

Distributionally Robust Convex Optimization

Wolfram Wiesemann¹, Daniel Kuhn², and Melvyn Sim³

¹Imperial College Business School, Imperial College London, United Kingdom

²College of Management and Technology, École Polytechnique Fédérale de Lausanne, Switzerland

³Department of Decision Sciences, Business School, National University of Singapore, Singapore

September 22, 2013

Abstract

Distributionally robust optimization is a paradigm for decision-making under uncertainty where the uncertain problem data is governed by a probability distribution that is itself subject to uncertainty. The distribution is then assumed to belong to an ambiguity set comprising all distributions that are compatible with the decision maker's prior information. In this paper, we propose a unifying framework for modeling and solving distributionally robust optimization problems. We introduce standardized ambiguity sets that contain all distributions with prescribed conic representable confidence sets and with mean values residing on an affine manifold. These ambiguity sets are highly expressive and encompass many ambiguity sets from the recent literature as special cases. They also allow us to characterize distributional families in terms of several classical and/or robust statistical indicators that have not yet been studied in the context of robust optimization. We determine conditions under which distributionally robust optimization problems based on our standardized ambiguity sets are computationally tractable. We also provide tractable conservative approximations for problems that violate these conditions.

Keywords. Robust optimization, ambiguous probability distributions, conic optimization.

1 Introduction

In recent years, robust optimization has witnessed an explosive growth and has now become a dominant approach to address practical optimization problems affected by uncertainty. Robust

optimization offers a computationally viable methodology for immunizing mathematical optimization models against parameter uncertainty by replacing probability distributions with *uncertainty sets* as fundamental primitives. One of the core enabling techniques in robust optimization is the tractable representation of the so-called *robust counterpart*, which is given by the following semi-infinite constraint:

$$v(\mathbf{x}, \mathbf{z}) \leq w \quad \forall \mathbf{z} \in \mathcal{C} \tag{1}$$

or equivalently,

$$\sup_{\mathbf{z} \in \mathcal{C}} v(\mathbf{x}, \mathbf{z}) \leq w,$$

where $\mathbf{z} \in \mathbb{R}^P$ is a vector of uncertain problem parameters, $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ represents a vector of here-and-now decisions taken before the realization of \mathbf{z} is known, $\mathcal{C} \subseteq \mathbb{R}^P$ is the uncertainty set, $v : \mathbb{R}^N \times \mathbb{R}^P \rightarrow \mathbb{R}$ is a given constraint function and $w \in \mathbb{R}$ constitutes a prescribed threshold. Intuitively, the constraint (1) ‘robustifies’ the solution of an optimization problem by requiring the decision \mathbf{x} to be feasible for all anticipated realizations of the uncertain parameters \mathbf{z} . The tractability of the robust counterpart (1) depends on the beautiful interplay between the functional properties of v as well as the geometry of \mathcal{C} . We refer interested readers to [6, 7, 11] for comprehensive guides to reformulating robust counterparts.

Despite the simplicity of characterizing uncertainty through sets, robust optimization has been exceptionally successful in providing computationally scalable antidotes for a wide variety of challenging problems ranging from engineering design, finance and machine learning to policy making and business analytics [6, 11]. However, it has been observed that robust optimization models can lead to an underspecification of uncertainty as they do not exploit distributional knowledge that may be available. In such cases, robust optimization may propose overly conservative decisions.

Contrary to robust optimization, stochastic programming explicitly accounts for distributional information through expectation constraints of the form

$$\mathbb{E}_{\mathbb{Q}^0} [v(\mathbf{x}, \tilde{\mathbf{z}})] \leq w, \tag{2}$$

where $\tilde{\mathbf{z}} \in \mathbb{R}^P$ is now a random vector, and the expectation is taken with respect to the distribution \mathbb{Q}^0 of $\tilde{\mathbf{z}}$. Expectation constraints of the type (2) display great modeling power. For instance, they emerge from epigraphical reformulations of single-stage stochastic programs such as the newsvendor problem. They may also arise from stochastic programming-based representations of polyhedral

risk measures such as the conditional value-at-risk [16, 52, 53]. Expectation constraints further serve as basic building blocks for various more sophisticated decision criteria such as optimized certainty equivalents, shortfall aspiration levels and satisficing measures [18, 19, 22, 44]. Finally, expectation constraints can also emerge from reformulations of chance constraints of the form $\mathbb{Q}^0 [v(\mathbf{x}, \tilde{\mathbf{z}}) \leq w] \geq 1 - \epsilon$ based on the identity $\mathbb{Q}^0 [v(\mathbf{x}, \tilde{\mathbf{z}}) \leq w] = \mathbb{E}_{\mathbb{Q}^0} [\mathbb{I}_{[v(\mathbf{x}, \tilde{\mathbf{z}}) \leq w]}]$, see [32, 51, 64].

Going back to the seminal works [40, 41], decision theory distinguishes between the concepts of *risk* (exposure to uncertain outcomes whose probability distribution is known) and *ambiguity* (exposure to uncertainty about the probability distribution of the outcomes). If we identify the uncertainty set \mathcal{C} in the robust counterpart (1) with the support of the probability distribution \mathbb{Q}^0 in (2), then we see that both robust optimization and stochastic programming provide complementary approaches to formulating a decision maker's risk attitude. However, from the perspective of decision theory, neither robust optimization nor stochastic programming addresses ambiguity.

In the era of modern business analytics, one of the biggest challenges in operations research concerns the development of highly scalable optimization problems that can accommodate vast amounts of noisy and incomplete data whilst at the same time truthfully capturing the decision maker's attitude towards both risk and ambiguity. We call this the *distributionally robust optimization* approach. In distributionally robust optimization, we study a variant of the stochastic constraint (2) where the probability distribution \mathbb{Q}^0 is itself subject to uncertainty. In particular, we are concerned with the following *distributionally robust counterpart*,

$$\mathbb{E}_{\mathbb{P}} [v(\mathbf{x}, \tilde{\mathbf{z}})] \leq w \quad \forall \mathbb{P} \in \mathcal{P}$$

or equivalently,

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [v(\mathbf{x}, \tilde{\mathbf{z}})] \leq w, \tag{3}$$

where the probability distribution \mathbb{Q}^0 is merely known to belong to an *ambiguity set* \mathcal{P} of probability distributions. In fact, while \mathbb{Q}^0 is often unknown, decision makers can typically deduce specific properties of \mathbb{Q}^0 from existing domain knowledge (*e.g.*, bounds on the customer demands or symmetry in the deviations of manufacturing processes) or from statistical analysis (*e.g.*, estimation of means and covariances from historical data).

Contrary to classical robust optimization and stochastic programming, the distributionally robust counterpart (3) captures both the decision maker's risk attitude (*e.g.* through the choice of

appropriate disutility functions for v) and an aversion towards ambiguity (through the consideration of the worst probability distribution within \mathcal{P}). Ambiguity aversion enjoys strong justification from decision-theory, where it has been argued that most decision makers have a low tolerance towards uncertainty in the distribution \mathbb{Q}^0 [29, 33]. For these decision makers it is rational to take decisions in view of the worst probability distribution that is deemed possible under the existing information. There is also strong empirical evidence in favor of distributional robustness. In fact, it has been frequently observed that fitting a single candidate distribution to the available information leads to biased optimization results with poor out-of-sample performance. In the context of portfolio management, this phenomenon is known as the “error maximization effect” of optimization [48].

It has been shown recently that under specific assumptions about the ambiguity set \mathcal{P} and the constraint function v , the distributionally robust counterpart (3) inherits computational tractability from the classical robust counterpart (1). This is surprising as the evaluation of the seemingly simpler expectation constraint (2) requires numerical integration over a multidimensional space, which becomes computationally intractable for high-dimensional random vectors. Moreover, if we consider ambiguity sets of the form $\mathcal{P} = \{\mathbb{Q} : \mathbb{Q}[\tilde{z} \in \mathcal{C}] = 1\}$, then the distributionally robust counterpart recovers the classical robust counterpart (1). Distributionally robust optimization therefore constitutes a true generalization of the classical robust optimization paradigm.

While there has been significant recent progress in distributionally robust optimization, there is no unifying framework for modeling and solving distributionally robust optimization problems. The situation is comparable to classical robust optimization, where prior to the papers [7, 15] there were no methods for reformulating generic classes of robust counterparts. The goal of this paper is to propose a similar methodology for distributionally robust optimization. In particular, we aim to develop a canonical form for distributionally robust optimization problems which is restrictive enough to give rise to tractable optimization problems, but which at the same time is expressive enough to cater for a large variety of relevant ambiguity sets and constraint functions. From a theoretical perspective, an optimization problem is considered to be *tractable* if it can be solved in polynomial time (*e.g.* by the ellipsoid method). However, our main interest is in optimization problems that are tractable in practice. In our experience, this is the case if the problems can be formulated as linear or conic-quadratic programs (or, to a lesser degree, semidefinite programs) that can be solved using mature off-the-shelf software tools.

To achieve these goals, we focus on ambiguity sets that contain all distributions with prescribed conic representable confidence sets and with mean values residing on an affine manifold. While conceptually simple, it turns out that this class of standardized ambiguity sets is rich enough to encompass and extend several ambiguity sets considered in the recent literature. They also allow us to model information about statistical indicators that have not yet been considered in the robust optimization literature. Examples are higher-order moments and the marginal median as well as variability measures based on the mean-absolute deviation and the Huber loss function studied in robust statistics. We remark that our framework does not cover ambiguity sets that impose infinitely many moment restrictions, as would be required to describe symmetry, independence or unimodality characteristics of the distributions contained in \mathcal{P} [36, 50].

We demonstrate that the distributionally robust expectation constraints arising from our framework can be solved in polynomial time under certain regularity conditions. In particular, the conditions are met if the constraint functions v is convex and piecewise affine in the decision variables and the random vector, and the confidence sets in the specification of the ambiguity set satisfy a strict nesting condition. For natural choices of the constraint functions and the ambiguity sets, these conditions hold, and (3) can be re-expressed in terms of linear, conic-quadratic or semidefinite constraints. Thus, the inclusion of distributionally robust constraints of the type (3) preserves the computational tractability of conic optimization problems. We also explain how the regularity conditions can be relaxed to accommodate for more general constraint functions, and we demonstrate that the nesting condition is necessary for the tractability of the distributionally robust optimization problem. For problems violating the nesting condition, we develop a tractable conservative approximation that strictly dominates a naïve benchmark approximation.

The contributions of the paper may be summarized as follows.

1. We develop a framework for distributionally robust optimization that uses expectation constraints as basic building blocks. Our framework unifies and generalizes several approaches from the literature. We identify conditions under which robust expectations constraints of the type (3) are tractable, and we derive explicit conic reformulations for these cases.
2. We show that distributionally robust expectation constraints that violate our nesting condition result in intractable optimization problems. We further develop a tractable conservative approximation for these irregular expectation constraints that significantly improves on a

naïve benchmark approximation.

3. We demonstrate that our standardized ambiguity sets are highly expressive in that they allow the modeler to prescribe a wide spectrum of distributional properties that have not yet been studied in robust optimization. This includes information about generalized and higher-order moments as well as selected indicators and metrics from robust statistics.

The history of distributionally robust optimization dates back to the 1950s. Much of the early research relies on ad hoc arguments to construct worst-case distributions for well-structured problem classes. For example, Scarf [54] studies a newsvendor problem where only the mean and variance of the demand are known, while Dupačová (as Žáčková) [3] derives tractable reformulations for stochastic linear programs where only the support and mean of the uncertain parameters are available. Distributionally robust expectation constraints can sometimes be reduced to ordinary expectation constraints involving a mixture distribution that is representable as a convex combination of only a few members of the ambiguity set. If this mixture distribution can be determined explicitly, the underlying expectation constraint becomes amenable to efficient Monte Carlo sampling techniques, see, *e.g.*, Lagoa and Barmish [43], Shapiro and Ahmed [57] and Shapiro and Kleywegt [58].

Most recent approaches to distributional robustness rely on the duality results for moment problems due to Isii [37], Shapiro [56] and Bertsimas and Popescu [14]. Among the first proponents of this idea are El Ghaoui *et al.* [32], who study distributionally robust quantile optimization problems. Their methods have later been extended to linear and conic chance constraints where only the mean, covariance matrix and support of the underlying probability distribution are specified, see, *e.g.* Calafiore and El Ghaoui [20], Chen *et al.* [23], Cheung *et al.* [25] and Zymler *et al.* [64].

Tractable reformulations for distributionally robust expectation constraints of the form (3) are studied by Delage [26] and Delage and Ye [27] under the assumption that the ambiguity set specifies the support as well as conic uncertainty sets for the mean and the covariance matrix of the uncertain parameters. The authors also provide a recipe for constructing ambiguity sets from historical data using McDiarmid’s inequality. Two-stage distributionally robust linear programs with first and second order moment information are investigated by Bertsimas *et al.* [12]. It is shown that these problems are \mathcal{NP} -hard if the uncertainty impacts the constraint right-hand side but reduce to tractable semidefinite programs if only the objective function is affected by the uncertainty. Tractable approximations to generic two-stage and multi-stage distributionally robust linear programs are derived

by Goh and Sim [34] and Kuhn *et al.* [42], assuming that only the support, the mean, the covariance matrix and/or the directional deviations of the uncertain problem parameters are known.

Ben-Tal *et al.* [4] extend the concepts of distributional robustness to parametric families of ambiguity sets $\mathcal{P}(\epsilon)$, $\epsilon \geq 0$, where the constraint (3) may be violated by ϵ for each ambiguity set $\mathcal{P}(\epsilon)$. There is also a rich literature on distributionally robust combinatorial and mixed-integer programming, see Li *et al.* [46]. In this setting a major goal is to calculate the persistence of the binary decision variables, that is, the probability that these variables adopt the value 1 in the optimal solution. Finally, there are deep and insightful connections between classical robust optimization, distributionally robust optimization and the theory of coherent risk measures, see, *e.g.* Bertsimas and Brown [10], Natarajan *et al.* [49] and Xu *et al.* [61].

The remainder of the paper is organized as follows. Section 2 develops tractable reformulations and conservative approximations for the robust expectation constraint (3). Section 3 explores the expressiveness of constraint (3). Section 4 discusses safeguarding constraints that account for both the ambiguity and the risk aversion of decision makers, and Section 5 presents numerical results. All proofs are relegated to the appendix. The electronic companion to this article generalizes our framework to accommodate any constraint functions that admit polynomial time separation oracles.

Notation. For a proper cone \mathcal{K} (*i.e.*, a closed, convex and pointed cone with nonempty interior), the relation $x \preceq_{\mathcal{K}} y$ indicates that $y - x \in \mathcal{K}$. We denote the cone of symmetric (positive semidefinite) matrices in $\mathbb{R}^{P \times P}$ by \mathbb{S}^P (\mathbb{S}_+^P). For $A, B \in \mathbb{S}^P$, we use $A \preceq B$ to abbreviate $A \preceq_{\mathbb{S}_+^P} B$. We denote by \mathcal{K}^* the dual cone of a proper cone \mathcal{K} . The sets $\mathcal{M}_+(\mathbb{R}^P)$ and $\mathcal{P}_0(\mathbb{R}^P)$ represent the spaces of nonnegative measures and probability distributions on \mathbb{R}^P , respectively. If $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ is a joint probability distribution of two random vectors $\tilde{z} \in \mathbb{R}^P$ and $\tilde{u} \in \mathbb{R}^Q$, then $\Pi_{\tilde{z}}\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P)$ denotes the marginal distribution of \tilde{z} under \mathbb{P} . We extend this definition to ambiguity sets $\mathcal{P} \subseteq \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ by setting $\Pi_{\tilde{z}}\mathcal{P} = \bigcup_{\mathbb{P} \in \mathcal{P}} \{\Pi_{\tilde{z}}\mathbb{P}\}$. Finally, we say that a set A is *strictly included* in a set B , or $A \Subset B$, if A is contained in the interior of B .

2 Distributionally Robust Optimization Problems

In this paper we study a class of distributionally robust optimization problems that accommodate a finite number of robust expectation constraints of the type (3). We require that these optimization

problems are tractable if they are stripped of all distributionally robust constraints. Clearly, if this were not the case, then there would be little hope that we could efficiently solve the more general problems involving constraints of the type (3). We now describe a set of regularity conditions that ensure the tractability of these distributionally robust optimization problems.

We assume that the *ambiguity set* \mathcal{P} in (3) is representable in the standard form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \\ \mathbb{P}[(\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{C}_i] \in [\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i] \quad \forall i \in \mathcal{I} \end{array} \right\}, \quad (4)$$

where \mathbb{P} represents a joint probability distribution of the random vector $\tilde{\mathbf{z}} \in \mathbb{R}^P$ appearing in the constraint function v in (3) and some auxiliary random vector $\tilde{\mathbf{u}} \in \mathbb{R}^Q$. We assume that $\mathbf{A} \in \mathbb{R}^{K \times P}$, $\mathbf{B} \in \mathbb{R}^{K \times Q}$, $\mathbf{b} \in \mathbb{R}^K$ and $\mathcal{I} = \{1, \dots, I\}$, while the confidence sets \mathcal{C}_i are defined as

$$\mathcal{C}_i = \{(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_i \mathbf{z} + \mathbf{D}_i \mathbf{u} \preceq_{\mathcal{K}_i} \mathbf{c}_i\} \quad (5)$$

with $\mathbf{C}_i \in \mathbb{R}^{L_i \times P}$, $\mathbf{D}_i \in \mathbb{R}^{L_i \times Q}$, $\mathbf{c}_i \in \mathbb{R}^{L_i}$ and \mathcal{K}_i being proper cones. We allow K or Q to be zero, in which case the expectation condition in (4) is void or the random vector $\tilde{\mathbf{u}}$ is absent, respectively. We also assume that $\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i \in [0, 1]$ and $\underline{\mathbf{p}}_i \leq \bar{\mathbf{p}}_i$ for all $i \in \mathcal{I}$.

We require that the ambiguity set \mathcal{P} satisfies the following two regularity conditions.

(C1) The confidence set \mathcal{C}_I is bounded and has probability one, that is, $\underline{\mathbf{p}}_I = \bar{\mathbf{p}}_I = 1$.

(C2) There is a distribution $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{P}[(\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{C}_i] \in (\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i)$ whenever $\underline{\mathbf{p}}_i < \bar{\mathbf{p}}_i$, $i \in \mathcal{I}$.

Condition **(C1)** ensures that the confidence set with the largest index, \mathcal{C}_I , contains the support of the joint random vector $(\tilde{\mathbf{z}}, \tilde{\mathbf{u}})$. The second condition stipulates that there is a probability distribution $\mathbb{P} \in \mathcal{P}$ that satisfies the probability bounds in (4) as strict inequalities whenever the corresponding probability interval $[\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i]$ is non-degenerate. This assumption allows us to exploit the strong duality results from [37, 56] to reformulate the distributionally robust counterpart (3).

The ambiguity set \mathcal{P} in (4) specifies joint probability distributions for the uncertain problem parameters $\tilde{\mathbf{z}}$ and an auxiliary random vector $\tilde{\mathbf{u}}$ that does not explicitly appear in (3). As we will see below, the inclusion of an auxiliary random vector $\tilde{\mathbf{u}}$ allows us to model a rich variety of structural information about the marginal distribution of $\tilde{\mathbf{z}}$ in a unified manner. The modeler should encode all available information about the true marginal distribution \mathbb{Q}^0 of $\tilde{\mathbf{z}}$ into the ambiguity set \mathcal{P} . In other words, she should choose \mathcal{P} as the smallest ambiguity set for which she knows with certainty

that $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$. Throughout the paper, the symbol \mathbb{P} denotes a joint probability distribution of $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{u}}$ from within $\mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$, whereas the symbol \mathbb{Q} refers to a marginal probability distribution of $\tilde{\mathbf{z}}$ from within $\mathcal{P}_0(\mathbb{R}^P)$. We denote by \mathbb{Q}^0 the true marginal distribution of $\tilde{\mathbf{z}}$.

We require that the constraint function v in (3) satisfies the following condition.

(C3) The constraint function $v(\mathbf{x}, \mathbf{z})$ can be written as

$$v(\mathbf{x}, \mathbf{z}) = \max_{l \in \mathcal{L}} v_l(\mathbf{x}, \mathbf{z})$$

where $\mathcal{L} = \{1, \dots, L\}$ and the auxiliary functions $v_l : \mathbb{R}^N \times \mathbb{R}^P \rightarrow \mathbb{R}$ are of the form

$$v_l(\mathbf{x}, \mathbf{z}) = \mathbf{s}_l(\mathbf{z})^\top \mathbf{x} + t_l(\mathbf{z})$$

with $\mathbf{s}_l(\mathbf{z}) = \mathbf{S}_l \mathbf{z} + \mathbf{s}_l$, $\mathbf{S}_l \in \mathbb{R}^{N \times P}$ and $\mathbf{s}_l \in \mathbb{R}^N$, $t_l(\mathbf{z}) = \mathbf{t}_l^\top \mathbf{z} + t_l$, $\mathbf{t}_l \in \mathbb{R}^P$ and $t_l \in \mathbb{R}$.

Thus, v must be convex and piecewise affine in the decision variables \mathbf{x} and the random vector \mathbf{z} . Condition (C3) will allow us to use robust optimization techniques to reformulate the semi-infinite constraints that arise from a dual reformulation of the distributionally robust constraint (3).

In the following, we show that the conditions (C1)–(C3) allow us to efficiently solve optimization problems involving distributionally robust constraints of the type (3) if and only if the ambiguity set \mathcal{P} satisfies a nesting condition. Afterwards, we develop a conservative approximation for problems that satisfy the conditions (C1)–(C3) but that violate the nesting condition.

Remark 1 (Generic Constraint Functions). *It is possible to relax (C3) so as to accommodate constraint functions that are convex in \mathbf{x} , that can be evaluated in polynomial time and that allow for a polynomial-time separation oracle with respect to $\max_{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i} v(\mathbf{x}, \mathbf{z})$, see [26, 27, 35]. This milder condition is satisfied, for example, by constraint functions that are convex in \mathbf{x} and convex and piecewise affine in \mathbf{z} . Moreover, if we assume that the confidence sets \mathcal{C}_i are described by ellipsoids, then we can accommodate constraint functions that are convex in \mathbf{x} and convex and piecewise (conic-)quadratic in \mathbf{z} . Furthermore, we can accommodate functions $v(\mathbf{x}, \mathbf{z})$ that are non-convex in \mathbf{z} as long as the number of confidence regions is small and all confidence regions constitute polyhedra. We relegate these extensions to the electronic companion of this paper.*

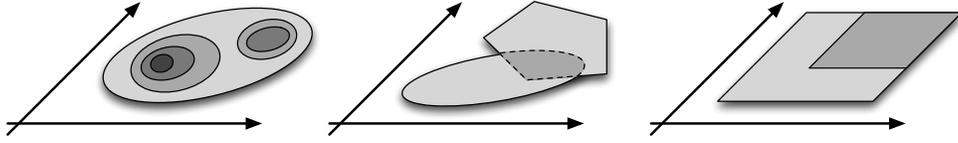


Figure 1. Illustration of the nesting condition **(N)**. The three charts show different arrangements of confidence sets \mathcal{C}_i in the (z, \mathbf{u}) -plane. The left arrangement satisfies the nesting condition, whereas the other two arrangements violate **(N)**.

2.1 Equivalent Reformulation under a Nesting Condition

The tractability of optimization problems with constraints of the type (3) critically depends on the following nesting condition for the confidence sets in the definition of \mathcal{P} :

(N) For all $i, i' \in \mathcal{I}$, $i \neq i'$, we have either $\mathcal{C}_i \subseteq \mathcal{C}_{i'}$, $\mathcal{C}_{i'} \subseteq \mathcal{C}_i$ or $\mathcal{C}_i \cap \mathcal{C}_{i'} = \emptyset$.

The nesting condition is illustrated in Figure 1. The condition implies a strict partial order on the confidence sets \mathcal{C}_i with respect to the \subseteq -relation, with the additional requirement that incomparable sets must be disjoint. The nesting condition is closely related to the notion of laminar families in combinatorial optimization [55], and a similar condition has been used recently to study distributionally robust Markov decision processes [62]. We remark that for two sets \mathcal{C}_i and $\mathcal{C}_{i'}$, the relation $\mathcal{C}_i \subseteq \mathcal{C}_{i'}$ can be checked efficiently if (i) \mathcal{C}_i is an ellipsoid and $\mathcal{C}_{i'}$ is the intersection of finitely many halfspaces and/or ellipsoids or (ii) both \mathcal{C}_i and $\mathcal{C}_{i'}$ constitute polyhedra. If both \mathcal{C}_i and $\mathcal{C}_{i'}$ are described by intersections of finitely many halfspaces and/or ellipsoids, then we can use the approximate \mathcal{S} -Lemma to efficiently verify a sufficient condition for $\mathcal{C}_i \subseteq \mathcal{C}_{i'}$. Section 3 investigates several ambiguity sets for which the nesting condition can be verified analytically.

For ease of notation, we denote by $\mathcal{A}(i) = \{i\} \cup \{i' \in \mathcal{I} : \mathcal{C}_i \subseteq \mathcal{C}_{i'}\}$ and $\mathcal{D}(i) = \{i' \in \mathcal{I} : \mathcal{C}_{i'} \subseteq \mathcal{C}_i\}$ the index sets of all supersets (antecedents) and all strict subsets (descendants) of \mathcal{C}_i , respectively. Our first main result shows that the distributionally robust constraint (3) has a tractable reformulation if the nesting condition **(N)** holds and the regularity conditions **(C1)**–**(C3)** are satisfied.

Theorem 1 (Equivalent Reformulation). *Assume that the conditions **(C1)**–**(C3)** and **(N)** hold. Then, the distributionally robust constraint (3) is satisfied for the ambiguity set (4) if and only if*

there is $\beta \in \mathbb{R}^K$, $\kappa, \lambda \in \mathbb{R}_+^I$ and $\phi_{il} \in \mathcal{K}_i^*$, $i \in \mathcal{I}$ and $l \in \mathcal{L}$, that satisfy the constraint system

$$\left. \begin{aligned} \mathbf{b}^\top \beta + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \kappa_i - \underline{\mathbf{p}}_i \lambda_i] &\leq w, \\ \mathbf{c}_i^\top \phi_{il} + \mathbf{s}_l^\top \mathbf{x} + t_l &\leq \sum_{i' \in \mathcal{A}(i)} [\kappa_{i'} - \lambda_{i'}] \\ \mathbf{C}_i^\top \phi_{il} + \mathbf{A}^\top \beta &= \mathbf{S}_l^\top \mathbf{x} + \mathbf{t}_l \\ \mathbf{D}_i^\top \phi_{il} + \mathbf{B}^\top \beta &= 0 \end{aligned} \right\} \forall i \in \mathcal{I}, \forall l \in \mathcal{L}.$$

If the confidence set \mathcal{C}_i is described by linear, conic-quadratic or semidefinite inequalities, then Theorem 1 provides a linear, conic-quadratic or semidefinite reformulation of the distributionally robust constraint (3), respectively. Thus, the nesting condition (N) is *sufficient* for the tractability of the constraint (3). We now prove that the nesting condition is also *necessary* for tractability.

Theorem 2. *Verifying whether the ambiguity set \mathcal{P} defined in (4) is empty is strongly \mathcal{NP} -hard even if the specification of \mathcal{P} does not involve any expectation conditions (i.e., $K = 0$) and there are only two confidence sets $\mathcal{C}_1, \mathcal{C}_2$ that satisfy $\mathcal{C}_1 \subseteq \mathcal{C}_2$ but $\mathcal{C}_1 \not\subseteq \text{int}(\mathcal{C}_2)$.*

Theorem 2 implies that if the nesting condition (N) is violated, then an optimization problem involving a constraint of type (3) can be strongly \mathcal{NP} -hard even if the problem without the constraint is tractable. Note that Theorem 2 addresses the ‘mildest possible’ violation of the nesting condition (N) in the sense that \mathcal{C}_1 is a subset of \mathcal{C}_2 , but fails to be contained in the interior of \mathcal{C}_2 .

2.2 Conservative Approximation for Generic Problems

We now assume that the ambiguity set \mathcal{P} violates the nesting condition (N). In this case, we know from Theorem 2 that distributionally robust constraints of the type (3) over \mathcal{P} are generically intractable. Our goal is thus to derive an approximation to the constraint (3) which is (i) *conservative*, that is, it does not introduce any spurious solutions that violate the original constraint (3); which is (ii) *tractable* in the sense that any optimization problem that can be solved efficiently without constraint (3) remains tractable when we include our approximation to the constraint (3); and which is (iii) *tight*, or at least not unduly conservative.

To achieve these objectives, we choose a partition $\{\mathcal{I}_j\}_{j \in \mathcal{J}}$ of the index set \mathcal{I} such that the following *weak nesting condition* is satisfied.

(N’) For all $j \in \mathcal{J}$ and all $i, i' \in \mathcal{I}_j$, $i \neq i'$, we have either $\mathcal{C}_i \subseteq \mathcal{C}_{i'}$, $\mathcal{C}_{i'} \subseteq \mathcal{C}_i$ or $\mathcal{C}_i \cap \mathcal{C}_{i'} = \emptyset$.

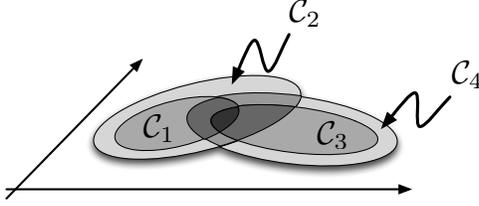


Figure 2. Illustration of the weak nesting condition (N'). The ambiguity set \mathcal{P} with the confidence regions $\mathcal{C}_1, \dots, \mathcal{C}_4$ violates the nesting condition (N), but each of the four partitions $\{\{1, 2\}, \{3, 4\}\}$, $\{\{1\}, \{2\}, \{3, 4\}\}$, $\{\{1, 2\}, \{3\}, \{4\}\}$ and $\{\{1\}, \{2\}, \{3\}, \{4\}\}$ satisfy the weak nesting condition (N').

In other words, the weak nesting condition (N') requires that the confidence sets \mathcal{C}_i , $i \in \mathcal{I}_j$, satisfy the nesting condition (N) for each $j \in \mathcal{J}$. This requirement is nonrestrictive since we can choose the sets \mathcal{I}_j to be singletons. The weak nesting condition is visualized in Figure 2.

We now define the following outer approximations of the ambiguity set \mathcal{P} :

$$\mathcal{P}^j = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\mathbf{A}\tilde{\mathbf{z}} + \mathbf{B}\tilde{\mathbf{u}}] = \mathbf{b}, \\ \mathbb{P}[(\tilde{\mathbf{z}}, \tilde{\mathbf{u}}) \in \mathcal{C}_i] \in [\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i] \quad \forall i \in \mathcal{I}_j \end{array} \right\} \quad \text{for } j \in \mathcal{J}.$$

By construction, \mathcal{P} is indeed a subset of each \mathcal{P}^j because the distributions in \mathcal{P}^j satisfy the condition associated with confidence set \mathcal{C}_i only if $i \in \mathcal{I}_j$. Hence, the following constraint constitutes a *naïve conservative approximation* of the distributionally robust expectation constraint (3).

$$\min_{j \in \mathcal{J}} \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}}[v(\mathbf{x}, \tilde{\mathbf{z}})] \leq w. \quad (6)$$

We further propose the following *infimal convolution bound* as an approximation to constraint (3),

$$\inf_{(\mathbf{y}, \boldsymbol{\delta}) \in \Gamma(\mathbf{x})} \sum_{j \in \mathcal{J}} \boldsymbol{\delta}_j \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}}[v(\mathbf{y}_j / \boldsymbol{\delta}_j, \tilde{\mathbf{z}})] \leq w, \quad (7)$$

where

$$\Gamma(\mathbf{x}) = \left\{ (\mathbf{y}, \boldsymbol{\delta}) : \mathbf{y} = (\mathbf{y}_j)_{j \in \mathcal{J}}, \mathbf{y}_j \in \mathbb{R}^N, \boldsymbol{\delta} = (\boldsymbol{\delta}_j)_{j \in \mathcal{J}}, \boldsymbol{\delta}_j \in \mathbb{R}, \sum_{j \in \mathcal{J}} \mathbf{y}_j = \mathbf{x}, \sum_{j \in \mathcal{J}} \boldsymbol{\delta}_j = 1, \boldsymbol{\delta} > \mathbf{0} \right\}.$$

Infimal convolution bounds have already been studied in the context of classical robust optimization in [22, 34]. The following theorem asserts that while both (6) and (7) constitute conservative approximations of the distributionally robust constraint (3), the infimal convolution bound (7) is preferable in terms of tightness.

Theorem 3. *The distributionally robust constraint (3), its naïve approximation (6) and the infimal convolution bound (7) satisfy the following chain of implications:*

$$(6) \text{ is satisfied} \implies (7) \text{ is satisfied} \implies (3) \text{ is satisfied.}$$

Moreover, the reverse implications hold if $J = 1$.

Note that the feasible set $\Gamma(\mathbf{x})$ of the auxiliary decision variables \mathbf{y} and δ is not closed. To circumvent this problem, we consider the following closed ϵ -approximation of $\Gamma(\mathbf{x})$:

$$\Gamma_\epsilon(\mathbf{x}) = \left\{ (\mathbf{y}, \delta) : \mathbf{y} = (\mathbf{y}_j)_{j \in \mathcal{J}}, \mathbf{y}_j \in \mathbb{R}^N, \delta = (\delta_j)_{j \in \mathcal{J}}, \delta_j \in \mathbb{R}, \sum_{j \in \mathcal{J}} \mathbf{y}_j = \mathbf{x}, \sum_{j \in \mathcal{J}} \delta_j = 1, \delta \geq \epsilon \mathbf{e} \right\}.$$

We henceforth denote by (7ϵ) the constraint (7) where $\Gamma(\mathbf{x})$ is replaced with its ϵ -approximation $\Gamma_\epsilon(\mathbf{x})$. The following tractability result is an immediate consequence of the proof of Theorem 1.

Observation 1. *Assume that the conditions (C1)–(C3) and (N') hold. Then, the infimal convolution bound (7ϵ) is satisfied if and only if there is $\tau \in \mathbb{R}^{|\mathcal{J}|}$, $(\mathbf{y}, \delta) \in \Gamma_\epsilon(\mathbf{x})$, $\beta_j \in \mathbb{R}^K$, κ_{ij} , $\lambda_{ij} \in \mathbb{R}_+$ and $\phi_{ijl} \in \mathcal{K}_i^*$, $i \in \mathcal{I}_j$, $j \in \mathcal{J}$ and $l \in \mathcal{L}$, that satisfy the constraint system*

$$\left. \begin{aligned} \sum_{j \in \mathcal{J}} \tau_j &\leq w, \\ \mathbf{b}^\top \beta_j + \sum_{i \in \mathcal{I}_j} [\bar{\mathbf{p}}_i \kappa_{ij} - \underline{\mathbf{p}}_i \lambda_{ij}] &\leq \tau_j \\ \mathbf{c}_i^\top \phi_{ijl} + \mathbf{s}_i^\top \mathbf{y}_j + \delta_j t_l &\leq \sum_{i' \in \mathcal{A}_j(i)} [\kappa_{i'j} - \lambda_{i'j}] \\ \mathbf{C}_i^\top \phi_{ijl} + \mathbf{A}^\top \beta_j &= \mathbf{S}_i^\top \mathbf{y}_j + \delta_j \mathbf{t}_l \\ \mathbf{D}_i^\top \phi_{ijl} + \mathbf{B}^\top \beta_j &= 0 \end{aligned} \right\} \quad \forall i \in \mathcal{I}_j, \forall l \in \mathcal{L} \quad \forall j \in \mathcal{J},$$

where $\mathcal{A}_j(i) = \mathcal{A}(i) \cap \mathcal{I}_j$ denotes the index set of all supersets of \mathcal{C}_i in \mathcal{I}_j .

Maybe surprisingly, it turns out that the naïve approximation (6) is not only inferior to the approximation (7) in terms of tightness, but also in terms of tractability.

Theorem 4. *Consider the optimization problem*

$$\begin{aligned} &\text{minimize} && \mathbf{d}^\top \mathbf{x} \\ &\text{subject to} && \min_{j \in \mathcal{J}} \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})] \leq \mathbf{w}_s \quad \forall s \in \mathcal{S}, \\ &&& \mathbf{x} \in \mathcal{X}, \end{aligned}$$

where the ambiguity set \mathcal{P} satisfies the conditions **(C1)**, **(C2)** and **(N')**, \mathcal{S} is a finite index set, $v_s(\mathbf{x}, \mathbf{z})$ is linear in \mathbf{x} and linear in \mathbf{z} for all $s \in \mathcal{S}$, and \mathcal{X} constitutes a polyhedron. This problem can be solved in polynomial time if $|\mathcal{S}| = 1$. Otherwise, the problem is strongly \mathcal{NP} -hard.

Assume that $\{\mathcal{I}_j^1\}_{j \in \mathcal{J}_1}$ and $\{\mathcal{I}_j^2\}_{j \in \mathcal{J}_2}$ are two partitions of the confidence regions of an ambiguity set \mathcal{P} that both satisfy the weak nesting condition **(N')**. We say that $\{\mathcal{I}_j^1\}_{j \in \mathcal{J}_1}$ is a *refinement* of $\{\mathcal{I}_j^2\}_{j \in \mathcal{J}_2}$ if for each set \mathcal{I}_j^1 , $j \in \mathcal{J}_1$, there is a set $\mathcal{I}_{j'}^2$, $j' \in \mathcal{J}_2$, such that $\mathcal{I}_j^1 \subseteq \mathcal{I}_{j'}^2$. In particular, the singleton partition $\{\{1\}, \dots, \{I\}\}$ is a refinement of any other partition. The following result shows that coarser partitions lead to tighter approximations of the distributionally robust constraint (3).

Proposition 1. *Let $\{\mathcal{I}_j^1\}_{j \in \mathcal{J}_1}$ and $\{\mathcal{I}_j^2\}_{j \in \mathcal{J}_2}$ be two partitions of the confidence regions of an ambiguity set \mathcal{P} that both satisfy the condition **(N')**, and let $\{\mathcal{P}_1^j\}_{j \in \mathcal{J}_1}$ and $\{\mathcal{P}_2^j\}_{j \in \mathcal{J}_2}$ be the associated sets of outer approximations. If $\{\mathcal{I}_j^1\}_{j \in \mathcal{J}_1}$ is a refinement of $\{\mathcal{I}_j^2\}_{j \in \mathcal{J}_2}$, then the infimal convolution bound (7) is satisfied for $\{\mathcal{P}_2^j\}_{j \in \mathcal{J}_2}$ whenever it is satisfied for $\{\mathcal{P}_1^j\}_{j \in \mathcal{J}_1}$.*

Proposition 1 implies that among all partitions of the confidence regions of an ambiguity set \mathcal{P} satisfying the weak nesting condition **(N')** we should endeavor to find one that is ‘maximally coarse’. We remark that there may be multiple maximal partitions. In this case it is a priori unclear which of these partitions entails the tightest bound of the distributionally robust constraint (3).

3 Expressiveness of the Ambiguity Set

In spite of the apparent simplicity of expectation conditions and probability bounds, ambiguity sets of the type (4) offer striking modeling power. For example, they allow us to impose constraints on the support of the random vector $\tilde{\mathbf{z}}$ by tailoring the confidence set \mathcal{C}_I . Due to the general structure of the confidence sets (5), we can model supports that emerge from finite intersections of halfspaces and generalized ellipsoids, such as flat ellipsoids embedded into subspaces of \mathbb{R}^P or ellipsoidal cylinders given by Minkowski sums of ellipsoids and linear manifolds [7]. Thus, ambiguity sets of the form (4) generalize many of the uncertainty sets that are used in classical robust optimization. In particular, they allow us to model distributionally robust constraints involving *discrete* random variables whose probability vectors range over uncertainty regions defined via ϕ -divergences. In this setting, we interpret $\tilde{\mathbf{z}}$ as the uncertain probability vector, while the ambiguity set \mathcal{P} contains all distributions of $\tilde{\mathbf{z}}$ supported on the corresponding ϕ -divergence-constrained uncertainty region. It

has been shown in [5] that such uncertainty regions admit conic quadratic representations for many popular ϕ -divergences such as the χ^2 -distance, the variation distance or the Hellinger distance etc. Note that \mathcal{P} is of the form (4), and it trivially satisfies the nesting condition (N).

By selecting the shapes and relative arrangement of the confidence sets \mathcal{C}_i , $i < I$, we can further encode information about the modality structure of the random vector $\tilde{\mathbf{z}}$. Such information could be gathered, for example, by applying clustering algorithms to an initial primitive data set.

While it is clear that the expected value of $\tilde{\mathbf{z}}$ can be set to any prescribed constant through an appropriate instantiation of the expectation condition in (4), it turns out that our standardized ambiguity set even allows us to encode (full or partial) information about certain higher-order moments of $\tilde{\mathbf{z}}$ by using a lifting technique. Before formalizing this method, we first introduce some terminology. We say that the \mathcal{K} -epigraph of a function $\mathbf{f} : \mathbb{R}^M \rightarrow \mathbb{R}^N$ and a proper cone \mathcal{K} is *conic representable* if the set $\{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^M \times \mathbb{R}^N : \mathbf{f}(\mathbf{x}) \preceq_{\mathcal{K}} \mathbf{y}\}$ can be expressed via conic inequalities, possibly involving a cone different from \mathcal{K} and additional auxiliary variables.

Theorem 5 (Lifting Theorem). *Let $\mathbf{f} \in \mathbb{R}^M$ and $\mathbf{g} : \mathbb{R}^P \rightarrow \mathbb{R}^M$ be a function with a conic representable \mathcal{K} -epigraph, and consider the ambiguity set*

$$\mathcal{P}' = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\mathbf{g}(\tilde{\mathbf{z}})] \preceq_{\mathcal{K}} \mathbf{f}, \\ \mathbb{P}[\tilde{\mathbf{z}} \in \mathcal{C}_i] \in [\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i] \quad \forall i \in \mathcal{I} \end{array} \right\},$$

as well as the lifted ambiguity set

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^M) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f}, \\ \mathbb{P}[\mathbf{g}(\tilde{\mathbf{z}}) \preceq_{\mathcal{K}} \tilde{\mathbf{u}}] = 1, \\ \mathbb{P}[\tilde{\mathbf{z}} \in \mathcal{C}_i] \in [\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i] \quad \forall i \in \mathcal{I} \end{array} \right\},$$

which involves the auxiliary random vector $\tilde{\mathbf{u}} \in \mathbb{R}^M$. We then have that (i) $\mathcal{P}' = \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$ and (ii) \mathcal{P} can be reformulated as an instance of the standardized ambiguity set (4).

By virtue of Theorem 5 we can recognize several ambiguity sets from the literature as special cases of the ambiguity set (4). For example, defining the function $\mathbf{g}(\mathbf{z})$ in Theorem 5 to be linear allows us to specify ambiguity sets that impose conic constraints on the mean value of $\tilde{\mathbf{z}}$.

Example 1 (Mean). *Assume that $\mathbf{G}\mathbb{E}_{\mathbb{Q}^0}[\tilde{\mathbf{z}}] \preceq_{\mathcal{K}} \mathbf{f}$ for a proper cone \mathcal{K} and $\mathbf{G} \in \mathbb{R}^{M \times P}$, $\mathbf{f} \in \mathbb{R}^M$, and consider the following instance of the ambiguity set (4), which involves the auxiliary random*

vector $\tilde{\mathbf{u}} \in \mathbb{R}^M$.

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^M) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f}, \mathbb{P}[\mathbf{G}\tilde{\mathbf{z}} \preceq_{\mathcal{K}} \tilde{\mathbf{u}}] = 1 \right\}$$

We then have $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbf{G}\mathbb{E}_{\mathbb{Q}}[\tilde{\mathbf{z}}] \preceq_{\mathcal{K}} \mathbf{f} \right\}$.

Example 1 enables us to design confidence sets for the mean value of $\tilde{\mathbf{z}}$ when only a noisy empirical estimator of the exact mean is available. Since the ambiguity set \mathcal{P} in Example 1 satisfies the nesting condition **(N)**, Theorem 1 provides an exact reformulation for distributionally robust expectation constraints over such ambiguity sets that results in linear, conic-quadratic or semidefinite programs whenever the cone \mathcal{K} in the ambiguity set \mathcal{P} is polyhedral, conic-quadratic or semidefinite, respectively.

Theorem 5 further facilitates the construction of ambiguity sets that impose conditions on the covariance matrix of $\tilde{\mathbf{z}}$. In this case we define $\mathbf{g}(\mathbf{z}) = (\mathbf{z} - \boldsymbol{\mu})(\mathbf{z} - \boldsymbol{\mu})^\top$, where $\boldsymbol{\mu} = \mathbb{E}_{\mathbb{Q}^0}[\tilde{\mathbf{z}}]$. Using Schur's complement, one readily shows that $\mathbf{g}(\mathbf{z})$ has a conic representable \mathbb{S}_+^P -epigraph [17].

Example 2 (Variance). Set $\boldsymbol{\mu} = \mathbb{E}_{\mathbb{Q}^0}[\tilde{\mathbf{z}}]$ and assume that $\mathbb{E}_{\mathbb{Q}^0}[(\tilde{\mathbf{z}} - \boldsymbol{\mu})(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma}$ for $\boldsymbol{\Sigma} \in \mathbb{S}_+^P$. Consider the following instance of (4), which involves the auxiliary random matrix $\tilde{\mathbf{U}} \in \mathbb{R}^{P \times P}$.

$$\mathcal{P} = \left\{ \mathbb{P} \in (\mathbb{R}^P \times \mathbb{R}^{P \times P}) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{U}}] = \boldsymbol{\Sigma}, \mathbb{P} \left(\begin{bmatrix} 1 & (\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \\ (\tilde{\mathbf{z}} - \boldsymbol{\mu}) & \tilde{\mathbf{U}} \end{bmatrix} \succeq \mathbf{0} \right) = 1 \right\}$$

We then have $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{Q}}[(\tilde{\mathbf{z}} - \boldsymbol{\mu})(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma} \right\}$.

The ambiguity set \mathcal{P} in Example 2 satisfies the nesting condition **(N)**, and Theorem 1 provides an exact reformulation for distributionally robust expectation constraints over such ambiguity sets that results in a semidefinite program. Example 2 can be extended in several directions. For example, if the upper bound $\boldsymbol{\Sigma}$ on the variance is only known to belong to some set $\mathcal{S} \subseteq \mathbb{S}_+^P$ described by conic inequalities, then we obtain an ambiguity set that is robust with respect to misspecifications of $\boldsymbol{\Sigma}$ as follows.

$$\mathcal{P} = \left\{ \mathbb{P} \in (\mathbb{R}^P \times \mathbb{R}^{P \times P}) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{U}}] \in \mathcal{S}, \mathbb{P} \left(\begin{bmatrix} 1 & (\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \\ (\tilde{\mathbf{z}} - \boldsymbol{\mu}) & \tilde{\mathbf{U}} \end{bmatrix} \succeq \mathbf{0} \right) = 1 \right\}$$

We then have $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{Q}}[(\tilde{\mathbf{z}} - \boldsymbol{\mu})(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top] \preceq \boldsymbol{\Sigma} \text{ for some } \boldsymbol{\Sigma} \in \mathcal{S} \right\}$. The expectation constraint in this ambiguity set can again be standardized using Theorem 5. As long as membership in \mathcal{S} can be expressed via semidefinite constraints, Theorem 1

implies that distributionally robust expectation constraints over the ambiguity set \mathcal{P} have exact reformulations as semidefinite programs. We remark that lower bounds on the covariance matrix $\mathbb{E}_{\mathbb{Q}} \left[(\tilde{\mathbf{z}} - \boldsymbol{\mu})(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \right]$ generically lead to nonconvex optimization problems, see [27, Remark 2]. However, as demonstrated in [27, Proposition 3], lower bounds on the covariance matrix of $\tilde{\mathbf{z}}$ can often be relaxed without affecting the feasible region of the distributionally robust constraint (3).

Sometimes it is natural to relate the variability of a random vector $\tilde{\mathbf{z}}$ to its mean value $\mathbb{E}_{\mathbb{Q}^0} [\tilde{\mathbf{z}}]$. This may be convenient, for example, if the components of $\tilde{\mathbf{z}}$ relate to quantities whose mean values vary widely, for example due to different units of measurement [2]. In such cases, one may impose bounds on the coefficient of variation, which is defined as the inverse of the signal-to-noise ratio known from information theory. We can again apply Theorem 5 to construct ambiguity sets of the form (4) that reflect bounds on the coefficient of variation.

Example 3 (Coefficient of Variation). *Assume that $\sqrt{\mathbb{E}_{\mathbb{Q}^0} \left[(\mathbf{f}^\top \tilde{\mathbf{z}} - \mathbf{f}^\top \boldsymbol{\mu})^2 \right]} / \mathbf{f}^\top \boldsymbol{\mu} \leq \vartheta$ for $\mathbf{f} \in \mathbb{R}^P$, $\vartheta \in \mathbb{R}_+$ and $\boldsymbol{\mu} = \mathbb{E}_{\mathbb{Q}^0} [\tilde{\mathbf{z}}]$ such that $\mathbf{f}^\top \boldsymbol{\mu} > 0$. Consider the following instance of the ambiguity set (4), which involves the auxiliary random vector $\tilde{u} \in \mathbb{R}$.*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}) : \mathbb{E}_{\mathbb{P}} [\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{P}} [\tilde{u}] = \vartheta (\mathbf{f}^\top \boldsymbol{\mu})^2, \mathbb{P} \left[\tilde{u} \geq (\mathbf{f}^\top \tilde{\mathbf{z}} - \mathbf{f}^\top \boldsymbol{\mu})^2 \right] = 1 \right\}$$

We then have $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}} \mathcal{P} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{P}} [\tilde{\mathbf{z}}] = \boldsymbol{\mu}, \sqrt{\mathbb{E}_{\mathbb{Q}} \left[(\mathbf{f}^\top \tilde{\mathbf{z}} - \mathbf{f}^\top \boldsymbol{\mu})^2 \right]} / \mathbf{f}^\top \boldsymbol{\mu} \leq \vartheta \right\}$.

The ambiguity set \mathcal{P} in Example 3 satisfies the nesting condition (N), and Theorem 1 implies that distributionally robust expectation constraints over this ambiguity set have exact reformulations as conic-quadratic programs. As an alternative to the variance and the coefficient of variation, we can describe the dispersion of a univariate random variable \tilde{z} through its absolute mean spread, which quantifies the difference between the expectation of \tilde{z} conditional on \tilde{z} being higher and lower than a given threshold, respectively, see [45]. In contrast to the previous examples, ambiguity sets involving absolute mean spread information can no longer be constructed via straightforward application of Theorem 5.

Proposition 2 (Absolute Mean Spread). *Let $\mathbb{E}_{\mathbb{Q}^0} [\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{Q}^0} [\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} < \theta] \leq \sigma$ and $\mathbb{Q}^0 [\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] = \rho$ for $\mathbf{f} \in \mathbb{R}^P$, $\theta \in \mathbb{R}$, $\sigma \in \mathbb{R}_+$ and $\rho \in (0, 1)$, and consider the following*

Since the ambiguity set \mathcal{P} in Example 4 satisfies the nesting condition **(N)**, Theorem 1 provides exact reformulations for distributionally robust expectation constraints over such ambiguity sets that result in conic-quadratic programs. Setting $f(\tilde{\mathbf{z}}) = |\mathbf{f}^\top \tilde{\mathbf{z}}|$ in Example 4 yields ambiguity sets that impose upper bounds on the expected value of $|\mathbf{f}^\top \tilde{\mathbf{z}}|^{m/n}$. Since $|x|^k = x^k$ for any $x \in \mathbb{R}$ and even k , this implies that we can prescribe upper bounds on all even moments of $\mathbf{f}^\top \tilde{\mathbf{z}}$. Likewise, setting $f(\tilde{\mathbf{z}}) = \max\{\mathbf{f}^\top \tilde{\mathbf{z}}, 0\}$ yields ambiguity sets that impose upper bounds on the (even and/or odd) partial moments of $\mathbf{f}^\top \tilde{\mathbf{z}}$. More generally, we can bound the moments of any piecewise linear functions $\max_{j \in \mathcal{J}} \{\mathbf{f}_j^\top \tilde{\mathbf{z}} + \mathbf{g}_j\}$, where $\mathbf{f}_j \in \mathbb{R}^P$ for $j \in \mathcal{J} = \{1, \dots, J\}$ and $\mathbf{g} \in \mathbb{R}^J$, as long as these functions are guaranteed to be nonnegative or if we focus exclusively on even moments.

To our knowledge, Example 4 presents the first conic-quadratic representation of ambiguity sets incorporating higher-order moment information. Previously studied ambiguity sets with higher-order moment information were either tied to specific problem classes [13], gave rise to reformulations that require solution algorithms akin to the ellipsoid method [26], or they resulted in semidefinite programs [14]. Note that our construction cannot be generalized to odd univariate moments without sacrificing the convexity of the almost sure constraints in the definition of \mathcal{P} .

Bounding the higher-order moments of the random vector $\tilde{\mathbf{z}}$ allows us to capture asymmetries in the probability distribution \mathbb{Q}^0 . This helps to reduce the conservatism of the distributionally robust constraint (3) when the constraint function v is nonlinear in $\tilde{\mathbf{z}}$, or when we consider any of the safeguarding constraints presented in Section 4.

In the remainder we demonstrate that our framework can even capture information that originates from robust statistics. Robust statistics provides descriptive measures of the location and dispersion of a random variable that are reminiscent of standard statistical indicators (such as mean or variance), but which are less affected by outliers and deviations from the model assumptions under which the traditional statistical measures are usually derived (*e.g.*, normality). Recently, robust statistics has received attention in portfolio optimization, where robust estimators help to immunize the portfolio weights against outliers in the historical return samples [28]. In the following, we consider three popular measures from robust statistics: the median, the mean absolute deviation and the expected Huber Loss function.

A univariate random variable \tilde{z} has *median* m under the distribution \mathbb{Q}^0 if $\mathbb{Q}^0(\tilde{z} \leq m) \geq 1/2$ and $\mathbb{Q}^0(\tilde{z} \geq m) \geq 1/2$. Likewise, a multivariate random variable $\tilde{\mathbf{z}} \in \mathbb{R}^P$ has *marginal median* $\mathbf{m} \in \mathbb{R}^P$

if the median of $\tilde{\mathbf{z}}_p$ is \mathbf{m}_p for all $p = 1, \dots, P$. The (marginal) median can be regarded as a robust counterpart of the expected value. Unlike the expected value, the median attaches less importance to the tails of the distribution, which makes it more robust against outliers if the distribution is estimated from historical data. It can be shown that in terms of asymptotic relative efficiency, the sample median is a better estimator of the expected value than the sample mean if the sample distribution is symmetric and has fat tails or if it is contaminated with another distribution [21].

In analogy to the median, we can define the mean absolute deviation as a robust counterpart of the standard deviation. For a univariate random variable \tilde{z} , the *mean absolute deviation* around the value m is given by $\mathbb{E}_{\mathbb{Q}^0} [|\tilde{z} - m|]$. The mean absolute deviation can be generalized to multivariate random variables $\tilde{\mathbf{z}}$ by considering the marginal mean absolute deviation $\mathbb{E}_{\mathbb{Q}^0} [|\tilde{\mathbf{z}} - \mathbf{m}|]$, where the absolute value is understood to apply component-wise. Compared to the standard deviation, the mean absolute deviation enjoys similar robustness properties as the median does in comparison to the expected value, see [21]. Next, for a scalar $z \in \mathbb{R}$ we define the *Huber Loss function* as

$$H_\delta(z) = \begin{cases} \frac{1}{2}z^2 & \text{if } |z| \leq \delta, \\ \delta \left(|z| - \frac{1}{2}\delta \right) & \text{otherwise,} \end{cases} \quad (8)$$

where $\delta > 0$ is a prescribed robustness parameter. We are interested in the Huber loss function because its expected value $\mathbb{E}_{\mathbb{Q}^0} [H_\delta(\tilde{z} - \mu)]$ represents a robust counterpart of the variance $\mathbb{E}_{\mathbb{Q}^0} [(\tilde{z} - \mu)^2]$. Figure 3 illustrates that the Huber Loss function $H_\delta(z)$ can be viewed as the concatenation of a quadratic function for $z \in [-\delta, +\delta]$ (reminiscent of the variance) and a shifted absolute value function for $z \notin [-\delta, +\delta]$ (reminiscent of the mean absolute deviation). The intercept and slope of the absolute value function are chosen to ensure continuity and smoothness at $\pm\delta$. In analogy to the median and the mean absolute deviation, the expected Huber Loss function displays favorable properties in the presence of outliers and distributions with fat tails or contaminations.

The following proposition describes how information about robust estimators can be reflected in our standardized ambiguity set \mathcal{P} .

Proposition 3 (Robust Statistics).

1. **Marginal Median.** Assume that the marginal median of $\tilde{\mathbf{z}}$ under the probability distribution \mathbb{Q}^0 is given by \mathbf{m} , and consider the following instance of the ambiguity set (4).

$$\mathcal{P} = \{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{P}[\tilde{\mathbf{z}}_p \leq \mathbf{m}_p] \geq 1/2, \mathbb{P}[\tilde{\mathbf{z}}_p \geq \mathbf{m}_p] \geq 1/2 \text{ for } p = 1, \dots, P \}$$

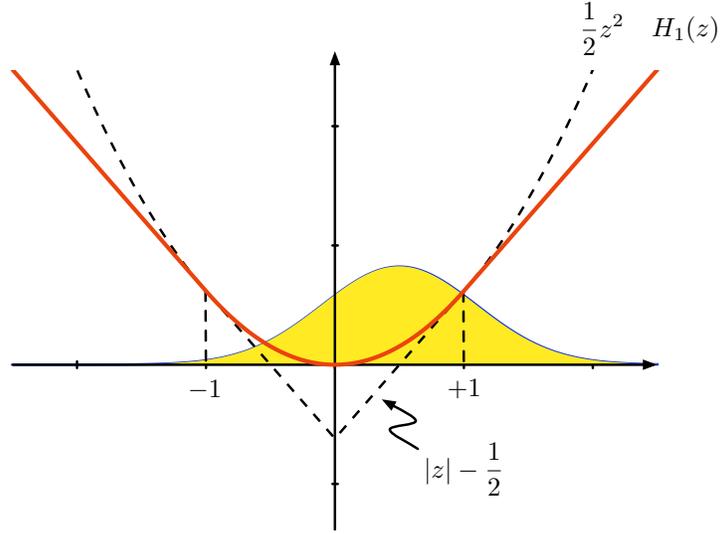


Figure 3. Huber Loss function $H_1(z)$. The chart shows that the Huber loss function is composed of the quadratic function $\frac{1}{2}z^2$ for $z \in [-1, +1]$ and the shifted absolute value function $|z| - \frac{1}{2}$ for $z \notin [-1, +1]$.

We have (i) $\mathbb{Q}^0 \in \mathcal{P}$ and (ii) the marginal median of $\tilde{\mathbf{z}}$ coincides with \mathbf{m} for all $\mathbb{P} \in \mathcal{P}$.

2. **Mean Absolute Deviation.** Assume that $\mathbb{E}_{\mathbb{Q}^0} [|\tilde{\mathbf{z}} - \mathbf{m}|] \leq \mathbf{f}$ for $\mathbf{m}, \mathbf{f} \in \mathbb{R}^P$, and consider the following instance of the ambiguity set (4) involving the auxiliary random vector $\tilde{\mathbf{u}} \in \mathbb{R}^P$.

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^P) : \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f}, \mathbb{P}[\tilde{\mathbf{u}} \geq \tilde{\mathbf{z}} - \mathbf{m}, \tilde{\mathbf{u}} \geq \mathbf{m} - \tilde{\mathbf{z}}] = 1 \right\}$$

We then have $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[|\tilde{\mathbf{z}} - \mathbf{m}|] \leq \mathbf{f} \right\}$.

3. **Expected Huber Loss Function.** Assume that $\mathbb{E}_{\mathbb{Q}^0} [H_{\delta}(\mathbf{f}^{\top} \tilde{\mathbf{z}})] \leq g$ for $\mathbf{f} \in \mathbb{R}^P$, $g \in \mathbb{R}_+$ and H_{δ} is defined as in (8), and consider the following instance of the ambiguity set (4), which involves the auxiliary random variables $\tilde{u}, \tilde{v}, \tilde{w}, \tilde{s}, \tilde{t} \in \mathbb{R}_+$.

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}_+^5) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}[\tilde{w}] = g, \\ \mathbb{P} \left[\begin{array}{l} \delta(\tilde{u} - \tilde{s}) + \frac{\tilde{s}^2}{2} + \delta(\tilde{v} - \tilde{t}) + \frac{\tilde{t}^2}{2} \leq \tilde{w}, \\ \tilde{u} \geq \tilde{s}, \tilde{v} \geq \tilde{t}, \mathbf{f}^{\top} \tilde{\mathbf{z}} = \tilde{u} - \tilde{v} \end{array} \right] = 1 \end{array} \right\}$$

We then have $\mathbb{Q}^0 \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}} [H_{\delta}(\mathbf{f}^{\top} \tilde{\mathbf{z}})] \leq g \right\}$.¹

Note that the ambiguity set \mathcal{P} in the first statement in Proposition 3 violates the nesting condition (N). Thus, we have to conservatively approximate distributionally robust expectation

¹We would like to thank an anonymous referee for pointing out this elegant reformulation.

constraints over such ambiguity sets using the results from Theorem 3 and Observation 1. The resulting approximations constitute linear programs. The first statement in Proposition 3 can be readily extended to situations in which only lower and upper bounds $\underline{\mathbf{m}}_p$ and $\overline{\mathbf{m}}_p$ on the marginal median are available. In this case, the corresponding ambiguity set can be redefined as

$$\mathcal{P} = \{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{P}[\tilde{\mathbf{z}}_p \leq \overline{\mathbf{m}}_p] \geq 1/2, \mathbb{P}[\tilde{\mathbf{z}}_p \geq \underline{\mathbf{m}}_p] \geq 1/2 \text{ for } p = 1, \dots, P \}.$$

More generally, by replacing $1/2$ with $\mathbf{q}_p \in [0, 1]$ in the ambiguity set, we can specify lower and upper bounds on any marginal quantile \mathbf{q}_p of $\tilde{\mathbf{z}}$. Also, we can extend the second statement in Proposition 3 to piecewise linear functions of the random vector $\tilde{\mathbf{z}}$. The ambiguity sets in the second and third statement of Proposition 3 both satisfy the nesting condition **(N)**, which implies that we can derive exact reformulations of distributionally robust expectation constraints over such ambiguity sets using Theorem 1. The resulting reformulations consist of linear programs (mean absolute deviation) or conic-quadratic programs (expected Huber Loss function).

An attractive feature of our approach is its modularity. In fact, several ambiguity sets of the type (4), each reflecting a different piece of information about the distribution of $\tilde{\mathbf{z}}$, can be amalgamated to a master ambiguity set that contains all distributions compatible with every piece of information available to us. The master ambiguity set is still of the form (4). We note that ambiguity sets involving only a single confidence set \mathcal{C}_1 , which has necessarily probability 1, and any combinations of such sets satisfy the nesting condition **(N)**. However, condition **(N)** is generally not preserved under combinations of ambiguity sets involving more than one confidence set.

In practice, the only directly observable data generally are historical realizations of $\tilde{\mathbf{z}}$, while distributional information such as location or dispersion measures or the support of the random vector $\tilde{\mathbf{z}}$ are not directly observable. However, the support of $\tilde{\mathbf{z}}$ can often be inferred from domain-specific knowledge (*e.g.*, customer demands are nonnegative and can be assumed to be bounded from above), or it can be constructed from historical data (*e.g.*, the convex hull of all historical realizations of $\tilde{\mathbf{z}}$ or some approximation thereof). Likewise, confidence regions for the mean and covariance matrix of $\tilde{\mathbf{z}}$ can be derived analytically from historical samples, see [27]. Confidence regions for the other indicators discussed in this section can be constructed from historical observations using resampling techniques such as jackknifing or bootstrapping [24].

Care must be taken, however, when we combine several statistical measures in the definition of the master ambiguity set \mathcal{P} . For example, if we restrict the mean and the variance of the admissible

distributions $\mathbb{Q} \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$ using 0.95-confidence regions for both measures, then the resulting master ambiguity set will contain the unknown true distribution \mathbb{Q}^0 with a confidence of *less* than 0.95. In order to guarantee a specified confidence level for the ambiguity set \mathcal{P} , we can adapt the confidence levels of the individual measures using Bonferroni’s inequality [16]. This implies that the individual confidence levels have to be increased. Hence, we should carefully select the measures to be included in the master ambiguity set. If there is evidence, *e.g.*, from historical data, that the random vector $\tilde{\mathbf{z}}$ is approximately normally distributed, then we expect mean and covariance information to provide a good description of the location and dispersion of \mathbb{Q}^0 . If, on the other hand, we have reason to suspect that the distribution of $\tilde{\mathbf{z}}$ displays fat tails or deviates significantly from normality, we should expect the median, mean absolute deviation and/or the expected Huber loss function to describe \mathbb{Q}^0 more accurately. Finally, if the distributionally robust optimization problem involves nonlinear constraint functions v or any of the safeguarding constraints of Section 4, then we anticipate that the inclusion of higher order moment information effectively reduces the conservatism of the model.

4 Safeguarding Constraints

The distributionally robust constraint (3) captures the ambiguity aversion of the decision maker, but it assumes risk neutrality since it requires the stochastic constraint $v(\mathbf{x}, \tilde{\mathbf{z}}) \leq w$ to be satisfied in the expected sense. In this section, we generalize (3) to various classes of *safeguarding constraints* that account for the decision maker’s risk aversion. We begin with safeguarding constraints based on Gilboa and Schmeidler’s minmax expected utility criterion.

Example 5 (Minimax Expected Utility [33]). *Consider the constraint*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [U(v(\mathbf{x}, \tilde{\mathbf{z}}))] \leq w, \tag{9}$$

where $U : \mathbb{R} \rightarrow \mathbb{R}$ is a non-decreasing convex piecewise affine disutility function of the form $U(y) = \max_{u \in \mathcal{U}} \{\gamma_u y + \delta_u\}$ for a finite index set \mathcal{U} and $\gamma \geq \mathbf{0}$. Under the conditions of Theorem 1, constraint (9) is satisfied if and only if there is $\beta \in \mathbb{R}^K$, $\kappa, \lambda \in \mathbb{R}_+^I$ and $\phi_{ilu} \in \mathcal{K}_i^*$, $i \in \mathcal{I}$, $l \in \mathcal{L}$

and $u \in \mathcal{U}$, that satisfy the constraint system

$$\left. \begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq w, \\ \mathbf{c}_i^\top \boldsymbol{\phi}_{ilu} + \gamma_u \mathbf{s}_l^\top \mathbf{x} + \gamma_u t_l + \boldsymbol{\delta}_u &\leq \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] \\ \mathbf{C}_i^\top \boldsymbol{\phi}_{ilu} + \mathbf{A}^\top \boldsymbol{\beta} &= \gamma_u \mathbf{S}_l^\top \mathbf{x} + \gamma_u t_l \\ \mathbf{D}_i^\top \boldsymbol{\phi}_{ilu} + \mathbf{B}^\top \boldsymbol{\beta} &= 0 \end{aligned} \right\} \quad \forall i \in \mathcal{I}, \forall l \in \mathcal{L}, \forall u \in \mathcal{U}.$$

In a similar way, we can generalize the distributionally robust constraint (3) to safeguarding constraints based on the shortfall risk measure [30] and the optimized certainty equivalent [9].

The shortfall risk measure is defined through

$$\rho^{\text{SR}} [v(\mathbf{x}, \tilde{\mathbf{z}})] = \inf_{\eta \in \mathbb{R}} \left\{ \eta : \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [U(v(\mathbf{x}, \tilde{\mathbf{z}}) - \eta)] \leq 0 \right\}, \quad (10)$$

where $U : \mathbb{R} \rightarrow \mathbb{R}$ is a non-decreasing convex disutility function that is normalized, that is, $U(0) = 0$ and the subdifferential map of U satisfies $\partial U(0) = \{1\}$. For a risky position $v(\mathbf{x}, \tilde{\mathbf{z}})$, the shortfall risk measure $\rho^{\text{SR}} [v(\mathbf{x}, \tilde{\mathbf{z}})]$ can be interpreted as the smallest amount of cash that needs to be injected in order to make the position ‘acceptable’, that is, to achieve a nonpositive expected disutility.

The optimized certainty equivalent risk measure is defined through

$$\rho^{\text{OCE}} [v(\mathbf{x}, \tilde{\mathbf{z}})] = \inf_{\eta \in \mathbb{R}} \left\{ \eta + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [U(v(\mathbf{x}, \tilde{\mathbf{z}}) - \eta)] \right\}, \quad (11)$$

where $U : \mathbb{R} \rightarrow \mathbb{R}$ is again assumed to be a normalized non-decreasing convex disutility function. The optimized certainty equivalent determines an optimized payment schedule of an uncertain future liability $v(\mathbf{x}, \tilde{\mathbf{z}})$ into a fraction η that is paid today and a remainder $v(\mathbf{x}, \tilde{\mathbf{z}}) - \eta$ that is paid after the uncertainty has been revealed. Optimized certainty equivalents generalize mean-variance and conditional value-at-risk measures, see [9].

Example 6 (Shortfall Risk and Optimized Certainty Equivalent). *Assume that $U : \mathbb{R} \rightarrow \mathbb{R}$ is a normalized non-decreasing convex piecewise affine disutility function of the form $U(y) = \max_{u \in \mathcal{U}} \{\gamma_u y + \boldsymbol{\delta}_u\}$ for a finite index set \mathcal{U} and $\boldsymbol{\gamma} \geq \mathbf{0}$. Then $\rho^{\text{SR}} [v(\mathbf{x}, \tilde{\mathbf{z}})] \leq w$ is equivalent to*

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [U(v(\mathbf{x}, \tilde{\mathbf{z}}) - w)] \leq 0,$$

which under the conditions of Theorem 1 is satisfied if and only if there is $\boldsymbol{\beta} \in \mathbb{R}^K$, $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I$ and

$\phi_{ilu} \in \mathcal{K}_i^*$, $i \in \mathcal{I}$, $l \in \mathcal{L}$ and $u \in \mathcal{U}$, that satisfy the constraint system

$$\left. \begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq 0, \\ \mathbf{c}_i^\top \phi_{ilu} + \gamma_u \mathbf{s}_l^\top \mathbf{x} + \gamma_u t_l - \gamma_u w + \delta_u &\leq \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] \\ \mathbf{C}_i^\top \phi_{ilu} + \mathbf{A}^\top \boldsymbol{\beta} &= \gamma_u \mathbf{S}_l^\top \mathbf{x} + \gamma_u \mathbf{t}_l \\ \mathbf{D}_i^\top \phi_{ilu} + \mathbf{B}^\top \boldsymbol{\beta} &= 0 \end{aligned} \right\} \forall i \in \mathcal{I}, \forall l \in \mathcal{L}, \forall u \in \mathcal{U}.$$

Likewise, $\rho^{OCE}[v(\mathbf{x}, \tilde{\mathbf{z}})] \leq w$ is satisfied if and only if there is $\eta \in \mathbb{R}$ such that

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [U(v(\mathbf{x}, \tilde{\mathbf{z}}) - \eta)] \leq w - \eta,$$

which under the conditions of Theorem 1 is satisfied if and only if there is $\eta \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^K$, $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I$ and $\phi_{ilu} \in \mathcal{K}_i^*$, $i \in \mathcal{I}$, $l \in \mathcal{L}$ and $u \in \mathcal{U}$, that satisfy the constraint system

$$\left. \begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq w - \eta, \\ \mathbf{c}_i^\top \phi_{ilu} + \gamma_u \mathbf{s}_l^\top \mathbf{x} + \gamma_u t_l - \gamma_u \eta + \delta_u &\leq \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] \\ \mathbf{C}_i^\top \phi_{ilu} + \mathbf{A}^\top \boldsymbol{\beta} &= \gamma_u \mathbf{S}_l^\top \mathbf{x} + \gamma_u \mathbf{t}_l \\ \mathbf{D}_i^\top \phi_{ilu} + \mathbf{B}^\top \boldsymbol{\beta} &= 0 \end{aligned} \right\} \forall i \in \mathcal{I}, \forall l \in \mathcal{L}, \forall u \in \mathcal{U}.$$

5 Numerical Example

To get a deeper grasp of our distributionally robust optimization framework, we compare the performance of some of the statistical indicators of Section 3 on a stylized newsvendor problem. All optimization problems are solved using the YALMIP modeling language and SDPT3 3.0 [38, 59].

We assume that a newsvendor trades in $i = 1, \dots, n$ products. Before observing the uncertain product demands $\tilde{\mathbf{z}}_i$, the newsvendor orders \mathbf{x}_i units of product i at the wholesale price \mathbf{c}_i . Once $\tilde{\mathbf{z}}_i$ is observed, she can sell the quantity $\mathbf{y}_i(\tilde{\mathbf{z}}_i) \in [0, \min\{\mathbf{x}_i, \tilde{\mathbf{z}}_i\}]$ at the retail price \mathbf{v}_i . Any unsold stock $[\mathbf{x}_i - \mathbf{y}_i(\tilde{\mathbf{z}}_i)]_+$ is cleared at the salvage price \mathbf{g}_i , and any unsatisfied demand $[\tilde{\mathbf{z}}_i - \mathbf{y}_i(\tilde{\mathbf{z}}_i)]_+$ is lost. Here we use the shorthand notation $[\cdot]_+ = \max\{\cdot, 0\}$. We assume that $\mathbf{c}_i < \mathbf{v}_i$ and $\mathbf{g}_i < \mathbf{v}_i$, which implies that the optimal sales decisions are of the form $\mathbf{y}_i(\tilde{\mathbf{z}}_i) = \min\{\mathbf{x}_i, \tilde{\mathbf{z}}_i\}$. We can thus

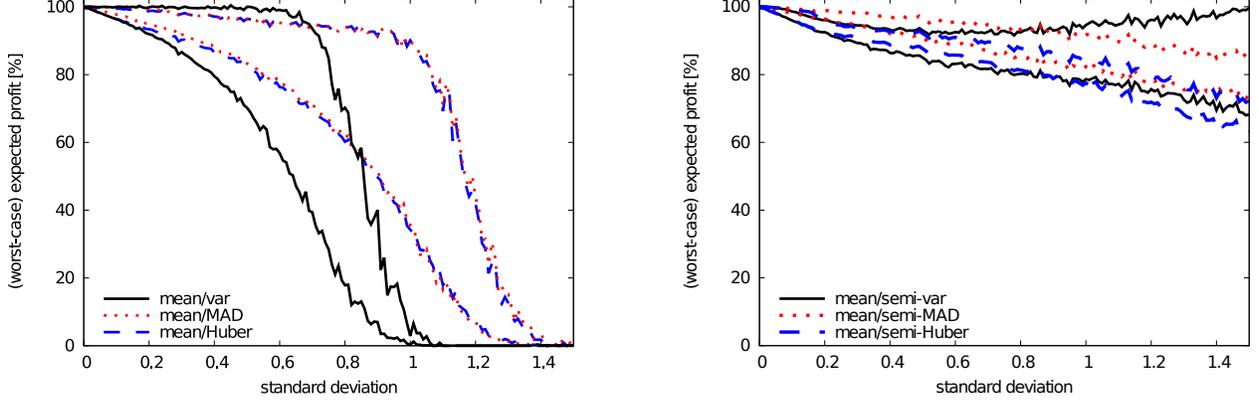


Figure 4. Single-product results for different symmetric (left) and asymmetric (right) statistical indicators. All results are averaged over 100 instances. For each indicator, the upper and lower curves represent the out-of-sample and in-sample results.

describe the newsvendor’s minimum losses as a function of the order decision \mathbf{x} :

$$\begin{aligned} L(\mathbf{x}, \tilde{\mathbf{z}}) &= \mathbf{c}^\top \mathbf{x} - \mathbf{v}^\top \min\{\mathbf{x}, \tilde{\mathbf{z}}\} - \mathbf{g}^\top [\mathbf{x} - \tilde{\mathbf{z}}]_+ \\ &= (\mathbf{c} - \mathbf{v})^\top \mathbf{x} + (\mathbf{v} - \mathbf{g})^\top [\mathbf{x} - \tilde{\mathbf{z}}]_+. \end{aligned}$$

Here, the minimum and the nonnegativity operator are applied component-wise.

We assume that the probability distribution \mathbb{Q}^0 governing the product demands $\tilde{\mathbf{z}}$ is unknown. Instead, the newsvendor has access to a limited number of i.i.d. samples of $\tilde{\mathbf{z}}$. Assuming stationarity, she can then construct an ambiguity set \mathcal{P} using the statistical indicators of Section 3. We first assume that the newsvendor solves the risk-neutral optimization problem

$$\begin{aligned} &\text{minimize} && \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [L(\mathbf{x}, \tilde{\mathbf{z}})] \\ &\text{subject to} && \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

Using the results of Sections 2 and 4, we can readily reformulate and solve this problem as a conic optimization problem for the ambiguity sets presented in Section 3.

In our first experiment, we assume that the newsvendor trades in a single product ($n = 1$) with a wholesale price $\mathbf{c}_1 = 5$, retail price $\mathbf{v}_1 = 10$ and salvage price $\mathbf{g}_1 = 2.5$. The demand \tilde{z}_1 follows a lognormal distribution with mean 5 and a varying standard deviation. Since the mean is fixed, higher standard deviations correspond to increasingly right-skewed demand distributions. The newsvendor has access to 250 i.i.d. samples of $\tilde{\mathbf{z}}$. Applying standard resampling techniques and

Bonferroni's inequality, she uses these samples to construct ambiguity sets that contain (a lifted version of) the unknown distribution \mathbb{Q}^0 with a confidence of at least 95%.

The left chart in Figure 4 shows the results for ambiguity sets \mathcal{P} constructed from the mean and variance (**mean/var**), the mean and mean absolute deviation (**mean/MAD**) and the mean and Huber loss function with $\delta = 1$ (**mean/Huber**). The right graph reports the performance of the corresponding asymmetric indicators, that is, ambiguity sets constructed from the mean and semi-variances $\mathbb{E}_{\mathbb{P}} [\tilde{z}_1 - \mu_1]_+^2$ and $\mathbb{E}_{\mathbb{P}} [\mu_1 - \tilde{z}_1]_+^2$ (**mean/semi-var**), the mean and semi-mean absolute deviations $\mathbb{E}_{\mathbb{P}} [\tilde{z}_1 - \mu_1]_+$ and $\mathbb{E}_{\mathbb{P}} [\mu_1 - \tilde{z}_1]_+$ (**mean/semi-MAD**), as well as the mean and semi-Huber loss functions $H_1([\tilde{z}_1 - \mu_1]_+)$ and $H_1([\mu_1 - \tilde{z}_1]_+)$ (**mean/semi-Huber**). For ease of exposition, all results are presented as expected profits relative to the optimal solution that could be achieved under full knowledge of the demand distribution \mathbb{Q}^0 . For each ambiguity set, the lower curve presents the in-sample results, that is, the worst-case expected profits predicted by the objective function of the respective optimization problem, and the upper curve reports the out-of-sample results from a backtest under the true distribution \mathbb{Q}^0 . By construction of our ambiguity sets, the out-of-sample results exceed the in-sample results with a probability of at least 95%.

Figure 4 shows that with increasing standard deviation, both the in-sample and the out-of-sample results tend to deteriorate. This effect is much more pronounced for the symmetric indicators, which are unable to capture the asymmetry of \mathbb{Q}^0 . Among these indicators, the two robust measures significantly outperform the variance if the demand is skewed. This confirms our intuition that robust indicators are preferable to classical indicators when the distributions deviate from normality. Among the asymmetric indicators, **mean/semi-var** has the best out-of-sample performance, whereas **mean/semi-MAD** combines a good out-of-sample performance with an accurate in-sample prediction. It is worth noting that **mean/semi-MAD** gives rises to linear programming problems that can be solved very efficiently. We do not show the curves for the ambiguity set that only bounds the mean demand $\mathbb{E}_{\mathbb{P}} [\tilde{z}]$ as it results in zero order quantities in all of our experiments. Also, the inclusion of higher-order moments beyond the (semi-)variance does not lead to significant improvements. The same holds true for ambiguity sets constructed from combinations of several statistical indicators, such as the mean, variance and the semi-variances or the mean, semi-variances and the semi-mean absolute deviations.

We now fix the standard deviation to 0.75 and investigate the impact of the number of available

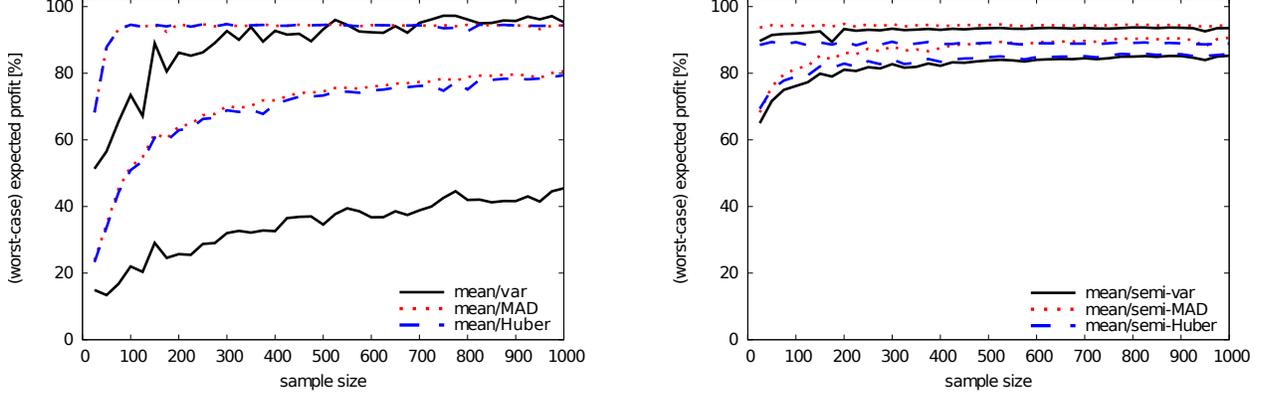


Figure 5. Single-product results for different symmetric (left) and asymmetric (right) statistical indicators. All results are averaged over 100 instances. The curves have the same meaning as in Figure 4.

demand samples on the performance of the ambiguity sets. Figure 5 shows that for increasing sample sizes, the in-sample and out-of-sample results improve for all ambiguity sets. The figure also shows that smaller sample sizes are sufficient for the ambiguity sets constructed from asymmetric measures, even though they require twice as many indicators to be estimated. We also observe that the robust indicators require less samples than their classical counterparts.

We now study the following risk-averse variant of the multi-product newsvendor problem:

$$\begin{aligned} & \text{minimize} && \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [U(L(\mathbf{x}, \tilde{\mathbf{z}}))] \\ & \text{subject to} && \mathbf{x} \geq \mathbf{0} \end{aligned}$$

Here, $U(y)$ approximates the exponential disutility function $e^{y/10}$ with 25 affine functions. We consider instances of this problem with $n = 3$ products and identical wholesale, retail and salvage prices of $\mathbf{c}_i = 5$, $\mathbf{v}_i = 10$ and $\mathbf{g}_i = 2.5$. The product demands are characterized by identical lognormal marginal distributions with mean 5 and a standard deviation of 0.75. The demands for the first two products are coupled by a copula whose parameter θ specifies the probability that both products exhibit an identical demand. Thus, the settings $\theta = 0$ and $\theta = 1$ correspond to the special cases of independent and perfectly dependent product demands, respectively. The demand for the third product is independent of the other two demands. We construct ambiguity sets from 750 i.i.d. samples of $\tilde{\mathbf{z}}$.

In the multi-product setting, the most elementary class of ambiguity sets only specifies bounds on component-wise indicators of the form $\mathbb{E}_{\mathbb{P}} [f_i(\tilde{\mathbf{z}}_i)]$, $i = 1, \dots, n$. For our problem, however, such

ambiguity sets do not capture the stochastic dependence between the product demands. In fact, the associated optimization problems hedge against a worst-case distribution in which all product demands exhibit strong dependences, and the resulting order quantities fall significantly below the optimal order quantities derived under full knowledge of the demand distribution \mathbb{Q}^0 . For the asymmetric indicators from the previous experiments, the component-wise ambiguity sets result in average order quantities of 2.0 (mean/semi-var and mean/semi-Huber) and 1.56 (mean/semi-MAD). The certainty equivalent of the associated out-of-sample expected utility ranges between 70.2% (for independent demands) and 77.9% (for perfectly dependent demands) of the certainty equivalent corresponding to the expected utility of the optimal order quantities.

To capture the dependence between the product demands, we now construct ambiguity sets from the component-wise semi-mean absolute deviations $\mathbb{E}_{\mathbb{P}}[\tilde{z}_i - \boldsymbol{\mu}_i]_+$ and $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\mu}_i - \tilde{z}_i]_+$, as well as all pairs of semi-mean absolute deviations $\mathbb{E}_{\mathbb{P}}[(\tilde{z}_i \pm \tilde{z}_j) - (\boldsymbol{\mu}_i \pm \boldsymbol{\mu}_j)]_+$ and $\mathbb{E}_{\mathbb{P}}[(\boldsymbol{\mu}_i \pm \boldsymbol{\mu}_j) - (\tilde{z}_i \pm \tilde{z}_j)]_+$. Figure 6 shows the resulting order quantities as functions of the copula parameter θ . We observe that the newsvendor orders less of the first two products when θ increases. This is intuitive as with increasingly dependent product demands, the newsvendor becomes exposed to the risk of low demand for both products. The certainty equivalents of the order quantities in Figure 6 range between 90.2% and 99.7% of the certainty equivalents of the optimal order quantities under full knowledge of \mathbb{Q}^0 . We do not show the results for the ambiguity sets constructed from the mean, variance and semi-variances, as well as the mean and the semi-Huber loss functions, since they are very similar to the ones in Figure 6.

Acknowledgments

The authors wish to express their gratitude to the referees for their constructive criticism, which led to substantial improvements of the paper. The first two authors also gratefully acknowledge financial support from the Engineering and Physical Sciences Research Council (EP/I014640/1).

References

- [1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.
- [2] D. Anderson, T. A. Williams, and D. J. Sweeney. *Fundamentals of Business Statistics*. South-Western College Publishing, 6th edition, 2011.

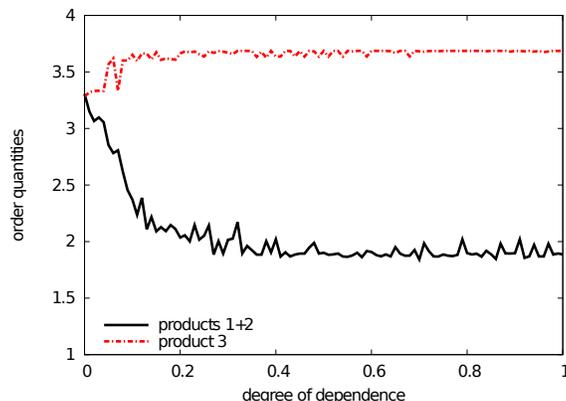


Figure 6. Three-product results for the ambiguity set specifying the mean and all component-wise and pairwise semi-mean absolute deviations. The curves show the average order quantities for each of the first two products and the third product as a function of the dependence parameter θ . All results are averaged over 100 instances.

- [3] J. Dupačová (as Žáčková). On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky*, 91(4):423–430, 1966.
- [4] A. Ben-Tal, D. Bertsimas, and D. B. Brown. A soft robust model for optimization under ambiguity. *Operations Research*, 58(4):1220–1234, 2010.
- [5] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.
- [6] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
- [7] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [8] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, 2001.
- [9] A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476, 2007.
- [10] D. Bertsimas and D. B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.
- [11] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.
- [12] D. Bertsimas, X. Vinh Doan, K. Natarajan, and C.-P. Teo. Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602, 2010.
- [13] D. Bertsimas, K. Natarajan, and C.-P. Teo. Persistence in discrete optimization under data uncertainty. *Mathematical Programming*, 108(2–3):251–274, 2006.
- [14] D. Bertsimas and I. Popescu. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal of Optimization*, 15(3):780–804, 2004.
- [15] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1–2):5–36, 2006.

- [16] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research. Springer, 1997.
- [17] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [18] D. B. Brown, E. G. De Giorgi, and M. Sim. Aspirational preferences and their representation by risk measures. *Accepted for publication in Management Science*, 2012.
- [19] D. B. Brown and M. Sim. Satisficing measures for analysis of risky positions. *Management Science*, 55(1):71–84, 2009.
- [20] G. C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.
- [21] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Thomson Learning, 2nd edition, 2002.
- [22] W. Chen and M. Sim. Goal driven optimization. *Operations Research*, 57(2):342–357, 2009.
- [23] W. Chen, M. Sim, J. Sun, and C.-P. Teo. From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Operations Research*, 58(2):470–485, 2010.
- [24] M. R. Chernick. *Bootstrap Methods: A Guide for Practitioners and Researchers*. Wiley-Blackwell, 2nd edition, 2007.
- [25] S.-S. Cheung, A. Man-Cho So, and K. Wang. Linear matrix inequalities with stochastically dependent perturbations and applications to chance-constrained semidefinite optimization. *SIAM Journal on Optimization*, 22(4):1394–1430, 2012.
- [26] E. Delage. *Distributionally Robust Optimization in context of Data-driven Problem*. PhD thesis, Stanford University, USA, 2009.
- [27] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):596–612, 2010.
- [28] V. DeMiguel and F. J. Nogales. Portfolio selection with robust estimation. *Operations Research*, 57(3):560–577, 2009.
- [29] L. G. Epstein. A definition of uncertainty aversion. *Review of Economic Studies*, 66(3):579–608, 1999.
- [30] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002.
- [31] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [32] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case Value-at-Risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.
- [33] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.
- [34] J. Goh and M. Sim. Distributionally robust optimization and its tractable approximations. *Operations Research*, 58(4):902–917, 2010.
- [35] M. Grötschel, L. Lovasz, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, 2nd edition, 1993.
- [36] Z. Hu and J. Hong. Kullback-Leibler divergence constrained distributionally robust optimization. *Optimization Online*, 2012.
- [37] K. Isii. On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14(1):185–197, 1962.

- [38] year = 2004 pages = 284–289 J. Löfberg title = YALMIP : A toolbox for modeling and optimization in MATLAB, booktitle = IEEE International Symposium on Computer Aided Control Systems Design.
- [39] B. Kawas and A. Thiele. A log-robust optimization approach to portfolio management. *OR Spectrum*, 33(1):207–233, 2011.
- [40] J. M. Keynes. *A treatise on probability*. MacMillan, 1921.
- [41] F. H. Knight. Risk, uncertainty and profit. *Hart, Schaffner and Marx*, 1921.
- [42] D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.
- [43] C. M. Lagoa and B. R. Barmish. Distributionally robust Monte Carlo simulation: A tutorial survey. In L. Basañez and J. A. de la Puente, editors, *Proceedings of the 15th IFAC World Congress*, volume 15, pages 1–12, 2002.
- [44] S.-W. Lam, T. S. Ng, M. Sim, and J.-H. Song. Multiple objectives satisficing under uncertainty. *Accepted for publication in Operations Research*, 2012.
- [45] R. Levi, G. Perakis, and J. Uichanco. The data-driven newsvendor problem: new bounds and insights. Technical report, MIT Sloan School of Management, USA, 2012.
- [46] X. Li, K. Natarajan, C.-P. Teo, and Z. Zheng. Distributionally robust mixed integer linear programs: Persistency models with applications. *Accepted for Publication in European Journal on Operational Research*, 2013.
- [47] M. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284(1–3):193–228, 1998.
- [48] R. Michaud. The Markowitz optimization enigma: Is ‘optimized’ optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.
- [49] K. Natarajan, D. Pachamanova, and M. Sim. Constructing risk measures from uncertainty sets. *Operations Research*, 57(5):1129–1141, 2009.
- [50] I. Popescu. A semidefinite programming approach to optimal-moment bounds for convex classes of distributions. *Mathematics of Operations Research*, 30(3):632–657, 2005.
- [51] A. Prékopa. *Stochastic Programming*. Kluwer Academic Publishers, 1995.
- [52] R. T. Rockafellar and S. Uryasev. Optimization of conditional Value-at-Risk. *Journal of Risk*, 2(3):21–41, 2000.
- [53] A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, 2003.
- [54] H. E. Scarf. A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, and H. E. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, 1958.
- [55] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 2003.
- [56] A. Shapiro. On duality theory of conic linear problems. In *Semi-Infinite Programming*, chapter 7, pages 135–165. Kluwer Academic Publishers, 2001.
- [57] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14(4):1237–1249, 2004.
- [58] A. Shapiro and A. Kleywegt. Minimax analysis of stochastic problems. *Optimization Methods and Software*, 17(3):523–542, 2002.

- [59] K. C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3 – a matlab software package for semidefinite programming, version 1.3. *Optimization Methods and Software*, 11(1–4):545–581, 1999.
- [60] W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov decision processes. *Accepted for publication in Mathematics of Operations Research*, 2012.
- [61] H. Xu, C. Caramanis, and S. Mannor. A distributional interpretation of robust optimization. *Accepted for Publication in Mathematics of Operations Research*, page DOI: 10.1287/moor.1110.0531, 2012.
- [62] H. Xu and S. Mannor. Distributionally robust Markov decision processes. *Mathematics of Operations Research*, 37(2):288–300, 2012.
- [63] S. Zymler, D. Kuhn, and B. Rustem. Worst-case Value-at-Risk of non-linear portfolios. *Accepted for publication in Management Science*, 2012.
- [64] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1–2):167–198, 2013.

Appendix A: Proofs

The proof of Theorem 1 requires the following auxiliary result.

Lemma 1. *Assume that $v(\mathbf{x}, \mathbf{z})$ satisfies (C3). Then the semi-infinite constraint*

$$v(\mathbf{x}, \mathbf{z}) + \mathbf{f}^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} \leq h \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i \quad (12)$$

is satisfied if and only if there is $\phi_l \in \mathcal{K}_i^$, $l \in \mathcal{L}$, such that $\mathbf{c}_i^\top \phi_l + \mathbf{s}_l^\top \mathbf{x} + t_l \leq h$, $\mathbf{C}_i^\top \phi_l = \mathbf{S}_l^\top \mathbf{x} + \mathbf{t}_l + \mathbf{f}$ and $\mathbf{D}_i^\top \phi_l = \mathbf{g}$ for all $l \in \mathcal{L}$.*

Proof. The semi-infinite constraint (12) is equivalent to

$$(\mathbf{S}_l^\top \mathbf{x} + \mathbf{t}_l + \mathbf{f})^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} \leq h - \mathbf{s}_l^\top \mathbf{x} - t_l \quad \forall l \in \mathcal{L}, \forall (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_i \mathbf{z} + \mathbf{D}_i \mathbf{u} \preceq_{\mathcal{K}_i} \mathbf{c}_i.$$

This constraint is satisfied if and only if the optimal values of the L strictly feasible problems

$$\begin{aligned} & \text{maximize} && (\mathbf{S}_l^\top \mathbf{x} + \mathbf{t}_l + \mathbf{f})^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} \\ & \text{subject to} && \mathbf{z} \in \mathbb{R}^P, \mathbf{u} \in \mathbb{R}^Q \\ & && \mathbf{C}_i \mathbf{z} + \mathbf{D}_i \mathbf{u} \preceq_{\mathcal{K}_i} \mathbf{c}_i \end{aligned}$$

do not exceed the values $h - \mathbf{s}_l^\top \mathbf{x} - t_l$, respectively. This is the case if and only if the optimal values of the L dual problems

$$\begin{aligned} & \text{minimize} && \mathbf{c}_i^\top \phi_l \\ & \text{subject to} && \phi_l \in \mathcal{K}_i^* \\ & && \mathbf{C}_i^\top \phi_l = \mathbf{S}_l^\top \mathbf{x} + \mathbf{t}_l + \mathbf{f} \\ & && \mathbf{D}_i^\top \phi_l = \mathbf{g}. \end{aligned}$$

do not exceed the values $h - \mathbf{s}_l^\top \mathbf{x} - t_l$, respectively, which is what the statement requires. \square

We are now ready to prove Theorem 1.

Proof of Theorem 1. The left-hand side of the distributionally robust constraint (3) coincides

with the optimal value of the following moment problem.

$$\begin{aligned}
& \text{maximize} && \int_{\mathcal{C}_I} v(\mathbf{x}, \mathbf{z}) \, d\mu(\mathbf{z}, \mathbf{u}) \\
& \text{subject to} && \mu \in \mathcal{M}_+(\mathbb{R}^P \times \mathbb{R}^Q) \\
& && \int_{\mathcal{C}_I} [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}] \, d\mu(\mathbf{z}, \mathbf{u}) = \mathbf{b} \\
& && \left. \begin{aligned} & \int_{\mathcal{C}_I} \mathbf{1}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} \, d\mu(\mathbf{z}, \mathbf{u}) \geq \underline{\mathbf{p}}_i \\ & \int_{\mathcal{C}_I} \mathbf{1}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} \, d\mu(\mathbf{z}, \mathbf{u}) \leq \bar{\mathbf{p}}_i \end{aligned} \right\} \quad \forall i \in \mathcal{I}
\end{aligned}$$

By assumption, we have $\underline{\mathbf{p}}_I = \bar{\mathbf{p}}_I = 1$. Hence, every feasible measure μ in this problem is naturally identified with a probability measure $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^Q)$ that is supported on \mathcal{C}_I . The dual of the moment problem is given by

$$\begin{aligned}
& \text{minimize} && \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] \\
& \text{subject to} && \boldsymbol{\beta} \in \mathbb{R}^K, \quad \boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I \\
& && [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} \mathbf{1}_{[(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i]} [\boldsymbol{\kappa}_i - \boldsymbol{\lambda}_i] \geq v(\mathbf{x}, \mathbf{z}) \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_I.
\end{aligned}$$

Strong duality is guaranteed by Proposition 3.4 in [56], which is applicable due to condition **(C2)**. Next, the nesting condition **(N)** allows us to partition the support \mathcal{C}_I into I nonempty and disjoint sets $\bar{\mathcal{C}}_i = \mathcal{C}_i \setminus \bigcup_{i' \in \mathcal{D}(i)} \mathcal{C}_{i'}$, $i = 1, \dots, I$, where $\mathcal{D}(i)$ denotes the index set of strict subsets of \mathcal{C}_i . The constraint in the dual problem is therefore equivalent to the constraint set

$$[\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] \geq v(\mathbf{x}, \mathbf{z}) \quad \forall (\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i, \quad \forall i \in \mathcal{I}.$$

We can then reformulate the i -th constraint as

$$\max_{(\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i} \left\{ v(\mathbf{x}, \mathbf{z}) - [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} - \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] \right\} \leq 0.$$

The expression inside the maximization inherits convexity from v , which implies that it is maximized on the boundary of $\bar{\mathcal{C}}_i$. Due to the nesting condition **(N)**, the boundary of $\bar{\mathcal{C}}_i$ coincides with the

boundary of \mathcal{C}_i . Thus, the robust expectation constraint (3) is satisfied if and only if

$$\begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq w \\ [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] &\geq v(\mathbf{x}, \mathbf{z}) \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i, \forall i \in \mathcal{I} \end{aligned}$$

is satisfied by some $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I$. The assertion now follows if we apply Lemma 1 to the second constraint by setting $\mathbf{f} = -\mathbf{A}^\top \boldsymbol{\beta}$, $\mathbf{g} = -\mathbf{B}^\top \boldsymbol{\beta}$ and $h = \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}]$. \square

For the proof of Theorem 2, we recall that the strongly \mathcal{NP} -hard 0/1 Integer Programming (IP) problem [31] is defined as follows.

0/1 INTEGER PROGRAMMING.

Instance. Given are $\mathbf{E} \in \mathbb{Z}^{M \times P}$, $\mathbf{f} \in \mathbb{Z}^M$, $\mathbf{g} \in \mathbb{Z}^P$, $\zeta \in \mathbb{Z}$.

Question. Is there a vector $\mathbf{y} \in \{0, 1\}^P$ such that $\mathbf{E}\mathbf{y} \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{y} \leq \zeta$?

Assume that $\mathbf{y}' \in [0, 1]^P$ constitutes a fractional vector that satisfies $\mathbf{E}\mathbf{y}' \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{y}' \leq \zeta$. The following lemma shows that we can obtain an integral vector $\mathbf{y} \in \{0, 1\}^P$ that satisfies $\mathbf{E}\mathbf{y} \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{y} \leq \zeta$ by rounding \mathbf{y}' if its components are ‘close enough’ to zero or one.

Lemma 2. *Let $0 < \epsilon \leq \min\{\epsilon_E, \epsilon_g\}$, where $0 < \epsilon_E \leq \min_m \left\{ \left(\sum_p |\mathbf{E}_{mp}| \right)^{-1} \right\}$ and $0 < \epsilon_g \leq \left(\sum_p |\mathbf{g}_p| \right)^{-1}$. Assume that $\mathbf{y}' \in ([0, \epsilon] \cup (1 - \epsilon, 1])^P$ satisfies $\mathbf{E}\mathbf{y}' \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{y}' \leq \zeta$. Then $\mathbf{E}\mathbf{y} \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{y} \leq \zeta$ for $\mathbf{y} \in \{0, 1\}^P$, where $\mathbf{y}_p = 1$ if $\mathbf{y}'_p > 1 - \epsilon$ and $\mathbf{y}_p = 0$ otherwise.*

Remark 2. *A proof of Lemma 2 can be found in [60]. To keep the paper self-contained, we repeat the proof here.*

Proof of Lemma 2. By construction, we have that $\mathbf{E}_m^\top \mathbf{y} \leq \mathbf{E}_m^\top \mathbf{y}' + \sum_p |\mathbf{E}_{mp}| \epsilon_E < \mathbf{E}_m^\top \mathbf{y}' + 1 \leq \mathbf{f}_m + 1$ for all $m \in \{1, \dots, M\}$. Similarly, we have that $\mathbf{g}^\top \mathbf{y} \leq \mathbf{g}^\top \mathbf{y}' + \sum_p |\mathbf{g}_p| \epsilon_g < \mathbf{g}^\top \mathbf{y}' + 1 \leq \zeta + 1$. Due to the integrality of \mathbf{E} , \mathbf{f} , \mathbf{g} , ζ and \mathbf{y} , we therefore conclude that $\mathbf{E}\mathbf{y} \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{y} \leq \zeta$. \square

Proof of Theorem 2. Fix an instance $(\mathbf{E}, \mathbf{f}, \mathbf{g}, \zeta)$ of the IP problem and consider the following instance of the ambiguity set \mathcal{P} defined in (4), where $\epsilon < \frac{1}{2}$ is chosen as prescribed in Lemma 2.

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P) : \begin{aligned} &\mathbb{P} \left(\begin{aligned} &\tilde{\mathbf{z}} \in [0, 1]^P, \quad \mathbf{E}\tilde{\mathbf{z}} \leq \mathbf{f}, \quad \mathbf{g}^\top \tilde{\mathbf{z}} \leq \zeta \\ &\|\tilde{\mathbf{z}} - \frac{1}{2}\mathbf{e}\|_2 \leq \sqrt{P} \left(\frac{1}{2} - \epsilon \right) \end{aligned} \right) = 0 \\ &\mathbb{P} \left(\begin{aligned} &\tilde{\mathbf{z}} \in [0, 1]^P, \quad \mathbf{E}\tilde{\mathbf{z}} \leq \mathbf{f}, \quad \mathbf{g}^\top \tilde{\mathbf{z}} \leq \zeta \end{aligned} \right) = 1 \end{aligned} \right\}$$

By construction, the first confidence set in the specification of \mathcal{P} (which is assigned probability 0) is a subset of the second confidence set (which is assigned probability 1).

Assume first that the ambiguity set \mathcal{P} is not empty. Then, fix some $\mathbb{P} \in \mathcal{P}$ and choose any vector $\mathbf{z} \in \mathbb{R}^P$ in the support of \mathbb{P} . By construction, we have $\mathbf{z} \in [0, 1]^P$, $\mathbf{E}\mathbf{z} \leq \mathbf{f}$, $\mathbf{g}^\top \mathbf{z} \leq \zeta$ and $\|\mathbf{z} - \frac{1}{2}\mathbf{e}\|_2 > \sqrt{P}(\frac{1}{2} - \epsilon)$, that is, $z_p \in [0, \epsilon) \cup (1 - \epsilon, 1]$ for all $p = 1, \dots, P$. We can then use Lemma 2 to round \mathbf{z} to a solution of the IP problem. Assume now that the instance of the IP problem is feasible, that is, there is $\mathbf{z} \in \{0, 1\}^P$ such that $\mathbf{E}\mathbf{z} \leq \mathbf{f}$ and $\mathbf{g}^\top \mathbf{z} \leq \zeta$. By construction, \mathbf{z} satisfies $\|\mathbf{z} - \frac{1}{2}\mathbf{e}\|_2 = \frac{1}{2}\sqrt{P} > \sqrt{P}(\frac{1}{2} - \epsilon)$. Thus, we have $\delta_{\mathbf{z}} \in \mathcal{P}$, where $\delta_{\mathbf{z}}$ represents the Dirac distribution that concentrates unit mass at \mathbf{z} . \square

Proof of Theorem 3. By construction, any $(\mathbf{y}, \boldsymbol{\delta}) \in \Gamma(\mathbf{x})$ satisfies

$$\begin{aligned} \sum_{j \in \mathcal{J}} \delta_j \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v(\mathbf{y}_j / \delta_j, \tilde{\mathbf{z}})] &\geq \sum_{j \in \mathcal{J}} \delta_j \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [v(\mathbf{y}_j / \delta_j, \tilde{\mathbf{z}})] \\ &\geq \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[v \left(\sum_{j \in \mathcal{J}} \mathbf{y}_j, \tilde{\mathbf{z}} \right) \right] = \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [v(\mathbf{x}, \tilde{\mathbf{z}})]. \end{aligned}$$

Here, the first inequality holds because $\mathcal{P} \subseteq \mathcal{P}^j$ for all $j \in \mathcal{J}$, the second inequality follows from the subadditivity of the supremum operator and the convexity of v in its first argument, and the identity is due to the definition of $\Gamma(\mathbf{x})$. Thus, the robust constraint (3) is implied by the infimal convolution constraint (7).

We now show that the infimal convolution constraint (7) is satisfied whenever the naïve approximation (6) is satisfied. Due to the strict non-negativity requirement $\boldsymbol{\delta} > \mathbf{0}$, we need a limiting argument to prove this implication. Assume that the naïve approximation (6) is satisfied and that the minimum in (6) is attained at $j^* \in \mathcal{J}$. Fix any $k \geq 2$ and set $\mathbf{y}_{j^*} = \mathbf{x}$ and $\delta_{j^*}(k) = 1 - \frac{1}{k}$, as well as $\mathbf{y}_j = \mathbf{0}$ and $\delta_j(k) = \frac{1}{k(J-1)}$ for $j \in \mathcal{J} \setminus \{j^*\}$. Then $(\mathbf{y}, \boldsymbol{\delta}(k)) \in \Gamma(\mathbf{x})$ and

$$\sum_{j \in \mathcal{J}} \delta_j(k) \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v(\mathbf{y}_j / \delta_j(k), \tilde{\mathbf{z}})] \xrightarrow{k \rightarrow \infty} \sup_{\mathbb{P} \in \mathcal{P}^{j^*}} \mathbb{E}_{\mathbb{P}} [v(\mathbf{x}, \tilde{\mathbf{z}})]$$

because $\sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v(\mathbf{0}, \tilde{\mathbf{z}})]$ is finite for all $j \in \mathcal{J}$. Hence, the infimal convolution constraint (7) is implied by the naïve approximation (6).

Finally, if $J = 1$, then we have $\mathcal{P} = \mathcal{P}^1$, and the distributionally robust constraint (3) and the naïve approximation (6) are equivalent. From the first part of the proof we can then conclude that all three constraints (3), (6) and (7) are equivalent. \square

Proof of Theorem 4. Assume first that $|\mathcal{S}| = 1$. In this case, the optimal value of the optimization problem coincides with the minimum of the optimal values of the J optimization problems

$$\begin{aligned} & \text{minimize} && \mathbf{d}^\top \mathbf{x} \\ & \text{subject to} && \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v_1(\mathbf{x}, \tilde{\mathbf{z}})] \leq \mathbf{w}_1 \\ & && \mathbf{x} \in \mathcal{X}. \end{aligned}$$

The assumptions of Theorem 1 are satisfied, and we conclude that each of these J optimization problems can be solved efficiently.

We now show that for $|\mathcal{S}| > 1$ the IP problem reduces to the problem in the theorem statement. To this end, fix an instance $(\mathbf{E}, \mathbf{f}, \mathbf{g}, \zeta)$ of the IP problem and consider the ambiguity set

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{2P}) : \tilde{\mathbf{z}} \in [0, 1]^{2P} \text{ } \mathbb{P}\text{-a.s.}, \tilde{\mathbf{z}}_p = 1 \text{ } \mathbb{P}\text{-a.s. } \forall p = 1, \dots, 2P \right\},$$

as well as the singleton partition $\mathcal{J} = \{\{1\}, \dots, \{2P\}\}$ with the associated $2P$ outer approximations

$$\mathcal{P}^j = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{2P}) : \tilde{\mathbf{z}} \in [0, 1]^{2P} \text{ } \mathbb{P}\text{-a.s.}, \tilde{\mathbf{z}}_j = 1 \text{ } \mathbb{P}\text{-a.s.} \right\} \quad \text{for } j = 1, \dots, 2P.$$

Note that the box constraints in the definition of the ambiguity set \mathcal{P} are redundant, and that \mathcal{P} satisfies the nesting condition (N). For the feasible region $\mathcal{X} = \{\mathbf{x} \in [0, 1]^P : \mathbf{E}\mathbf{x} \leq \mathbf{f}\}$, objective function coefficients $\mathbf{d} = \mathbf{g}$, index set $\mathcal{S} = \{1, \dots, P\}$, constraint functions $v_s(\mathbf{x}, \mathbf{z}) = -\mathbf{z}_s \mathbf{x}_s - \mathbf{z}_{s+P}(1 - \mathbf{x}_s)$ and right-hand sides $\mathbf{w}_s = -1$, $s \in \mathcal{S}$, the optimization problem in the theorem statement reads as follows.

$$\begin{aligned} & \text{minimize} && \mathbf{g}^\top \mathbf{x} \\ & \text{subject to} && \min \left\{ \sup_{\mathbb{P} \in \mathcal{P}^s} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})], \sup_{\mathbb{P} \in \mathcal{P}^{s+P}} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})], \right. \\ & && \left. \min_{j \in \mathcal{J} \setminus \{s, s+P\}} \sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})] \right\} \leq -1 \quad \forall s \in \mathcal{S} \\ & && \mathbf{E}\mathbf{x} \leq \mathbf{f}, \mathbf{x} \in [0, 1]^P. \end{aligned}$$

Note that the third argument of the outmost minimization operator in each distributionally robust constraints is redundant. Indeed, for each $j \in \mathcal{J} \setminus \{s, s+P\}$ there exists a probability distribution $\hat{\mathbb{P}} \in \mathcal{P}^j$ with $\hat{\mathbb{P}}[\tilde{\mathbf{z}}_s = \tilde{\mathbf{z}}_{s+P} = 0] = 1$, and thus we have

$$\sup_{\mathbb{P} \in \mathcal{P}^j} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})] \geq \mathbb{E}_{\hat{\mathbb{P}}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})] = 0 > -1.$$

The optimization problem therefore simplifies to

$$\begin{aligned}
& \text{minimize} && \mathbf{g}^\top \mathbf{x} \\
& \text{subject to} && \min \left\{ \sup_{\mathbb{P} \in \mathcal{P}^s} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})], \sup_{\mathbb{P} \in \mathcal{P}^{s+P}} \mathbb{E}_{\mathbb{P}} [v_s(\mathbf{x}, \tilde{\mathbf{z}})] \right\} \leq -1 \quad \forall s \in \mathcal{S} \\
& && \mathbf{E}\mathbf{x} \leq \mathbf{f}, \mathbf{x} \in [0, 1]^P.
\end{aligned} \tag{13}$$

The first argument of the minimization operator evaluates to

$$\sup_{\mathbb{P} \in \mathcal{P}^s} \mathbb{E}_{\mathbb{P}} [-\tilde{\mathbf{z}}_s \mathbf{x}_s - \tilde{\mathbf{z}}_{s+P}(1 - \mathbf{x}_s)] = -\mathbf{x}_s - (1 - \mathbf{x}_s) \inf_{\mathbb{P} \in \mathcal{P}^s} \mathbb{E}_{\mathbb{P}} [\tilde{\mathbf{z}}_{s+P}].$$

Because there is $\hat{\mathbb{P}} \in \mathcal{P}^s$ with $\hat{\mathbb{P}}[\tilde{\mathbf{z}}_{s+P} = 0] = 1$, this term is less or equal to -1 if and only if $\mathbf{x}_s = 1$.

Similarly, the second term inside the minimization expression in (13) simplifies to

$$\sup_{\mathbb{P} \in \mathcal{P}^{s+P}} \mathbb{E}_{\mathbb{P}} [-\tilde{\mathbf{z}}_s \mathbf{x}_s - \tilde{\mathbf{z}}_{s+P}(1 - \mathbf{x}_s)] = -(1 - \mathbf{x}_s) - \mathbf{x}_s \inf_{\mathbb{P} \in \mathcal{P}^{s+P}} \mathbb{E}_{\mathbb{P}} [\tilde{\mathbf{z}}_s].$$

Because there is $\hat{\mathbb{P}} \in \mathcal{P}^{s+P}$ with $\hat{\mathbb{P}}[\tilde{\mathbf{z}}_s = 0] = 1$, this term is less or equal to -1 if and only if $\mathbf{x}_s = 0$.

Hence, the optimization problem (13) is equivalent to

$$\begin{aligned}
& \text{minimize} && \mathbf{g}^\top \mathbf{x} \\
& \text{subject to} && \mathbf{E}\mathbf{x} \leq \mathbf{f}, \mathbf{x} \in \{0, 1\}^P,
\end{aligned}$$

which we readily identify as the strongly \mathcal{NP} -hard IP problem. \square

Proof of Proposition 1. For $j \in \mathcal{J}_2$, we define $\mathcal{J}_1(j) = \{j' \in \mathcal{J}_1 : \mathcal{I}_{j'}^1 \subseteq \mathcal{I}_j^2\}$ as the set of indices corresponding to elements of the partition $\{\mathcal{I}_{j'}^1\}_{j' \in \mathcal{J}_1}$ that are contained in \mathcal{I}_j^2 . Similarly, for $j \in \mathcal{J}_1$ we define $\mathcal{J}_2(j)$ as the index of the element of $\{\mathcal{I}_{j'}^2\}_{j' \in \mathcal{J}_2}$ that contains \mathcal{I}_j^1 , that is, $\mathcal{J}_2(j) = j'$ if and only if $j' \in \mathcal{J}_2$ and $\mathcal{I}_j^1 \subseteq \mathcal{I}_{j'}^2$. Let $\Gamma_1(\mathbf{x})$ and $\Gamma_2(\mathbf{x})$ denote the sets of feasible vectors $(\mathbf{y}^1, \boldsymbol{\delta}^1)$ and $(\mathbf{y}^2, \boldsymbol{\delta}^2)$ associated with the two partitions. Fix any $(\mathbf{y}^1, \boldsymbol{\delta}^1) \in \Gamma_1(\mathbf{x})$ and define $(\mathbf{y}^2, \boldsymbol{\delta}^2) \in \Gamma_2(\mathbf{x})$ through $\mathbf{y}_j^2 = \sum_{j' \in \mathcal{J}_1(j)} \mathbf{y}_{j'}^1$ and $\boldsymbol{\delta}_j^2 = \sum_{j' \in \mathcal{J}_1(j)} \boldsymbol{\delta}_{j'}^1$. We then obtain

$$\begin{aligned}
\sum_{j \in \mathcal{J}_1} \boldsymbol{\delta}_j^1 \sup_{\mathbb{P} \in \mathcal{P}_1^j} \mathbb{E}_{\mathbb{P}} [v(\mathbf{y}_j^1 / \boldsymbol{\delta}_j^1, \tilde{\mathbf{z}})] &\geq \sum_{j \in \mathcal{J}_1} \boldsymbol{\delta}_j^1 \sup_{\mathbb{P} \in \mathcal{P}_2^{\mathcal{J}_2(j)}} \mathbb{E}_{\mathbb{P}} [v(\mathbf{y}_j^1 / \boldsymbol{\delta}_j^1, \tilde{\mathbf{z}})] \\
&\geq \sum_{j \in \mathcal{J}_2} \left[\sum_{j' \in \mathcal{J}_1(j)} \boldsymbol{\delta}_{j'}^1 \right] \sup_{\mathbb{P} \in \mathcal{P}_2^j} \mathbb{E}_{\mathbb{P}} \left[v \left(\left[\sum_{j' \in \mathcal{J}_1(j)} \mathbf{y}_{j'}^1 \right] / \left[\sum_{j' \in \mathcal{J}_1(j)} \boldsymbol{\delta}_{j'}^1 \right], \tilde{\mathbf{z}} \right) \right] \\
&= \sum_{j \in \mathcal{J}_2} \boldsymbol{\delta}_j^2 \sup_{\mathbb{P} \in \mathcal{P}_2^j} \mathbb{E}_{\mathbb{P}} [v(\mathbf{y}_j^2 / \boldsymbol{\delta}_j^2, \tilde{\mathbf{z}})],
\end{aligned}$$

where the first inequality follows from the fact that $\mathcal{P}_2^{\mathcal{J}_2(j)} \subseteq \mathcal{P}_1^j$, which is a consequence of the assumption that $\mathcal{I}_j^1 \subseteq \mathcal{I}_{\mathcal{J}_2(j)}^2$. The second inequality holds because v is convex in its first argument, and the identity is due to the definition of $(\mathbf{y}^2, \boldsymbol{\delta}^2)$. Thus, the infimal convolution constraint (7) is satisfied for $\left\{ \mathcal{P}_2^j \right\}_{j \in \mathcal{J}_2}$ whenever it is satisfied for $\left\{ \mathcal{P}_1^j \right\}_{j \in \mathcal{J}_1}$. \square

Proof of Theorem 5. To prove statement (i), we first show that $\mathcal{P}' \subseteq \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$. To this end, fix any probability distribution $\mathbb{P}' \in \mathcal{P}_0(\mathbb{R}^P)$ that satisfies $\mathbb{E}_{\mathbb{P}'}[\mathbf{g}(\tilde{\mathbf{z}})] \preceq_{\mathcal{K}} \mathbf{f}$ and $\mathbb{P}'[\tilde{\mathbf{z}} \in \mathcal{C}_i] \in [\underline{\mathbf{p}}_i, \bar{\mathbf{p}}_i]$, $i \in \mathcal{I}$. We can then construct a probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}^M)$ such that $\mathbb{P}' = \Pi_{\tilde{\mathbf{z}}}\mathbb{P}$ and $\tilde{\mathbf{u}} = \mathbf{g}(\tilde{\mathbf{z}}) - \mathbb{E}_{\mathbb{P}}[\mathbf{g}(\tilde{\mathbf{z}})] + \mathbf{f}$ \mathbb{P} -a.s. By construction, \mathbb{P} satisfies $\mathbf{g}(\tilde{\mathbf{z}}) \preceq_{\mathcal{K}} \tilde{\mathbf{u}}$ \mathbb{P} -a.s. and $\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f}$. To prove that $\mathcal{P}' \supseteq \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$, fix any probability distribution $\mathbb{P} \in \mathcal{P}$. Since \mathcal{K} is a convex cone, $\mathbf{g}(\tilde{\mathbf{z}}) \preceq_{\mathcal{K}} \tilde{\mathbf{u}}$ \mathbb{P} -a.s. implies that $\mathbb{E}_{\mathbb{P}}[\mathbf{g}(\tilde{\mathbf{z}})] \preceq_{\mathcal{K}} \mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}]$. The statement now follows since $\mathbb{E}_{\mathbb{P}}[\tilde{\mathbf{u}}] = \mathbf{f}$ for all $\mathbb{P} \in \mathcal{P}$.

Statement (ii) follows immediately from our definition of conic representable \mathcal{K} -epigraphs. \square

Proof of Proposition 2. We first show that

$$\Pi_{\tilde{\mathbf{z}}}\mathcal{P} \subseteq \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{Q}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} < \theta] \leq \sigma, \mathbb{Q}[\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] = \rho \right\}.$$

To this end, fix any $\mathbb{P} \in \mathcal{P}$. By construction, we have that $\mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] = \rho$. Next, we show that $\mathbb{E}_{\mathbb{P}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{P}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} < \theta] \leq \sigma$. To this end, we note that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] &= \mathbb{E}_{\mathbb{P}}[\tilde{u} - \tilde{v} \mid \tilde{u} \geq \tilde{v}] \mathbb{P}[\tilde{u} \geq \tilde{v}] + \theta \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] \\ &\leq \mathbb{E}_{\mathbb{P}}[\tilde{u} \mid \tilde{u} \geq \tilde{v}] \mathbb{P}[\tilde{u} \geq \tilde{v}] + \theta \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] \\ &\leq \mathbb{E}_{\mathbb{P}}[\tilde{u}] + \theta \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta]. \end{aligned}$$

Here, the identity follows from the fact that $\tilde{u} - \tilde{v} = \mathbf{f}^\top \tilde{\mathbf{z}} - \theta$ \mathbb{P} -a.s. The first inequality holds because $\tilde{v} \geq 0$ \mathbb{P} -a.s., and the second inequality is due to the law of total expectation and the fact that $\mathbb{E}_{\mathbb{P}}[\tilde{u} \mid \tilde{u} < \tilde{v}] \geq 0$ since $\tilde{u} \geq 0$ \mathbb{P} -a.s. We thus conclude that $\mathbb{E}_{\mathbb{P}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} \geq \theta] \leq \rho^{-1} \mathbb{E}_{\mathbb{P}}[\tilde{u}] + \theta$. Using an analogous argument, we observe that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\mathbf{f}^\top \tilde{\mathbf{z}} \mid \mathbf{f}^\top \tilde{\mathbf{z}} < \theta] \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} < \theta] &= \mathbb{E}_{\mathbb{P}}[\tilde{u} - \tilde{v} \mid \tilde{u} < \tilde{v}] \mathbb{P}[\tilde{u} < \tilde{v}] + \theta \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} < \theta] \\ &\geq \mathbb{E}_{\mathbb{P}}[-\tilde{v} \mid \tilde{u} < \tilde{v}] \mathbb{P}[\tilde{u} < \tilde{v}] + \theta \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} < \theta] \\ &\geq \mathbb{E}_{\mathbb{P}}[-\tilde{v}] + \theta \mathbb{P}[\mathbf{f}^\top \tilde{\mathbf{z}} < \theta], \end{aligned}$$

that is, $\mathbb{E}_{\mathbb{P}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} < \theta] \geq (1 - \rho)^{-1} \mathbb{E}_{\mathbb{P}}[-\tilde{v}] + \theta$. From the definition of the ambiguity set \mathcal{P} we now conclude that

$$\mathbb{E}_{\mathbb{P}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{P}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} < \theta] \leq \rho^{-1} \mathbb{E}_{\mathbb{P}}[\tilde{u}] + (1 - \rho)^{-1} \mathbb{E}_{\mathbb{P}}[\tilde{v}] \leq \mathbb{E}_{\mathbb{P}}[\tilde{w}] = \sigma.$$

The claim now follows from the observation that functions of $\tilde{\mathbf{z}}$ (but not of $\tilde{u}, \tilde{v}, \tilde{w}$) have the same expected value under \mathbb{P} and $\Pi_{\tilde{\mathbf{z}}}\mathbb{P}$. We now show that

$$\Pi_{\tilde{\mathbf{z}}}\mathcal{P} \supseteq \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{Q}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} < \theta] \leq \sigma, \mathbb{Q}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] = \rho \right\}.$$

To this end, fix any probability distribution $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P)$ that satisfies $\mathbb{Q}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] = \rho$ and $\mathbb{E}_{\mathbb{Q}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{Q}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} < \theta] \leq \sigma$. We show that there is a probability distribution $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{Q} = \Pi_{\tilde{\mathbf{z}}}\mathbb{P}$, $\tilde{u} = [\mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta]^+$ and $\tilde{v} = [\theta - \mathbf{f}^{\top} \tilde{\mathbf{z}}]^+$ \mathbb{P} -a.s. Note that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{u}] &= \mathbb{E}_{\mathbb{P}}\left([\mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta]^+\right) = \mathbb{E}_{\mathbb{P}}\left([\mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta]^+ \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta \geq 0\right) \mathbb{P}[\mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta \geq 0] \\ &= \rho \mathbb{E}_{\mathbb{P}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] - \rho\theta. \end{aligned}$$

Here, the first identity follows from the definition of \tilde{u} . The second identity is due to the law of total expectation and the fact that $\mathbb{E}_{\mathbb{P}}([\mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta]^+ \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta < 0) = 0$, and the last identity follows from the fact that $\mathbb{P}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] = \mathbb{Q}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] = \rho$. An analogous argument shows that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\tilde{v}] &= \mathbb{E}_{\mathbb{P}}\left([\theta - \mathbf{f}^{\top} \tilde{\mathbf{z}}]^+\right) = \mathbb{E}_{\mathbb{P}}\left([\theta - \mathbf{f}^{\top} \tilde{\mathbf{z}}]^+ \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta < 0\right) \mathbb{P}[\mathbf{f}^{\top} \tilde{\mathbf{z}} - \theta < 0] \\ &= (1 - \rho) \mathbb{E}_{\mathbb{P}}[-\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} < \theta] + (1 - \rho)\theta. \end{aligned}$$

We thus conclude that $\mathbb{E}_{\mathbb{P}}[\rho^{-1}\tilde{u} + (1 - \rho)^{-1}\tilde{v}] = \mathbb{E}_{\mathbb{P}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} \geq \theta] - \mathbb{E}_{\mathbb{P}}[\mathbf{f}^{\top} \tilde{\mathbf{z}} \mid \mathbf{f}^{\top} \tilde{\mathbf{z}} < \theta] \leq \sigma$, which implies that there is indeed a probability distribution $\mathbb{P} \in \mathcal{P}$ such that $\mathbb{Q} = \Pi_{\tilde{\mathbf{z}}}\mathbb{P}$. \square

Proof of Proposition 3. Assertion 1 follows directly from the definition of the marginal median, while assertion 2 is an immediate consequence of Theorem 5. For the remainder of the proof, we introduce the shorthand notation $[x]_+ = \max\{x, 0\}$ for $x \in \mathbb{R}$.

To prove assertion 3, we first show that $\{\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[H_{\delta}(\mathbf{f}^{\top} \tilde{\mathbf{z}})] \leq g\} \subseteq \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$. To this end, fix any probability distribution $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P)$ that satisfies $\mathbb{E}_{\mathbb{Q}}[H_{\delta}(\mathbf{f}^{\top} \tilde{\mathbf{z}})] \leq g$. Next we construct a probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}_+^5)$ such that $\Pi_{\tilde{\mathbf{z}}}\mathbb{P} = \mathbb{Q}$, $\tilde{u} = [\mathbf{f}^{\top} \tilde{\mathbf{z}}]_+$, $\tilde{v} =$

$[-\mathbf{f}^\top \tilde{\mathbf{z}}]_+$, $\tilde{s} = \min\{\tilde{u}, \delta\}$ and $\tilde{t} = \min\{\tilde{v}, \delta\}$ \mathbb{P} -a.s. By construction, we have \mathbb{P} -a.s. that

$$\begin{aligned} \delta(\tilde{u} - \tilde{s}) + \frac{\tilde{s}^2}{2} + \delta(\tilde{v} - \tilde{t}) + \frac{\tilde{t}^2}{2} &= \delta \left| \mathbf{f}^\top \tilde{\mathbf{z}} \right| - \delta \min\left\{ \left| \mathbf{f}^\top \tilde{\mathbf{z}} \right|, \delta \right\} + \frac{\min\left\{ \left| \mathbf{f}^\top \tilde{\mathbf{z}} \right|, \delta \right\}^2}{2} \\ &= \begin{cases} \frac{1}{2} \left| \mathbf{f}^\top \tilde{\mathbf{z}} \right|^2 & \text{if } \left| \mathbf{f}^\top \tilde{\mathbf{z}} \right| \leq \delta, \\ \delta \left(\left| \mathbf{f}^\top \tilde{\mathbf{z}} \right| - \frac{1}{2} \delta \right) & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, we have $\mathbb{E}_{\mathbb{P}} \left[\delta(\tilde{u} - \tilde{s}) + \frac{\tilde{s}^2}{2} + \delta(\tilde{v} - \tilde{t}) + \frac{\tilde{t}^2}{2} \right] = \mathbb{E}_{\mathbb{P}} [H_\delta(\mathbf{f}^\top \tilde{\mathbf{z}})] \leq g$, which allows us to construct a probability measure $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^P \times \mathbb{R}_+^5)$ such that $\tilde{w} \geq \delta(\tilde{u} - \tilde{s}) + \frac{\tilde{s}^2}{2} + \delta(\tilde{v} - \tilde{t}) + \frac{\tilde{t}^2}{2}$ \mathbb{P} -a.s. and $\mathbb{E}_{\mathbb{P}}[\tilde{w}] \leq g$. We then conclude that $\mathbb{P} \in \mathcal{P}$, that is, we indeed have $\mathbb{Q} \in \Pi_{\tilde{\mathbf{z}}}\mathcal{P}$ as claimed.

We now show that $\Pi_{\tilde{\mathbf{z}}}\mathcal{P} \subseteq \{\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^P) : \mathbb{E}_{\mathbb{Q}}[H_\delta(\mathbf{f}^\top \tilde{\mathbf{z}})] \leq g\}$. To this end, fix any probability distribution $\mathbb{P} \in \mathcal{P}$. By construction, we have that

$$\begin{aligned} g &= \mathbb{E}_{\mathbb{P}}[\tilde{w}] \geq \mathbb{E}_{\mathbb{P}} \left[\delta(\tilde{u} - \tilde{s}) + \frac{\tilde{s}^2}{2} + \delta(\tilde{v} - \tilde{t}) + \frac{\tilde{t}^2}{2} \right] \\ &\geq \mathbb{E}_{\mathbb{P}} \left[\delta(\tilde{u} - \min\{\tilde{u}, \delta\}) + \frac{\min\{\tilde{u}, \delta\}^2}{2} + \delta(\tilde{v} - \min\{\tilde{v}, \delta\}) + \frac{\min\{\tilde{v}, \delta\}^2}{2} \right] \\ &= \mathbb{E}_{\mathbb{P}} \left[\mathbb{I}_{[\tilde{u} + \tilde{v} \leq \delta]} \cdot \frac{1}{2}(\tilde{u} + \tilde{v})^2 + \mathbb{I}_{[\tilde{u} + \tilde{v} > \delta]} \cdot \delta \left(\tilde{u} + \tilde{v} - \frac{1}{2} \delta \right) \right] \\ &\geq \mathbb{E}_{\mathbb{P}} \left[\mathbb{I}_{[|\mathbf{f}^\top \tilde{\mathbf{z}}| \leq \delta]} \cdot \frac{1}{2} \left| \mathbf{f}^\top \tilde{\mathbf{z}} \right|^2 + \mathbb{I}_{[|\mathbf{f}^\top \tilde{\mathbf{z}}| > \delta]} \cdot \delta \left(\left| \mathbf{f}^\top \tilde{\mathbf{z}} \right| - \frac{1}{2} \delta \right) \right] = \mathbb{E}_{\mathbb{P}} [H_\delta(\mathbf{f}^\top \tilde{\mathbf{z}})]. \end{aligned}$$

Here, the second inequality holds since for fixed $u \in \mathbb{R}$, the function $\delta(u-s) + \frac{s^2}{2}$ attains its minimum over the interval $[0, u]$ at $\min\{u, \delta\}$; the same holds true for v and t . The identity in the next to last row follows from a case distinction. Since the expression in this row is increasing in $\tilde{u} + \tilde{v}$, the definition of \mathcal{P} implies that the expression is minimized when $\tilde{u} = [\mathbf{f}^\top \tilde{\mathbf{z}}]_+$ and $\tilde{v} = [-\mathbf{f}^\top \tilde{\mathbf{z}}]_+$ \mathbb{P} -a.s., which leads to the expression in the last row. We thus conclude that $\mathbb{E}_{\mathbb{P}} [H_\delta(\mathbf{f}^\top \tilde{\mathbf{z}})] \leq g$ whenever $\mathbb{P} \in \mathcal{P}$, which completes the proof. \square

Appendix B: E-Companion

We keep the regularity conditions **(C1)** and **(C2)** regarding the ambiguity set \mathcal{P} , but we replace the condition **(C3)** concerning the constraint function v with the following two conditions.

(C3a) The function $v(\mathbf{x}, \mathbf{z})$ is convex in \mathbf{x} for all $\mathbf{z} \in \mathbb{R}^P$ and can be evaluated in polynomial time.²

(C3b) For $i \in \mathcal{I}$, $\mathbf{x} \in \mathcal{X}$ and $\theta \in \mathbb{R}$, it can be decided in polynomial time whether

$$\max_{(\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i} v(\mathbf{x}, \mathbf{z}) \leq \theta. \quad (14)$$

Moreover, if (14) is not satisfied, then a separating hyperplane $(\boldsymbol{\pi}, \phi) \in \mathbb{R}^N \times \mathbb{R}$ can be constructed in polynomial time such that $\boldsymbol{\pi}^\top \mathbf{x} > \phi$ and $\boldsymbol{\pi}^\top \hat{\mathbf{x}} \leq \phi$ for all $\hat{\mathbf{x}} \in \mathcal{X}$ satisfying (14).

One readily verifies that the condition **(C3)** from the main paper implies both **(C3a)** and **(C3b)**.

If **(C3a)** fails to hold, then the distributionally robust expectation constraint (3) may have a non-convex feasible set. Condition **(C3b)** will enable us to efficiently separate the semi-infinite constraints that arise from the dual reformulation of constraint (3). The conditions **(C3a)** and **(C3b)** are satisfied by a wide class of constraint functions v that are convex in \mathbf{x} and convex and piecewise affine in \mathbf{z} . In the following, we will show that both conditions are also satisfied for constraint functions that are convex in \mathbf{x} and convex and piecewise (conic-)quadratic in \mathbf{z} , provided that the confidence set \mathcal{C}_i in the ambiguity set \mathcal{P} are described by ellipsoids. We will also show that both conditions are satisfied by certain constraint functions that are non-convex in \mathbf{z} as long as the number of confidence regions is small and that all confidence regions constitute polyhedra.

We first provide tractable reformulations of the distributionally robust expectation constraint (3) under the relaxed conditions **(C3a)** and **(C3b)**. Afterwards, we present various classes of constraint functions that satisfy the conditions **(C3a)** and **(C3b)** and that give rise to conic optimization problems that can be solved with standard optimization software.

B.1 Tractable Reformulation for Generic Constraints

Theorem 1 in the main paper provides a tractable reformulation for the distributionally robust expectation constraint (3). In its proof, we first re-express constraint (3) in terms of semi-infinite

²Here and in the following, “polynomial time” is understood relative to the length of the input data (*i.e.*, the constraint function v and the ambiguity set \mathcal{P}) and $\log \epsilon^{-1}$ for a pre-specified approximation tolerance ϵ .

constraints, and we afterwards apply robust optimization techniques to obtain a tractable reformulation of these semi-infinite constraints. Condition **(C3)** is not needed for the first step if the constraint function $v(\mathbf{x}, \mathbf{z})$ is convex in \mathbf{z} .

Theorem 6 (Convex Constraint Functions). *Assume that the conditions **(C1)**, **(C2)** and **(N)** hold and that the constraint function $v(\mathbf{x}, \mathbf{z})$ is convex in \mathbf{z} . Then, the distributionally robust constraint (3) is satisfied for the ambiguity set (4) if and only if the semi-infinite constraint system*

$$\begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq w \\ [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] &\geq v(\mathbf{x}, \mathbf{z}) \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i, \forall i \in \mathcal{I} \end{aligned}$$

is satisfied by some $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I$.

Proof. The statement follows immediately from the proof of Theorem 1 in the main paper. \square

The semi-infinite constraints in Theorem 6 are tractable if and only if **(C3a)** and **(C3b)** hold. This follows from Theorem 3.1 in the celebrated treatise on the ellipsoid method by Grötschel *et al.* [35]. Next, we investigate situations in which $v(\mathbf{x}, \mathbf{z})$ fails to be convex in \mathbf{z} .

Theorem 7 (Nonconvex Constraint Functions). *Assume that the conditions **(C1)**, **(C2)** and **(N)** hold and that the confidence regions \mathcal{C}_i constitute polyhedra, that is, $\mathcal{K}_i = \mathbb{R}_+^{L_i}$ for all $i \in \mathcal{I}$. Then, the distributionally robust constraint (3) is satisfied for the ambiguity set (4) if and only if the semi-infinite constraint system*

$$\begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq w \\ [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] &\geq v(\mathbf{x}, \mathbf{z}) \quad \forall (\mathbf{z}, \mathbf{u}) \in \left[\bigcap_{i' \in \mathcal{D}(i)} \bar{\mathcal{C}}_{i', l_{i'}} \right] \cap \mathcal{C}_i, \forall i \in \mathcal{I}, \\ \forall l &= (l_{i'})_{i' \in \mathcal{D}(i)} : l_{i'} \in \{1, \dots, L_{i'}\} \quad \forall i' \in \mathcal{D}(i) \end{aligned}$$

is satisfied by some $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I$. Here, $\bar{\mathcal{C}}_{i, l} = \{(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : \mathbf{C}_{il}^\top \mathbf{z} + \mathbf{D}_{il}^\top \mathbf{u} \geq \mathbf{c}_{il}\}$ denotes the closed complement of the l -th halfspace defining the confidence region \mathcal{C}_i .

Proof. Following the argument in Theorem 1 in the main paper, the distributionally robust constraint (3) is satisfied if and only if there is $\boldsymbol{\beta} \in \mathbb{R}^K$ and $\boldsymbol{\kappa}, \boldsymbol{\lambda} \in \mathbb{R}_+^I$ such that

$$\begin{aligned} \mathbf{b}^\top \boldsymbol{\beta} + \sum_{i \in \mathcal{I}} [\bar{\mathbf{p}}_i \boldsymbol{\kappa}_i - \underline{\mathbf{p}}_i \boldsymbol{\lambda}_i] &\leq w \\ [\mathbf{A}\mathbf{z} + \mathbf{B}\mathbf{u}]^\top \boldsymbol{\beta} + \sum_{i' \in \mathcal{A}(i)} [\boldsymbol{\kappa}_{i'} - \boldsymbol{\lambda}_{i'}] &\geq v(\mathbf{x}, \mathbf{z}) \quad \forall (\mathbf{z}, \mathbf{u}) \in \bar{\mathcal{C}}_i, \forall i \in \mathcal{I}. \end{aligned}$$

Contrary to the constraints in the statement of Theorem 6, the semi-infinite constraints only have to be satisfied for all vectors (\mathbf{z}, \mathbf{u}) belonging to the non-convex sets $\bar{\mathcal{C}}_i$.

Using De Morgan's laws and elementary set algebraic transformations, we can express $\bar{\mathcal{C}}_i$ as

$$\begin{aligned} \bar{\mathcal{C}}_i &= \mathcal{C}_i \setminus \bigcup_{i' \in \mathcal{D}(i)} \mathcal{C}_{i'} = \bigcap_{i' \in \mathcal{D}(i)} \mathcal{C}_i \setminus \mathcal{C}_{i'} = \bigcap_{i' \in \mathcal{D}(i)} \bigcup_{l_{i'} \in \{1, \dots, L_{i'}\}} \mathcal{C}_i \cap \text{int} \bar{\mathcal{C}}_{i', l_{i'}} \\ &= \bigcup_{\substack{l=(l_{i'})_{i' \in \mathcal{D}(i)}: \\ l_{i'} \in \{1, \dots, L_{i'}\} \forall i' \in \mathcal{D}(i)}} \bigcap_{i' \in \mathcal{D}(i)} \mathcal{C}_i \cap \text{int} \bar{\mathcal{C}}_{i', l_{i'}}. \end{aligned}$$

The statement of the theorem now follows from the continuity of v in \mathbf{z} , which allows us to replace $\text{int} \bar{\mathcal{C}}_{i', l_{i'}}$ with its closure $\bar{\mathcal{C}}_{i', l_{i'}}$. \square

Individually, each semi-infinite constraint in Theorem 7 is tractable if and only if the conditions **(C3a)** and **(C3b)** are satisfied, see [35]. Note, however, that the number of semi-infinite constraints grows exponentially with the number I of confidence regions. Thus, Theorem 7 is primarily of interest for ambiguity sets with a small number of confidence regions. We remark that apart from the absolute mean spread and the marginal median, all statistical indicators of Section 3 result in ambiguity sets with a single confidence region. In those cases, Theorem 7 provides a tractable reformulation of the distributionally robust constraint (3).

B.2 Reformulation as Conic Optimization Problems

Theorems 6 and 7 demonstrate that the distributionally robust constraint (3) with ambiguity set (4) admits an equivalent dual representation that involves several robust constraints of the form

$$v(\mathbf{x}, \mathbf{z}) + \mathbf{f}^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} \leq h \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i, \quad (15)$$

where $\mathbf{f} \in \mathbb{R}^P$, $\mathbf{g} \in \mathbb{R}^Q$ and $h \in \mathbb{R}$ are interpreted as auxiliary decision variables, while \mathcal{C}_i is defined as in (5). Robust constraints of the form (15) are tractable if and only if the conditions **(C3a)** and **(C3b)** are satisfied. Unlike **(C3a)**, condition **(C3b)** may not be easy to check. In the following, we provide a more elementary condition that implies **(C3b)**.

Observation 2. *If $v(\mathbf{x}, \mathbf{z})$ is concave in \mathbf{z} for all admissible \mathbf{x} and if one can compute subgradients with respect to \mathbf{x} and supergradients with respect to \mathbf{z} in polynomial time, then **(C3b)** holds.*

The proof of Observation 2 is standard and thus omitted, see *e.g.* [27, 35]. We emphasize that a wide range of convex-concave constraint functions satisfy the conditions of Observation 2.

For example, $v(\mathbf{x}, \tilde{\mathbf{z}})$ could represent the losses (negative gains) of a portfolio involving long-only positions in European stock options [63]. In this situation \mathbf{x} denotes the nonnegative asset weights, which enter the loss function linearly, while $\tilde{\mathbf{z}}$ reflects the random stock returns, which enter the loss function in a concave manner due to the convexity of the option payoffs. The conditions of Observation 2 also hold, for instance, if $v(\mathbf{x}, \tilde{\mathbf{z}})$ represents the losses of a long-only asset portfolio where the primitive uncertain variables $\tilde{\mathbf{z}}$ reflect the assets' log returns [39].

Although (C3a) and (C3b) represent the weakest conditions to guarantee tractability of (15), the methods required to solve the resulting optimization problems (*e.g.* the ellipsoid method [26]) can still be slow and suffer from numerical instabilities. Therefore, we now characterize classes of constraint functions for which the robust constraint (15) admits an equivalent reformulation or a conservative approximation in terms of polynomially many conic inequalities. The resulting conic problem can then be solved efficiently and reliably with standard optimization software.

From the proof of Theorem 1 in the main text, we can directly extract the following result.

Observation 3 (Bi-Affine Functions). *Assume that $v(\mathbf{x}, \mathbf{z}) = \mathbf{s}(\mathbf{z})^\top \mathbf{x} + t(\mathbf{z})$ with $\mathbf{s}(\mathbf{z}) = \mathbf{S}\mathbf{z} + \mathbf{s}$, $\mathbf{S} \in \mathbb{R}^{N \times P}$ and $\mathbf{s} \in \mathbb{R}^N$, as well as $t(\mathbf{z}) = \mathbf{t}^\top \mathbf{z} + t$, $\mathbf{t} \in \mathbb{R}^P$ and $t \in \mathbb{R}$. Then the following two statements are equivalent.*

- (i) *The semi-infinite constraint (15) is satisfied, and*
- (ii) *there is $\boldsymbol{\lambda} \in \mathcal{K}_i^*$ such that $\mathbf{c}_i^\top \boldsymbol{\lambda} + \mathbf{s}^\top \mathbf{x} + t \leq h$, $\mathbf{C}_i^\top \boldsymbol{\lambda} = \mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f}$ and $\mathbf{D}_i^\top \boldsymbol{\lambda} = \mathbf{g}$.*

Here, \mathcal{K}_i^* represents the cone dual to \mathcal{K}_i .

If the confidence set \mathcal{C}_i is described by linear, conic quadratic or semidefinite inequalities, then Observation 3 provides a linear, conic quadratic or semidefinite reformulation of (15), respectively. Bi-affine constraint functions can represent the portfolio losses in asset allocation models with uncertain returns or the total costs in revenue management models with demand uncertainty.

Next, we study a richer class of functions $v(\mathbf{x}, \mathbf{z})$ that are quadratic in \mathbf{x} and affine in \mathbf{z} .

Proposition 4 (Quadratic-Affine Functions). *Assume that $v(\mathbf{x}, \mathbf{z}) = \mathbf{x}^\top \mathbf{S}(\mathbf{z}) \mathbf{x} + \mathbf{s}(\mathbf{z})^\top \mathbf{x} + t(\mathbf{z})$ with $\mathbf{S}(\mathbf{z}) = \sum_{p=1}^P \mathbf{S}_p \mathbf{z}_p + \mathbf{S}_0$, $\mathbf{S}_p \in \mathbb{S}^N$ for $p = 0, \dots, P$ and $\mathbf{S}(\mathbf{z}) \succeq \mathbf{0}$ for all $\mathbf{z} \in \mathcal{C}_i$, $\mathbf{s}(\mathbf{z}) = \mathbf{S}\mathbf{z} + \mathbf{s}$, $\mathbf{S} \in \mathbb{R}^{N \times P}$ and $\mathbf{s} \in \mathbb{R}^N$, as well as $t(\mathbf{z}) = \mathbf{t}^\top \mathbf{z} + t$, $\mathbf{t} \in \mathbb{R}^P$ and $t \in \mathbb{R}$. Then the following two statements are equivalent.*

(i) The semi-infinite constraint (15) is satisfied, and

(ii) there is $\boldsymbol{\lambda} \in \mathcal{K}_i^*$ and $\boldsymbol{\Gamma} \in \mathbb{S}^N$ such that $\mathbf{c}_i^\top \boldsymbol{\lambda} + \langle \mathbf{S}_0, \boldsymbol{\Gamma} \rangle + \mathbf{s}^\top \mathbf{x} + t \leq h$, $\mathbf{D}_i^\top \boldsymbol{\lambda} = \mathbf{g}$ and

$$\mathbf{C}_i^\top \boldsymbol{\lambda} - \begin{pmatrix} \langle \mathbf{S}_1, \boldsymbol{\Gamma} \rangle \\ \vdots \\ \langle \mathbf{S}_P, \boldsymbol{\Gamma} \rangle \end{pmatrix} = \mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f}, \quad \begin{pmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \boldsymbol{\Gamma} \end{pmatrix} \succcurlyeq \mathbf{0}.$$

Here, \mathcal{K}_i^* represents the cone dual to \mathcal{K}_i , and $\langle \cdot, \cdot \rangle$ denotes the trace product.

Proof. Since $\mathbf{S}(\mathbf{z}) \succcurlyeq \mathbf{0}$ over \mathcal{C}_i , the semi-infinite constraint (15) is equivalent to

$$\begin{aligned} \mathbf{x}^\top \mathbf{T} \mathbf{x} + \left[\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f} \right]^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} &\leq h - \mathbf{s}^\top \mathbf{x} - t \quad \forall (\mathbf{z}, \mathbf{u}, \mathbf{T}) \in \mathbb{R}^P \times \mathbb{R}^Q \times \mathbb{S}_+^N : \\ \mathbf{C}_i \mathbf{z} + \mathbf{D}_i \mathbf{u} &\preccurlyeq_{\mathcal{K}_i} \mathbf{c}_i, \\ \mathbf{T} &= \sum_{p=1}^P \mathbf{S}_p \mathbf{z}_p + \mathbf{S}_0. \end{aligned}$$

This constraint is equivalent to the requirement that the optimal value of the optimization problem

$$\begin{aligned} &\text{maximize} \quad \langle \mathbf{x} \mathbf{x}^\top, \mathbf{T} \rangle + \left[\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f} \right]^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} \\ &\text{subject to} \quad \mathbf{z} \in \mathbb{R}^P, \quad \mathbf{u} \in \mathbb{R}^Q, \quad \mathbf{T} \in \mathbb{S}_+^N \\ &\quad \mathbf{C}_i \mathbf{z} + \mathbf{D}_i \mathbf{u} \preccurlyeq_{\mathcal{K}_i} \mathbf{c}_i \\ &\quad \mathbf{T} = \sum_{p=1}^P \mathbf{S}_p \mathbf{z}_p + \mathbf{S}_0 \end{aligned}$$

does not exceed $h - \mathbf{s}^\top \mathbf{x} - t$. This is the case if and only if the optimal value of the dual problem

$$\begin{aligned} &\text{minimize} \quad \mathbf{c}_i^\top \boldsymbol{\lambda} + \langle \mathbf{S}_0, \boldsymbol{\Gamma} \rangle \\ &\text{subject to} \quad \boldsymbol{\lambda} \in \mathcal{K}_i^*, \quad \boldsymbol{\Gamma} \in \mathbb{S}^N \\ &\quad \mathbf{C}_i^\top \boldsymbol{\lambda} - \begin{pmatrix} \langle \mathbf{S}_1, \boldsymbol{\Gamma} \rangle \\ \vdots \\ \langle \mathbf{S}_P, \boldsymbol{\Gamma} \rangle \end{pmatrix} = \mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f} \\ &\quad \mathbf{D}_i^\top \boldsymbol{\lambda} = \mathbf{g}, \quad \boldsymbol{\Gamma} \succcurlyeq \mathbf{x} \mathbf{x}^\top \end{aligned}$$

does not exceed $h - \mathbf{s}^\top \mathbf{x} - t$. Applying Schur's complement [17] to the last constraint, we see that this is precisely what the second statement requires. \square

In contrast to Observation 3, Proposition 4 results in a semidefinite reformulation of the constraint (15) even if the confidence sets \mathcal{C}_i are described by linear or conic quadratic inequalities.

Next, we consider two classes of constraint functions that are quadratic in \mathbf{z} . In order to maintain tractability, we now assume that the confidence set \mathcal{C}_i is representable as an intersections of ellipsoids, that is,

$$\mathcal{C}_i = \left\{ \boldsymbol{\xi} = (\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q : (\boldsymbol{\xi} - \boldsymbol{\mu}_j)^\top \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}_j) \leq 1 \quad \forall j = 1, \dots, J_i \right\}, \quad (16)$$

where $J_i \in \mathbb{N}$, $\boldsymbol{\mu}_j \in \mathbb{R}^{P+Q}$ and $\boldsymbol{\Sigma}_j^{-1} \in \mathbb{S}^{P+Q}$ positive definite. Propositions 5 and 6 below provide conservative approximations for the robust constraint (15). These approximations become exact if \mathcal{C}_i reduces to a single ellipsoid.

Proposition 5 (Affine-Quadratic Functions). *Let $v(\mathbf{x}, \mathbf{z}) = \mathbf{z}^\top \mathbf{S}(\mathbf{x}) \mathbf{z} + s(\mathbf{z})^\top \mathbf{x} + t(\mathbf{z})$ with $\mathbf{S}(\mathbf{x}) = \sum_{n=1}^N \mathbf{S}_n \mathbf{x}_n + \mathbf{S}_0$, $\mathbf{S}_n \in \mathbb{S}^P$ for $s = 0, \dots, N$, $s(\mathbf{z}) = \mathbf{S} \mathbf{z} + \mathbf{s}$, $\mathbf{S} \in \mathbb{R}^{N \times P}$ and $\mathbf{s} \in \mathbb{R}^N$, as well as $t(\mathbf{z}) = \mathbf{t}^\top \mathbf{z} + t$, $\mathbf{t} \in \mathbb{R}^P$ and $t \in \mathbb{R}$. Assume that the confidence set \mathcal{C}_i is defined as in (16), and consider the following two statements.*

(i) *The semi-infinite constraint (15) is satisfied, and*

(ii) *there is $\boldsymbol{\lambda} \in \mathbb{R}_+^{J_i}$ such that*

$$\begin{pmatrix} \gamma(\mathbf{x}) - \sum_j \lambda_j (1 - \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j) & -\frac{1}{2} \boldsymbol{\pi}(\mathbf{x})^\top - \boldsymbol{\mu}(\boldsymbol{\lambda})^\top \\ -\frac{1}{2} \boldsymbol{\pi}(\mathbf{x}) - \boldsymbol{\mu}(\boldsymbol{\lambda}) & \boldsymbol{\Sigma}(\boldsymbol{\lambda}) - \begin{bmatrix} \mathbf{S}(\mathbf{x}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{pmatrix} \succcurlyeq \mathbf{0},$$

where $\gamma(\mathbf{x}) = h - t - \mathbf{s}^\top \mathbf{x}$, $\boldsymbol{\pi}(\mathbf{x}) = ([\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f}]^\top, \mathbf{g}^\top)^\top$, $\boldsymbol{\mu}(\boldsymbol{\lambda}) = \sum_j \lambda_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}(\boldsymbol{\lambda}) = \sum_j \lambda_j \boldsymbol{\Sigma}_j^{-1}$.

For any $J_i \in \mathbb{N}$, (ii) implies (i). The reverse implication holds if $J_i = 1$ or $\mathbf{S}(\mathbf{x}) \preccurlyeq \mathbf{0}$ for all $\mathbf{x} \in \mathcal{X}$.

Proof. The semi-infinite constraint (15) is equivalent to

$$\mathbf{z}^\top \mathbf{S}(\mathbf{x}) \mathbf{z} + (\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f})^\top \mathbf{z} + \mathbf{g}^\top \mathbf{u} \leq h - \mathbf{s}^\top \mathbf{x} - t \quad \forall (\mathbf{z}, \mathbf{u}) \in \mathcal{C}_i.$$

This constraint is satisfied if and only if for all $(\mathbf{z}, \mathbf{u}) \in \mathbb{R}^P \times \mathbb{R}^Q$, we have

$$\begin{bmatrix} 1 \\ \mathbf{z} \\ \mathbf{u} \end{bmatrix}^\top \begin{pmatrix} h - \mathbf{s}^\top \mathbf{x} - t & -\frac{1}{2} (\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f})^\top & -\frac{1}{2} \mathbf{g}^\top \\ -\frac{1}{2} (\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f}) & -\mathbf{S}(\mathbf{x}) & \mathbf{0} \\ -\frac{1}{2} \mathbf{g} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{bmatrix} 1 \\ \mathbf{z} \\ \mathbf{u} \end{bmatrix} \geq 0$$

whenever (\mathbf{z}, \mathbf{u}) satisfies

$$\begin{bmatrix} 1 \\ \mathbf{z} \\ \mathbf{u} \end{bmatrix}^\top \begin{pmatrix} 1 - \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j & \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1} \\ \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j & -\boldsymbol{\Sigma}_j^{-1} \end{pmatrix} \begin{bmatrix} 1 \\ \mathbf{z} \\ \mathbf{u} \end{bmatrix} \geq 0 \quad \forall j = 1, \dots, J_i.$$

The statement now follows from the exact and approximate \mathcal{S} -Lemma, see *e.g.* [42], as well as from Farkas' Theorem, see *e.g.* [60]. \square

Proposition 6 (Bi-Quadratic Functions). *Let $v(\mathbf{x}, \mathbf{z}) = \mathbf{x}^\top \mathbf{S}(\mathbf{z}) \mathbf{S}(\mathbf{z})^\top \mathbf{x} + \mathbf{s}(\mathbf{z})^\top \mathbf{x} + t(\mathbf{z})$ with $\mathbf{S}(\mathbf{z}) = \sum_{p=1}^P \mathbf{S}_p \mathbf{z}_p + \mathbf{S}_0$, $\mathbf{S}_p \in \mathbb{R}^{N \times S}$ for $p = 0, \dots, P$, $\mathbf{s}(\mathbf{z}) = \mathbf{S} \mathbf{z} + \mathbf{s}$, $\mathbf{S} \in \mathbb{R}^{N \times P}$ and $\mathbf{s} \in \mathbb{R}^N$, as well as $t(\mathbf{z}) = \mathbf{t}^\top \mathbf{z} + t$, $\mathbf{t} \in \mathbb{R}^P$ and $t \in \mathbb{R}$. Assume that the confidence set \mathcal{C}_i is defined as in (16), and consider the following two statements.*

(i) *The semi-infinite constraint (15) is satisfied, and*

(ii) *there is $\boldsymbol{\lambda} \in \mathbb{R}_+^{J_i}$ such that*

$$\begin{pmatrix} \mathbf{I} & \mathbf{S}_0^\top \mathbf{x} & \hat{\mathbf{S}}(\mathbf{x})^\top \\ \mathbf{x}^\top \mathbf{S}_0 & \gamma(\mathbf{x}) - \sum_j \lambda_j (1 - \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j) & -\frac{1}{2} \boldsymbol{\pi}(\mathbf{x})^\top - \boldsymbol{\mu}(\boldsymbol{\lambda})^\top \\ \hat{\mathbf{S}}(\mathbf{x}) & -\frac{1}{2} \boldsymbol{\pi}(\mathbf{x}) - \boldsymbol{\mu}(\boldsymbol{\lambda}) & \boldsymbol{\Sigma}(\boldsymbol{\lambda}) \end{pmatrix} \succcurlyeq \mathbf{0},$$

where \mathbf{I} is the $S \times S$ -identity matrix, $\hat{\mathbf{S}}(\mathbf{x}) \in \mathbb{R}^{(P+Q) \times S}$ with $\hat{\mathbf{S}}(\mathbf{x}) = [\mathbf{S}_1^\top \mathbf{x}, \dots, \mathbf{S}_P^\top \mathbf{x}, \mathbf{0}]^\top$, $\gamma(\mathbf{x}) = h - t - \mathbf{s}^\top \mathbf{x}$, $\boldsymbol{\pi}(\mathbf{x}) = ([\mathbf{S}^\top \mathbf{x} + \mathbf{t} + \mathbf{f}]^\top, \mathbf{g}^\top)^\top$, $\boldsymbol{\mu}(\boldsymbol{\lambda}) = \sum_j \lambda_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}(\boldsymbol{\lambda}) = \sum_j \lambda_j \boldsymbol{\Sigma}_j^{-1}$.

For any $J_i \in \mathbb{N}$, (ii) implies (i). The reverse implication holds if $J_i = 1$.

Proof. The proof follows closely the argument in Theorem 3.2 of [7]. \square

Bi-quadratic constraint functions of the type considered in Proposition 6 arise, for example, in mean-variance portfolio optimization when bounds on the portfolio variance are imposed.

Proposition 7 (Conic-Quadratic Functions). *Assume that $v(\mathbf{x}, \mathbf{z}) = \|\mathbf{S}(\mathbf{z}) \mathbf{x} + \mathbf{t}(\mathbf{z})\|_2$ with $\mathbf{S}(\mathbf{z}) = \sum_{p=1}^P \mathbf{S}_p \mathbf{z}_p + \mathbf{S}_0$, $\mathbf{S}_p \in \mathbb{R}^{S \times N}$ for $p = 0, \dots, P$, as well as $\mathbf{t}(\mathbf{z}) = \mathbf{T} \mathbf{z} + \mathbf{t}$, $\mathbf{T} \in \mathbb{R}^{S \times P}$ and $\mathbf{t} \in \mathbb{R}^S$. Moreover, assume that the confidence set \mathcal{C}_i is defined as in (16) and that $\mathbf{f} = \mathbf{0}$ and $\mathbf{g} = \mathbf{0}$ in (15). Consider the following two statements.*

(i) The semi-infinite constraint (15) is satisfied, and

(ii) there is $\alpha \in \mathbb{R}_+$ and $\boldsymbol{\lambda} \in \mathbb{R}_+^{J_i}$ such that $\alpha \leq h - t$ and

$$\begin{pmatrix} \alpha \mathbf{I} & \mathbf{S}_0 \mathbf{x} + \mathbf{t} & \hat{\mathbf{S}}(\mathbf{x})^\top + [\mathbf{T}, \mathbf{0}] \\ \mathbf{x}^\top \mathbf{S}_0^\top + \mathbf{t}^\top & \alpha - \sum_j \lambda_j (1 - \boldsymbol{\mu}_j^\top \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j) & -\boldsymbol{\mu}(\boldsymbol{\lambda})^\top \\ \hat{\mathbf{S}}(\mathbf{x}) + [\mathbf{T}, \mathbf{0}]^\top & -\boldsymbol{\mu}(\boldsymbol{\lambda}) & \boldsymbol{\Sigma}(\boldsymbol{\lambda}) \end{pmatrix} \succcurlyeq \mathbf{0},$$

where \mathbf{I} is the $S \times S$ -identity matrix, $\hat{\mathbf{S}}(\mathbf{x}) \in \mathbb{R}^{(P+Q) \times S}$ with $\hat{\mathbf{S}}(\mathbf{x}) = [\mathbf{S}_1 \mathbf{x}, \dots, \mathbf{S}_P \mathbf{x}, \mathbf{0}]^\top$, $\gamma(\mathbf{x}) = h - t - \mathbf{s}^\top \mathbf{x}$, $\boldsymbol{\mu}(\boldsymbol{\lambda}) = \sum_j \lambda_j \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}(\boldsymbol{\lambda}) = \sum_j \lambda_j \boldsymbol{\Sigma}_j$.

For any $J_i \in \mathbb{N}$, (ii) implies (i). The reverse implication holds if $J_i = 1$.

Proof. The proof follows closely the argument in Theorem 3.3 of [7]. \square

In contrast to the previous results, Proposition 7 requires that $\mathbf{f} = \mathbf{0}$ and $\mathbf{g} = \mathbf{0}$ in the semi-infinite constraint (15). An inspection of Theorems 6 and 7 reveals that this implies $\mathbf{A} = \mathbf{B} = \mathbf{0}$ in (4), that is, the ambiguity set \mathcal{P} must not contain any expectation constraints.

Note that the constraint functions in Propositions 5–7 are convex or indefinite in \mathbf{z} , which implies that the conditions of Observation 2 fail to hold. This is in contrast to the constraint functions in Observation 3 and Proposition 4, which satisfy the conditions of Observation 2.

The following result allows us to further expand the class of admissible constraint functions.

Observation 4 (Maxima of Tractable Functions). *Let the constraint function be representable as $v(\mathbf{x}, \mathbf{z}) = \max_{l \in \mathcal{L}} v_l(\mathbf{x}, \mathbf{z})$ for a finite index set \mathcal{L} and constituent functions $v_l : \mathbb{R}^N \times \mathbb{R}^P \rightarrow \mathbb{R}$.*

(i) *If every function v_l satisfies condition (C3b), then v satisfies (C3b) as well.*

(ii) *If the robust constraint (15) admits a tractable conic reformulation or conservative approximation for every constituent function v_l , then the same is true for v .*

This result exploits the fact that inequality (14) of condition (C3b) and the robust constraint (15) are cast as less-than-or-equal constraints. For a proof of Observation 4, we refer to [12, 26, 27].

Table 1 consolidates the results of the main paper and the electronic companion.

Constraint function	Restrictions on ambiguity set	Solution method	Source
bi-affine	condition (\mathbf{N}) ; \mathcal{K}_i 's linear	linear program	Theorem 1
bi-affine	condition (\mathbf{N}) ; \mathcal{K}_i 's conic-quadratic	conic-quadratic program	Theorem 1
bi-affine	condition (\mathbf{N}) ; \mathcal{K}_i 's semidefinite	semidefinite program	Theorem 1
quadratic-affine	condition (\mathbf{N})	semidefinite program	Theorem 6, Proposition 4
affine-quadratic, concave in \mathbf{z}	condition (\mathbf{N}) , C_i 's polyhedral, I small	semidefinite program	Theorem 7, Proposition 5
affine-quadratic, convex in \mathbf{z}	condition (\mathbf{N}) , C_i 's ellipsoids	semidefinite program	Theorem 6, Proposition 5
bi-quadratic, convex in \mathbf{z}	condition (\mathbf{N}) , C_i 's ellipsoids	semidefinite program	Theorem 6, Proposition 6
conic-quadratic	condition (\mathbf{N}) , C_i 's ellipsoids, $\mathbf{A} = \mathbf{B} = \mathbf{0}$	semidefinite program	Theorem 6, Proposition 7
conditions (C3a) , (C3b)	condition (\mathbf{N}) , I small and C_i 's polyhedral if v nonconvex in \mathbf{z}	Ellipsoid method	Theorem 6 or 7
bi-affine	condition (\mathbf{N}^*) ; \mathcal{K}_i 's linear	linear program	Theorem 3, Observation 1
bi-affine	condition (\mathbf{N}^*) ; \mathcal{K}_i 's conic-quadratic	conic-quadratic program	Theorem 3, Observation 1
bi-affine	condition (\mathbf{N}^*) ; \mathcal{K}_i 's semidefinite	semidefinite program	Theorem 3, Observation 1
quadratic-affine	condition (\mathbf{N}^*)	semidefinite program	Theorems 3, 6, Proposition 4
affine-quadratic	condition (\mathbf{N}) , C_i 's intersections of ellipsoids, I small and C_i 's polyhedral if v nonconvex in \mathbf{z}	semidefinite program	Theorem 6 or 7, Proposition 5
affine-quadratic	condition (\mathbf{N}^*) , C_i 's intersections of ellipsoids, I small and C_i 's polyhedral if v nonconvex in \mathbf{z}	semidefinite program	Theorems 3 and 6 or 7, Proposition 5
bi-quadratic	condition (\mathbf{N}) , C_i 's intersections of ellipsoids, I small and C_i 's polyhedral if v nonconvex in \mathbf{z}	semidefinite program	Theorem 6 or 7, Proposition 6
bi-quadratic	condition (\mathbf{N}^*) , C_i 's intersections of ellipsoids, I small and C_i 's polyhedral if v nonconvex in \mathbf{z}	semidefinite program	Theorems 3 and 6 or 7, Proposition 6
conic-quadratic	condition (\mathbf{N}) , C_i 's intersections of ellipsoids, $\mathbf{A} = \mathbf{B} = \mathbf{0}$	semidefinite program	Theorem 6, Proposition 7
conic-quadratic	condition (\mathbf{N}^*) , C_i 's intersections of ellipsoids, $\mathbf{A} = \mathbf{B} = \mathbf{0}$	semidefinite program	Theorems 3, 6, Proposition 7
conditions (C3a) , (C3b)	condition (\mathbf{N}^*) , I small and C_i 's polyhedral if v nonconvex in \mathbf{z}	Ellipsoid method	Theorems 3 and 6 or 7

Table 1. Summary of the results of the paper and its companion. The results above (below) the middle line constitute exact reformulations (conservative reformulations).

All constraint functions can be combined to piecewise functions using Observation 4.