in \mathcal{P}} \left\{ -\kappa - \sum_{n \in \mathcal{N}} p_n(u(X_n(w) - \kappa)) \right\}.$$

Using this result, the inf term in (9) can be removed, and the following constraint, with κ as a variable, is equivalent to (9):

$$f(p, w) = -\kappa - \sum_{n \in \mathcal{N}} p_n(u(X_n(w) - \kappa)) \leq \beta, \quad \forall p \in \mathcal{P}.$$

This formulation is already in the form of Theorem 1 and the concave conjugate of $f(.)$ with respect to its first argument is given by (8) with $Z_0 = -\kappa$ and $Z_n(w) = -u(X_n(w) - \kappa)$. If $V(.)$ is linear in w, then this formulation involves convex inequalities in the decision variables. For the Conditional Value-at-Risk, as a special case of the OCE, we have $Z_0 = -\kappa$ and $Z_n(w) = -\frac{1}{\alpha} \min\{X_n(w) - \kappa, 0\}$. If $V(.)$ is linear in w, the inequalities included in this formulation are representable as a system of linear inequalities in the decision variables.

Certainty Equivalent. For general $u(.)$ the formulation of a conjugate function would involve inequalities that are nonconvex in the decision variables. If one assumes that β is a fixed number, then a more tractable way to include a constraint on the CE:

$$F(p, w) = -u^{-1}\left(\sum_{n \in \mathcal{N}} p_n u(X(w)) \right) \leq \beta, \quad \forall p \in \mathcal{P}$$

is to multiply both sides by -1, then apply the function $u(.)$ to both sides to arrive at an equivalent constraint

$$\tilde{F}(p, w) = - \sum_{n \in \mathcal{N}} p_n u(X(w)) \leq -u(-\beta), \quad \forall p \in \mathcal{P}.$$

This constraint is of the same type as the robust constraint for the Shortfall risk. Therefore, the result for Shortfall risk can be used to obtain the relevant concave conjugate. In this case one cannot combine the CE with other risk measures via using the β as a variable.

Mean absolute deviation from the median. The constraint for this risk measure is given by:

$$F(p, w) = \sum_{n \in \mathcal{N}} p_n \left| X_n(w) - G^{-1}_{X(w)}(0.5) \right| \leq \beta, \quad \forall p \in \mathcal{P}.$$

Because of the median, $G_{X(w)}^{-1}(0.5)$, the function above is nonlinear in p and its concavity status is difficult to determine. However, we have:

$$F(p, w) = \sum_{n \in \mathcal{N}} p_n \left| X_n(w) - G_{X(w)}^{-1}(0.5) \right| = \inf_{\kappa \in \mathbb{R}} \sum_{n \in \mathcal{N}} p_n \left| X_n(w) - \kappa \right|.$$

The conditions of Lemma 2 (see Appendix A.1) are satisfied so that, similar to the Optimized Certainty Equivalent, we can remove the inf term to study equivalently the robust constraint on the following function:

$$f(p, w) = \sum_{n \in \mathcal{N}} p_n \left| X_n(w) - \kappa \right|,$$

where κ is a variable. Its conjugate is given by (8) with $Z_0 = 0$ and $Z_n(w) = |X_n(w) - \kappa|$. If $V(.)$ is linear in w, the inequalities included in the formulation above are representable as a system of linear inequalities in the decision variables.

Variance less the mean. The constraint for this risk measure is given by:

$$F(p, w) = \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w) \leq \beta, \quad \forall p \in \mathcal{P}.$$

Even though this formulation is concave in p, the results obtained in [5] for the variance in this form are difficult to implement. We propose to use, similar to the case of mean absolute deviation from the median, the following fact:

$$
\begin{aligned}
F(p, w) &= \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w) \\
&= \inf_{\kappa \in \mathbb{R}} \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \kappa \right)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w).
\end{aligned}
\tag{10}
$$

The conditions of Lemma 2 (see Appendix A.1) are satisfied, thus we can remove the inf term to study equivalently the robust constraint on the following function:

$$f(p, w) = \sum_{n \in \mathcal{N}} p_n \left((X_n(w) - \kappa)^2 - \alpha X_n(w) \right).$$

Its concave conjugate is given by (8) with $Z_0 = 0$ and $Z_n(w) = (X_n(w) - \kappa)^2 - \alpha X_n(w)$. The result for the variance is obtained by setting $\alpha = 0$. If $V(.)$ is linear in w, then this formulation involves convex quadratic inequalities in the decision variables.

Entropic Value-at-Risk. A robust constraint on the EVaR is given by

$$F(q, w) = \sup_{\tilde{p} \in \mathcal{P}_q} \mathbb{E}^{\tilde{p}}(-X(w)) \leq \beta, \quad \forall q \in \mathcal{Q}$$

with

$$\mathcal{P}_q = \left\{ \tilde{p} : \tilde{p} \geq 0, \quad 1^T \tilde{p} = 1, \quad \sum_{n \in \mathcal{N}} \tilde{p}_n \log \left(\frac{\tilde{p}_n}{q_n} \right) \leq -\log \alpha \right\},$$

and $\mathcal{Q}$ defined as in Table 3. The derivation of the concave conjugate with such a definition is troublesome since the function $F(.)$ is formulated as a supremum. Because of this we introduce the notion of a combined uncertainty set to include

13

the formulations of $\mathcal{P}_q$ and $\mathcal{Q}$ in the definition of a joint uncertainty set $\mathcal{U}^{\mathrm{C}}$ and to construct a relevant matrix A.

Then, the robust constraint on the EVaR is:

$$f(p, w) = \sum_{n \in \mathcal{N}} p_n \left(-X(w) \right), \quad \forall p \in \mathcal{P}^{\mathrm{C}},$$

where

$$\mathcal{P}^{\mathrm{C}} = \left\{ p : \quad p = A^{\mathrm{C}} p' \right\}, \quad A^{\mathrm{C}} = [I | 0_{N \times N}], \quad p' \in \mathcal{U}^{\mathrm{C}},$$

and

$$\mathcal{U}^{\mathrm{C}} = \left\{ p' = \begin{bmatrix} p \\ q \end{bmatrix} : p' \geq 0, \quad 1^T p = 1, \quad \sum_{n \in \mathcal{N}} p_n \log \left(\frac{p_n}{q_n} \right) \leq \rho, \quad h_i(q) \leq 0, \quad i = 1, ..., Q \right\}.$$

The function $f(.)$ for which the concave conjugate is to be derived, is the same as for the negative mean return, for which (8) holds with $Z_n(w) = -X_n(w)$ and $Z_0 = 0$. The only thing left is the derivation of the support function for $\mathcal{U}^{\mathrm{C}}$, which is done in Section 3. The approach developed here for the EVaR could also be used for other types of uncertainty sets $\mathcal{P}_q$.

Case 3: Both $F(p, w)$ and $f(p, w)$ nonlinear in p

In this subsection we analyze the risk measures for which both $F(.)$ and $f(.)$ are nonlinear in p.

Standard deviation less the mean. The constraint on this risk measure is given by:

$$F(p, w) = \sqrt{\sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2} - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w) \leq \beta, \quad \forall p \in \mathcal{P}.$$

The function $F(.)$ is nonlinear in p and a derivation of its conjugate would be troublesome. We propose to use the fact that:

$$
\begin{aligned}
F(p, w) &= \sqrt{\sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2} - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w) \\
&= \inf_{\kappa \in \mathbb{R}} \sqrt{\sum_{n \in \mathcal{N}} p_n (X_n(w) - \kappa)^2} - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w).
\end{aligned}
$$

The conditions of Lemma 2 (see Appendix A.1) are satisfied and, similar to the Optimized Certainty Equivalent, one can remove the inf term to reformulate equivalently the robust constraint on the following function:

$$f(p, w) = \sqrt{\sum_{n \in \mathcal{N}} p_n (X_n(w) - \kappa)^2} - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w).$$

14

The function $f(.)$ is concave in p and we can use Theorem 1. The conjugate of $f(.)$ is equal to (for sake of readability we switch to a problem-like notation):

$$
\begin{aligned}
f_*(v, w) = \sup_{y} \quad & -\tfrac{y}{4} \\
\text{s.t.} \quad & \left\| \begin{bmatrix} X_n(w) - \kappa \\ \left(\frac{v_n + \alpha X_n(w) - y}{2} \right) \end{bmatrix} \right\|_2 \leq \frac{v_n + \alpha X_n(w) + y}{2}, \quad \forall n \in \mathcal{N} \\
& v_n + \alpha X_n(w) \geq 0, \quad \forall n \in \mathcal{N} \\
& y \geq 0.
\end{aligned}
\tag{11}
$$

The derivation can be found in Appendix A. If $V(.)$ is linear in w, the above formulation involves second-order conic inequalities in the decision variables. The result for the standard deviation is obtained by setting $\alpha = 0$.

Sharpe ratio. A robust constraint on the Sharpe ratio risk measure is:

$$
F(p, w) = \frac{-\sum\limits_{n \in \mathcal{N}} p_n \left(X_n(w) \right)}{\sqrt{\sum\limits_{n \in \mathcal{N}} p_n \left(X_n(w) - \sum\limits_{n' \in \mathcal{N}} p_{n'} X_{n'}(w) \right)^2}} \leq \beta, \quad \forall p \in \mathcal{P}.
$$

The left-hand side function is neither convex, nor concave in the probabilities and we did not find a more tractable function $f(.)$ for it. If one assumes that β is a fixed number, then the constraint can be reformulated equivalently to:

$$
\sqrt{\sum_{n \in \mathcal{N}} p_n (X_n(w) - \sum_{n' \in \mathcal{N}} p_{n'} X_{n'}(w))^2} - \frac{1}{\beta} \sum_{n \in \mathcal{N}} p_n \left(X_n(w) \right) \leq 0, \quad \forall p \in \mathcal{P}.
$$

This constraint is equivalent to a robust constraint on the standard deviation less the mean with $\alpha = 1/\beta$ and the right hand side equal to 0. Thus, the corresponding result can be used for the conjugate function. In this case one cannot combine the Sharpe ratio with other risk measures using β as a variable.

In the case of VaR we did not find a formulation of the risk measure that would allow us to find a closed-form concave conjugate. A similar situation occurred for the general distortion, spectral, and coherent risk measures. We found the structure of their definitions intractable unless, for example, a coherent risk measure can be analyzed using a combined uncertainty set, as in the case of EVaR. The mean absolute deviation from the mean is nonconvex and nonconcave in the probabilities. For that reason we could not obtain a closed-form or inf-form for its concave conjugate.

5   Support functions of the uncertainty sets

In this section, the formulations of the support functions are given for the sets $\mathcal{U}$ corresponding to the uncertainty sets listed in Table 3. Most of the uncertainty sets have been obtained using the following lemma, taken from [5]:

Lemma 1. Let $Z \subset \mathbb{R}^L$ be of the form $Z = \{\zeta : \quad h_i(\zeta) \leq 0, \quad i = 1, ..., H\}$, where the $h_i(.)$ is convex for each i. If it holds that $\cap_{i=1}^H \text{ri}\,(\text{dom} h_i) \neq \emptyset$, then:

$$
\delta^* (v | Z) = \min_{u \geq 0} \left\{ \sum_{i=1}^H u_i h_i^* \left(\frac{v^i}{u_i} \right) \Bigg| \sum_{i=1}^H v^i = v \right\}. \qquad \square
$$

For each of the support functions we proceed in the same way. First, we give the necessary parameters, assuming that $A = I$ and $\mathcal{P} = \mathcal{U}$ unless stated otherwise. Then the support function is given, referring to Appendix B for the derivations.

ϕ-divergence functions. For the uncertainty set defined using the ϕ-divergence the support function is:

$$\delta^* \left(v \left| \mathcal{P}_q^\phi \right. \right) = \inf_{u \geq 0, \eta} \left\{ \eta + u\rho + u \sum_{n \in \mathcal{N}} q_n \phi^* \left(\frac{v_n - \eta}{u} \right) \right\}. \tag{12}$$

This result has also been obtained in [6]. In the general case the right-hand side expression between the brackets is a nonlinear convex function of the decision variables. However, for specific choices (see Table 5 in Appendix B) it can have more tractable forms - for instance, for the Variation distance it is linear. Result (12) holds also for the Pearson and likelihood ratio sets since they are specific cases of the ϕ-divergence set.

Kolmogorov-Smirnov. For an uncertainty set defined using the Kolmogorov-Smirnov test we take a matrix $D \in \mathbb{R}^{(2N+2) \times N}$ and a vector $d \in \mathbb{R}^{2N+2}$ whose components are:

$$
\begin{array}{lll}
D_{1n} = 1, & d_1 = 1, & \forall n \in \mathcal{N} \\
D_{2n} = -1, & d_2 = -1, & \forall n \in \mathcal{N} \\
D_{2+n,i} = 1, & d_{2+n} = \rho + q^T 1^n, & \forall i \leq n, \quad n \in \mathcal{N} \\
D_{2+N+n,i} = -1, & d_{2+N+n} = \rho - q^T 1^n, & \forall i \leq n, \quad n \in \mathcal{N},
\end{array}
$$

with the other components equal to 0. Under such a parametrization, the support function is equal to:

$$
\begin{aligned}
\delta^* \left(v \left| \mathcal{P}^{\text{KS}} \right. \right) = \inf_u \quad & u^T d \\
\text{s.t.} \quad & v \leq D^T u \\
& u \geq 0.
\end{aligned} \tag{13}
$$

The 'optimization problem' in (13) is linear.

Wasserstein. For an uncertainty set defined using the Wasserstein distance we take $A^{\text{W}} = [I \,|\, 0_{N \times N^2}]$. This choice is motivated in the derivation in Appendix B. Also, a matrix $D \in \mathbb{R}^{(4N+3) \times (N^2+N)}$ and a vector $d \in \mathbb{R}^{4N+3}$ are needed, whose components are:

$$
\begin{array}{lll}
D_{1n} = 1, & d_1 = 1, & \forall n \in \mathcal{N} \\
D_{2n} = -1, & d_2 = -1, & \forall n \in \mathcal{N} \\
D_{3,Ni+n} = \|Y_i - Y_n\|^d, & d_3 = \rho, & \forall i, n \in \mathcal{N} \\
D_{3+n,n} = -1, & D_{3+n,Nn+i} = 1, & \forall i, n \in \mathcal{N} \\
D_{3+N+n,n} = 1, & D_{3+N+n,Nn+i} = -1, & \forall i, n \in \mathcal{N} \\
D_{3+2N+n,Ni+n} = 1, & d_{3+2N+n} = q_n, & \forall i, n \in \mathcal{N} \\
D_{3+3N+n,Ni+n} = -1, & d_{3+3N+n} = -q_n, & \forall i, n \in \mathcal{N},
\end{array}
$$

with all other components of D and d equal to 0. The corresponding support function is equal to:

$$
\begin{aligned}
\delta^* \left(\left(A^{\text{W}} \right)^T v \left| \mathcal{U}_q^{\text{W}} \right. \right) = \inf_u \quad & u^T d \\
\text{s.t.} \quad & (A^{\text{W}})^T v \leq D^T u \\
& u \geq 0.
\end{aligned} \tag{14}
$$

The 'optimization problem' in (14) is linear.

Combined set. We assume that the uncertainty set $\mathcal{P}_q$ is defined as a ϕ-divergence set around q (being the Kullback-Leibler divergence for the EVaR). We take a matrix $A^{\mathrm{C}} = [I|0_{N \times N}]$, motivated in the corresponding section of Appendix B. The support function is equal to:

$$\delta^* \left(\left(A^{\mathrm{C}} \right)^T v \middle| \mathcal{U}^{\mathrm{C}} \right) = \inf_{\substack{\{u_i, v^i\}, \\ i=1,...,Q+3}} \quad u_1 - u_2 + u_3 \rho + \sum_{i=1}^{Q} u_{i+3} h_i^* \left(\frac{v_{N+1:2N}^{i+3}}{u_{i+3}} \right)$$

$$\begin{aligned}
\text{s.t.} \quad & v_1 \leq u_1 1 \\
& v_{1:N}^2 \leq -u_2 1 \\
& v_{N+1:2N}^i = 0, \quad i = 1, 2, 3 \\
& v_{1:N}^i = 0, \quad i = 4, ..., Q+3 \\
& v_{N+n}^3 + u_3 \phi^* \left(\frac{v_n^3}{u_3} \right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=1}^{Q+3} v^i = \left(A^{\mathrm{C}} \right)^T v \\
& u_i \geq 0, \quad i = 1, ..., Q+3.
\end{aligned} \tag{15}$$

For all ϕ-divergence functions listed in Table 5 the 'optimization problem' in (15) is convex. If the ϕ-divergence is the Variation distance or the modified χ^2 distance and the functions $h_i(.)$ are all linear or convex quadratic, then the 'optimization problem' in (15) is linear or convex quadratic, respectively.

Anderson-Darling. For an uncertainty set defined using the Anderson-Darling test the support function $\delta^* \left(v \middle| \mathcal{P}_{\mathrm{emp}}^{\mathrm{AD}} \right)$ is equal to:

$$\begin{aligned}
\inf_{\eta, u, \{w^{n+}, w^{n-}\}, n \in \mathcal{N}} \quad & -\sum_{n \in \mathcal{N}} \frac{(2n-1)u}{N} \left[2 + \log \left(\frac{-N z_n^+}{(2n-1)u} \right) + \log \left(\frac{-N z_n^-}{(2n-1)u} \right) \right] \\
& + u (\rho + N) + \eta \\
\text{s.t.} \quad & v \leq \sum_{n \in \mathcal{N}} \left(z_n^+ 1^n + z_n^- 1^{-n} \right) + \eta 1 \\
& z_n^+, z_n^- \leq 0 \quad \forall n \in \mathcal{N} \\
& u \geq 0.
\end{aligned} \tag{16}$$

This result has also been obtained in [5]. The 'optimization problem' in (16) is convex.

Cramer-von Mises. For an uncertainty set defined using the Cramer-von Mises test we use the following parameters:

$$c = -\rho + \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} \right)^2, \quad b = \begin{bmatrix} -2 \sum_{j=1}^{N} \frac{2j-1}{N} \\ -2 \sum_{j=2}^{N} \frac{2j-1}{N} \\ \vdots \\ -2 \sum_{j=N}^{N} \frac{2j-1}{N} \end{bmatrix},$$

a matrix $E \in \mathbb{R}^{N \times N}$ such that $E_{ij} = N + 1 - \max\{i, j\}$ for $i, j \in \mathcal{N}$ and a unique matrix P such that $P^T P = E^{-1}$. With such a parametrization, the support function

is equal to:

$$\delta^* \left(v \left| \mathcal{P}_{\text{emp}}^{\text{CvM}} \right. \right) = \inf_{z,t,\{u_i,v^i\},i=1,\dots,3} \quad u_1 - u_2 + \tfrac{1}{4}t - u_3 c$$

$$\text{s.t.} \quad \left\| \left[\begin{array}{c} Pz \\ \frac{t-u_3}{2} \end{array} \right] \right\|_2 \le \frac{t+u_3}{2} \tag{17}$$

$$z = u_3 b - v^3$$

$$u_1 - u_2 + v_n^3 - v_n \ge 0, \quad \forall n \in \mathcal{N}$$

$$u_1, u_2, u_3 \ge 0.$$

The 'optimization problem' in (17) is convex quadratic.

Watson. For an uncertainty set defined using the Watson test we use the following parameters:

$$c = -\rho + \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} \right)^2 - \frac{N}{4}, \quad b = \left[\begin{array}{c} -2\sum_{j=1}^{N} \frac{2j-1}{N} + N \\ -2\sum_{j=2}^{N} \frac{2j-1}{N} + (N-1) \\ \vdots \\ -2\sum_{j=N}^{N} \frac{2j-1}{N} + 1 \end{array} \right],$$

a matrix $E \in \mathbb{R}^{N \times N}$ such that:

$$E_{i,j} = N + 1 - \max\{i,j\} - \frac{(N+1-i)(N+1-j)}{N}, \quad \forall i, j \in \mathcal{N}$$

and a matrix P such that $P^T P = E$. With such a parametrization, the support function is given by:

$$\delta^* \left(v \left| \mathcal{P}_{\text{emp}}^{\text{Wa}} \right. \right) = \inf_{\substack{z,t,\lambda,\{u_i,v^i\}, \\ i=1,\dots,3}} \quad u_1 - u_2 + \tfrac{1}{4}t - u_3 c$$

$$\text{s.t.} \quad \left\| \left[\begin{array}{c} Pz \\ \frac{t-u_3}{2} \end{array} \right] \right\|_2 \le \frac{t+u_3}{2} \tag{18}$$

$$z = u_3 b - v^3$$

$$u_1 - u_2 + v_n^3 - v_n \ge 0, \quad n \in \mathcal{N}$$

$$u_1, u_2, u_3, t \ge 0$$

$$E\lambda = z.$$

The 'optimization problem' in (18) is convex quadratic.

Kuiper. For the uncertainty set defined using the Kuiper test we take $A^{\text{K}} = [I \,|\, 0_{N \times 2}]$. Also, a matrix $D \in \mathbb{R}^{(2N+3) \times (N+2)}$ and a vector $d \in \mathbb{R}^{2N+3}$ are used, whose components are:

$$
\begin{array}{llll}
D_{1,n} = 1, & d_1 = 1, & & \forall n \in \mathcal{N} \\
D_{2,n} = -1, & d_2 = -1, & & \forall n \in \mathcal{N} \\
D_{2+n,i} = -1, & D_{2+n,N+1} = -1, & d_{n+2} = -n/N, & \forall i \le n, n \in \mathcal{N} \\
D_{N+2+n,i} = 1, & D_{N+2+n,N+2} = -1, & d_{N+2+n} = (n-1)/N, & \forall i \le n-1, n \in \mathcal{N} \\
D_{2N+3,N+1} = 1, & D_{2N+3,N+2} = 1, & d_{2N+3} = \rho, &
\end{array}
$$

with all other components of the matrix D and vector d equal to 0. Under such a parametrization, the support function is

$$\delta^* \left(\left(A^{\text{K}} \right)^T v \left| \mathcal{U}_{\text{emp}}^{\text{K}} \right. \right) = \inf_{u} \quad u^T d$$

$$\text{s.t.} \quad \left(A^{\text{K}} \right)^T v \le D^T u \tag{19}$$

$$u \ge 0.$$

The 'optimization problem' in (19) is linear.

6 Examples

6.1 Portfolio management

We consider as first application of our methodology a portfolio optimization problem. In this problem, the aim is to maximize the (worst-case) mean return subject to a maximum risk measure level, in both a nominal and robust setting. We choose the risk measure to be the Entropic Value-at-Risk for its importance as an upper bound on both the Value-at-Risk and the Conditional Value-at-Risk.

6.1.1 Formulation and derivations of the robust counterparts

There are M available assets and N joint return scenarios for these assets, where Y_i^n denotes the gross return on the i-th asset in the n-th scenario. The decision vector $w \in \mathcal{W} = \left\{ w \in \mathbb{R}^M, \quad 1^T w = 1, \quad w \geq 0 \right\}$ consists of the portfolio weights of assets where we assume that shortselling is not allowed. The portfolio return in the n-th scenario is $X_n(w) = \sum_{i=1}^M w_i Y_i^n$. The maximum (robust) EVaR level is z. The nominal optimization problem is then:

$$
\begin{aligned}
\max \quad & \mu \\
\text{s.t.} \quad & \sum_{n \in \mathcal{N}} q_n \left(-X_n(w) \right) \leq -\mu \\
& \sup_{\tilde{p} \in \mathcal{P}_q} \sum_{n \in \mathcal{N}} \tilde{p}_n \left(-X_n(w) \right) \leq z \\
& w \in \mathcal{W},
\end{aligned}
\tag{20}
$$

where $\mathcal{P}_q$ is defined in the row of Table 2 corresponding to the EVaR. Problem (20) includes a constraint involving a sup term, which requires a reformulation to a tractable form. In the terminology of this paper, this constraint is equivalent to a robust constraint on the negative mean return with uncertainty set $\mathcal{P}_q$ defined by the Kullback-Leibler divergence, and can be reformulated using the results of Sections 4 and 5.

We proceed to the more difficult and, hence, more illustrative robust problem. The uncertainty set for the nominal probability distribution q is defined as the Pearson set around a vector r (see Table 5):

$$
\mathcal{Q} = \left\{ q \geq 0 : \quad 1^T q = 1, \quad \sum_{n \in \mathcal{N}} \frac{(q_n - r_n)^2}{r_n} \leq \rho_{\mathcal{Q}} \right\}.
$$

This formulation satisfies the conditions for the set $\mathcal{Q}$ in Table 3 for the combined uncertainty set since all the defining constraints can be formulated as constraints

19

on convex functions in q. The portfolio optimization problem is then:

$$\begin{aligned}
\max \quad & \mu \\
\text{s.t.} \quad & \sum_{n \in \mathcal{N}} q_n \left(-X_n(w)\right) \leq -\mu, \quad \forall q \in \mathcal{Q} \qquad \text{(21a)} \\
& \sup_{\tilde{p} \in \mathcal{P}_q} \sum_{n \in \mathcal{N}} \tilde{p}_n \left(-X_n(w)\right) \leq z, \quad \forall q \in \mathcal{Q} \quad \text{(21b)} \\
& w \in \mathcal{W}.
\end{aligned} \qquad (21)$$

We shall reformulate the two constraints in problem (21) to their tractable forms using the results of Sections 4 and 5.

Constraint (21a). This is a robust constraint on the negative mean return with uncertainty set $\mathcal{Q}$ being the Pearson set. The corresponding conjugate function (Section 4) is given by:

$$f_*(v, w) = \begin{cases} 0 & \text{if } -X_n(w) \leq v_n^1, \quad \forall n \in \mathcal{N} \\ -\infty & \text{otherwise.} \end{cases}$$

The support function of the Pearson set is:

$$\delta^* \left(v^1 \,\Big|\, \mathcal{P}^{\chi^2}\right) = \inf_{u_1 \geq 0, \eta} \; \eta + u_1 \left(\rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max\left\{-1, \frac{v_n^1 - \eta}{u_1} + \frac{1}{4}\left(\frac{v_n^1 - \eta}{u_1}\right)^2\right\}\right).$$

Inserting the results on the conjugate and the support into (7) yields the tractable robust counterpart of (22b):

$$\begin{cases} \eta + u_1 \rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max\left\{-u_1, v_n^1 - \eta + \frac{1}{4}\frac{\left(v_n^1 - \eta\right)^2}{u_1}\right\} \leq -\mu \\ u_1 \geq 0 \\ -X_n(w) \leq v_n^1, \quad \forall n \in \mathcal{N}. \end{cases}$$

Constraint (21b). This is a robust constraint on the EVaR with $\mathcal{Q}$ defined as the Pearson set. We shall use the results for EVaR (Section 4) and the combined uncertainty set (Section 5). The conjugate function $f_*(v, w)$ is the same as in the case of constraint (22a). To obtain the support function of the set $\mathcal{U}^{\mathrm{C}}$ we use the fact that $\mathcal{Q}$ is a ϕ-divergence set. The conjugate functions obtained in the part of Appendix B corresponding to the ϕ-divergence sets can be used as functions $h_i^*(.)$ needed in (15). Then, the support function $\delta^* \left(\left(A^{\mathrm{C}}\right)^T v \,\Big|\, \mathcal{U}^{\mathrm{C}}\right)$ is equal to:

$$\inf_{\substack{\{u_i, v^i\} \\ i=2,\ldots,7}} \; u_2 - u_3 - u_4 \log \alpha + u_5 - u_6 + u_7 \left(\rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max\left\{-1, \frac{v_{N+n}^7}{u_7} + \frac{1}{4}\left(\frac{v_{N+n}^7}{u_7}\right)^2\right\}\right)$$

$$\begin{aligned}
\text{s.t.} \quad & v_{1:N}^2 \leq u_2 1 \\
& v_{1:N}^3 \leq -u_3 1 \\
& v_{N+1:2N}^5 \leq u_5 1 \\
& v_{N+1:2N}^6 \leq -u_6 1 \\
& v_{N+1:2N}^i = 0, \quad i = 2, 3, 4 \\
& v_{1:N}^i = 0, \quad i = 5, 6, 7 \\
& v_{N+n}^4 + u_4 \left(\exp\left(\frac{v_n^4}{u_4}\right) - 1\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=2}^{7} v^i = \left(A^{\mathrm{C}}\right)^T v \\
& u_i \geq 0, \quad i = 2, \ldots, 7.
\end{aligned}$$

Inserting the results on the conjugate and the support function into (7) yields the tractable robust counterpart of (22a):

$$
\begin{cases}
u_2 - u_3 - u_4 \log \alpha + u_5 - u_6 + u_7 \rho_{\mathcal{Q}} + \sum\limits_{n \in \mathcal{N}} r_n \max \left\{ -u_7, v_{N+n}^7 + \frac{1}{4} \frac{\left(v_{N+n}^7\right)^2}{u_7} \right\} \leq z \\
v_{1:N}^2 \leq u_2 1 \\
v_{1:N}^3 \leq -u_3 1 \\
v_{N+1:2N}^5 \leq u_5 1 \\
v_{N+1:2N}^6 \leq -u_6 1 \\
v_{N+1:2N}^i = 0, \quad i = 2, 3, 4 \\
v_{1:N}^i = 0, \quad i = 5, 6, 7 \\
v_{N+n}^4 + u_4 \left(\exp\left(\frac{v_n^4}{u_4}\right) - 1 \right) \leq 0, \quad \forall n \in \mathcal{N} \\
\sum\limits_{i=2}^{7} v^i = \left(A^{\text{C}}\right)^T v \\
u_i \geq 0, \quad i = 2, \ldots, 7 \\
-X_n(w) \leq v_n, \quad \forall n \in \mathcal{N}.
\end{cases}
$$

It was possible to remove the inf term in the support function formulation because it occurs on the left-hand side of the constraint. All the constraints in the above counterpart are convex in the decision variables. To our best knowledge, this paper is the first to obtain a computationally tractable robust counterpart of a constraint on the EVaR with general uncertainty sets.

Combining the tractable robust counterparts of the constraints with the rest of the problem formulation, we obtain that (21) is equivalent to:

$$
\begin{aligned}
\max_{\substack{v, u_i, v^i, i=1,\ldots,7 \\ w, \eta, \mu}} \quad & \mu \\
\text{s.t.} \quad & \eta + u_1 \rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max \left\{ -u_1, v_n^1 - \eta + \frac{1}{4} \frac{\left(v_n^1 - \eta\right)^2}{u_1} \right\} \leq -\mu \\
& u_2 - u_3 - u_4 \log \alpha + u_5 - u_6 + u_7 \rho_{\mathcal{Q}} + \\
& \quad + \sum_{n \in \mathcal{N}} r_n \max \left\{ -u_7, v_{N+n}^7 + \frac{1}{4} \frac{\left(v_{N+n}^7\right)^2}{u_7} \right\} \leq z \\
& v_{1:N}^2 \leq u_2 1 \\
& v_{1:N}^3 \leq -u_3 1 \\
& v_{N+1:2N}^5 \leq u_5 1 \\
& v_{N+1:2N}^6 \leq -u_6 1 \\
& v_{N+1:2N}^i = 0, \quad i = 2, 3, 4 \\
& v_{1:N}^i = 0, \quad i = 5, 6, 7 \\
& v_{N+n}^4 + u_4 \left(\exp\left(\frac{v_n^4}{u_4}\right) - 1 \right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=2}^{7} v^i = \left(A^{\text{C}}\right)^T v \\
& -X_n(w) \leq v_n, \quad \forall n \in \mathcal{N} \\
& -X_n(w) \leq v_n^1, \quad \forall n \in \mathcal{N} \\
& u_i \geq 0, \quad i = 1, \ldots, 7 \\
& w \in \mathcal{W}.
\end{aligned}
\tag{22}
$$

This problem involves linear, convex quadratic, and convex constraints in the decision variables.
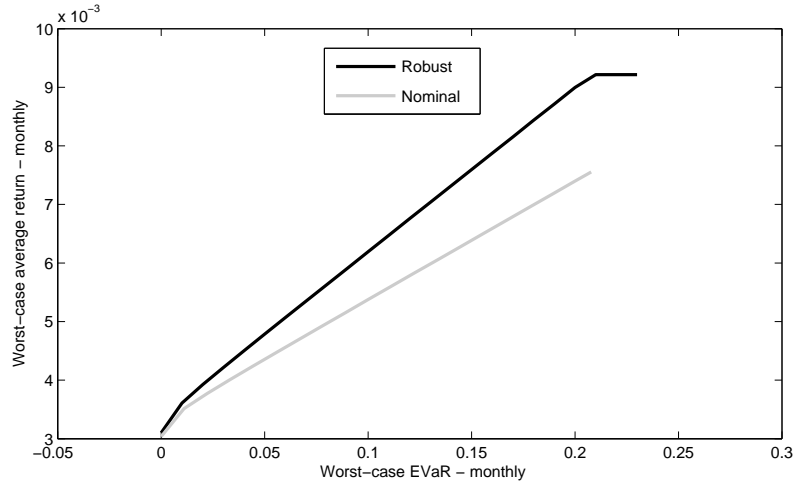
Figure 1: The EVaR-mean return frontier for the robust and the nominal portfolios.

6.1.2 Numerical illustration

As a numerical illustration, we use 6 risky assets and 1 riskless asset, with data taken from the website of Kenneth M. French.[1] The risky portfolios, constructed at the end of each June, are the intersections of 2 portfolios formed on size (market equity, ME) and 3 portfolios formed on the ratio of book equity to market equity (BE/ME). The size breakpoint for year t is the median NYSE market equity at the end of June of year t. BE/ME for June of year t is the book equity for the last fiscal year end in $t - 1$ divided by ME for December of $t - 1$. The BE/ME breakpoints are the 30th and 70th NYSE percentiles. The riskless asset is the one-month US Treasury bill rate. The monthly data on all assets includes 360 observations from February 1984 to January 2014.

The nominal distribution of the return scenarios assigns probability $r_n = \frac{1}{360}$ to each of the scenarios. We take $\alpha = 0.05$, which makes the EVaR an upper bound for the Value-at-Risk and Conditional Value-at-Risk at level 0.05. The degree of uncertainty about the distribution of q in the robust model is defined by $\rho_Q = 0.005$. The value of this parameter has been chosen to allow possibly many robust portfolios to be feasible for various values of z.

First, we investigate how the optimal (worst-case) mean return changes when we impose different EVaR limits. To do this, we solve problems (20) and (22) for $z = 0, 0.01, \ldots, 0.23$. For the robust portfolio, we plot its worst-case EVaR - worst-case mean return curve. For each of the nominal portfolios we compute the most pessimistic EVaR and the most pessimistic mean return with $q \in Q$ as a possible probability measure. Then, we plot the worst-case EVaR - worst-case mean return frontier for the nominal portfolios. Figure 1 depicts the worst-case mean - worst-case EVaR frontier for the nominal and the robust case.

For each worst-case EVaR value, the robust portfolio outperforms the nominal portfolio in terms of the worst-case outcome. The break of the both curves around EVaR

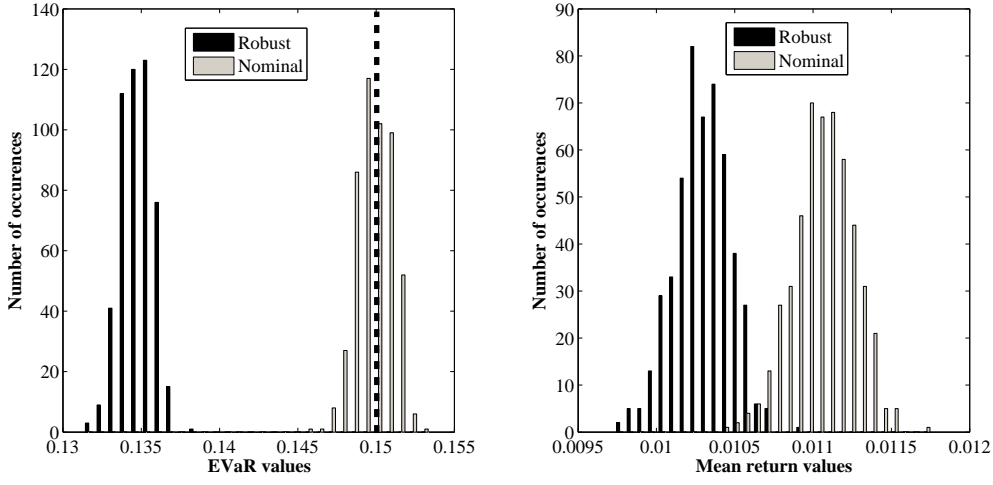[1]Available at: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

Figure 2: Histograms of the mean return and EVaR value of the sampled portfolios. The dashed line in the left panel denotes the $z = 0.15$ constraint on the portfolio that was used in optimization problems.

close to 0.01 is due to the return variability of the riskless asset used (thus, it is not fully riskless since its risk is nonzero). The second kink of the robust frontier corresponds to the no-shortselling constraint - for all $z \geq 0.21$ the optimal robust portfolio is identical. Similarly, for all $z \geq 0.19$ the optimal nominal portfolio is identical, and its worst-case EVaR is 0.2078, hence the nominal frontier covers only values of EVaR less than or equal to 0.2078.

To test the performance of robust and nominal portfolios, we conduct the following bootstrap experiment. We take the nominal and robust portfolios for the maximum EVaR value 0.15. Then, we sample 500 probability distributions q around the nominal distribution r as follows: for $n = 1, ..., N - 1$ the value r_n is sampled from a normal distribution with mean $r_n = \frac{1}{360}$ and standard deviation $\sqrt{\frac{\rho_{\mathcal{Q}}}{N^2}}$ and the last element is set $q_N = 1 - \sum_{n=1}^{N-1} q_n$. If it holds that $q \geq 0$, then the given vector is accepted. Out of this sample, 85% belonged to $\mathcal{Q}$. For each such q, we compute the EVaR and the mean return on the nominal and the robust portfolios. Figure 2 shows the results of the experiment.

The portfolios show significant differences in the distribution of their return and the EVaR value. In the left panel, the nominal portfolio violates the 0.15 upper bound (the dashed vertical line) in a large number of cases, whereas the robust portfolio's EVaR values oscillate in a region relatively far from 0.15. The robust portfolio does not reveal any overconservatism - it is possible to find such q and $\tilde{p}$ that the EVaR of the robust portfolio is equal to 0.15. In the right panel we can see that on average the nominal portfolio has a significantly higher mean return. The differences between the means of EVaR and the return distributions are statistically significant at the 99% level.

All problems have been solved using the convex programming toolbox cvx for problem formulation and the Mosek solver. Solving a single robust optimization problem for a given ρ took on average 43.1 seconds on an Intel Core 2.66GHz computer. This

23

time is a result mostly of the sequential approximation method used by cvx for problems involving exponential constraints.

6.2 Multi-item newsvendor problem

In this subsection we consider the application of our methodology to a multi-item newsvendor problem with a mean-variance objective function.

6.2.1 Formulation and derivations of the robust counterparts

We follow the formulation given in [6]. The newsvendor problem is how many units of a product (item) to order, taking into account that the demand for the product is stochastic. Due to uncertainty, the newsvendor can face both unsold items or unmet demand. The unsold items will return a loss because their salvage value is lower than the purchase price. In the case of unmet demand the newsvendor incurs a cost of lost sales, which may include a penalty for the lost customer goodwill.

We assume that there are M products and N joint demand scenarios for the products. If the newsvendor chooses to buy w_i items of the i-th product, then his net profit in the n-th scenario from the i-th product is given by:

$$V_i^n(w_i) = v_i \min \{Y_i^n, w_i\} + s_i (w_i - Y_i^n)^+ - l_i (Y_i^n - w_i)^+ - c_i w_i,$$

where $Y_i^n \geq 0$ is the uncertain demand for the i-th product in scenario n, v_i is the unit selling price, s_i is the salvage value per unsold item, l_i is the shortage cost per unit of unsatisfied demand, and c_i is the purchasing price per unit. Then, the total net profit in the n-th scenario is given by $X_n(w) = \sum_{i=1}^M V_i^n(w_i)$. A standard assumption for this problem is $v_i + l_i \geq s_i$ for each i. We assume that each of the scenarios occurs with probability p_n. We solve the nominal problem for a fixed p and the robust problem, with an uncertainty set $\mathcal{P}^\phi$ for p defined with the Variation distance around q. Such a set $\mathcal{P}^\phi$ is LP-representable, which improves the speed of solving an instance.

The nominal problem to be solved is:

$$\max_{w \in \mathbb{Z}_+^M} \sum_{n \in \mathcal{N}} q_n X_n(w) - \frac{1}{\alpha} \inf_{\kappa \in \mathbb{R}} \sum_{n \in \mathcal{N}} q_n (X_n(w) - \kappa)^2, \quad \alpha > 0. \tag{23}$$

Its robust version is:

$$\max_{w \in \mathbb{Z}_+^M, \kappa} \min_{p \in \mathcal{P}^\phi} \left\{ \sum_{n \in \mathcal{N}} p_n X_n(w) - \frac{1}{\alpha} \inf_{\kappa \in \mathbb{R}} \sum_{n \in \mathcal{N}} p_n (X_n(w) - \kappa)^2 \right\}, \quad \alpha > 0, \tag{24}$$

where

$$\mathcal{P}^\phi = \left\{ p : \quad 1^T p = 1, \quad \sum_{n \in \mathcal{N}} |p_n - q_n| \leq \rho \right\}.$$

Problem (24) is equivalent to minimizing the variance less the mean:

$$\min_{w \in \mathbb{Z}_+^M, \kappa} \quad z$$
$$\text{s.t.} \quad \sum_{n \in \mathcal{N}} p_n (X_n(w) - \kappa)^2 - \alpha \sum_{n \in \mathcal{N}} p_n X(w) \leq z, \quad \forall p \in \mathcal{P}^\phi.$$

24

Table 4: Product parameters

Product	1	2	3	4	5	6	7	8	9	10	11	12
c_i	4	5	6	4	5	6	4	5	6	4	5	6
v_i	6	8	9	5	9	8	6	8	9	6.5	7	8
s_i	2	2.5	1.5	1.5	2.5	2	2.5	1.5	2	2	1.5	1
l_i	4	3	5	4	3.5	4.5	3.5	3	5	3.5	3	5

Using results of Sections 4 and 5 we obtain that (24) is equivalent to:

$$
\begin{aligned}
\min_{u,v,w,\eta,\kappa} \quad & z \\
\text{s.t.} \quad & \eta + u\rho + \sum_{n \in \mathcal{N}} q_n \max\{-u, v_n - \eta\} \\
& \left(\sum_{i=1}^{M} V_i^n(w_i) - \kappa\right)^2 - \alpha \sum_{i=1}^{M} V_i^n(w_i) \leq v_n, \quad \forall n \in \mathcal{N} \\
& v_n - \eta \leq u, \quad \forall n \in \mathcal{N} \\
& w \in \mathbb{Z}_+^M.
\end{aligned}
\tag{25}
$$

Due to the concavity of the functions $V_i^n(w_i)$, the above problem is nonconvex. One way way to deal with this issue would be to use a global solver. Another way, used by us in the numerical experiment, is a brute-force approach by splitting the problem into $(N+1)^M$ problems over per-item intervals $w_i \in [Y_i^{n_i-1}, Y_i^{n_i}]$, where $0 = Y_i^0 \leq \ldots \leq Y_i^{N_i+1} = +\infty$. Thus, we solve (25) for each $(n_1, n_2, \ldots, n_M) \in \{1, \ldots, N+1\}^M$. Then each $X_n(w)$ is linear in the decision variable w over the domain of a single problem. We choose the subproblem with the best objective to be the solution w^*.

6.2.2 Numerical illustration

In the numerical experiment we solve 50 newsvendor problems sampled as follows. First, out of 12 available products (see Table 4) we randomly choose 3 for the given problem instance. The product parameters are taken from [6]. We assign value $\alpha = 10$ to the mean-variance parameter. We assume each of the products to have three demand scenarios: 4, 8, and 10. Because of that, in a given problem there are $3^3 = 27$ joint demand scenarios for the three products. To these 27 scenarios we assign randomly a nominal probability vector q by sampling first $k_n, n = 1, \ldots, 27$ from the uniform distribution on $[0,1]$ and assigning $q_n = k_n / \left(\sum_{j=1}^{27} k_j\right), n = 1, \ldots, 27$.

First, we are interested in the sensitivity of the optimal solutions to changes in ρ. To investigate this, we solve each of the 50 problems for $\rho = 0, 0.05, 0.1, \ldots, 0.5$, where $\rho = 0$ corresponds to a nominal version without uncertainty. Figure 3 shows the results on changes in the w vector for different values of ρ in a sample problem. As we can see in this case, as the degree of uncertainty grows, the decision maker decides to buy less of each product. Overall, the changes are not big compared to the decision for $\rho = 0$. In 8 out of 50 problems the nominal solution is the same as the robust solution for $\rho = 0.5$. The monotonic pattern in Figure 3 is not typical for
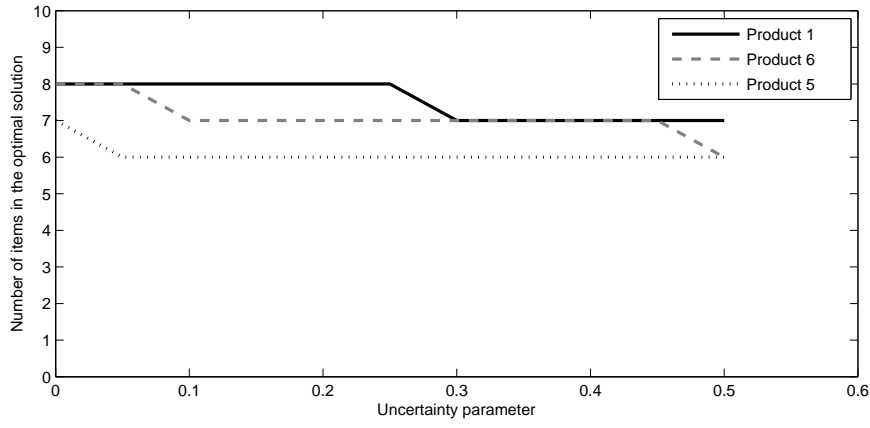
Figure 3: Newsvendor's strategies in a sample problem for different values of ρ. The nominal solution corresponds to $\rho = 0$.
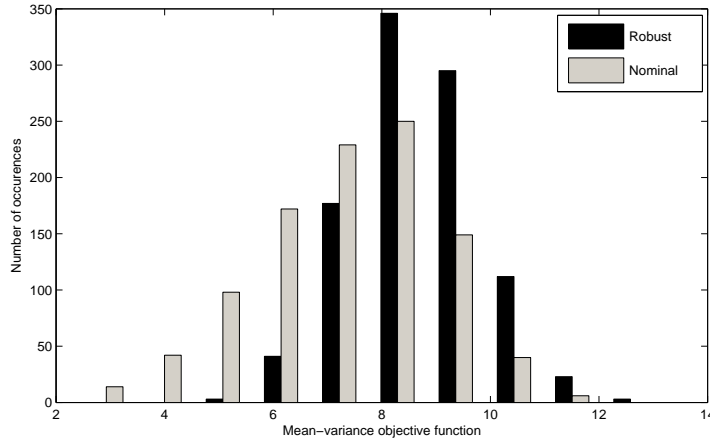

Figure 4: Histogram of the mean-variance objective function on a sample of 1000 probability distributions for a sample problem.

all the sampled problems - sometimes when ρ becomes larger, the decision maker chooses to buy more items of a given product.

To compare the nominal and the robust solutions, for each of the 50 problems we take the solutions for $\rho = 0$ and $\rho = 0.5$ and conduct a bootstrap test of their performance. For each of the problems, we sample 1000 sample probability vectors p around q: for $n = 1, ..., 26$ the value p_n is sampled from a normal distribution with mean q_n and standard deviation $\frac{1}{2}\sqrt{\frac{\rho q_n}{N}}$ and for $n = 27$ we assign $p_{27} = 1 - \sum_{n=1}^{26} p_n$. If it holds that $p \geq 0$, then a given vector is accepted. For each problem, around 98% of sampled probability distributions belonged to the corresponding uncertainty set for $\rho = 0.5$. For each of the sampled probability vectors, we compute the original mean-variance objective function of the nominal and the robust solution.

Figure 4 shows the bootstrapped performance of the robust and nominal newsvendor strategy of a sample problem. The distribution of the sample objective values for the robust solution is more concentrated. Also, the mean outcome is greater than in the case of the nominal solution. Out of the 42 problems where the nominal and robust solutions differed, 41 show a better average-case performance of the
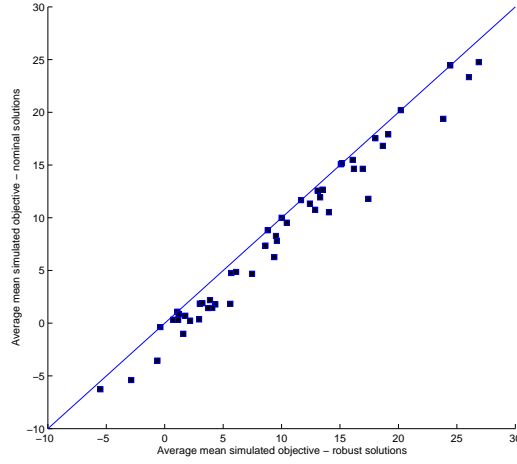
Figure 5: Scatterplot of simulated mean objective values for the robust and nominal solutions to problems.

robust solution at the 95% significance level. The scatterplot of the simulated mean objective values for the robust and nominal solutions is given in Figure 5.

All problems have been solved using the convex programming toolbox cvx for problem formulation and the Gurobi solver. Solving a single newsvendor problem for a fixed ρ took on average 23.1 seconds on an Intel Core 2.66GHz computer.

7   Conclusions

In this paper we have shown that for many risk measures and statistically based uncertainty sets the distributionally robust constraints on risk measures with discrete probabilities can be reformulated to a computationally tractable form. In particular, components corresponding to the risk measure and to the uncertainty set can be separated. We also demonstrated that our approach can be applied to risk measures that are nonlinear in the probability vector. Our results can be used in finance, economics, and other fields.

We now give potential directions of further research. Following the work of Wozabal (2011), where the Wasserstein distance was analyzed, it is interesting to investigate whether the results of our paper can be extended to the case with continuous probability distributions, without conversion of continuous probability distributions into discrete ones.

Second, it is important to check the differences in the practical performance of different types of uncertainty for the risk measures. If some uncertainty sets, yielding a better computational status of the tractable counterpart, can credibly substitute for others, then our methodology could be applied to larger instances.

Finally, for the risk measures that we have not been able to analyze successfully one could investigate their sensitivity to the uncertainty considered in this paper. It may turn out that these risk measures themselves are sufficiently robust or that

27

different tools are needed to develop computationally tractable robust constraints in terms of these risk measures.

References

[1] Ben-Tal, A., Ben-Israel, A. & Teboulle, M. (1991). Certainty equivalents and information measures: duality and extremal principles. *Journal of Mathematical Analysis and Applications*, Vol. 157(1), pp. 211-236.

[2] Ben-Tal, A. & Nemirovski, A. (2001) Lectures on modern convex optimization: analysis, algorithms, and engineering applications. (SIAM).

[3] Ben-Tal, A. & Teboulle, M. (2007). An old-new concept of convex risk measures: the Optimized Certainty Equivalent. *Mathematical Finance*, Vol. 17(3), pp. 449-476.

[4] Ben-Tal, A., El Ghaoui, L. & Nemirovski, A. (2009). *Robust optimization.* (Princeton University Press).

[5] Ben-Tal, A., Den Hertog, D. & Vial, J.-Ph. (2012) Deriving Robust Counterparts of nonlinear uncertain inequalities. *CentER Discussion Paper Series No. 2012-053.* Available at SSRN: `https://pure.uvt.nl/portal/files/1436907/2012-053.pdf`. To appear in *Mathematical Programming.*

[6] Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B. & Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, Vol. 59(2), pp. 341-357.

[7] Bertsimas, D., Gupta, V. & Kallus, N. (2013). Data-driven robust optimization. Available online at:
`http://www.mit.edu/~vgupta1/Papers/DataDrivenRobOptv1.pdf`.

[8] Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, Vol. 18(3), pp, 853-877.

[9] Chen, L., He, S., & Zhang, S. (2011). Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, Vol. 59(4), pp. 847-865.

[10] El Ghaoui, L., Oks, M. & Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, Vol. 51(4), pp. 543-556.

[11] Fertis, A., Baes, M. & Laethi, H. J. (2012). Robust risk management. *European Journal of Operational Research*, Vol. 222(3), pp. 663-672.

[12] Föllmer, H. & Schied, A. (2010). Convex risk measures. *Encyclopedia of Quantitative Finance,* pp. 355-363. (John Wiley & Sons, Ltd)

[13] Goldfarb, D. & Iyengar, G. (2003). Robust convex quadratically constrained programs. *Mathematical Programming*, Vol. 97(3), pp. 495-515.

[14] Grant, M., & Boyd, S., *CVX: Matlab software for disciplined convex programming, version 2.0 beta.* Available at: `http://cvxr.com/cvx`. September 2013.

[15] Gulpinar, N. & Rustem, B. (2007). Worst-case robust decisions for multi-period mean-variance portfolio optimization. *European Journal of Operational Research*, Vol. 183(3), pp. 981-1000.

[16] Gurobi Optimization, Inc. (2014) *Gurobi Optimizer Reference Manual.* `http://www.gurobi.com`.

[17] Hu, Z., Hong, L. J. & So, A. M. C. (2013a) Ambiguous probabilistic programs. Available online at:
`http://www.optimization-online.org/DB_FILE/2013/09/4039.pdf`.

[18] Hu, Z., & Hong, L. J. (2013b). Kullback-Leibler divergence constrained distributionally robust optimization. Available online at: `http://www.optimization-online.org/DB_FILE/2012/11/3677.pdf`.

[19] Huang, D., Zhu, S., Fabozzi, F. J. & Fukushima, M. (2010). Portfolio selection under distributional uncertainty: A relative robust CVaR approach. *European Journal of Operational Research*, Vol. 203(1), pp. 185-194.

[20] Jiang, R., & Guan, Y. (2013). Data-driven chance constrained stochastic program. Technical report, University of Florida. Available online at: `http://www.optimization-online.org/DB_FILE/2012/07/3525.pdf`.

[21] Klabjan, D., Simchi-Levi, D., & Song, M. (2013). Robust stochastic lot sizing by means of histograms. *Production and Operations Management*, Vol. 22(3), pp. 691-710.

[22] Mosler, K., & Bazovkin, P. (2012). Stochastic linear programming with a distortion risk constraint. *Discussion Papers in Statistics and Econometrics (No. 6/11).* Available online at: `http://arxiv.org/pdf/1208.2113.pdf`.

[23] Natarajan, K., Pachamanova, D. & Sim, M. (2009). Constructing risk measures from uncertainty sets. *Operations Research*, Vol. 57(5), pp. 1129-1141.

[24] Pichler, A. (2013) Evaluations of risk measures for different probability measures. *SIAM Journal on Optimization*, Vol. 23(1), pp. 530-551.

[25] Rockafellar, R. T. (1970) *Convex analysis.* (Princeton University Press)

[26] Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, Vol. 2, pp. 21-42.

[27] Ruehlicke, R. (2013) Robust risk management in the context of Solvency II Regulations. PhD Dissertation submitted at the Universität Duisburg-Essen.

[28] Thas, O.. (2010) *Comparing distributions.* (Springer)

[29] Tutuncu, R. H., & Koenig, M. (2004). Robust asset allocation. *Annals of Operations Research*, Vol. 132(1-4), pp. 157-187.

[30] Wang, Z., Glynn, P., & Ye, Y. (2013). Likelihood robust optimization for data-driven problems. Available online at:
`http://arxiv.org/pdf/1307.6279v2.pdf`.

[31] Wiesemann, W., Kuhn, D., & Sim, M. (2013). Distributionally robust convex optimization. Available online at:
http://www.optimization-online.org/DB_FILE/2013/02/3757.pdf.

[32] Wozabal, D. (2012). Robustifying convex risk measures: a non-parametric approach. Available online at:
http://www.optimization-online.org/DB_FILE/2011/11/3238.pdf

[33] Zhu, S., & Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, Vol. 57(5), pp. 1155-1168.

[34] Zymler, S., Kuhn, D., & Rustem, B. (2013). Worst-case Value-at-Risk of non-linear portfolios. *Management Science*, Vol. 59(1), pp. 172-188.

A Conjugates of the risk measures

A.1 Necessary lemmas

First result presented here is taken from [25] (see his Corollary 37.3.2). It allows us to interchange the inf and sup terms in the worst-case formulations of the Optimized Certainty Equivalent, mean absolute deviation from the median, variance less the mean, and standard deviation less the mean.

Lemma 2. [25, Corollary 37.3.2] Let C and D be nonempty closed convex sets in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively and let K be a continuous finite concave-convex function on $C \times D$. Then, if either C or D is bounded, one has:

$$\inf_{v \in D} \sup_{u \in C} K(u, v) = \sup_{u \in C} \inf_{v \in D} K(u, v). \qquad \square$$

For the derivation of the conjugate function of the standard deviation less the mean we also need the following result.

Lemma 3. [25, Theorem 16.3] Let B be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}$ be a concave function. Assume there exists an x such that $Bx \in \mathrm{ri}(\mathrm{dom} g)$. Then, it holds that:

$$(gB)_*(z) = \sup_y \left\{ g_*(y) \,\middle|\, B^T y = z \right\},$$

where for each z the supremum is attained, and where the function gB is defined by $(gB)(x) = g(Bx)$. $\qquad \square$

A.2 Standard deviation less the mean

In the case of the standard deviation less the mean we study the function:

$$f(p, w) = \sqrt{\sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \kappa \right)^2} - \alpha \sum_{n \in \mathcal{N}} p_n X_n(w).$$

We use Lemma 3 to obtain (after several transformations) that the concave conjugate is equal to (where for sake of readability the formulation is in a problem-like notation):

$$f_*(v, w) = \sup_{y} \quad -\tfrac{y}{4}$$

$$\text{s.t.} \quad \tfrac{1}{y} \begin{bmatrix} (X_1(w) - \kappa)^2 \\ \vdots \\ (X_N(w) - \kappa)^2 \end{bmatrix} \leq v - u$$

$$u_n = -\alpha X_n(w), \quad n \in \mathcal{N}$$

$$y \geq 0.$$

The first constraint can be reformulated using the results of [2] to obtain the following:

$$f_*(v, w) = \sup_{y} \quad -\tfrac{y}{4}$$

$$\text{s.t.} \quad \left\| \begin{bmatrix} X_n(w) - \kappa \\ \left(\frac{v_n - u_n - y}{2}\right) \end{bmatrix} \right\|_2 \leq \frac{v_n - u_n + y}{2}, \quad n \in \mathcal{N}$$

$$v_n - u_n \geq 0, \quad n \in \mathcal{N}$$

$$u_n = -\alpha X_n(w), \quad n \in \mathcal{N}$$

$$y \geq 0.$$

To obtain the final result (11) in the main text, the equality constraints are eliminated by inserting the equalities involving u_n into other expressions. This result is also obtained in Example 28 in [5].

B   Support functions of the uncertainty sets

B.1   Examples of ϕ-divergence functions

One of the types of uncertainty sets for the probabilities is defined using so-called ϕ-divergence functions. For the statistical background behind this tool we refer the reader to [6]. Table 5, adopted from [6], presents potential choices for the function $\phi(.)$ and its conjugate $\phi^*(.)$. Two of specific cases are commonly known. These are: (1) the Kullback-Leibler divergence which defines an uncertainty set based on the likelihood ratio statistical test, (2) the χ^2-distance which defines an uncertainty set based on the χ^2 goodness of fit test, also known as the Pearson test.

B.2   Derivations

ϕ-**divergence.**   For the ϕ-divergence function the uncertainty region is defined as $\mathcal{P}_q^\phi = \{p : p \geq 0, \quad g_i(p) \leq 0, \quad i = 1, 2, 3\}$, where

$$
\begin{aligned}
g_1(p) &= 1^T p - 1 \\
g_2(p) &= -1^T p + 1 \\
g_3(p) &= \sum_{n \in \mathcal{N}} q_n \phi\left(\tfrac{p_n}{q_n}\right) - \rho.
\end{aligned}
$$

31

Name	$\phi(t), \quad t \geq 0$	$\phi^*(s)$
Kullback-Leibler	$t \log t - t + 1$	$e^s - 1$
Burg entropy	$-\log t + t - 1$	$-\log(1 - s), \quad s < 1$
χ^2 distance	$\frac{1}{t}(t-1)^2$	$2 - 2\sqrt{1-s}, \quad s < 1$
Modified χ^2 distance	$(t-1)^2$	$\begin{cases} -1 & s < -2 \\ s + s^2/4 & s \geq -2 \end{cases}$
Hellinger distance	$(\sqrt{t} - 1)^2$	$\frac{s}{1-s}, \quad s < 1$
χ-divergence	$\|t - 1\|^\theta$	$s + (\theta - 1)\left(\frac{\|s\|}{\theta}\right)^{\theta/(\theta-1)}$
Variation distance	$\|t - 1\|$	$\max\{-1, s\}, \quad s \leq 1$
Cressie-Read	$\frac{1-\theta+\theta t - t^\theta}{\theta(1-\theta)}, \quad t \neq 0, 1$	$\frac{1}{\theta}(1 - s(1-\theta))^{\theta/(1-\theta)} - \frac{1}{\theta}, \quad s < \frac{1}{1-\theta}$

Now, the convex conjugates of these three functions over the domain $p \geq 0$ are needed.[2] We start with the function $g_1(.)$:

$$
\begin{aligned}
g_1^*(y) &= \sup_{p \geq 0} \left\{ y^T p - 1^T p + 1 \right\} \\
&= \sup_{p \geq 0} \left\{ (y-1)^T p + 1 \right\} \\
&= \begin{cases} 1 & \text{if } y - 1 \leq 0 \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

Analogously:

$$
g_2^*(y) = \begin{cases} -1 & \text{if } y + 1 \leq 0 \\ +\infty & \text{otherwise.} \end{cases}
$$

For the third function the derivation is:

$$
\begin{aligned}
g_3^*(y) &= \sup_{p \geq 0} \left\{ y^T p - \sum_{n \in \mathcal{N}} q_n \phi\left(\frac{p_n}{q_n}\right) + \rho \right\} \\
&= \sup_{p \geq 0} \left\{ \sum_{n \in \mathcal{N}} y_n p_n - q_n \phi\left(\frac{p_n}{q_n}\right) \right\} + \rho \\
&= \rho + \sum_{n \in \mathcal{N}} \sup_{p_n \geq 0} \left\{ y_n p_n - q_n \phi\left(\frac{p_n}{q_n}\right) \right\} \\
&= \rho + \sum_{n \in \mathcal{N}} \sup_{t \geq 0} q_n \left\{ y_n t - \phi(t) \right\} \\
&= \rho + \sum_{n \in \mathcal{N}} q_n \phi^*(y_n).
\end{aligned}
$$

---

[2]This part requires a separate remark because an equivalent way to derive the conjugate function would be not to reduce the domains of $g_i(.)$ to $p \geq 0$, but to include a functional constraint $p_n \geq 0$ for each $n \in \mathcal{N}$. However, the first way saves us some notation.

Lemma 1 gives:

$$\delta^* \left(v \left| \mathcal{P}_q^\phi \right. \right) = \inf_{\{u_i, v^i\}, i=1,2,3} \; u_1 - u_2 + u_3 \left(\rho + \sum_{n \in \mathcal{N}} q_n \phi^* \left(\frac{v_n^3}{u_3} \right) \right)$$

$$\text{s.t.} \quad \frac{v^1}{u_1} - 1 \le 0$$
$$\frac{v^2}{u_2} + 1 \le 0$$
$$\sum_{i=1}^{3} v^i = v$$
$$u_i \ge 0, \quad i = 1, 2, 3.$$

The equality constraint can be eliminated by inserting $v_n^3 = v_n - v_n^1 - v_n^2$ for each $n \in \mathcal{N}$. Together with a slight reformulation of the inequalities we get:

$$\delta^* \left(v \left| \mathcal{P}_q^\phi \right. \right) = \inf_{u_1, u_2, u_3, v^1, v^2} \; u_1 - u_2 + u_3 \left(\rho + \sum_{n \in \mathcal{N}} q_n \phi^* \left(\frac{v_n - v_n^1 - v_n^2}{u_3} \right) \right)$$

$$\text{s.t.} \quad v^1 \le u_1 \mathbf{1}$$
$$v^2 \le -u_2 \mathbf{1}$$
$$u_i \ge 0, \quad i = 1, 2, 3.$$

Since the functions $\phi^*(.)$ are nondecreasing, one can substitute $\eta = u_1 - u_2$ to obtain result (12) in the main text.

Kolmogorov-Smirnov. The relevant uncertainty set is:

$$\mathcal{P}_q^{\mathrm{KS}} = \left\{ p : p \ge 0, \quad \mathbf{1}^T p = 1, \quad \max_{n \in \mathcal{N}} \left| p^T \mathbf{1}^n - q^T \mathbf{1}^n \right| \le \rho \right\}.$$

Since all the constraints in the definition of $\mathcal{P}_q^{\mathrm{KS}}$ are linear in p, the Kolmogorov-Smirnov set can be defined as:

$$\mathcal{P}_q^{\mathrm{KS}} = \{ p : \quad p \ge 0, \quad Dp \le d \},$$

where $D \in \mathbb{R}^{(2N+2) \times N}$, $d \in \mathbb{R}^{2N+2}$ with:

$$
\begin{aligned}
&D_{1n} = 1, &&d_1 = 1, &&\forall n \in \mathcal{N} \\
&D_{2n} = -1, &&d_2 = -1, &&\forall n \in \mathcal{N} \\
&D_{2+n,i} = 1, &&d_{2+n} = \rho + q^T \mathbf{1}^n, &&\forall i \le n, \quad n \in \mathcal{N} \\
&D_{2+N+n,i} = -1, &&d_{2+N+n} = \rho - q^T \mathbf{1}^n, &&\forall i \le n, \quad n \in \mathcal{N},
\end{aligned}
$$

with the other components equal to 0. The support function is equal to:

$$\delta^* \left(v \left| \mathcal{P}^{\mathrm{KS}} \right. \right) = \sup_p \; v^T p$$

$$\text{s.t.} \quad Dp \le d$$
$$p \ge 0.$$

The final result (13) in the main text is obtained via strong LP duality.

Wasserstein. The definition of the Wasserstein set involves a variable matrix K, so that the set $\mathcal{U}$ is actually a set both in K and q. For that reason, we use an extended vector p' consisting of both these variables and 'extract' the vector p out of p' using a relevant A matrix. We take the extended vector to be:

$$p' = \left[p^T, K_1^T, K_2^T, ..., K_N^T \right]^T,$$

33

where $K_1, ..., K_N$ are the subsequent columns of K. A matrix A^W such that $A^W p' = p$ is given by $A^W = [I \,|\, 0_{N \times N^2}]$. Since the constraints in the definition of $\mathcal{P}_q^W$ are linear in (p, K), the Wasserstein set can be defined as:

$$\mathcal{U}_q^W = \{p' : \quad p' \geq 0, \quad Dp' \leq d\},$$

where $D \in \mathbb{R}^{(4N+3) \times N(N+1)}$, $d \in \mathbb{R}^{4N+3}$ and their entries are:

$$
\begin{array}{lll}
D_{1n} = 1, & d_1 = 1, & \forall n \in \mathcal{N} \\
D_{2n} = -1, & d_2 = -1, & \forall n \in \mathcal{N} \\
D_{3,Ni+j} = \|Y_i - Y_j\|^d, & d_3 = \rho, & \forall i, j \in \mathcal{N} \\
D_{3+n,n} = -1, & D_{3+n,Nn+i} = 1, & \forall i, n \in \mathcal{N} \\
D_{3+N+n,n} = 1, & D_{3+N+n,Nn+i} = -1, & \forall i, n \in \mathcal{N} \\
D_{3+2N+n,Ni+n} = 1, & d_{3+2N+n} = -q_n, & \forall i, n \in \mathcal{N} \\
D_{3+3N+n,Ni+n} = -1, & d_{3+3N+n} = q_n, & \forall i, n \in \mathcal{N},
\end{array}
$$

with the other components equal to 0. The support function is equal to:

$$
\begin{aligned}
\delta^* \left(\left(A^W\right)^T v \,\Big|\, \mathcal{U}_q^W \right) = \sup_{p'} \quad & v^T A^W p' \\
\text{s.t.} \quad & Dp' \leq d \\
& p' \geq 0.
\end{aligned}
$$

From here, the final result (14) is obtained via strong LP duality.

Combined set. We substitute $p' = \left[p^T, q^T\right]^T$ so that $p = A^C p'$, where $A^C = [I \,|\, 0_{N \times N}]$. The set $\mathcal{U}^C$ is then:

$$\mathcal{U}^C = \{p' : p' \geq 0, \quad g_i(p') \leq 0, \quad i = 1, 2, 3, \quad h_i(q) \leq 0, \quad i = 1, ..., Q\}.$$

The first three convex functions from formulation of $\mathcal{U}^C$ are:

$$
\begin{aligned}
g_1(p') &= 1^T p - 1 \\
g_2(p') &= -1^T p + 1 \\
g_3(p') &= \sum_{n \in \mathcal{N}} q_n \phi\left(\frac{p_n}{q_n}\right) - \rho.
\end{aligned}
$$

The conjugates of the first two have been obtained for the ϕ-divergence set. Thus, only the third one remains:

$$
\begin{aligned}
g_3^*(y) &= \sup_{p' \geq 0} \left\{y^T p' - g_3(p')\right\} \\
&= \sup_{p, q \geq 0} \left\{y_{1:N}^T p + y_{N+1:2N}^T q - \sum_{n \in \mathcal{N}} q_n \phi\left(\frac{p_n}{q_n}\right) + \rho\right\} \\
&= \sup_{q \geq 0} \left\{y_{N+1:2N}^T q + \sup_{p \geq 0} \left\{y_{1:N}^T p - \sum_{n \in \mathcal{N}} q_n \phi\left(\frac{p_n}{q_n}\right) + \rho\right\}\right\} \\
&= \sup_{q \geq 0} \left\{y_{N+1:2N}^T q + \sum_{n \in \mathcal{N}} q_n \sup_{u_n \geq 0} \left\{y_n u_n - \phi(u_n)\right\} + \rho\right\} \\
&= \sup_{q \geq 0} \left\{\sum_{n \in \mathcal{N}} q_n \left(y_{N+n} + \phi^*(y_n)\right) + \rho\right\} \\
&= \begin{cases} \rho & \text{for } y_{N+n} + \phi^*(y_n) \leq 0 \quad \forall n \in \mathcal{N} \\ +\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

Since all $h_i(.)$ depend only on q, the support function of $\mathcal{U}^C$ is given by (Lemma 1):

$$\delta^*\left(\left(A^C\right)^T v \,\Big|\, \mathcal{U}^C\right) = \inf \quad u_1 - u_2 + u_3\rho + \sum_{i=1}^{Q} u_{i+3}h_i^*\left(\frac{v_{N+1:2N}^{i+3}}{u_{i+3}}\right)$$

$$\text{s.t.} \quad v_{1:N}^1 \le u_1 1$$
$$v_{1:N}^2 \le -u_2 1$$
$$v_{N+1:2N}^i = 0, \quad i = 1,2,3$$
$$v_{1:N}^i = 0, \quad i = 4,...,Q+3$$
$$\frac{v_{N+n}^3}{u_3} + \phi^*\left(\frac{v_n^3}{u_3}\right) \le 0, \quad \forall n \in \mathcal{N}$$
$$\sum_{i=1}^{Q+3} v^i = v$$
$$u_i \ge 0, \quad i = 1,...,Q+3.$$

The only thing left is to remove nonconvexity from the constraint $\frac{v_{N+n}^3}{u_3} + \phi^*\left(\frac{v_n^3}{u_3}\right) \le 0$. One can do that by multiplying both sides by u_3 to obtain the final result.

Anderson-Darling. The relevant set formulation is (see Table 3):

$$\mathcal{P}_{\text{emp}}^{AD} = \{p : p \ge 0, \quad g_i(p) \le 0, \quad i = 1,2,3\},$$

where

$$g_1(p) = 1^T p - 1$$
$$g_2(p) = -1^T p + 1$$
$$g_3(p) = -N - \sum_{n \in \mathcal{N}} \frac{2n-1}{N}\left[\log\left(p^T 1^n\right) + \log\left(p^T 1^{-n}\right)\right] - \rho.$$

It is only necessary to derive the conjugate of $g_3(.)$. Let us write $g_3(.)$ as:

$$g_3(p) = \sum_{n \in \mathcal{N}}\left[-\left[\frac{2n-1}{N}\log\left(p^T 1^n\right) + \frac{\rho+N}{2N}\right] - \left[\frac{2n-1}{N}\log\left(p^T 1^{-n}\right) + \frac{\rho+N}{2N}\right]\right].$$

By results of [5], it is only needed to derive the convex conjugate of the function

$$H_n(t) = -\frac{2n-1}{N}\log(t) - \frac{\rho+N}{2N}, \quad t \ge 0.$$

It is given by:

$$H_n^*(s) = \sup_{t \ge 0}\left\{st + \frac{2n-1}{N}\log(t) + \frac{\rho+N}{N}\right\}$$
$$= \begin{cases} -\frac{2n-1}{N} - \frac{2n-1}{N}\log\left(\frac{-Ns}{2n-1}\right) + \frac{\rho+N}{2N} & \text{if } s < 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Using Lemma 1, we obtain:

$$\delta^*\left(v \,\Big|\, \mathcal{P}_{\text{emp}}^{AD}\right) = \inf_{\substack{\{w^{n+},w^{n-}\},n\in\mathcal{N}; \\ \{z_n^+,z_n^-\},n\in\mathcal{N}; \\ u_1,u_2,u_3,v^1,v^2}} \quad -\sum_{n \in \mathcal{N}} \frac{(2n-1)u_3}{N}\left[2 + \log\left(\frac{-Nz_n^+}{(2n-1)u_3}\right) + \log\left(\frac{-Nz_n^-}{(2n-1)u_3}\right)\right]$$
$$+ u_3(\rho + N) + u_1 - u_2$$

$$\text{s.t.} \quad z_n^+ 1^n = w^{n+}, \quad \forall n \in \mathcal{N}$$
$$z_n^- 1^{-n} = w^{n-}, \quad \forall n \in \mathcal{N}$$
$$v^1 \le u_1 1$$
$$v^2 \le -u_2 1$$
$$\sum_{n \in \mathcal{N}}(w^{n+} + w^{n-}) + v^1 + v^2 = v$$
$$z_n^+, z_n^- \le 0, \quad \forall n \in \mathcal{N}$$
$$u_1, u_2, u_3 \ge 0.$$

We eliminate the equalities involving w_n^+ and w_n^- to obtain:

$$\inf_{\substack{\{w^{n+},w^{n-}\},n\in\mathcal{N}, \\ u_1,u_2,u_3,v^1,v^2}} \quad -\sum_{n\in\mathcal{N}} \frac{(2n-1)u_3}{N}\left[2+\log\left(\frac{-Nz_n^+}{(2n-1)u_3}\right)+\log\left(\frac{-Nz_n^-}{(2n-1)u_3}\right)\right]$$

$$+u_3\left(\rho+N\right)+u_1-u_2$$

$$\text{s.t.} \quad v^1 \leq u_1 1$$

$$v^2 \leq -u_2 1$$

$$\sum_{n\in\mathcal{N}}(z_n^+ 1^n + z_n^- 1^{-n}) + v^1 + v^2 = v$$

$$z_n^+, z_n^- \leq 0, \quad \forall n \in \mathcal{N}.$$

$$u_1, u_2, u_3 \geq 0.$$

In the third constraint it is possible to change the equality into inequality because of the properties of the other constraints and the 'objective function'. Also, by the properties of the formulation above one can substitute $\eta = u_1 - u_2$ and remove the variables v^1, v^2. In this way result (16) in the main text is obtained.

Cramer-von Mises. The set definition is:

$$\mathcal{P}_{\text{emp}}^{\text{CvM}} = \left\{ p : p \geq 0, \quad 1^T p = 1, \quad \frac{1}{12N} + \sum_{n\in\mathcal{N}}\left[\frac{2n-1}{2N} - p^T 1^n\right]^2 \leq \rho \right\},$$

which can be reformulated as $\mathcal{P}_{\text{emp}}^{\text{CvM}} = \{p : \quad g_i(p) \leq 0, \quad i = 1, ..., N+3\}$, where

$$\begin{aligned}
g_1(p) &= 1^T p - 1 \\
g_2(p) &= -1^T p + 1 \\
g_3(p) &= p^T E p + b^T p + c \\
g_{3+n}(p) &= -p^T e^n, \quad \forall n \in \mathcal{N},
\end{aligned}$$

where

$$c = -\rho + \frac{1}{12N} + \sum_{n\in\mathcal{N}}\left(\frac{2n-1}{2N}\right)^2, \qquad b = \begin{bmatrix} -\sum_{j=1}^{N}\frac{2j-1}{N} \\ -\sum_{j=2}^{N}\frac{2j-1}{N} \\ \vdots \\ -\sum_{j=N}^{N}\frac{2j-1}{N} \end{bmatrix},$$

and where E denotes a matrix such $E_{ij} = N+1-\max\{i,j\}$ for $i,j \in \mathcal{N}$.[3] It is important to note that the matrix E is positive definite for all N and that its inverse has a tridiagonal structure, allowing for efficient computations.

We proceed to the derivations of the conjugates. These are:

$$\begin{aligned}
g_3^*(y) &= \sup_p \left\{ y^T p - p^T E p - b^T p - c \right\} \\
&= \sup_p \left\{ -p^T E p - (b-y)^T p - c \right\} \\
&= \tfrac{1}{4}(b-y)E^{-1}(b-y) - c,
\end{aligned}$$

---

[3]In this case we include the nonnegativity constraints for p as functions and theoretically allow p to be unconstrained. This approach makes the derivation of $g_3^*(.)$ easier.

and

$$g^*_{3+n}(y) = \sup_p \left\{ y^T p + p^T e^n \right\}$$

$$= \begin{cases} 0 & \text{if } y + e^n = 0 \\ +\infty & \text{otherwise} \end{cases}$$

for all $n \in \mathcal{N}$. The support function is equal to:

$$\delta^* \left(v \,\middle|\, \mathcal{P}^{\mathrm{CvM}}_{\mathrm{emp}} \right) = \inf_{\{u_i, v^i\}, i=1,\ldots,N+3} \quad u_1 - u_2 + \tfrac{1}{4} u_3 \left(b - \tfrac{v^3}{u_3} \right)^T E^{-1} \left(b - \tfrac{v^3}{u_3} \right) - u_3 c$$

$$\text{s.t.} \quad v^1 = u_1 \mathbf{1}$$
$$v^2 = -u_2 \mathbf{1}$$
$$v^{3+n} = -u_{3+n} e^n, \quad n \in \mathcal{N}$$
$$\sum_{i=1}^{N+3} v^i = v$$
$$u_i \geq 0, \quad i = 1, \ldots, N+3.$$

The 'objective function' in the above formulation, already convex in its arguments, can be transformed into a system of linear and second-order conic constraints. Indeed, one may introduce an extra variable $t \geq 0$ such that

$$u_3 \left(b - \tfrac{v^3}{u_3} \right)^T E^{-1} \left(b - \tfrac{v^3}{u_3} \right) \leq t \quad \Leftrightarrow \quad \frac{\left(u_3 b - v^3 \right)^T E^{-1} \left(u_3 b - v^3 \right)}{u_3} \leq t.$$

Then, introducing $z = u_3 b - v^3$ and $E^{-1} = P^T P$ (where P is a $N \times N$ matrix because of the positive definiteness of E) we obtain

$$\frac{\left(u_3 b - v^3 \right)^T E^{-1} \left(u_3 b - v^3 \right)}{u_3} \leq t \quad \Leftrightarrow \quad \frac{(Pz)^T (Pz)}{u_3} \leq t.$$

This can be transformed, using the results from [2], to:

$$\left\| \begin{bmatrix} Pz \\ \frac{t-u_3}{2} \end{bmatrix} \right\|_2 \leq \frac{t + u_3}{2}.$$

Implementing this and eliminating the equality constraints by inserting the equalities involving u_{3+n} into other places yields result (17) in the main text.

Watson test. The set definition is:

$$\mathcal{P}^{\mathrm{Wa}}_{\mathrm{emp}} = \left\{ p : p \geq 0, \mathbf{1}^T p = 1, \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} - p^T \mathbf{1}^n \right)^2 - N \left(\frac{1}{N} \sum_{n \in \mathcal{N}} p^T \mathbf{1}^n - \frac{1}{2} \right)^2 \leq \rho \right\},$$

where the last constraint can be formulated as in the case of the Cramer-von Mises set, with parameter values:

$$c = -\rho + \frac{1}{12N} + \sum_{n \in \mathcal{N}} \left(\frac{2n-1}{2N} \right)^2 - \frac{N}{4}, \qquad b = \begin{bmatrix} -\sum_{j=1}^{N} \frac{2j-1}{N} + N \\ -\sum_{j=2}^{N} \frac{2j-1}{N} + (N-1) \\ \vdots \\ -\sum_{j=N}^{N} \frac{2j-1}{N} + 1 \end{bmatrix},$$

37

and a $N \times N$ matrix E such that $E_{i,j} = N + 1 - \max\{i, j\} - \frac{(N+1-i)(N+1-j)}{N}$ for all $i, j \in \mathcal{N}$. One can prove that the matrix E is positive semidefinite with a one-dimensional nullspace.

We proceed to the derivation of the support function $g_3^*(.)$. It is:

$$
\begin{aligned}
g_3^*(y) &= \sup_p \left\{ y^T p - p^T E p - b^T p - c \right\} \\
&= \sup_p \left\{ -p^T E p - (b - y)^T p - c \right\} \\
&= \begin{cases} \frac{1}{4}(b - y) E^\dagger (b - y) - c & \text{if } (b - y) \in \mathrm{Im} E \\ +\infty & \text{otherwise,} \end{cases}
\end{aligned}
$$

where $E^\dagger$ denotes a pseudo-inverse of E and $\mathrm{Im} E$ denotes the subspace spanned by the columns of E. From here on, the derivation is analogous to the case of the Cramer-von Mises test, with an extra constraint $(b - y) \in \mathrm{Im} E$, implemented as $\exists \lambda$ s.t. $b - y = E\lambda$.

Kuiper test. The Kuiper set is defined by

$$
\mathcal{P}_{\mathrm{emp}}^{\mathrm{K}} = \left\{ \max_{n \in \mathcal{N}} \left(\frac{n}{N} - p^T 1^n \right) + \max_{n \in \mathcal{N}} \left(p^T 1^{n-1} - \frac{n-1}{N} \right) \le \rho \right\}.
$$

Using additional variables z_1, z_2 it can be transformed to

$$
\begin{aligned}
\mathcal{U}_{\mathrm{emp}}^{\mathrm{K}} = \Big\{ (p, z_1, z_2) : 1^T p = 1, \quad & z_1 + z_2 \le \rho, \\
\max_{n \in \mathcal{N}} \left(\frac{n}{N} - p^T 1^n \right) \le z_1, \quad & \max_{n \in \mathcal{N}} \left(p^T 1^{n-1} - \frac{n-1}{N} \right) \le z_2 \Big\}.
\end{aligned}
$$

Thus, we use a vector $p' = \left[p^T, z_1, z_2 \right]^T$ and a matrix $A^{\mathrm{K}} = [I \,|\, 0_{N \times 2}]$. The set $\mathcal{U}_{\mathrm{emp}}^{\mathrm{K}}$ is then:

$$
\mathcal{U}_{\mathrm{emp}}^{\mathrm{K}} = \left\{ p' : p' \ge 0, \quad Dp' \le d \right\},
$$

where $D \in \mathbb{R}^{(2N+3) \times (N+2)}$, $d \in \mathbb{R}^{2N+3}$ are defined by:

$$
\begin{array}{llll}
D_{1,n} = 1, & d_1 = 1, & & \forall n \in \mathcal{N} \\
D_{2,n} = -1, & d_2 = -1, & & \forall n \in \mathcal{N} \\
D_{2+n,i} = -1, & D_{2+n,N+1} = -1, & d_{n+2} = -n/N, & \forall i \le n, n \in \mathcal{N} \\
D_{N+2+n,i} = 1, & D_{N+2+n,N+2} = -1, & d_{N+2+n} = (n-1)/N, & \forall i \le n-1, n \in \mathcal{N} \\
D_{2N+3,N+1} = 1, & D_{2N+3,N+2} = 1, & d_{2N+3} = \rho, &
\end{array}
$$

with all other components equal to 0. The final form (19) in the main text is obtained via strong LP duality.