Computationally Tractable Counterparts of Distributionally Robust Constraints on Risk Measures (revision of CentER DP 2014-031)

Postek, K.S.; den Hertog, D.; Melenberg, B.

Publication date:
2015

Document Version
Early version, also known as pre-print

CentER

Discussion Paper

No. 2015-047

COMPUTATIONALLY TRACTABLE COUNTERPARTS OF DISTRIBUTIONALLY ROBUST CONSTRAINTS ON RISK MEASURES

By

Krzysztof Postek, Dick den Hertog,
Bertrand Melenberg

28 September 2015

This is a revised version of CentER Discussion Paper

No. 2014-031

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

Computationally tractable counterparts of distributionally robust constraints on risk measures

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

September 24, 2015

Abstract

In optimization problems appearing in fields such as economics, finance, or engineering, it is often important that a risk measure of a decision-dependent random variable stays below a prescribed level. At the same time, the underlying probability distribution determining the risk measure's value is typically known only up to a certain degree and the constraint should hold for a reasonably wide class of probability distributions. In addition to that, the constraint should be computationally tractable. In this paper we review and generalize results on the derivation of tractable counterparts of such constraints for discrete probability distributions. Using established techniques in robust optimization, 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. This holds for a wide range of risk measures and uncertainty sets for probability distributions defined using statistical goodness-of-fit tests or probability metrics. In this way, we provide a unified framework of reformulating this class of constraints, extending the number of solvable risk measure-uncertainty set combinations considerably, including also risk measures that are nonlinear in the probabilities. To provide a clear overview for the user, we give the computational tractability status for each of the uncertainty set-risk measure pairs of which some have been solved in the literature. Examples, including portfolio optimization and antenna array design, illustrate the proposed approach in a theoretical and numerical setting.

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

1 Introduction

Robust Optimization (RO, see Ben-Tal et al. [9]) has become one of the main approaches to optimization under uncertainty. Since its introduction it has already

[*]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

found numerous applications. Its paradigm lies in assuming that some or all of the optimization problem's parameters are uncertain. For these parameters an uncertainty set is specified. This uncertainty set includes parameter values for which the solution should be feasible. The robust optimization problem is then solved in such a way that the solution is best possible for the worst-case parameter values from the uncertainty set. For an introduction and overview of techniques used in RO, we refer the reader to the work by Bertsimas et al. [13] and references therein. A more recent survey of the developments and applications of RO is Gabrel et al. [26].

A particular application field for RO is keeping risk measures of decision-dependent random variables below pre-specified limits. Such constraints typically appear 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 estimated. Such an estimate is typically based on a number of past observations. Due to sampling error, this estimate approximates the true distribution only with a limited accuracy. A confidence set around the estimate gives rise to a natural uncertainty set of admissible probability distributions (at a given confidence level). The key difficulty lies in reformulating the problem's constraints in a way that allows for the application of efficient optimization algorithms. Such a reformulation is referred to as a tractable robust counterpart of the constraint (see Ben-Tal et al. [9]). Many authors study this type of distributional uncertainty and the corresponding tractable robust counterparts. In Section 2 we present a review of existing results. Typically, they focus on a specific combination of risk measure and uncertainty set.

In this paper, apart from providing an overview of the results in the literature, we propose a unified approach to derive computationally tractable robust counterparts of this kind of constraints. Our approach allows us to deal with many more risk measure-uncertainty set combinations than have been considered up to now. The unifying approach of our paper consists of the following three parts.

First, using Fenchel duality and results of Ben-Tal et al. [10] we show that the derivation of the tractable robust counterpart can be separated in terms of the components corresponding to the risk measure and the uncertainty set. 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 provides the decision maker with a unified structure to reformulate this type of constraints, allowing to cover many more risk measure-uncertainty set combinations than is captured up to now in the literature. The first building block includes the following risk measures: 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 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.

Secondly, we address a common feature rendering many optimization constraints computationally difficult, namely, nonlinearity of some of the risk measures in the underlying probability distribution. In our setting, this is the case for 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 use different, equivalent reformulations 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.

Thirdly, 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.

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 reformulate the relevant constraints into a tractable form within our framework. Section 5 contains a brief discussion of reasons why our approach is not applicable to such cases.

We remark that another type of distributional uncertainty in RO problems is uncertainty in the moments of some key random variables, as studied, e.g., by El Ghaoui et al. [21] and Delage and Ye [19]. However, in our paper we focus on uncertainty in the discrete probabilities of these random variables.

The composition of the remainder of the paper is as follows. Section 2 provides a survey of the results obtained in the literature so far. Section 3 introduces the definitions and the main tool for deriving the computationally tractable robust counterparts. Section 4 lists the risk measures and uncertainty sets for the probability distribution that we investigate. Sections 5 and 6 include the results on the building blocks of the robust counterparts of the constraints on the risk measures. In Section 7, we provide examples of combining the blocks, including a numerical study. Section 8 concludes and lists the potential directions for future research.

2 Literature review

In this section we present the results of existing research on deriving tractable forms of distributionally robust constraints on risk measures, grouped according to the type of uncertainty sets, which is in line with Table 1. To provide the reader with a better understanding of the literature review, we first present as example a constraint on the standard deviation.

In our setting, it is assumed that the decisions form a vector w and the decision-dependent random variable $X(w)$, representing the decision maker's reward, has possible outcomes $X_1(w), \dots, X_N(w)$ with probabilities $p_1, \dots, p_N$, respectively. The probability vector $p = [p_1, \dots, p_N]^T$ is known to belong to a confidence region $\mathcal{P}$. For such a random variable, we want its standard deviation not to be greater

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, or that the tractable robust counterpart can be obtained from authors' results with some additional work. 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 (2)) 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 3).

Uncertainty set → **Risk measure** ↓	ϕ-divergences	Pearson	Likelihood ratio	Kolmogorov-Smirnov	Wasserstein (Kantorovich)	Anderson-Darling	Cramer-von Mises	Watson	Kuiper
Negative mean return	• CP [10],[11],[16]	• SOCP [10][11][38][55]	• CP [10][11][34][54]	∘ LP [55]	• LP [45][55][56]	• CP [10]	∘ SOCP [55]	∘ SOCP [55]	∘ LP [55]
Standard deviation less the mean	CP	SOCP	CP	SOCP	• SOCP [56]	CP	SOCP	SOCP	SOCP
Standard deviation	CP	SOCP	CP	SOCP	• SOCP [56]	CP	SOCP	SOCP	SOCP
Variance less the mean	CP	SOCP	∘ CP [17]	QP	QP	CP	SOCP	SOCP	QP
Variance	∘ CP [10]	∘ SOCP [10]	∘ CP [10][17]	QP	QP	∘ CP [10]	SOCP	SOCP	QP
Mean absolute deviation from the median	CP	SOCP	CP	LP	• LP [56]	CP	SOCP	SOCP	LP
Sharpe ratio	CP$*$	SOCP$*$	CP$*$	SOCP$*$	SOCP$*$	CP$*$	SOCP$*$	SOCP $*$	SOCP $*$
Lower partial moment $\alpha = 1$	∘ CP [10],[11]	∘ SOCP [10][11][38][55]	∘ CP [10][11][34][54]	∘ LP [55]	∘ LP [55]	∘ CP [10]	∘ SOCP [55]	∘ SOCP [55]	∘ LP [55]
Lower partial moment $\alpha = 2$	∘ CP [10],[11]	∘ SOCP [10][11][38]	∘ CP [10][11][34][54]	QP	QP	∘ CP [10]	SOCP	SOCP	QP
Certainty Equivalent	CP$*$	CP$*$	∘ CP $*$	CP$*$	CP$*$	CP$*$	CP$*$	CP$*$	CP$*$
Shortfall risk	∘ CP [10],[11]	∘ CP [10][11][38]	∘ CP [10][11][34][54]	CP	CP	∘ CP [10]	CP	CP	CP
Optimized Certainty Equivalent	CP	CP	CP	CP	CP	CP	CP	CP	CP
Value-at-Risk	∘ [35]	∘ [35]	∘ [34][35][37]						
Conditional Value-at-Risk	CP	∘ SOCP [14][55]	• CP [14][34]	∘ LP [14][55]	• LP [45][55][56]	∘ CP [14]	∘ SOCP [14][55]	∘ SOCP [14][55]	∘ LP [14][55]
Entropic Value-at-Risk	CP	CP	CP	CP	CP	CP	CP	CP	CP
Mean absolute deviation from the mean			∘ [17]						
Distortion risk measures					• [45][56]				
Coherent risk measures									
Spectral risk measures									

than β for all possible probability distributions in $\mathcal{P}$. The constraint is then:

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

The key question is whether (1) can be reformulated into an equivalent form that does not have the semi-infinite 'for all' form. Such combinations of uncertain discrete probabilities and risk measures have already been investigated by several authors.

Calafiore [17] studies a portfolio optimization problem with risk defined as the variance less the mean and mean absolute deviation from the mean, under distributional uncertainty defined with the use of Kullback-Leibler divergence. He notices that the step of finding the worst-case probability distribution for a given portfolio can be conducted efficiently. Combining it with the generation of cutting planes for the general robust optimization problem, he proposes an algorithm that finds the optimal portfolios in polynomial time.

Jiang and Guan [37] consider ambiguous chance constraints under the Kullback-Leibler divergence, reducing the robust chance-constrained problem to a problem under the nominal probability measure, with modified violation probability. In our setting, these results have applications to constraints on the Value-at-Risk (VaR), whose equivalence to chance constraints has been noted already by Nemirovski and Shapiro [44]. Hu and Hong [34] also consider optimization problems with convex expectation constraints under distributional uncertainty defined with the Kullback-Leibler divergence. They provide closed-form distributionally robust counterparts of constraints on expectations of general convex performance measures. Their results apply, for example, to Conditional Value-at-Risk (CVaR) and, as an approximation, to VaR. Results of Hu and Hong [34] are partly generalized in Hu et al. [35], who consider chance-constrained problems with distributional uncertainty sets defined by general ϕ-divergence functions. They also show that the robust constraints are equivalent to nominal constraints with modified confidence levels. Another work, that does not concentrate on risk measures as such, including derivations of tractable robust counterparts under distributional uncertainty, is Klabjan et al. [38]. In this work, a lot-sizing problem with uncertainty defined with the χ^2-test statistic is solved. After an appropriate reformulation, the problem solved is a SOCP problem.

Wang et al. [54] derive tractable counterparts of constraints involving linear functions of the probability vector, with uncertainty defined by the likelihood ratio test. They also provide various interpretations of the obtained result, for example, from the Bayesian perspective. Ben-Tal et al. [11] study constraints on general convex functions under distributional uncertainty defined by ϕ-divergence functions, generalizing the result of Wang et al. [54]. They derive tractable robust counterparts, showing that for several of the divergences the resulting constraint allows for a self-concordant barrier function. Their results, as a specific case, apply to such risk measures as the negative mean return. Methods for obtaining worst-case probability distributions under ϕ-divergence uncertainty are given also in Breuer and Csiszár [16].

Wozabal [56] studies portfolio optimization with risk measures such as expectation, standard deviation less the mean, mean absolute deviation from the median, CVaR, distortion risk measure, Wang transform, proportional hazards transform, and the Gini measure, under distributional uncertainty defined with the Wasserstein distance. Using the so-called subdifferential representation of risk measures, he derives closed-form worst-case values of risk measures. For the first three risk measures the resulting worst-case expressions are linear, convex quadratic, or piecewise linear in the decision variables. Pichler [45] focuses on the worst-case values of general spectral and distortion risk measures under distributional uncertainty defined with the Wasserstein distance. He provides expressions for the so-called transport maps that define the worst-case probability distributions for given values of the decision variables. However, in a general case these formulas cannot be implemented easily because of their nonlinear (nonconvex) forms.

There are also several works studying risk measures or uncertainty sets different from the ones we consider. Zhu and Fukushima [57] analyze the CVaR under box and ellipsoidal uncertainty sets for discrete probabilities. Using the min-formulation for CVaR from Rockafellar and Uryasev [47] and minmax results from convex analysis, they show that the problem of minimizing the worst-case CVaR in such a case can be formulated as LP (for box uncertainty) or SOCP (for ellipsoidal uncertainty). Fertis et al. [24] study the CVaR under an uncertainty set with a two-stage structure. In this structure, the uncertainty set is defined as a ball with an arbitrary norm around a reference probability distribution. This reference probability distribution is allowed to be a convex combination of a finite number of known probability distributions. In this way, the authors generalize the results of Zhu and Fukushima [57] to the continuous case, showing that a constraint on the CVaR can be reformulated tractably to a system of constraints involving dual norms.

Huang et al. [36] propose a framework replacing the standard CVaR by a less conservative measure, namely the Relative Robust CVaR, and show that under a multiple-expert uncertainty set the resulting optimization problem can be reformulated either as LP or as SOCP. Bazovkin and Mosler [42] 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.

Wiesemann et al. [55] propose a general framework of distributionally robust convex optimization. They require that the function to be constrained is bilinear in the decision variables and the random vector, imposing an uncertainty set that must possess a conic representation, with some regularity conditions. The reformulation they provide applies to multiple risk measures and conic-representable uncertainty sets for the probabilities, see Table 1 for an overview.

Ben-Tal et al. [10] provide the mathematical framework used in our paper. They study general nonlinear robust constraints reformulated using Fenchel duality. Results of their paper allow to obtain tractable constraints for the variance with distributional uncertainty defined by ϕ-divergences and the Anderson-Darling goodness of fit tests. In both cases the resulting system of constraints is convex, and for some of the ϕ-divergence functions it is second-order conic. In our paper we extend their framework to other types of uncertainty sets and risk measures.

Bertsimas et al. [14] show how in a data-driven setting one can construct uncertainty sets based on statistical tests such as Kolmogorov-Smirnov, χ^2, Anderson-Darling, Watson and likelihood ratio, used to obtain conservative bounds on the VaR via CVaR. They utilize a cutting plane algorithm with an efficient method of evaluating the worst-case values of the decision-dependent random variables.

Finally, two works, Natarajan et al. [43] and Bertsimas and Brown [12] provide a more general insight into the relation between robust optimization and risk measurement. They show that there is a one-to-one relationship between coherent risk measures (see Artzner et al. [4] for an introduction) and uncertainty sets for general uncertain parameters in the case of constraints that are linear in this uncertainty.

3 Preliminaries

As already introduced in Section 2, we study constraints on risk measures of decision-dependent random variables, where $w \in \mathbb{R}^M$ is the decision vector. The random variable $X(w)$, representing the decision maker's reward, and whose risk is measured, takes a value $X_n(w)$ with probability p_n for each $n \in \mathcal{N} = \{1, \ldots, N\}$. The uncertain parameter is the probability vector $p = [p_1, ..., p_N]^T \in \mathbb{R}^N_+$, representing the discrete distribution of $X(w)$. The reference probability vector, around which the uncertainty set for p may be specified, is denoted by $q \in \mathbb{R}^N_+$.

Let the risk measure of the random variable $X(w)$ under the probability distribution represented by the vector p be given by $\overline{F}(p, X(w))$, with $\overline{F} : \mathbb{R}^N_+ \times \mathbb{R}^N \to \mathbb{R}$. The robust constraint on the risk measure that we shall reformulate to a tractable form is given by:

$$F(p, w) = \overline{F}(p, X(w)) \leq \beta, \quad \forall p \in \mathcal{P}, \tag{2}$$

where $F : \mathbb{R}^N_+ \times \mathbb{R}^M \to \mathbb{R}$ and $\mathcal{P}$ is the uncertainty set for the probabilities defined as:

$$\mathcal{P} = \{p : \quad p = Ap', \quad p' \in \mathcal{U}\}, \tag{3}$$

where the set $\mathcal{U} \subseteq \mathbb{R}^L_+$ is a nonempty, compact convex set, and $A \in \mathbb{R}^{N \times L}$ such that $\mathcal{P} \subseteq \mathbb{R}^N_+$. The 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 standard deviation and the uncertainty set is defined using a ϕ-divergence function around the reference probability vector q (see Table 3), then the constraint is:

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

with $A = I$ and

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

The robust constraint of Example 1 is reformulated to a computationally tractable form in Section 7.1.

We now introduce the key theorem used in this paper to construct a unified framework for tackling constraints involving a risk measure and an uncertainty set for probabilities. First, we give the definition of the conjugate functions and the support function, adapted to our context. The *concave conjugate* $f_*(\cdot)$ of a function $f : \mathbb{R}^N_+ \to \mathbb{R}$ is defined as a function $f_* : \mathbb{R}^N \to \mathbb{R} \cup \{-\infty\}$:

$$f_*(v) = \inf_{p \in \mathbb{R}^N_+} \left\{ v^T p - f(p) \right\}. \tag{4}$$

The *convex conjugate* $g^*(\cdot)$ of a function $g : \mathbb{R}^N_+ \to \mathbb{R}$ is defined as a function $g^* : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$:

$$g^*(v) = \sup_{p \in \mathbb{R}^N_+} \left\{ v^T p - g(p) \right\}. \tag{5}$$

Remark 1. In the above definitions, the domains of $f(\cdot)$ and $g(\cdot)$ are given by $\mathbb{R}^N_+$, and the corresponding conjugates are defined as an infimum/supremum over $\mathbb{R}^N_+$, instead of $\mathbb{R}^N$. Our approach can easily be adapted in a way that fits the standard definitions by setting the values of $f(\cdot)$ and $g(\cdot)$ equal to $-\infty$ and $+\infty$ outside $\mathbb{R}^N_+$, respectively, so that $\mathbb{R}^N_+$ is the effective domain. □

The indicator function $\delta(\cdot|\mathcal{U})$ of a nonempty set $\mathcal{U} \subseteq \mathbb{R}^L_+$ is defined as

$$\delta(p'|\mathcal{U}) = \begin{cases} 0 & \text{if } p' \in \mathcal{U} \\ +\infty & \text{otherwise,} \end{cases}$$

and the *support function* $\delta^*(\cdot|\mathcal{U}) : \mathbb{R}^L \to \mathbb{R} \cup \{\infty\}$ of a nonempty set $\mathcal{U} \subseteq \mathbb{R}^L_+$ is defined as a convex conjugate of the indicator function:

$$\delta^*(v|\mathcal{U}) = \sup_{p' \in \mathbb{R}^L_+} (v^T p' - \delta(p'|\mathcal{U})) = \sup_{p' \in \mathcal{U}} v^T p'. \tag{6}$$

The following theorem, adapted from Ben-Tal et al. [10], 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 $f(\cdot, w)$ is closed and concave for each $w \in \mathbb{R}^M$. Consider a constraint of the form:

$$f(p, w) \leq \beta, \quad \forall p \in \mathcal{P}, \tag{7}$$

where $\mathcal{P}$ is defined by (3) and where it holds that:

$$\mathrm{ri}(\mathcal{P}) \cap \mathbb{R}^N_{++} \neq \emptyset. \tag{8}$$

Then (7) holds for a given w if and only if:

$$\exists v \in \mathbb{R}^N : \quad \delta^* \left(A^T v \middle| \mathcal{U} \right) - f_*(v, w) \leq \beta, \tag{9}$$

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

Proof. The proof relies on the Fenchel duality theorem, included as Theorem 2 in Appendix A. Constraint (7) is equivalent to:

$$G(w) = \sup_{p \in \mathcal{P}} \{ f(p,w) - \delta(p\,|\mathcal{P}) \} \leq \beta.$$

We have that

$$\begin{aligned} G(w) &= \sup_{p \in \mathcal{P}} \{ f(p,w) - \delta(p\,|\mathcal{P}) \} \\ &= \inf_{v} \{ \delta^*(v|\,\mathcal{P}) - f_*(v,w) \} \\ &= \inf_{v} \left\{ \delta^*\left(A^T v \middle| \mathcal{U}\right) - f_*(v,w) \right\}, \end{aligned}$$

where the second equality follows from Fenchel duality. Moreover, due to condition (8) the infimum is attained, see Appendix A. Because of this, the constraint

$$\inf_{v} \left\{ \delta^*\left(A^T v \middle| \mathcal{U}\right) - f_*(v,w) \right\} \leq \beta$$

is equivalent to (9), obtained by removing the inf term. □

We argue in Section 4.2 that for the uncertainty sets considered in this paper condition (8) holds under mild conditions on the vector q. For cases where $f(\cdot, w)$ is not closed or concave, (9) is a conservative approximation of (7), which follows from weak Fenchel duality.

Theorem 1 allows for a separation of the derivation of two components: (i) the support function of the uncertainty set $\mathcal{U}$ at the point $A^T v$, and (ii) the concave conjugate of $f(\cdot, w)$. Moreover, it is not necessary to have closed-form formulations of the two components. For example, if one can express the support function as an infimum of some convex function over a set of parameter values, then the inf symbol can be removed after inserting such a formulation into (9), due to its position on the left hand side of the constraint. As we shall see in further sections, we often make use of this property.

If the function $F(\cdot, w)$ in (2) satisfies the concavity assumption with respect to p and we can obtain its conjugate directly from (5), then we can use Theorem 1 to reformulate the semi-infinite constraint (2) as a finite constraint of the type (9), setting $f(p, w) = F(p, w)$.

On the other hand, if the concavity assumption is not satisfied or the standard form of $F(\cdot, w)$ is too difficult to obtain a tractable conjugate, then we try to choose another function $f(\cdot, \cdot)$ such that (2) and (7) are equivalent, and Theorem 1 can be used.

Remark 2. The form of the left-hand side in (9) is particularly useful when the conjugate and support functions are not available as closed-form expressions, but instead, are formulated as supremums and infimums, respectively:

$$f_*(v,w) = \sup_{\lambda \in \Lambda(v,w)} \overline{f}(v,w,\lambda), \quad \delta^*(A^T v|\mathcal{U}) = \inf_{\theta \in \Theta(v)} \overline{g}(v,\theta),$$

where $\Lambda(v,w) \subseteq \mathbb{R}^{n_\Lambda}$ and $\Theta(v) \subseteq \mathbb{R}^{n_\Theta}$. Inserting these formulations into (9) yields:

$$\inf_{\theta \in \Theta(v)} \overline{g}(v,\theta) - \sup_{\lambda \in \Lambda(v,w)} \overline{f}(v,w,\lambda) \leq \beta.$$

Under the condition that the infimum of the left-hand side is attained, satisfied in examples considered in this paper, we obtain an equivalent formulation by removing the inf and sup symbols, including the relevant constraints as:

$$\exists (\lambda,\theta) \in \mathbb{R}^{n_\Lambda} \times \mathbb{R}^{n_\Theta} : \quad \begin{cases} \overline{g}(v,\theta) - \overline{f}(v,w,\lambda) \leq \beta \\ \lambda \in \Lambda(v,w) \\ \theta \in \Theta(v). \end{cases}$$

□

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

Additional notation

We distinguish the vectors by using the superscripts and the components of a vector using subscripts. For example, v^i_k 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 the components indexed 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.

For formulas such as $f(p) = \sup_{h_i(x) \leq 0,\ i \in \mathcal{I}} g(x,p)$, we make use of the following layout:

$$\begin{aligned} f(p) = \sup \quad & g(x,p) \\ \text{s.t.} \quad & h_i(x) \leq 0,\ i \in \mathcal{I}, \end{aligned}$$

in situations where the terms under the sup/inf symbol would make the formulation difficult to read.

4 Risk measures and uncertainty sets

4.1 Risk measures

Risk measures, as concise quantifiers of the riskiness of random variables, can be designed for a great variety of purposes. The most important fields of their application are finance and economics (see, e.g., Schied and Föllmer [51] and Dowd [20]), linear regression in statistics (see, e.g., Rockafellar et al. [48]), supply chain management (see, e.g., Ahmed et al. [3]), engineering (see, e.g., Rockafellar and Royset [49]), and medicine (see, e.g., Chan et al. [18]).

The choice of a risk measure depends on the risk characteristics one is interested in and is therefore likely to be application-specific. Also, each risk measure has its implications in terms of the tractable robust counterpart's complexity, as indicated in Table 1. In Table 2 we list a wide range of examples, exhausting a large share of practical applications. In their formulations, we follow the convention that the random variable $X(w)$ represents the reward and 'the smaller the risk measure, the better' (as a requirement for risk measures, known as the monotonicity axiom, see Artzner et al. [4]):

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

As an example, the first risk measure in Table 2 is the negative mean return, instead of its positive counterpart. In the remainder of this section we briefly introduce the included risk measures according to the aspect of risk they measure.

Some of the risk measures in Table 2 quantify the dispersion of the random variable $X(w)$ around a given 'central' level, such as the mean or the median. These risk measures include the variance, the standard deviation, and the mean absolute deviation from the median. The difference between, for example, the standard deviation and the variance lies in contributions made by deviations of different magnitude.

Variance or standard deviation minus the mean multiplied by a constant are popular in finance as they represent the classical 'return minus risk' performance measure, see Markowitz [41]. The Sharpe ratio (see Sharpe [52]), defined as the proportion of the mean of the random variable to its standard deviation is also popular in the financial context as a measure of the assets' riskiness. Typically, the mean used then is the mean excess return of an asset above the return of a riskfree asset. The combination of the standard deviation and the mean can also be used in engineering to ensure that the value of some random variable is not greater than a prescribed level by at least 'some number of standard deviations'.

Lower partial moments are useful when a one-sided deviation of the random variable around a specified level is important, as can be the case with losses in a financial setting. Lower partial moments with $\alpha = 1$ and $\alpha = 2$ differ in the contribution made by deviations of different magnitude. The name 'lower partial moment' is just a convention and one can analyze similarly the upper partial moments of the variables.

To study not only the deviation around some given level, but the overall riskiness of a variable $X(w)$ in an economic context, risk measures involving the agent's utility function $u(\cdot)$ are used as well. These are (1) the Certainty Equivalent - the negative of the 'sure amount for which a decision maker remains indifferent to the outcome of random variable $X(w)$' (see Ben-Tal and Teboulle [8]), (2) the Shortfall risk - the minimum amount of additional resources needed to make the expected utility of a decision maker from his portfolio nonnegative, and (3) the Optimized Certainty Equivalent - representing the optimal allocation of $X(w)$ between present and future consumption.

Yet another type of risk measures are the tail-oriented risk measures. The most popular of them, the Value-at-Risk, represents the negative of the left α-th quantile of the distribution of $X(w)$. A constraint imposed on the VaR is equivalent to a chance constraint, as noted in Nemirovski and Shapiro [44] and used in Bertsimas et al. [14], Jiang and Guan [37], Hu and Hong [34], and Hu et al. [35]. However, efficient optimization of the VaR is difficult unless distributional assumptions are made. This problem can be mitigated by using instead of it a special case of the Optimized Certainty Equivalent, the Conditional Value-at-Risk, which represents the negative of the average of the worst $100\alpha\%$ outcomes of $X(w)$. Nemirovski and Shapiro [44] have shown the CVaR to be the best conservative convex approximation of the VaR. In Table 2 we use the formulation of CVaR adopted from Rockafellar and Uryasev [47]. Another advantage of the CVaR compared to the VaR is that it provides information about the mean of the least $100\alpha\%$ worst positions, instead of only the largest of them as is the case for the VaR.

Recently, an upper bound on both the VaR and the CVaR has been proposed by Ahmadi-Javid [2] - the Entropic Value-at-Risk. Its definition in Table 2 requires a separate comment since it does not involve p. Instead, EVaR is defined as 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 $f(p, w)$ (given in Section 5) consistent with the terminology of Theorem 1.

Table 2 includes also general risk measure classes whose definitions are based on axioms that risk measures should satisfy. In particular, these are the spectral, coherent, and convex risk measures, which are increasing classes in the sense of set inclusion. For a discussion of the differences between these three types of risk measures we refer the reader to Acerbi [1] and Föllmer and Schied [25].

Some of the measures in Table 2 are specific cases of the other ones: for instance, the CVaR is an example of an Optimized Certainty Equivalent. Nevertheless, a distinction has been made because of the popularity of the use of the 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.

For a further reference, each of the papers mentioned in Table 1 includes also a discussion of the applications of the relevant risk measures. For a broader overview of risk measurement and possible choices for risk measures we refer the reader to Embrechts et al. [22], Dowd [20], and Rockafellar and Royset [49].

4.2 Uncertainty sets for the probabilities

Distributional uncertainty has been studied up to now mostly in finance (see, e.g. Calafiore [17] and other references in Table 1), insurance (see, e.g., Klugman et al. [39]), economics (denoted in this context as ambiguity, see, e.g., Epstein [23]), and machine learning (see, e.g., Gotoh and Uryasev [28]).

Since some of the uncertainty sets are constructed using information on the outcomes of an underlying random vector, we assume that $X_n(w)$ corresponds to the outcome $Y^n \in \mathbb{R}^{M_Y}$ of some underlying random vector Y. Table 3 presents the uncertainty sets for the discrete probabilities studied in this paper.

Table 2: Risk measures analyzed in this 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 $X(w)$ as induced by p. We define the left α-quantile of a distribution of $X(w)$ as $G^{-1}_{X(w)}(\alpha) = \inf\{\kappa \in \mathbb{R} : \mathbb{P}(X(w) \le \kappa) \ge \alpha\}$. The utility functions $u(\cdot)$ are assumed to be defined on the entire real line.

Risk measure	**Formulation** $F(p,w)$
Negative mean return	$-\mathbb{E}^p(X(w))$
Standard deviation less the mean	$\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2} - \alpha \mathbb{E}^p(X(w)), \quad \alpha \ge 0$
Standard deviation	$\sqrt{\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2}$
Variance less the mean	$\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2 - \alpha \mathbb{E}^p(X(w)), \quad \alpha \ge 0$
Variance	$\mathbb{E}^p(X(w) - \mathbb{E}^p X(w))^2$
Mean absolute deviation from the median	$\mathbb{E}^p \left\| X(w) - G^{-1}_{X(w)}(0.5) \right\|$
Sharpe ratio	$\frac{-\mathbb{E}^p(X(w))}{\sqrt{\mathbb{E}^p(X(w)-\mathbb{E}^p X(w))^2}}$
Lower partial moment	$\mathbb{E}^p\left(\max\{0, \bar{\kappa} - X(w)\}^\alpha\right)$ $\alpha = 1, 2, \quad \bar{\kappa}$ - any value
Certainty Equivalent (CE)	$-u^{-1}\left(\mathbb{E}^p u(X(w))\right)$ $u(\cdot)$ concave, invertible, with $-\frac{u'(t)}{u''(t)}$ concave
Shortfall risk	$\inf\{\kappa \in \mathbb{R} : \mathbb{E}^p\left(u(X(w) + \kappa\right) \ge 0\}$ $u(\cdot)$ concave
Optimized Certainty Equivalent (OCE)	$\inf\limits_{\kappa \in \mathbb{R}} -\kappa - \mathbb{E}^p(u(X(w) - \kappa))$, $u(\cdot)$ concave, nondecreasing
Value-at-Risk (VaR)	$-G^{-1}_{X(w)}(\alpha), \quad 0 < \alpha < 1$
Conditional Value-at-Risk (CVaR)	$\inf\limits_{\kappa \in \mathbb{R}} -\kappa - \mathbb{E}^p\left(\frac{1}{\alpha}\min\{X(w) - \kappa, 0\}\right), \quad 0 < \alpha < 1$
Entropic Value-at-Risk (EVaR) (see comments in Section 4.1)	$\sup\limits_{\tilde{p} \in \mathcal{P}_q} \mathbb{E}^{\tilde{p}}(-X(w)), \quad 0 < \alpha < 1$ $\mathcal{P}_q = \left\{\tilde{p} : \tilde{p} \ge 0, 1^T \tilde{p} = 1, \sum\limits_{n \in \mathcal{N}} \tilde{p}_n \log\left(\frac{\tilde{p}_n}{q_n}\right) \le -\log \alpha\right\}$
Mean deviation from the mean	$\mathbb{E}^p \left\|X(w) - \mathbb{E}^p(X(w))\right\|$
Distortion risk measures	$\int_0^{+\infty} g\left(1 - G_{X(w)}(t)\right) dt$, $X(w)$ 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	$-\int_0^1 G^{-1}_{X(w)}(t)\psi(t)dt$, $\psi(\cdot)$ nonnegative, non-increasing, right-continuous, integrable

We follow the view, motivated in Rühlicke [50], that the formulation of an uncertainty set for a probability distribution should be supported by results in statistics. For that reason, 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. Goodness-of-fit tests are typically used to test the hypothesis that a random sample with empirical probability distribution p has been sampled from an underlying distribution q. For that purpose, a test statistic is computed and compared to the critical value, based on the test type, sample size, and the chosen confidence level (see, for example, Thas [53]).

Example 2. The test statistic corresponding to the Pearson test is given by

$$D = \sum_{n \in \mathcal{N}} \frac{(p_n - q_n)^2}{q_n},$$

where p_n and q_n, are the empirical and the postulated (tested) probability of the n-th outcome, $n \in \mathcal{N}$, respectively. If, at the assumed confidence level, it holds that $D > D_0$, where D_0 is the critical level at the given confidence level, the hypothesis that the given random sample comes from the underlying distribution q, is rejected. Otherwise, there is no evidence to reject this hypothesis. □

For examples of ϕ-divergences we refer the reader to Table 5 in Appendix C.1. 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 because of their popularity. The likelihood ratio set, having some computational and statistical advantages due to its relation to information theory, is a common choice in studies considering distributional uncertainty, see, for example, Hu and Hong [34], Hu et al. [35], and Jiang and Guan [37].

The Wasserstein set definition, 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 $\|Y^i - Y^j\|^d$, where Y^i is the i-th observation of the underlying random variable Y, as specified in Section 3. This type of uncertainty is studied extensively in a robust setting in Wozabal [56] and the statistical advantages of its use are motivated in Rühlicke [50].

A separate explanation is also needed for the 'combined uncertainty set'. This class of uncertainty sets has been introduced here to derive the tractable robust counterpart of a constraint on the Entropic Value-at-Risk. Its definition in Table 3 says that $\mathcal{P}^{\mathrm{C}}$ 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. 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.

Some of the formulations in Table 3 include explicitly both the vectors p and q and the others only the vector p (with an abbreviation 'emp' next to set name in the last column of Table 3). The former case corresponds to the situation where the uncertainty set for p is defined with reference to a nominal distribution q that in principle can be chosen arbitrarily. In such a case, a typical choice for q will be the empirical distribution. The latter case corresponds to goodness-of-fit tests for one-dimensional random samples $Y^1 \leq Y^2 \leq \ldots \leq Y^N$. There, 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 uncertainty sets only for the case where Y is one-dimensional. Under the assumption that the dependence structure of the marginal distributions is unknown, a simple goodness-of-fit test (and hence, an uncertainty set) can be constructed by applying a given goodness-of-fit test for each marginal distribution, with modified confidence level. Such an approach is applied, for example, by Bertsimas et al. [15], including some of the uncertainty sets considered in Table 3.

For other dependence structures between the marginal distributions, however, construction of credible goodness-of-fit tests is still an area of research. Bertsimas et al. [14] show how to use the Kolmogorov-Smirnov test for each of the marginal distributions, under the assumption of their independence, to obtain approximations of constraints on the VaR.

If needed, continuous probability distributions can be transformed to discrete ones to fit into the framework of our survey, e.g., by dividing the support of a random variable into cells and using the per-cell probabilities. However, one has to be aware of potential shortcomings of such an approach in statistical testing, see, for example, Thas [53].

The choice of an uncertainty set for a particular application is related to the properties of the problem at hand, such as the dimensionality of data, number of observations, and properties of the goodness-of-fit test/probability metric on which a given set is based. Similar to the case of risk measures, the choice of a specific uncertainty set type has its implications for the complexity of the reformulated problem, see Table 1. For an overview and further discussion of statistical tests and probability metrics we refer the reader to Gibbs and Su [27] and Thas [53].

We now verify under which conditions the uncertainty sets listed in Table 3 satisfy condition (8) so that we can apply Theorem 1 to derive the tractable robust counterparts. For the set definitions with no q involved, condition (8) holds. For sets involving a vector q, (8) holds if we assume that (i) $q \in \mathcal{P}$, and (ii) $q \in \mathbb{R}^N_{++}$ (since then the sets $\mathrm{ri}(\mathcal{P})$ and $\mathbb{R}^N_{++}$ obviously have a point q in common). Condition (i) is reasonable since the best estimate of the uncertain probability vector should also belong to the uncertainty set. For (ii), if there were a scenario j such that $q_j = 0$, then such a scenario should not be taken into account, since an estimated probability equal to zero means that the scenario is empirically irrelevant. Thus, the assumptions (i) and (ii) can be expected to hold in applications of our methodology.

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	$\max_{n \in \mathcal{N}} \left\lvert p^T 1^n - q^T 1^n \right\rvert \leq \rho$	$\mathcal{P}_q^{\mathrm{KS}}$
Wasserstein (Kantorovich)	$\inf_{\substack{K: K_{ij} \geq 0, \forall i,j \\ K1 = q, K^T 1 = p}} \left(\sum_{i,j \in \mathcal{N}} K_{ij} \lVert Y^i - Y^j \rVert^d \right) \leq \rho, \quad d \geq 1$	$\mathcal{P}_q^{\mathrm{W}}$
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	$\max_{n \in \mathcal{N}} \left(\frac{n}{N} - p^T 1^n \right) + \max_{n \in \mathcal{N}} \left(p^T 1^{n-1} - \frac{n-1}{N} \right) \leq \rho$	$\mathcal{P}_{\mathrm{emp}}^{\mathrm{K}}$

5 Conjugates of the risk measures

In Table 2 each risk measure corresponds to a specific function $F(p, w)$, defining its value for w under the probability distribution induced by the vector p. However, not for all cases it is possible to apply Theorem 1 using in the forms presented in Table 2. For such cases, we may need an equivalent formulation of the risk measure using some new function $f(p, w)$, for which the assumptions of Theorem 1 are satisfied.

In this section we give the results on concave conjugates $f_*(v, w)$ of such relevant functions $f(p, w)$ corresponding to the risk measures from Table 2. For some cases we take $f(p, w) = F(p, w)$. For others, such as the Optimized Certainty Equivalent or the variance, $F(p, w)$ is reformulated to an equivalent form using a new function $f(p, w)$ linear in p:

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

for appropriate Z_0 and $Z_n(w)$. Linearity in p is a desirable property since then the conjugate $f_*(v, w)$ follows directly from (4):

$$f_*(v, w) = \begin{cases} -Z_0 & \text{if } Z_n(w) \leq v_n, \quad \forall n \in \mathcal{N} \\ -\infty & \text{otherwise.} \end{cases} \tag{10}$$

Derivations for the risk measures where even the $f(p, w)$ is nonlinear in p are given in Appendix B. The remainder of this section distinguishes three cases, depending on the type of the functions $F(p, w)$ and $f(p, w)$: (1) when $F(p, w) \equiv f(p, w)$ is linear in p, (2) when $F(p, w)$ is nonlinear in p, but $f(p, w)$ is linear in p, and (3) when both $F(p, w)$ and $f(p, w)$ are nonlinear in p. For each conjugate function we give the complexity of the system of inequalities in the formulation under the condition that each $X_n(\cdot)$ is linear.

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

In this subsection we analyze the risk measures for which $F(p, w) \equiv f(p, w)$ is 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 (10) with $Z_0 = 0$ and $Z_n(w) = -X_n(w)$. If each $X_n(\cdot)$ is linear, 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 $f(p, w)$ we take is:

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

Its conjugate is given by (10) with $Z_0 = 0$ and $Z_n(w) = -u(X_n(w) + \kappa)$. If each $X_n(\cdot)$ is linear, then, due to the concavity of $u(\cdot)$, the inequalities included in this formulation are convex in the decision variables.

Lower partial moment. In this case the function is:

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

Its conjugate is given by (10) with $Z_0 = 0$ and $Z_n(w) = \max\{0, \bar{\kappa} - X_n(w)\}^\alpha$. If each $X_n(\cdot)$ is linear, 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 $F(p, w)$ is nonlinear in p but $f(p, w)$ is linear in p.

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

$$F(p, w) = \inf_{\kappa \in \mathbb{R}} \left\{ -\kappa - \sum_{n \in \mathcal{N}} p_n (u(X_n(w) - \kappa)) \right\} \leq \beta, \quad \forall p \in \mathcal{P}. \tag{11}$$

Due to Lemma 2 (see Appendix B.1), for continuous and finite-valued functions $u(\cdot)$ 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 (11) can be removed, and the following constraint, with κ as a variable, is equivalent to (11):

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

This formulation is already in the form of Theorem 1 and the concave conjugate of $f(p, w)$ with respect to its first argument is given by (10) with $Z_0 = -\kappa$ and $Z_n(w) = -u(X_n(w) - \kappa)$. If each $X_n(\cdot)$ is linear, 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 each $X_n(\cdot)$ is linear, the inequalities included in this formulation are representable as a system of linear inequalities in the decision variables.

Certainty Equivalent. For general $u(\cdot)$ 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 $u(\cdot)$ to both sides to arrive at an equivalent constraint

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

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:

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

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

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

This result is obtained by considering the impact of changing κ on the value of the sum on the right hand-side, separately for the cases $\kappa > G^{-1}_{X(w)}(0.5)$ and $\kappa < G^{-1}_{X(w)}(0.5)$.

By formulating $F(p,w)$ as an infimum over linear functions in p, we immediately know that it is also concave in p. The conditions of Lemma 2 (see Appendix B.1) are therefore 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 (10) with $Z_0 = 0$ and $Z_n(w) = |X_n(w) - \kappa|$. If each $X_n(\cdot)$ is linear, 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:

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

Even though this formulation is concave in p, the results obtained in [10] 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, well-known fact:

$$\begin{aligned} F(p,w) &= \sum_{n\in\mathcal{N}} p_n \left(X_n(w) - \sum_{n'\in\mathcal{N}} p_{n'} X_{n'}(w) \right)^2 - \alpha \sum_{n\in\mathcal{N}} p_n X_n(w) \\ &= \inf_{\kappa\in\mathbb{R}} \sum_{n\in\mathcal{N}} p_n \left(X_n(w) - \kappa \right)^2 - \alpha \sum_{n\in\mathcal{N}} p_n X_n(w). \end{aligned} \tag{12}$$

Indeed, the minimized expression is strictly convex in κ. Deriving the first-order optimality condition results in $\kappa = \sum_{n \in \mathcal{N}} p_n X_n(w)$.

The conditions of Lemma 2 (see Appendix B.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 (10) 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 each $X_n(\cdot)$ is linear, 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 $F(q, w)$ is formulated as a supremum. Because of this we introduce the notion of a combined uncertainty set to include the formulations of $\mathcal{P}_q$ and $\mathcal{Q}$ in the definition of a joint uncertainty set for (p, q) and to construct a relevant matrix A^{C}. Then, the robust constraint on the EVaR is:

$$\begin{aligned} f(p, w) \quad &= \quad \sum_{n \in \mathcal{N}} p_n (-X(w)) \leq \beta, \quad p = A^{\mathrm{C}} p', \quad A^{\mathrm{C}} = [I | 0_{N \times N}], \qquad (13) \\ &\forall p' \in \left\{ p' = \begin{bmatrix} \overline{p} \\ q \end{bmatrix} : p' \geq 0, \quad 1^T \overline{p} = 1, \quad \sum_{n \in \mathcal{N}} \overline{p}_n \log \left(\frac{\overline{p}_n}{q_n} \right) \leq \rho, q \in \mathcal{Q} \right\}. \end{aligned}$$

The function $f(p, w)$ for which the concave conjugate is to be derived, is the same as for the negative mean return, for which (10) holds with $Z_n(w) = -X_n(w)$ and $Z_0 = 0$. What is left, is the derivation of the support function for the uncertainty set for p', which is done in Section 6. The approach developed here for the EVaR can also be applied to 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 $F(p, w)$ and $f(p, w)$ 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 $F(p, w)$ is nonlinear in p and a derivation of its conjugate would be troublesome. We use the fact that:

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

This formulation follows for the same reason as in the case of variance, since the minimized expression is an increasing function (square root) of the variance. The conditions of Lemma 2 (see Appendix B.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:

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

The function $f(p, w)$ is concave in p and we can use Theorem 1. The conjugate of $f(p, w)$ is equal to:

$$\begin{aligned} f_*(v,w) = \sup_{y} \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \left\| \begin{bmatrix} X_n(w)-\kappa \\ \left(\frac{v_n+\alpha X_n(w)-y}{2}\right) \end{bmatrix} \right\|_2 \le \frac{v_n+\alpha X_n(w)+y}{2}, \quad \forall n\in\mathcal{N} \\ & v_n + \alpha X_n(w) \ge 0, \quad \forall n \in \mathcal{N} \\ & y \ge 0. \end{aligned} \tag{14}$$

The derivation can be found in Appendix B. If each $X_n(\cdot)$ is linear, the above formulation involves second-order conic inequalities in the decision variables. The result for the standard deviation is obtained by setting $\alpha = 0$. One can note that the sup-formulation of the conjugate function fits well into (9) since the conjugate function appears there with a negative sign. Thus, one can omit the sup symbol after inserting (14) into (9) and still have an equivalent constraint.

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

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

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

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

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 and for that reason we could not obtain a closed-form or inf-form for its concave conjugate.

6 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. Our results of this section utilize heavily the property of (9), where it is sufficient to have the support function formulated as an infimum. Then, the inf of the support function symbol can be dropped after inserting the expression for the support function into (9), as explained in Remark 2. We need this property as most of the support functions of the uncertainty sets have been obtained using the following lemma, taken from [10]:

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 $h_i(\cdot)$ is convex for each i. If it holds that $\cap_{i=1}^{H} \mathrm{ri}\,(\mathrm{dom} h_i) \neq \emptyset$, then:

$$\delta^* \left(v|Z\right) = \inf_{u \geq 0} \left\{ \sum_{i=1}^{H} u_i h_i^* \left(\frac{v^i}{u_i} \right) \middle| \sum_{i=1}^{H} v^i = v \right\},$$

where we define $0 h_i^* \left(v^i/0\right) = \lim_{u_i \to 0^+} u_i h_i^* \left(v^i/u_i\right)$. □

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 follows, referring to Appendix C for the derivations.

ϕ-divergence functions. For the uncertainty set defined using the ϕ-divergence the support function is (see Appendix C.2 for a derivation):

$$\delta^* \left(v \middle| \mathcal{P}_q^{\phi} \right) = \inf_{u \geq 0, \eta} \left\{ \eta + u\rho + u \sum_{n \in \mathcal{N}} q_n \phi^* \left(\frac{v_n - \eta}{u} \right) \right\}. \tag{15}$$

This result has also been obtained in [11]. 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 C.1) it can have more tractable forms - for instance, for the Variation distance it is linear. Result (15) 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:

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

with the other components equal to 0. Under such a parametrization, the support function is equal to (see Appendix C.3 for a derivation):

$$
\begin{aligned}
\delta^*\left(v \middle| \mathcal{P}^{\mathrm{KS}}\right) = \inf_u \quad & u^T d \\
\text{s.t.} \quad & v \leq D^T u \\
& u \geq 0.
\end{aligned} \tag{16}
$$

The optimization problem in (16) is linear.

Wasserstein. For an uncertainty set defined using the Wasserstein distance we take $A^{\mathrm{W}} = [I \,|0_{N\times N^2}]$. This choice is motivated in the derivation in Appendix C.4. 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:

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

with all other components of D and d equal to 0. The corresponding support function is equal to (see Appendix C.4 for a derivation):

$$
\begin{aligned}
\delta^*\left(\left(A^{\mathrm{W}}\right)^T v \middle| \mathcal{U}_q^{\mathrm{W}}\right) = \inf_u \quad & u^T d \\
\text{s.t.} \quad & (A^{\mathrm{W}})^T v \leq D^T u \\
& u \geq 0.
\end{aligned} \tag{17}
$$

The optimization problem in (17) 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 C.5.

The support function is equal to (see Appendix C.5 for a derivation):

$$
\begin{aligned}
\delta^*\left(\left(A^{\mathrm{C}}\right)^T v \middle| \mathcal{U}^{\mathrm{C}}\right) = \inf_{\substack{\{u_i, v^i\},\\ i=1,\ldots,Q+3}} \quad & u_1 - u_2 + u_3\rho + \sum_{i=1}^{Q} u_{i+3} h_i^*\left(\frac{v^{i+3}_{N+1:2N}}{u_{i+3}}\right) \\
\text{s.t.} \quad & v_1 \leq u_1 1 \\
& v^2_{1:N} \leq -u_2 1 \\
& v^i_{N+1:2N} = 0, \quad i = 1, 2, 3 \\
& v^i_{1:N} = 0, \quad i = 4, \ldots, Q+3 \\
& v^3_{N+n} + u_3 \phi^*\left(\frac{v^3_n}{u_3}\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=1}^{Q+3} v^i = \left(A^{\mathrm{C}}\right)^T v \\
& u_i \geq 0, \quad i = 1, \ldots, Q+3.
\end{aligned} \tag{18}
$$

For all ϕ-divergence functions listed in Table 5 the optimization problem in (18) is convex. If the ϕ-divergence is the Variation distance or the modified χ^2 distance and the functions $h_i(\cdot)$ are all linear or convex quadratic, then the optimization problem in (18) is linear or convex quadratic, respectively.

Anderson-Darling. For an uncertainty set defined using the Anderson-Darling test the support function $\delta^*\left(v \middle| \mathcal{P}^{\mathrm{AD}}_{\mathrm{emp}}\right)$ is equal to (see Appendix C.6 for a derivation):

$$
\begin{aligned}
\inf_{\eta, u, \{w^{n+}, w^{n-}\}, n\in\mathcal{N}} \quad & -\sum_{n\in\mathcal{N}} \frac{(2n-1)u}{N}\left[2 + \log\left(\frac{-Nz_n^+}{(2n-1)u}\right) + \log\left(\frac{-Nz_n^-}{(2n-1)u}\right)\right] \\
& +u\left(\rho + N\right) + \eta \\
\text{s.t.} \quad & v \leq \sum_{n\in\mathcal{N}} \left(z_n^+ 1^n + z_n^- 1^{-n}\right) + \eta 1 \\
& z_n^+, z_n^- \leq 0 \quad \forall n \in \mathcal{N} \\
& u \geq 0.
\end{aligned} \tag{19}
$$

This result has also been obtained in [10]. The optimization problem in (19) 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 (see Appendix C.7 for a derivation):

$$
\begin{aligned}
\delta^*\left(v \left| \mathcal{P}^{\mathrm{CvM}}_{\mathrm{emp}} \right.\right) = \inf_{z,t,\{u_i,v^i\},i=1,\dots,3} \quad & u_1 - u_2 + \tfrac{1}{4}t - u_3 c \\
\text{s.t.} \quad & \left\| \begin{bmatrix} Pz \\ \frac{t-u_3}{2} \end{bmatrix} \right\|_2 \leq \tfrac{t+u_3}{2} \\
& z = u_3 b - v^3 \\
& u_1 - u_2 + v^3_n - v_n \geq 0, \quad \forall n \in \mathcal{N} \\
& u_1, u_2, u_3 \geq 0.
\end{aligned} \tag{20}
$$

The optimization problem in (20) is convex quadratic.

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

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

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 (see Appendix C.8 for a derivation):

$$
\begin{aligned}
\delta^*\left(v \left| \mathcal{P}^{\mathrm{Wa}}_{\mathrm{emp}} \right.\right) = \inf_{\substack{z,t,\lambda,\{u_i,v^i\},\\ i=1,\dots,3}} \quad & u_1 - u_2 + \tfrac{1}{4}t - u_3 c \\
\text{s.t.} \quad & \left\| \begin{bmatrix} Pz \\ \frac{t-u_3}{2} \end{bmatrix} \right\|_2 \leq \tfrac{t+u_3}{2} \\
& z = u_3 b - v^3 \\
& u_1 - u_2 + v^3_n - v_n \geq 0, \quad n \in \mathcal{N} \\
& u_1, u_2, u_3, t \geq 0 \\
& E\lambda = z.
\end{aligned} \tag{21}
$$

The optimization problem in (21) is convex quadratic.

Kuiper. For the uncertainty set defined using the Kuiper test we take $A^{\mathrm{K}} = [I \,| 0_{N\times 2}]$. 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:

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

with all other components of the matrix D and vector d equal to 0. Under such a parametrization, the support function is (see Appendix C.9 for a derivation):

$$
\begin{aligned}
\delta^* \left(\left(A^{\mathrm{K}}\right)^T v \middle| \mathcal{U}^{\mathrm{K}}_{\mathrm{emp}} \right) = \inf_u \quad & u^T d \\
\text{s.t.} \quad & \left(A^{\mathrm{K}}\right)^T v \leq D^T u \\
& u \geq 0.
\end{aligned} \tag{22}
$$

The optimization problem in (22) is linear.

7 Examples

In this section we present three examples of constraints or problems involving distributional uncertainty. The first example is a simple one where we demonstrate our unifying approach on a single constraint introduced earlier in the paper. The second example is in the field of finance and the third one is of industrial type - a data-driven antenna array design problem. The latter two examples are also studied numerically.

7.1 Standard deviation with ϕ-divergence uncertainty set

For a simple exposition of the advantages of our method, we shall derive a tractable counterpart of the constraint from Example 1 (see page 7), where the constraint is imposed on the standard deviation and the uncertainty set is defined by means of a ϕ-divergence uncertainty set:

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

with

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

In order to obtain a tractable robust counterpart of the form (9), we need to identify:

i the function $f(p, w)$ corresponding to the standard deviation and derive its conjugate, ii the conjugate of the support function of the ϕ-divergence set.

The relevant component for (i) is (14), see page 21. For (ii), we refer to (15), page 22.

Inserting the two components into (9), we obtain an equivalent constraint:

$$\left(\begin{array}{ll} \inf\limits_{\eta,u} & \eta + u\rho + u \sum\limits_{n\in\mathcal{N}} q_n \phi^*\left(\frac{v_n-\eta}{u}\right) \\ \text{s.t.} & u \geq 0 \\ & v_n + \alpha X_n(w) \geq 0, \quad \forall n \in \mathcal{N} \end{array}\right) - \left(\begin{array}{ll} \sup\limits_{y} & -\frac{y}{4} \\ \text{s.t.} & \left\| \begin{bmatrix} X_n(w) - \kappa \\ \left(\frac{v_n+\alpha X_n(w)-y}{2}\right) \end{bmatrix} \right\|_2 \leq \frac{v_n+\alpha X_n(w)+y}{2} \\ & \forall n \in \mathcal{N} \\ & y \geq 0 \end{array}\right) \leq \beta,$$

where η, u, v, w, y are the variables. Since the inf appears on the left hand side of the inequality and the sup is preceded by a negative sign, both can be dropped (see Remark 2), and the resulting equivalent constraint system is:

$$\begin{cases} \eta + u\rho + u \sum\limits_{n\in\mathcal{N}} q_n \phi^*\left(\frac{v_n-\eta}{u}\right) + \frac{y}{4} \leq \beta \\ \left\| \begin{bmatrix} X_n(w) - \kappa \\ \left(\frac{v_n+\alpha X_n(w)-y}{2}\right) \end{bmatrix} \right\|_2 \leq \frac{v_n+\alpha X_n(w)+y}{2}, \quad \forall n \in \mathcal{N} \\ v_n + \alpha X_n(w) \geq 0, \quad \forall n \in \mathcal{N} \\ u, y \geq 0. \end{cases}$$

Since the $\phi^*(\cdot)$ functions are convex, the resulting system of constraints is a system of convex, second-order conic, and linear constraints. Correspondingly, the complexity of the combination of the standard deviation and the ϕ-divergence set in Table 1 is denoted as CP.

7.2 Portfolio management

We consider as first numerical application of our methodology a stylized 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 EVaR for its importance as an upper bound on both the VaR and the CVaR. Additionally, the use of EVaR allows us to illustrate the power of our approach to tackle two-layer uncertainty sets.

7.2.1 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:

$$\begin{aligned}
\max \quad & \mu \\
\text{s.t.} \quad & \sum_{n \in \mathcal{N}} q_n \left(-X_n(w)\right) \leq -\mu \\
& \sup_{\tilde{p} \in \mathcal{P}_q} \sum_{n \in \mathcal{N}} \tilde{p}_n \left(-X_n(w)\right) \leq z \\
& w \in \mathcal{W},
\end{aligned} \tag{23}$$

where $\mathcal{P}_q$ is defined in the row of Table 2 corresponding to the EVaR. Problem (23) 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 5 and 6.

We proceed to the more difficult and, hence, more illustrative robust problem. It shows the unifying power of our approach, including the derivation of the support function of the combined uncertainty set. Moreover, the constraint on the worst-case portfolio return in the robust problem is of the same type as the constraint on the risk measure in the nominal problem and, thus, the corresponding reformulation is also similar.

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

$$\begin{aligned}
\max \quad & \mu \\
\text{s.t.} \quad & \sum_{n \in \mathcal{N}} q_n \left(-X_n(w)\right) \leq -\mu, \quad \forall q \in \mathcal{Q} \qquad \text{(24a)} \\
& \sup_{\tilde{p} \in \mathcal{P}_q} \sum_{n \in \mathcal{N}} \tilde{p}_n \left(-X_n(w)\right) \leq z, \quad \forall q \in \mathcal{Q} \quad \text{(24b)} \\
& w \in \mathcal{W}.
\end{aligned} \tag{24}$$

We shall reformulate the two constraints in problem (24) to their tractable forms using the results of Sections 4 and 5.

Constraint (24a). This is a robust constraint on the negative mean return with uncertainty set $\mathcal{Q}$ being the Pearson set. The corresponding conjugate function (page 20) is given by:

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

To obtain the support function of the Pearson set, we combine the result on the support functions of the ϕ-divergence sets (15) (page 22) with the definition of the modified χ^2 distance in Table 5 (page 44):

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

Inserting the results on the conjugate and the support into (9) yields the tractable robust counterpart of (24a):

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

where η, u_1, v^1, w are the variables.

Constraint (24b). This is a robust constraint on the EVaR with $\mathcal{Q}$ defined as the Pearson set. We shall use the results for the EVaR (page 20) and the combined uncertainty set (page 24). The conjugate function $f_*(v, w)$ is thus the same as in the case of (24a). For the support function of the combined uncertainty set, we insert into the formula (18) for the support function of a combined uncertainty set the following components:

i conjugates related to the condition that the components of q should sum up to 1, that is, convex conjugates of $h_1(q) = 1^T q - 1$ and $h_2(q) = 1 - 1^T q$,

ii the convex conjugate of the modified χ^2 distance (Table 5, page 44), that is, $h_3^*(s) = \max\{-1, s + s^2/4\}$.

As a result, $\delta^*\left(\left(A^{\mathrm{C}}\right)^T v \middle| \mathcal{U}^{\mathrm{C}}\right)$ is equal to:

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

Inserting the results on the conjugate and the support function into (9) yields the tractable robust counterpart of (24b):

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

with variables u_i, v^i, $i = 2, \ldots, 7$, v, w. We remark that it was possible to remove the inf term in the support function formulation due to its position on the left-hand side of the constraint. All the constraints in the above counterpart are convex in the decision variables, so in the terminology of Table 1 the complexity symbol of the system would be CP.

Combining the tractable robust counterparts of the constraints (24a) and (24b) with the rest of the problem formulation, we obtain that problem (24) is equivalent to:

$$\begin{aligned} \max_{\substack{v, u_i, v^i, i=1,\ldots,7 \\ w, \eta, \mu}} \quad & \mu \\ \text{s.t.} \quad & \eta + u_1 \rho_{\mathcal{Q}} + \sum_{n \in \mathcal{N}} r_n \max \left\{ -u_1, v^1_n - \eta + \frac{1}{4} \frac{\left(v^1_n - \eta\right)^2}{u_1} \right\} \leq -\mu \\ & u_2 - u_3 - u_4 \log \alpha + u_5 - u_6 + u_7 \rho_{\mathcal{Q}} + \\ & \qquad + \sum_{n \in \mathcal{N}} r_n \max \left\{ -u_7, v^7_{N+n} + \frac{1}{4} \frac{\left(v^7_{N+n}\right)^2}{u_7} \right\} \leq z \\ & v^2_{1:N} \leq u_2 \mathbf{1} \\ & v^3_{1:N} \leq -u_3 \mathbf{1} \\ & v^5_{N+1:2N} \leq u_5 \mathbf{1} \\ & v^6_{N+1:2N} \leq -u_6 \mathbf{1} \\ & v^i_{N+1:2N} \leq 0, \quad i = 2, 3 \\ & v^i_{1:N} \leq 0, \quad i = 5, 6, 7 \\ & v^4_{N+n} + u_4 \left(\exp\left(\frac{v^4_n}{u_4} \right) - 1 \right) \leq 0, \quad \forall n \in \mathcal{N} \\ & \sum_{i=2}^{7} v^i = \left(A^{\mathrm{C}}\right)^T v \\ & -X_n(w) \leq v_n, \quad \forall n \in \mathcal{N} \\ & -X_n(w) \leq v^1_n, \quad \forall n \in \mathcal{N} \\ & u_i \geq 0, \quad i = 1, \ldots, 7 \\ & w \in \mathcal{W}. \end{aligned} \tag{25}$$

This problem involves linear, convex quadratic, and convex constraints in the decision variables.

7.2.2 Numerical illustration

As a numerical illustration, we use 6 risky assets and 1 riskless asset, with data obtained from the website of Kenneth M. French.[1] The monthly data consists of 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 VaR and CVaR at level 0.05. The degree of uncertainty about the distribution of q in the robust model is defined by $\rho_Q = 0.005$. This value has been chosen for illustration purposes - for high values of ρ_Q the best worst-case return of portfolios involving substantial fractions of the risk assets is negative. For that reason, for higher ρ_Q, constraints on the EVaR will not be active as the optimal robust portfolio will consist of mostly the riskless asset in order to obtain a nonnegative worst-case return. Since the dataset includes two major crises (the dot.com crisis in 2000 and the financial crisis of 2007/2008), this indicates that on datasets that include such periods, portfolios whose key performance measure is the mean return, can be very conservative in a situation of substantial distributional uncertainty. Note also that in the case of the constraint on the EVaR, the size of the 'true' uncertainty set is much larger as it has a two layer combined set structure - the first layer defined with $\rho_Q = 0.005$ and the second one with $\rho = -\log(0.05) \approx 3$.

A potential drawback of this method (historical simulation) is the use a discretized set of portfolio returns. The actual outcomes of portfolio returns typically will not coincide with any of the observations used in portfolio construction. Nevertheless, some authors show that such an approach might yield good results. For instance, Hanasusanto [33] use a distributional uncertainty set modelled with a χ^2 distance in a data-driven dynamic programming setting and show that the decisions obtained under the assumption of distributional uncertainty of the data-driven sample of the uncertain parameter value under consideration, exhibit substantial stability with respect to the sample used.

We solve problem (25) for values of $z = 0, 0.01, \ldots, 0.25$. In this way, we obtain the worst-case EVaR - worst-case mean return frontier. In addition to that, we compute the worst-case EVaR and worst-case mean return of the risky asset. Figure 1 presents both the frontier and the points corresponding to the six individual risky assets.

As it turns out, there is a single asset that dominates all other five risky assets, and that lies on the efficient frontier. An implication of this is that for all values of z

[1] Available at: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The six risky assets are the '6 Portfolios Formed on Size and Book-to-Market (2 x 3)'. See the file with this name for a detailed description. As the riskless asset, we use the one-month US Treasury bill rate.

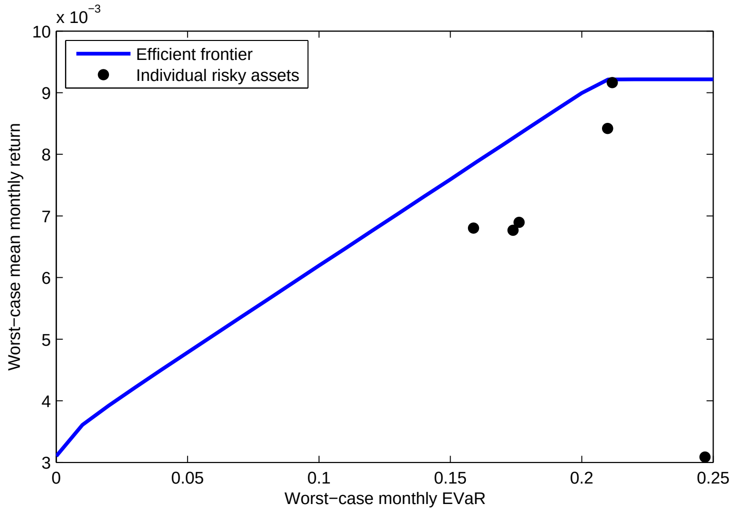

Figure 1: Worst-case EVaR - worst-case mean return efficient frontier in our experiment. The dots illustrate the worst-case EVaR/mean returns of individual risky assets.

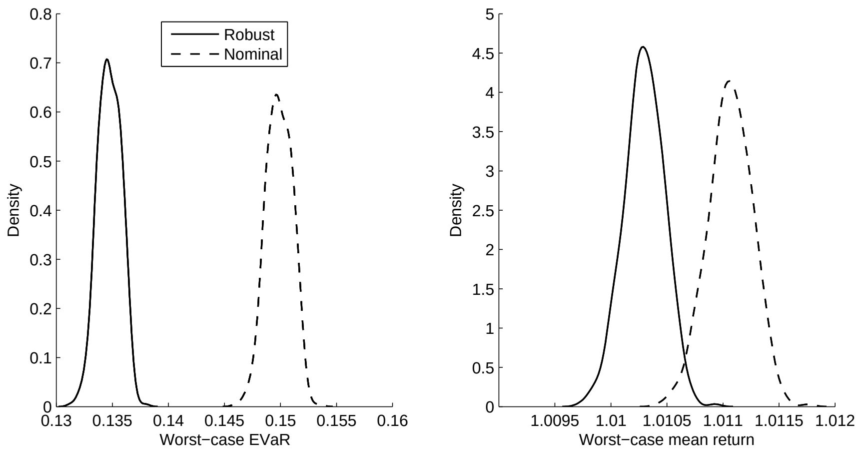

Figure 2: Kernel density plots of the mean return and EVaR value of the nominal and robust portfolios obtained for $z = 0.15$.

the optimal portfolio consists of a mixture of the riskfree asset and this 'best' risky asset. In such a setting, the portfolio optimizer has no incentive to diversify the risk among the six risky assets.

To compare the performance of robust and nominal portfolios under distributional uncertainty, we conduct the following bootstrap experiment. We take the nominal (obtained by solving problem (23)) and robust portfolios for the maximum EVaR value $z = 0.15$. 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 $\sqrt{\frac{\rho_Q}{N^2}}$ 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 in a large number of cases, whereas the robust portfolio's EVaR values oscillate in a region relatively far from 0.15. At the same time, 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.

7.3 Data-driven antenna array design

7.3.1 CVaR and variation distance

As next application, we consider the use of CVaR to approximate chance constraints in a distributionally uncertain antenna design problem. For that, we need the tractable robust counterpart of the constraint on the CVaR with uncertainty set $\mathcal{P}$ defined with the variation distance (one of the ϕ-divergence measures), see Table 5 in Appendix C.1. Such a constraint is given by:

$$\inf_{\kappa} -\kappa + \frac{1}{\alpha} \sum_{n=1}^{N} p_n \max\{0, \kappa - X_i(w)\} \leq \beta, \quad \forall p \in \mathcal{P}$$

where

$$\mathcal{P} = \left\{ p : \quad p \geq 0, \quad 1^T p = 1, \quad \sum_{n \in \mathcal{N}} |p_n - q_n| \leq \rho \right\}. \tag{26}$$

To construct the robust counterpart we use: (1) the expression for the support function of $\mathcal{P}$, given in Section 6, page 22, with the relevant convex conjugate provided in Table 5 in Appendix C.1; (2) the expression for the concave conjugate of the risk measure, discussed as a special case of the Optimized Certainty Equivalent in Section 5, page 18. Then, the robust counterpart is given by

$$\begin{cases} -\kappa + \eta + u\rho + \sum\limits_{n \in \mathcal{N}} q_n \max\{-u, v_n - \eta\} \leq \beta \\ v_n - \eta \leq u, \quad \forall n \in \mathcal{N} \\ v_n \geq \frac{1}{\alpha} \max\{0, \kappa - X_n(w)\}, \quad \forall n \in \mathcal{N} \\ u \geq 0. \end{cases} \tag{27}$$

7.3.2 Antenna design

In this section we consider an antenna design problem, adopted from [7]. The setting of the problem is as follows. There are N_A ring-shaped antennas belonging to the XY plane in $\mathbb{R}^3$. The radius of the k-th antenna is defined as k/N_A and the diagram $D(\phi)$ of the antenna array is defined as a sum of diagrams $D_k(\phi)$ of the antennas, with $D_k(\phi)$ given by:

$$D_k(\phi) = \frac{1}{2} \int_0^{2\pi} \cos\left(\frac{2\pi k}{N_A} \cos(\phi) \cos(\gamma) \right) d\gamma.$$

The objective of the problem is to minimize the maximum of the diagram modulus in the angle of interest $0 \leq \phi \leq 70°$:

$$\max_{0 \leq \phi \leq 70°} \left| \sum_{k=1}^{N_A} w_k D_k(\phi) \right|,$$

subject to the restrictions that:

- the diagram in the interval $77° \leq \phi \leq 90°$ is nearly uniform:

$$0.9 \leq \sum_{k=1}^{N_A} w_k D_k(\phi) \leq 1, 77° \leq \phi \leq 90°$$

- the diagram in other angles is not too large:

$$\left| \sum_{k=1}^{N_A} w_k D_k(\phi) \right| \leq 1, \quad 70° \leq \phi \leq 77°.$$

The problem is thus:

$$\begin{array}{ll} \min\limits_{\tau,w} & \tau \\ \text{s.t.} & -\tau \leq \sum\limits_{k=1}^{N_A} D_k(\phi_i) w_k \leq \tau, \quad 0° \leq \phi_i \leq 70° \\ & -1 \leq \sum\limits_{k=1}^{N_A} D_k(\phi_i) w_k \leq 1, \quad 70° \leq \phi_i \leq 77° \\ & 0.9 \leq \sum\limits_{k=1}^{N_A} D_k(\phi_i) w_k \leq 1, \quad 77° \leq \phi_i \leq 90°, \end{array} \tag{28}$$

$i = 1, \ldots, N_G$, where $\phi_1, ..., \phi_{N_G}$ is a 'fine grid' of equidistance placed points on $[0°, 90°]$.

However, we assume that a multiplicative implementation error affects the decision variable related to the k-th antenna:

$$w_k \mapsto \tilde{w}_k = (1 + z_k) w_k, \quad k = 1, \ldots, N_A.$$

The implementation error z_k consists of two parts: (1) a general error ζ affecting all antennas with the same power; (2) an idiosyncratic error δ_k, specific for each antenna:

$$z_k = \zeta + \delta_k,$$

where ζ and δ_k are independent, normally distributed with zero means and standard deviations σ_1 and σ_2 (σ_2 is the same for $k = 1, \ldots, N_A$), respectively. This is in line with the fact that in complex electrical systems, a part of the implementation error is common for all elements.

Assume there are N past observations of the errors, with the n-th sample denoted as $z^n = (\widehat{z}_1^{(n)}, \ldots, \widehat{z}_{N_A}^{(n)})$. Assume that z^n occurs with an uncertain probability, p_n, with nominal value $q_n = 1/N$ (each sample having equal probability), for each $n \in \mathcal{N}$. In such a setting, the random variables we consider are the diagrams of the antenna arrays at angles ϕ_i, as in the constraints in (28). For a given i, the n-th outcome of the random antenna array diagram is equal to $X_n^i(w) = \sum_k (1+\hat{z}_k^{(n)}) w_k D_k(\phi_i)$, and its probability of occurrence is p_n.

We want each of the constraints in (28) to hold with at least probability $(1-\alpha)100\%$ over the sample at hand. For the first constraint in (28) this is equivalent to:

$$\sum_{n: X_n^i(w) \leq \tau} p_n \geq 1-\alpha, \quad \forall p \in \mathcal{P}.$$

This constraint is in fact equivalent to a constraint on the Value-at-Risk, as noted by [44] and it requires an integer linear programming reformulation. A tractable way to safely approximate chance constraints is the use of the CVaR. Then, the constraint is given by

$$\left\{ \mathrm{CVaR}_{p,\alpha}(-X^i(w)) \leq -\tau \quad \Leftrightarrow \quad \inf_{\kappa} -\kappa + \sum_{n=1}^{N} p_n \max\left\{0, \kappa - X_n^i(w)\right\} \leq -\tau \right\}, \quad \forall p \in \mathcal{P}.$$

We assume $\mathcal{P}$ to be specified as a variation distance set around q, as in (26). This choice is motivated by the fact that such $\mathcal{P}$ is linearly representable in p and q, and the resulting robust counterpart is a system of LP-representable constraints. Since in practice the past samples may be very large, computational tractability is one of the primary criteria for the set to choose. The other constraints are reformulated in a similar fashion. In this way, the problem to be solved becomes:

$$\begin{array}{lll}
\min\limits_{\tau, w} & \tau & \\
\text{s.t.} & \mathrm{CVaR}_{\alpha}(\tau - \sum\limits_k D_k(\phi_i)\tilde{w}_k) \leq 0, & \forall p \in \mathcal{P}, \quad 0^\circ \leq \phi_i \leq 70^\circ \\
& \mathrm{CVaR}_{p,\alpha}(\sum\limits_k D_k(\phi_i)\tilde{w}_k + \tau) \leq 0, & \forall p \in \mathcal{P}, \quad 0^\circ \leq \phi_i \leq 70^\circ \\
& \mathrm{CVaR}_{p,\alpha}(1 - \sum\limits_k D_k(\phi_i)\tilde{w}_k) \leq 0, & \forall p \in \mathcal{P}, \quad 70^\circ \leq \phi_i \leq 77^\circ \\
& \mathrm{CVaR}_{p,\alpha}(\sum\limits_k D_k(\phi_i)\tilde{w}_k + 1) \leq 0, & \forall p \in \mathcal{P}, \quad 70^\circ \leq \phi_i \leq 77^\circ \\
& \mathrm{CVaR}_{p,\alpha}(1 - \sum\limits_k D_k(\phi_i)\tilde{w}_k) \leq 0, & \forall p \in \mathcal{P}, \quad 77^\circ \leq \phi_i \leq 90^\circ \\
& \mathrm{CVaR}_{p,\alpha}(\sum\limits_k D_k(\phi_i)\tilde{w}_k - 0.9) \leq 0, & \forall p \in \mathcal{P}, \quad 77^\circ \leq \phi_i \leq 90^\circ,
\end{array} \tag{29}$$

where $i = 1, \ldots, N_G$. We consider $N_A = 40$ antennas and a sample of past $N = 200$ error vectors sampled with $\sigma_1 = 0.005$ and $\sigma_2 = 0.0025$, which implies dominance of the common error over the idiosyncratic error. In such a setting, we solve the distributionally robust problem (29) with $\alpha = 0.1$, and uncertainty levels $\rho \in \{0, 0.01, \ldots, 0.1\}$. The value $\rho = 0$ corresponds to the problem with no distributional uncertainty. Higher values of ρ would not change the solution as already $\rho = 0.1$ implies that the constraint on CVaR with $\alpha = 0.1$ is in fact a constraint that is to hold for every sampled $z^{(n)}$.

We investigate two questions:

1. On the given random sample of past errors, what is the impact of distributional uncertainty on the probability that at least one of the constraints is violated, for solutions assuming and not assuming distributional uncertainty?

2. How do solutions constructed with and without the assumption on distributional uncertainty perform out-of-sample, that is, with implementation error sampled from the original normal distribution?

To study the first question, for each of the solutions we conduct a simulation study where 10^5 error scenarios are bootstrapped from the sample in the following fashion. First, 100 probability distributions $\widehat{p}$ are sampled in such a way that $\widehat{p}_i$ is sampled from the normal distribution with mean q_i and standard deviation equal to $\rho/2$ for $i = 1, \dots, N-1$, and p_N is defined as $\widehat{p}_N = 1 - \sum_{n=1}^{N-1} \widehat{p}_n$. If it holds that $\widehat{p} \geq 0$, the given vector is accepted as the probability distribution. For each probability distribution we sample subsequently 1000 error scenarios from the sample used to construct a given solution.

To study the second question, for each solution we sample 10^5 error vectors from the normal distribution as the sample drawn used to solve the problem. For both samples, we compute then the average probability of violating at least one problem's constraint.

Table 4 presents the results on the optimal value of the objective function and probabilities of violating at least one constraint. What can be observed is that the differences in the optimal value of the objective function are relatively small, ranging from 6.66 for $\rho = 0$ to 6.85 for $\rho = 0.1$. At the same time the robust solutions exhibit much smaller probabilities of at least one constraint being violated. For example, for the in-sample bootstrap the difference between the nominal solution and the robust solution with $\rho = 0.1$ is 35.88% compared to 23.92%.

What is even more interesting is that the robust solutions perform also consistently better than the nominal solution on out-of-sample implementation errors, with the biggest difference being 34.70% compared to 42.73%. Comparing the first ($\rho = 0$) and the last solution ($\rho = 0.1$) we see that the worst-case objective value which is 2.85% worse provides probability guarantees better by 33% on the in-sample errors and by 17% on the out-of-sample errors.

8 Conclusions

Constraints on risk measures of decision-dependent random variables under distributional uncertainty arise in numerous fields, such as economics, finance, and engineering. In this paper we have reviewed the literature on the problem of reformulating such constraints into tractable forms. As our contribution, we have provided a unified framework for tackling this issue, showing that for many risk measures and statistically based uncertainty sets the constraints' components corresponding to the risk measure and to the uncertainty set can be separated. We have also demonstrated that this framework can be applied to risk measures that are nonlinear in the probability vector. In this way, for risk measures and uncertainty sets for which we provide a closed-form tractable robust counterpart and its complexity, our framework covers the results obtained up to now in the literature (see Table 1).

Table 4: Results of the antenna design experiment. 'In-sample bootstrap' denotes the probability of violating at least one constraint when the implementation error is sampled from the sample used for optimization. Similarly, 'Out-of-sample bootstrap' denotes the probability of violating at least one constraint when the implementation error is sampled from the original normal distribution for the implementation error.

	Measure		
ρ	Objective value ($\cdot 10^{-2}$)	In-sample bootstrap violation probability (%)	Out-of-sample bootstrap violation probability (%)
0 (no uncertainty)	6.66	35.88	42.73
0.01	6.66	29.82	41.19
0.02	6.70	28.89	39.49
0.03	6.73	30.90	39.09
0.04	6.76	26.91	37.84
0.05	6.78	27.85	35.95
0.06	6.80	24.87	34.70
0.07	6.81	27.84	35.04
0.08	6.83	26.91	34.97
0.09	6.84	23.96	35.26
0.1	6.85	23.92	35.47

To provide the decision maker with a clear overview of available techniques, we summarize the complexity results obtained with our framework in combination with results already obtained in the literature in Table 1. These results can provide a useful guideline for researchers and practitioners of various backgrounds.

There are two issues that we find of particular importance when applying robust optimization to risk measures. Following the work of Wozabal [56], who analyzes the Wasserstein distance, it is interesting to investigate whether our framework can be extended to the case with continuous probability distributions, without converting continuous probability distributions into discrete ones.

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

Acknowledgements

We are grateful to two anonymous referees whose comments contributed significantly to the improvement of the content and exposition of our paper. Also, we are grateful to Bram Gorissen for his critical reading of the paper and his remarks.

References

[1] Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, Vol. 26(7), pp. 1505-1518.

[2] Ahmadi-Javid, A. (2012). Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, Vol. 155(3), pp. 1105-1123.

[3] Ahmed, S., Cakmak, U., & Shapiro, A. (2007). Coherent risk measures in inventory problems. *European Journal of Operational Research*, Vol. 182(1), pp. 226-238.

[4] Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, Vol. 9(3), pp. 203-228.

[5] 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.

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

[7] Ben-Tal, A., & Nemirovski, A. (2002). Robust optimization - methodology and applications. *Mathematical Programming*, Vol. 92(3), pp. 453-480.

[8] 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.

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

[10] Ben-Tal, A., Hertog, D. den, & Vial, J.-Ph. (2014). Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming.* Vol. 149(1-2), pp. 265-299.

[11] 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.

[12] Bertsimas, D., & Brown, D. B. (2009). Constructing uncertainty sets for robust linear optimization. *Operations Research*, Vol. 57(6), pp. 1483-1495.

[13] Bertsimas, D., Brown, D. B., & Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, Vol. 53(3), pp. 464-501.

[14] Bertsimas, D., Gupta, V. & Kallus, N. (2013). Data-driven robust optimization. Available online at: `http://arxiv.org/pdf/1401.0212.pdf.`

[15] Bertsimas, D., Gupta, V., & Kallus, N. (2014). Robust SAA. Available online at: `http://arxiv.org/pdf/1408.4445v1.pdf`.

[16] Breuer, T., & Csiszár, I. (2013). Measuring distribution model risk. *Mathematical Finance.* Advance online publication. doi:10.1111/mafi.12050.

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

[18] Chan, T. C., Mahmoudzadeh, H., & Purdie, T. G. (2014). A robust-CVaR optimization approach with application to breast cancer therapy. *European Journal of Operational Research.* Vol. 238(3), pp. 876-885.

[19] Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, Vol. 58(3), pp. 595-612.

[20] Dowd, K. (2007). *Measuring market risk.* John Wiley & Sons.

[21] 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.

[22] Embrechts, P., Frey, R., & McNeil, A. (2005). *Quantitative risk management.* Princeton Series in Finance, Princeton.

[23] Epstein, L. G. (1999). A definition of uncertainty aversion. *The Review of Economic Studies*, Vol. 66(3), pp. 579-608.

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

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

[26] Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European Journal of Operational Research*, Vol. 235(3), pp. 471-483.

[27] Gibbs, A. L., & Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review*, Vol. 70(3), pp. 419-435.

[28] Gotoh, J. Y., & Uryasev, S. (2013). Support vector machines based on convex risk functionals and general norms. Technical Report 2013-6, Risk Management and Financial Engineering Lab, University of Florida. Available at: `http://goo.gl/J94Vsk`.

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

[30] Gupta, V.,(2014) *Data-driven models for uncertainty and behavior.* Ph. D. dissertation, Massachusetts Institute of Technology.

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

[32] Hadar, J., & Russell, W. R. (1969). Rules for ordering uncertain prospects. *American Economic Review*, Vol. 59, pp. 25-34.

[33] Hanasusanto, G. A., & Kuhn, D. (2013). Robust data-driven dynamic programming. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pp. 827-835.

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

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

[36] 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.

[37] 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.

[38] 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.

[39] Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: from data to decisions.* John Wiley & Sons.

[40] Lobo, M. S., Vandenberghe, L., Boyd, S., & Lebret, H. (1998). Applications of second-order cone programming. *Linear Algebra and its Applications*, Vol. 284(1), pp. 193-228.

[41] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, Vol. 7(1), pp. 77-91.

[42] Mosler, K., & Bazovkin, P. (2014). Stochastic linear programming with a distortion risk constraint. *OR Spectrum*, Vol. 36(4), pp. 949-969.

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

[44] Nemirovski, A., & Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, Vol. 17(4), pp. 969-996.

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

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

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

[48] Rockafellar, R. T., Uryasev, S., & Zabarankin, M. (2008). Risk tuning with generalized linear regression. *Mathematics of Operations Research*, Vol. 33(3), pp. 712-729.

[49] Rockafellar, R. T., & Royset, J. O. (2013). Engineering decisions under risk-averseness. Naval postgraduate school Monterey CA, Department of Operations Research. Available at:
http://faculty.nps.edu/joroyset/docs/risk7_withFigs.pdf

[50] Rühlicke, R. (2013) *Robust risk management in the context of Solvency II Regulations.* Ph.D. Dissertation submitted at the Universität Duisburg-Essen.

[51] Schied, A., & Föllmer, H. (2011). *Stochastic finance: an introduction in discrete time.* Walter De Gruyter.

[52] Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, Vol. 39, pp. 119-138.

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

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

[55] Wiesemann, W., Kuhn, D., & Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, Vol. 62(6), pp. 1358-1376.

[56] Wozabal, D. (2014). Robustifying Convex Risk Measures for Linear Portfolios: A Nonparametric Approach. *Operations Research*, Vol. 62(6), pp. 1302-1315.

[57] 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.

A Fenchel duality

Assume $f : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$ is a closed concave function and $g : \mathbb{R}^N \mapsto \mathbb{R} \cup \{+\infty\}$ is a closed convex function, with $f_*(\cdot)$ and $g^*(\cdot)$ being their concave and convex conjugates, respectively. Define the following primal problem

$$\sup \{ f(x) - g(x) \,|x \in \mathrm{dom}(f) \cap \mathrm{dom}(g) \} \tag{P}$$

and its dual problem

$$\inf \{ g^*(x) - f_*(x) \,|x \in \mathrm{dom}(f_*) \cap \mathrm{dom}(g^*) \} . \tag{D}$$

Then, the following theorem holds:

Theorem 2. If $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset$, then the optimal values of (P) and (D) are equal and the minimal value of (D) is attained.
If $\mathrm{ri}(\mathrm{dom}(f_*)) \cap \mathrm{ri}(\mathrm{dom}(g^*)) \neq \emptyset$, then the optimal values of (P) and (D) are equal and the maximal value of (P) is attained. □

B Conjugates of the risk measures

B.1 Necessary lemmas

First result presented here is taken from [46] (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. [46, 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 results.

Lemma 3. Assume that $f_i(\cdot), i = 1, \ldots, m$, are concave, and the intersection of the relative interiors of the effective domains of $f_i(\cdot), i = 1, \ldots, m$ is nonempty, i.e., $\cap_{i=1}^m \mathrm{ri}(\mathrm{dom} f_i) \neq \emptyset$. Then,

$$\left(\sum_{i=1}^m f_i \right)_* (v) = \sup_{v_i, i=1,\ldots,m} \left\{ \sum_{i=1}^m (f_i)_*(v^i) \,\middle|\, \sum_{i=1}^m v^i = v \right\}.$$

$\square$

Lemma 4. [46, Theorem 16.3] Let B be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\}$ be a concave function. Assume there exists an x such that $Bx \in \mathrm{ri}(\mathrm{dom} f)$. Then, it holds that:

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

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

B.2 Standard deviation less the mean

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

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

As an exception, we define the effective domain of $f(p, w)$ to be:

$$\mathrm{dom} f = \left\{ (p, w) : \quad \sum_{n \in \mathcal{N}} p_n \left(X_n(w) - \kappa \right)^2 \geq 0, \quad w \in \mathbb{R}^M \right\}.$$

For this particular case it is easier to operate in this setting and the results are still valid in combination with any uncertainty sets. We use Lemmas 3 and 4 with

$$f(p, w) = f_1(p, w) + f_2(p, w), \quad f_1(p, w) = -\alpha \sum_{n \in \mathcal{N}} p_n X_n(w), \quad f_2(p, w) = \sqrt{b^T p},$$

where

$$b = \left[\left(X_1(w) - \kappa \right)^2, \ldots, \left(X_N(w) - \kappa \right)^2 \right]^T.$$

We have that

$$\begin{aligned} (f_1)_*(v^1, w) &= \begin{cases} 0 & \text{for } v_n^1 = -\alpha X_n(w), \quad n \in \mathcal{N} \\ -\infty & \text{otherwise} \end{cases} \\ (f_2)_*(v^2, w) &= \sup \left\{ -\frac{1}{4s} \,\middle|\, bs = v^2, s \geq 0 \right\}. \end{aligned}$$

Then, substituting $v^1 = u$ and $s = 1/y$, by Lemma 3 we obtain:

$$\begin{aligned} f_*(v,w) = \sup_y \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \tfrac{1}{y}\begin{bmatrix} (X_1(w)-\kappa)^2 \\ \vdots \\ (X_N(w)-\kappa)^2 \end{bmatrix} + u = v \\ & u_n = -\alpha X_n(w), \quad n \in \mathcal{N} \\ & y \geq 0. \end{aligned}$$

Since the objective in the above formulation is decreasing in y and the left-hand side of the first constraint is increasing in y, we can change the '=' sign in the first constraint to '$\leq$' and arrive at the following, equivalent formulation:

$$\begin{aligned} f_*(v,w) = \sup_y \quad & -\tfrac{y}{4} \\ \text{s.t.} \quad & \tfrac{1}{y}\begin{bmatrix} (X_1(w)-\kappa)^2 \\ \vdots \\ (X_N(w)-\kappa)^2 \end{bmatrix} \leq v - u \\ & u_n = -\alpha X_n(w), \quad n \in \mathcal{N} \\ & y \geq 0. \end{aligned}$$

The first constraint can be reformulated using the result of [40] on hyperbolic constraints to obtain the following:

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

To obtain the final result (14) 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 [10].

C Support functions of the uncertainty sets

C.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

Table 5: Examples of ϕ-divergence functions and their convex conjugate functions. Table is taken from Ben-Tal et al. (2013).

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	$\lvert t-1 \rvert^\theta$	$s + (\theta - 1)\left(\frac{\lvert s \rvert}{\theta}\right)^{\theta/(\theta-1)}$
Variation distance	$\lvert t-1 \rvert$	$\max\{-1, s\}, \quad s \leq 1$
Cressie-Read	$\frac{1-\theta+\theta t - t^\theta}{\theta(1-\theta)}, \quad t \neq 0, 1$	$\frac{1}{\theta}(1 - s(1-\theta))^{\theta/(1-\theta)} - \frac{1}{\theta}, \quad s < \frac{1}{1-\theta}$

the reader to [11]. Table 5, adopted from [11], presents potential choices for the function $\phi(\cdot)$ and its conjugate $\phi^*(\cdot)$. Two 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 modified χ^2-distance which defines an uncertainty set based on the χ^2 goodness of fit test, also known as the Pearson test.

C.2 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

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

Now, the convex conjugates of these three functions over the domain $p \geq 0$ are needed.

We begin with the function $g_1(\cdot)$:

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

Analogously:

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

For the third function the derivation is:

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

Lemma 1 gives us:

$$\delta^*\left(v \,\middle|\, \mathcal{P}_q^\phi\right) = \inf_{\{u_i, v^i\}, i=1,2,3} \left\{ u_1 g_1^*\left(\frac{v^1}{u_1}\right) + u_2 g_2^*\left(\frac{v^2}{u_2}\right) + u_3 g_3^*\left(\frac{v^3}{u_3}\right), \,\middle|\, \sum_{i=1}^3 v^i = v,\; u_i \ge 0 \right\}. \tag{30}$$

Notice that by Lemma 1 we have

$$\begin{aligned} u_1 g_1^*\left(\frac{v^1}{u_1}\right) \quad &= \quad \begin{cases} u_1 & \text{for } v^1 \le u_1 1 \\ +\infty & \text{otherwise} \end{cases} \\ u_2 g_2^*\left(\frac{v^2}{u_2}\right) \quad &= \quad \begin{cases} -u_2 & \text{for } v^2 \le -u_2 1 \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

From here, we get:

$$\begin{aligned} \delta^*\left(v \,\middle|\, \mathcal{P}_q^\phi\right) = \inf_{\{u_i, v^i\}, i=1,2,3} \quad & u_1 - u_2 + u_3 \left(\rho + \sum_{n\in\mathcal{N}} q_n \phi^*\left(\tfrac{v_n^3}{u_3}\right)\right) \\ \text{s.t.} \quad & v^1 \le u_1 1 \\ & v^2 \le -u_2 1 \\ & \sum_{i=1}^3 v^i = v \\ & u_i \ge 0, \quad i = 1,2,3. \end{aligned} \tag{31}$$

The equality constraint can be eliminated by inserting $v_n^3 = v_n - v_n^1 - v_n^2$ for each $n \in \mathcal{N}$. We get:

$$\begin{aligned} \delta^*\left(v \,\middle|\, \mathcal{P}_q^\phi\right) = \inf_{u_1,u_2,u_3,v^1,v^2} \quad & u_1 - u_2 + u_3 \left(\rho + \sum_{n\in\mathcal{N}} q_n \phi^*\left(\tfrac{v_n - v_n^1 - v_n^2}{u_3}\right)\right) \\ \text{s.t.} \quad & v^1 \le u_1 1 \\ & v^2 \le -u_2 1 \\ & u_i \ge 0, \quad i = 1,2,3. \end{aligned} \tag{32}$$

Since the functions $\phi^*(\cdot)$ are nondecreasing, one can substitute $\eta = u_1 - u_2$ to obtain result (15) in the main text.

C.3 Kolmogorov-Smirnov

The relevant uncertainty set is:

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

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:

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

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

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

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

$$\begin{aligned}
\delta^* \left(v \middle| \mathcal{P}^{\mathrm{KS}} \right) = \sup_p \quad & v^T p \\
\text{s.t.} \quad & Dp \leq d \\
& p \geq 0.
\end{aligned}$$

The final result (16) in the main text is obtained via strong LP duality.

C.4 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:

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

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

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

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:

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

From here, the final result (17) is obtained via strong LP duality.

C.5 Combined set

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

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

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

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

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

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

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

$$\begin{aligned}
\delta^*\left(\left(A^{\mathrm{C}}\right)^T v \,\middle|\, \mathcal{U}^{\mathrm{C}}\right) = \inf \quad & u_1 - u_2 + u_3\rho + \sum_{i=1}^{Q} u_{i+3} h_i^*\left(\frac{v^{i+3}_{N+1:2N}}{u_{i+3}}\right) \\
\text{s.t.} \quad & v^1_{1:N} \leq u_1 1 \\
& v^2_{1:N} \leq -u_2 1 \\
& v^i_{N+1:2N} \leq 0, \quad i = 1,2 \\
& v^i_{1:N} \leq 0, \quad i = 4, ..., Q+3 \\
& \tfrac{v^3_{N+n}}{u_3} + \phi^*\left(\tfrac{v^3_n}{u_3}\right) \leq 0, \quad \forall n \in \mathcal{N} \\
& \sum_{i=1}^{Q+3} v^i = v \\
& u_i \geq 0, \quad i = 1, ..., Q+3.
\end{aligned}$$

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

C.6 Anderson-Darling

The relevant set formulation is (see Table 3):

$$\mathcal{P}^{\mathrm{AD}}_{\mathrm{emp}} = \{p : p \geq 0, \quad g_i(p) \leq 0, \quad i = 1, 2, 3\},$$

where

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

It is only necessary to derive the conjugate of $g_3(\cdot)$. Let us write $g_3(\cdot)$ 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 [10], it is only needed to derive the convex conjugate of the function

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

It is given by:

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

Using Lemma 1, we obtain:

$$\begin{aligned}
\delta^*\left(v\,\middle|\,\mathcal{P}^{\mathrm{AD}}_{\mathrm{emp}}\right) = \inf_{\substack{\{w^{n+},w^{n-}\},n\in\mathcal{N};\\ \{z_n^+,z_n^-\},n\in\mathcal{N};\\ u_1,u_2,u_3,v^1,v^2}} \quad & -\sum_{n\in\mathcal{N}}\tfrac{(2n-1)u_3}{N}\left[2+\log\left(\tfrac{-Nz_n^+}{(2n-1)u_3}\right)+\log\left(\tfrac{-Nz_n^-}{(2n-1)u_3}\right)\right]\\
& +u_3\left(\rho+N\right)+u_1-u_2\\
\text{s.t.} \quad & z_n^+1^n = w^{n+}, \quad \forall n\in\mathcal{N}\\
& z_n^-1^{-n} = w^{n-}, \quad \forall n\in\mathcal{N}\\
& v^1 \le u_1 1\\
& v^2 \le -u_2 1\\
& \sum_{n\in\mathcal{N}}(w^{n+}+w^{n-})+v^1+v^2 = v\\
& z_n^+, z_n^- \le 0, \quad \forall n\in\mathcal{N}\\
& u_1,u_2,u_3 \ge 0.
\end{aligned}$$

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

$$\begin{aligned}
\inf_{\substack{\{w^{n+},w^{n-}\},n\in\mathcal{N},\\ u_1,u_2,u_3,v^1,v^2}} \quad & -\sum_{n\in\mathcal{N}}\tfrac{(2n-1)u_3}{N}\left[2+\log\left(\tfrac{-Nz_n^+}{(2n-1)u_3}\right)+\log\left(\tfrac{-Nz_n^-}{(2n-1)u_3}\right)\right]\\
& +u_3\left(\rho+N\right)+u_1-u_2\\
\text{s.t.} \quad & v^1 \le u_1 1\\
& v^2 \le -u_2 1\\
& \sum_{n\in\mathcal{N}}(z_n^+1^n+z_n^-1^{-n})+v^1+v^2 = v\\
& z_n^+, z_n^- \le 0, \quad \forall n\in\mathcal{N}.\\
& u_1,u_2,u_3 \ge 0.
\end{aligned}$$

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 (19) in the main text is obtained.

C.7 Cramer-von Mises

The set definition is:

$$\mathcal{P}^{\mathrm{CvM}}_{\mathrm{emp}} = \left\{p : p\ge 0, \quad 1^Tp=1, \quad \frac{1}{12N}+\sum_{n\in\mathcal{N}}\left[\frac{2n-1}{2N}-p^T1^n\right]^2\le\rho\right\},$$

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

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

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 $E \in \mathbb{R}^{N \times N}$ is a positive definite matrix such that $E_{ij} = N + 1 - \max\{i, j\}$ for $i, j \in \mathcal{N}$.

Contrary to the previous cases, we assume the domains of the functions $g_i(\cdot)$ to be $\mathbb{R}^N$ and we include the nonnegativity constraints on p as explicit functional constraints $g_{3+n}(p) = -p^T e^n \leq 0$. Then, we derive the conjugates of $g_i(\cdot)$ as supremums over $p \in \mathbb{R}^N$, which makes the derivation of $g_3^*(\cdot)$ easier. The resulting formula for the support function is equivalent to the formula obtained using the standard assumption about the domains of $g_i(\cdot)$ which, however, would require much more algebra.

Remark 3. Positive definiteness of E follows from the following transformations. Denote by $E^{(k)}$ a matrix for which

$$E_{ij}^{(k)} = \begin{cases} 1 & \text{for } i, j \leq k \\ 0 & \text{otherwise.} \end{cases}$$

Consider $p \in \mathbb{R}^N$. We then have:

$$\begin{aligned} p^T E p &= p^T \left(\sum_{k=1}^{N} E^{(k)}\right) p \\ &= \sum_{k=1}^{N} p^T E^{(k)} p \\ &= \sum_{k=1}^{N} (p_1 + \cdots + p_k)^2 \geq 0, \end{aligned}$$

with 0 being attained if and only if $p_1 = \ldots = p_N = 0$. □

It is important to note that the inverse of E has a tridiagonal structure, allowing for efficient computations.

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

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

and

$$
\begin{aligned}
g_{3+n}^*(y) &= \sup_p \left\{ y^T p + p^T e^n \right\} \\
&= \begin{cases} 0 & \text{if } y + e^n = 0 \\ +\infty & \text{otherwise} \end{cases}
\end{aligned}
$$

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

$$
\begin{aligned}
\delta^*\left(v \,\middle|\, \mathcal{P}_{\text{emp}}^{\text{CvM}}\right) = \inf_{\{u_i, v^i\}, i=1,\dots,N+3} \quad & u_1 - u_2 + \tfrac{1}{4} u_3 \left(b - \tfrac{v^3}{u_3}\right)^T E^{-1} \left(b - \tfrac{v^3}{u_3}\right) - u_3 c \\
\text{s.t.} \quad & v^1 = u_1 1 \\
& v^2 = -u_2 1 \\
& v^{3+n} = -u_{3+n} e^n, \quad n \in \mathcal{N} \\
& \textstyle\sum_{i=1}^{N+3} v^i = v \\
& u_i \geq 0, \quad i = 1, \dots, N+3.
\end{aligned}
$$

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

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

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 [6], 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 (20) in the main text.

C.8 Watson

The set definition is:

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

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

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

and $E \in \mathbb{R}^{N\times N}$ such that $E_{i,j} = N+1-\max\{i,j\} - \frac{(N+1-i)(N+1-j)}{N}$ for all $i,j\in\mathcal{N}$. The matrix E is positive semidefinite with a one-dimensional nullspace, which we prove in the following remark.

Remark 4. Assume $p \in \mathbb{R}^N$ and $d_n = p_1 + \ldots + p_n$ for $n \in \mathcal{N}$. We have:

$$\begin{aligned} p^T E p &= \sum_{n=1}^{N} \left(p^T 1^n\right)^2 - \frac{1}{N}\left(\sum_{n=1}^{N} p^T 1^n\right)^2 \\ &= \sum_{n=1}^{N} d_n^2 - \frac{1}{N}\left(\sum_{n=1}^{N} d_n\right)^2 \\ &= N\left(\frac{\sum_{n=1}^{N} d_n^2}{N} - \left(\frac{\sum_{n=1}^{N} d_n}{N}\right)^2\right) \geq 0, \end{aligned}$$

where the first equality follows from the definition of $\mathcal{P}^{\text{Wa}}_{\text{emp}}$ and the inequality follows from the inequality between arithmetic and quadratic means, and 0 is attained if and only if $d_1 = \ldots = d_N$, that is, when $p_2 = p_3 = \ldots = p_N$ with arbitrary p_1. □

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

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

where $E^\dagger$ denotes a pseudo-inverse of E and ImE 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 $(b-y) \in \text{Im}E$, implemented as $\exists\lambda$ s.t. $b - y = E\lambda$.

C.9 Kuiper

The Kuiper set is defined by

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

Using additional variables z_1,z_2 it can be transformed to

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

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 $\mathcal{U}^{\mathrm{K}}_{\mathrm{emp}}$ is then:

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

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 (22) in the main text is obtained via strong LP duality.