manuscript

A Practically Efficient Approach for Solving Adaptive Distributionally Robust Linear Optimization Problems

Dimitris Bertsimas
Massachusetts Institute of Technology

Melvyn Sim, Meilin Zhang
NUS Business School, National University of Singapore

We develop a modular and tractable framework for solving an adaptive distributionally robust linear optimization problem, where we minimize the worst-case expected cost over an *ambiguity set* of probability distributions. The adaptive distrbutaionally robust optimization framework caters for dynamic decision making, where decisions can adapt to the uncertain outcomes as they unfold in stages. For tractability considerations, we focus on a class of second-order conic (SOC) representable ambiguity set, though our results can easily be extended to more general conic representations. We show that the adaptive distributionally robust linear optimization problem can be formulated as a classical robust optimization problem. To obtain tractable formulation, we approximate the adaptive distributionally robust optimization problem using linear decision rule (LDR) techniques. More interestingly, by incorporating the primary and auxiliary random variables of the lifted ambiguity set in the LDR approximation, we can significantly improve the solutions and for a class of adaptive distributionally robust optimization problems, exact solutions can also be obtained. Using the new LDR approximation, we can transform the distributionally adaptive robust optimization problem to a classical robust optimization problem with an SOC representable uncertainty set. Finally, to demonstrate the potential for solving management decision problems, we develop an algebraic modeling package and illustrate how it can be used to facilitate modeling and obtain high quality solutions for addressing a medical appointment scheduling problem and a multiperiod inventory control problem.

Key words: robust optimization, risk, ambiguity

---

1. Introduction

Addressing uncertainty in many real world optimization problems has often lead to computationally intractable models. As a result, uncertainty is often ignored in optimization models and this may lead to poor or even unacceptable decisions when implementing them in practice. We characterize uncertainty as *risk*, whenever the probability distribution is known or otherwise as *ambiguity* (Knight 1921). Traditionally, mathematical optimization models such as stochastic programming (see, for instance, Birge and Louveaux 1997, Ruszczyńskii and Shapiro 2003) are based on the paradigm of risk and they do not incorporate ambiguity in their decision criteria for optimization. However, with the growing importance of ambiguity in decision making (see, for instance, Ellsberg 1961, Hsu et al. 2005), research on ambiguity has garnered considerable interest in various fields including economics, mathematical finance and management science. In particular, robust optimization is a relatively new approach that deals with ambiguity in mathematical optimization problems. In classical robust optimization, uncertainty is described by a distribution free uncertainty set, which is typically a conic representable bounded convex set (see, for instance, El Ghaoui and Lebret 1997, El Ghaoui et al. 1998, Ben-Tal and Nemirovski 1998, 1999, 2000, Bertsimas and Sim 2004, Bertsimas and Brown 2009, Bertsimas et al. 2011). The key advantage of a robust optimization model is its computational tractability and it has been successful in providing computationally scalable solutions for a wide variety of management inspired optimization problems.

In evaluating preferences over risk and ambiguity, Gilboa and Schmeidler (1989) propose a decision criterion by taking the worst-case expected utility or disutility over an ambiguity set of probability distributions. Scarf (1958) is first to study a single-product Newsvendor problem where the precise demand distribution is unknown but is only characterized by its mean and variance. Subsequently, such models have been extended to minimax stochastic optimization models (see, for instance, Žáčková 1966, Dupacova 1987, Breton and EI Hachem 1995, Shapiro and Kleywegt 2002, Shapiro and Ahmed 2004), and recently to distributionally robust optimization models (see, for instance, Chen et al. 2007, Chen and Sim 2009, Popescu 2007, Delage and Ye 2010, Xu and Mannor 2012). In terms of tractable formulations for a wide variety of single stage convex optimization problems, Wiesemann et al. (2014) propose a broad class of ambiguity set where the family of probability distributions are characterized by conic representable expectation constraints and nested conic representable confidence sets.

Dynamic optimization models, where decisions adapt to the uncertain outcomes as they unfold in stages, are typically computationally intractable due to the "curse of dimensionality" (see, for instance, Shapiro and Nemirovski 2005, Dyer and Stougie 2006, Ben-Tal et al. 2004). A common approximation technique is to restrict the dynamic or adaptive decisions to being affinely dependent of the uncertain parameters, an approach known as linear decision rule (LDR). LDR approximation was discussed in the early literature of stochastic programming but the technique had been abandoned due to suboptimality (see Garstka and Wets 1974). The interest in LDR approximation has been rekindled by Ben-Tal et al. (2004) in their seminal work on adjustable robust optimization that extends classical robust optimization to encompass adaptive decisions. In this paper, we prefer the term "adaptive" over "adjustable" in describing robust optimization problem with adaptive decisions or recourse. Subsequently, Bertsimas et al. (2010) establish the optimality of LDR approximation in some important classes of adaptive robust optimization problems under full ambiguity. Chen and Zhang (2009) further improve the LDR approximation by extending the affine dependency to the auxiliary variables associated with the support set. For solving adaptive distributionally robust optimization problems, Chen et al. (2007) adopt LDR approximation to provide tractable formulations. Henceforth, variants of piecewise-linear decision rule approximation have been proposed to improve the approximation while maintaining the tractability of the adaptive distributionally robust optimization models. Such approaches include the deflected and segregated LDR approximation of Chen et al. (2008), the truncated LDR approximation of See and Sim (2009), and the bideflected and (generalized) segregated LDR approximation of Goh and Sim (2010). Interestingly, there is also a revival of using LDR approximation for solving multistage stochastic optimization problems (Kuhn et al. 2011).

For broader impact, a general purpose optimization framework should be implementable in software packages where reliable solutions can be obtained with reasonable computational effort to address a wide variety of optimization problems. Classical stochastic programming and dynamic programing have been less successful as a general purpose framework than linear programming. Software packages that facilitate robust robust optimization modeling have begun to surface in recent years. Existing toolboxes include YALMIP (Löberg 2012), AIMMS (`http://www.aimms.com/`), ROME (Goh and Sim 2009). and JuMPeR (`http://jumper.readthedocs.org`). Of those, ROME, AIMMS and JuMPeR have provisions for decision rule approximation and hence, they are capable of solving adaptive distributionally robust optimization problems approximately as well. In particular, ROME is a toolbox built in the MATLAB environment that implements the adaptive distributionally robust optimization framework of Goh and Sim (2010). However, due to the computationally demanding reformulation approach of Goh and Sim (2010), its scalability is limited.

Our contributions to this paper are as follows:

1. We propose a tractable and scalable framework for solving an adaptive distributionally robust linear optimization problem, where we minimize the worst-case expected cost over a second-order conic (SOC) representable ambiguity set. We show that adaptive distributionally robust linear optimization problem can be formulated as a classical robust optimization problem.

2. To obtain tractable formulation, we approximate the adaptive distributionally robust linear optimization problem using LDR techniques. Depending on the choice of ambiguity set, the resulting framework is either a linear optimization problem or a second order conic optimization problem (SOCP), which can be solved efficiently by general purpose commercial grade solvers such as CPLEX and Gurobi.

3. We can obtain significant improvement by simply incorporating the auxiliary random variable associated with the lifted ambiguity set in the LDR approximation. In some cases, we can also

recover the exact solutions. This approach also improves upon more sophisticated decision rule approximations developed in Chen and Zhang (2009), Chen et al. (2008), See and Sim (2009), Goh and Sim (2010). Using the new LDR approximation, we can transform the adaptive distributionally robust optimization problem to a classical robust optimization problem with an SOC representable uncertainty set.

4. We demonstrate our approach for addressing a medical appointment scheduling problem as well as a multiperiod inventory control problem. In these problems, we also show that by incorporating partial cross moments information in the ambiguity set, we can significantly improve the solutions over alternatives found in recent literature where the ambiguity set is only characterized by marginal moments.

Notations. We use $[N]$, $N \in \mathbb{N}$ to denote the set of running indices, $\{1, \dots, N\}$. We generally use bold faced characters such as $\boldsymbol{x} \in \mathbb{R}^N$ and $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ to represent vectors and matrices, respectively and $[\boldsymbol{x}]_i$ or x_i to denote the ith element of the vector $\boldsymbol{x}$. We use $(x)^+$ to denote $\max\{x, 0\}$. Special vectors include $\mathbf{0}$, $\mathbf{1}$ and $\boldsymbol{e}_i$ which are respectively the vector of zeros, the vector of ones and the standard unit basis vector. We denote $\mathcal{R}^{N,M}$ as the space of all measurable functions from $\mathbb{R}^N$ to $\mathbb{R}^M$ that are bounded on compact sets. We use tilde to denote a random variable without associating it with a particular probability distribution. We use $\tilde{\boldsymbol{z}} \in \mathbb{R}^I$ to represent an I dimensional random variable and it can be associated with a probability distribution $\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I)$, where $\mathcal{P}_0(\mathbb{R}^I)$ represents the set of all probability distributions on $\mathbb{R}^I$. We denote $\mathbb{E}_{\mathbb{P}}(\cdot)$ as the expectation over the probability distribution $\mathbb{P}$. For a set $\mathcal{W} \subseteq \mathbb{R}^I$, $\mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W})$ represents the probability of $\tilde{\boldsymbol{z}}$ being in the set $\mathcal{W}$ evaluated on the distribution $\mathbb{P}$.

2. An adaptive distributionally robust linear optimization problem

We first focus on a two-stage adaptive distributionally robust linear optimization problem where the first stage or *here-and-now* decision is a vector $\boldsymbol{x} \in \mathbb{R}^{N_1}$ chosen over the feasible set X. The cost incurred during the first stage in association with the decision $\boldsymbol{x}$ is deterministic and given by $\boldsymbol{c}'\boldsymbol{x}$, $\boldsymbol{c} \in \mathbb{R}^{N_1}$. In progressing to the next stage, the random variable $\tilde{\boldsymbol{z}} \in \mathbb{R}^{I_1}$ with support $\mathcal{W} \subseteq \mathbb{R}^{I_1}$ is realized; thereafter, we could determine the cost incurred at the second stage. Similar to a typical stochastic programming model, for a given decision vector, $\boldsymbol{x}$ and a realization of the random variable, $\boldsymbol{z} \in \mathcal{W}$, we evaluate the second stage cost via the following linear optimization problem,

$$
\begin{aligned}
Q(\boldsymbol{x}, \boldsymbol{z}) = \min \ & \boldsymbol{d}'\boldsymbol{y} \\
\text{s.t.} \ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b}(\boldsymbol{z}) \\
& \boldsymbol{y} \in \mathbb{R}^{N_2}
\end{aligned} \tag{1}
$$

Here, $\boldsymbol{A} \in \mathcal{R}^{I_1, M \times N_1}$, $\boldsymbol{b} \in \mathcal{R}^{I_1, M}$ are functions that maps from the vector $\boldsymbol{z} \in \mathcal{W}$ to the input parameters of the linear optimization problem. Adopting the common assumptions in the robust optimization literature, these functions are affinely dependent on $\boldsymbol{z} \in \mathbb{R}^{I_1}$ and are given by,

$$\boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{A}^0 + \sum_{k \in [I_1]} \boldsymbol{A}^k z_k, \quad \boldsymbol{b}(\boldsymbol{z}) = \boldsymbol{b}^0 + \sum_{k \in [I_1]} \boldsymbol{b}^k z_k,$$

with $\boldsymbol{A}^0, \boldsymbol{A}^1, ..., \boldsymbol{A}^{I_1} \in \mathbb{R}^{M \times N_1}$ and $\boldsymbol{b}^0, \boldsymbol{b}^1, ..., \boldsymbol{b}^{I_1} \in \mathbb{R}^M$. The matrix $\boldsymbol{B} \in \mathbb{R}^{M \times N_2}$, also known in stochastic programming as the *recourse matrix*, and the vector $\boldsymbol{d} \in \mathbb{R}^{N_2}$ are constants, which correspond to the stochastic programming format known as *fixed recourse*. Note that Problem (1) may not aways be feasible and, as in the case of *complete recourse*, the recourse matrix can influence the feasibility of the second stage problem.

Definition 1. The second stage problem (1) has *complete recourse* if and only if the recourse matrix, $\boldsymbol{B}$ satisfies for any $\boldsymbol{t} \in \mathbb{R}^M$, there exists $\boldsymbol{y} \in \mathbb{R}^{N_2}$ such that $\boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{t}$.

Complete recourse is a strong sufficient condition that guarantees the feasibility of the second stage problem for all $\boldsymbol{x} \in \mathbb{R}^{N_1}$ and $\boldsymbol{z} \in \mathbb{R}^{I_1}$. Many real-life stochastic programming models, including the newsvendor problem and its variants are modeled as complete recourse problems, which ensure that no outcome can produce infeasible results. However, there are also problems that would generally not satisfy complete recourse, such as a production planning problem where a manager determines a production plan today to satisfy uncertain demands for tomorrow. For more information, we refer interested readers to Birge and Louveaux (1997).

Typically, a weaker condition is assumed in stochastic programming to ensure that the second stage problem is essentially feasible.

Definition 2. The second stage problem (1) has *relatively complete recourse* if and only if the problem is feasible for all $\boldsymbol{x} \in X$ and $\boldsymbol{z} \in \mathcal{W}$.

Given an ambiguity set of probability distributions, $\mathbb{F}$, the second stage cost is evaluated based on the worst-case expectation over the ambiguity set given by

$$\beta(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(Q(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right). \tag{2}$$

Corresponding, the *here-and-now* decision, $\boldsymbol{x}$ is determined by minimizing the sum of the deterministic first stage cost and the worst-case expected second stage cost over the ambiguity set as follows:

$$\begin{aligned} Z^* = \min \ & \boldsymbol{c}'\boldsymbol{x} + \beta(\boldsymbol{x}) \\ \text{s.t.} \ & \boldsymbol{x} \in X. \end{aligned} \tag{3}$$

The SOC ambiguity set

It is well known that the tractability of a robust linear optimization problem is highly dependent on the choice of the uncertainty set. In particular, an SOC representable set, which encompasses a polyhedron as a special case, can describe a wide variety of common uncertainty sets including the Euclidean ball and the budgeted uncertainty set of Bertsimas and Sim (2004). More importantly, the robust counterpart corresponding to an SOC uncertainty set is at most an SOCP, which is an established optimization format supported by commercial solvers that scale well with the number of variables in the problem. Similar to the classical robust linear optimization problems, the tractability of a distributionally robust linear optimization problem is also dependent on the choice of the ambiguity set. In particular the ambiguity set based on information of moments, notwithstanding its popularity, may not necessarily yield tractable distributionally robust counterparts. We propose an SOC representable ambiguity set where we restrict only to SOC representation. For generalization to the ambiguity set of Wiesemann et al. (2014), we refer interested readers to Appendix B.

DEFINITION 3. An SOC ambiguity set, $\mathbb{F}$ is an ambiguity set of probability distributions that can be expressed as

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1}\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \in \mathbb{R}^{I_1} \\ \mathbb{E}_{\mathbb{P}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}(g_i(\tilde{\boldsymbol{z}})) \leq \sigma_i \quad \forall i \in [I_2] \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1 \end{array} \right\} \tag{4}$$

with parameters $\boldsymbol{G} \in \mathbb{R}^{L_1 \times I_1}$, $\boldsymbol{\mu} \in \mathbb{R}^{L_1}$, $\boldsymbol{\sigma} \in \mathbb{R}^{I_2}$, support set $\mathcal{W} \in \mathbb{R}^{I_1}$ and functions $g_i \in \mathcal{R}^{I_1,1}$, $i \in [I_2]$.

ASSUMPTION 1. *The support set $\mathcal{W}$ is an SOC representable set and the functions g_i, $i \in [I_2]$ are also SOC representable functions. Furthermore, there exists $\boldsymbol{z}^\dagger$ in the relative interior of $\mathcal{W}$ such that $\boldsymbol{G}\boldsymbol{z}^\dagger = \boldsymbol{\mu}$ and $g_i(\boldsymbol{z}^\dagger) < \sigma_i$, for all $i \in [I_2]$.*

The equality expectation constraints of the SOC ambiguity set allow the modeler to specify the mean values of $\tilde{\boldsymbol{z}}$, while the inequality expectation constraints offer more flexible and interesting characterization of distributional ambiguity. Specifically, an SOC representable function, $g \in \mathcal{R}^{I_1,1}$ is defined as a function whose epigraph, i.e.,

$$\text{epi} g = \left\{ (\boldsymbol{z}, u) \in \mathbb{R}^{I_1} \times \mathbb{R} \;\middle|\; g(\boldsymbol{z}) \leq u \right\}$$

is an SOC representable set. As an illustration, the function $g(\boldsymbol{z}) = ((\boldsymbol{f}'\boldsymbol{z}+h)^+)^3$ for some given $(\boldsymbol{f},h) \in \mathbb{R}^{I_1+1}$ is an SOC representable function because its epigraph is SOC representable given by

$$\text{epi}g = \left\{ (\boldsymbol{z},u) \in \mathbb{R}^{I_1} \times \mathbb{R} \;\middle|\; \exists \boldsymbol{v} \in \mathbb{R}^2 : \begin{array}{c} v_1 \geq 0, v_1 \geq \boldsymbol{f}'\boldsymbol{z} + h \\ \sqrt{v_1^2 + \left(\frac{v_2-1}{2}\right)^2} \leq \frac{v_2+1}{2} \\ \sqrt{v_2^2 + \left(\frac{v_1-u}{2}\right)^2} \leq \frac{v_1+u}{2} \end{array} \right\}.$$

The formulation of SOC representable functions is a process that can be automated in an algebraic modeling software package. For more information, we refer interested readers to Ben-Tal and Nemirovski (2001a) for an excellent reference on the algebra of SOC representable functions. With regard to the SOC ambiguity set, SOC representable functions can provide useful and interesting characterization of distributions including,

- **Bounds on mean values:** $\mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) \in [\underline{\boldsymbol{\mu}}, \overline{\boldsymbol{\mu}}]$.
- **Upper bound on absolute deviation:** $\mathbb{E}_{\mathbb{P}}(|\boldsymbol{f}'\boldsymbol{z}+h|) \leq \sigma$, for some vector, $(\boldsymbol{f},h) \in \mathbb{R}^{I_1+1}$.
- **Upper bound on variance:** $\mathbb{E}_{\mathbb{P}}((\boldsymbol{f}'\boldsymbol{z}+h)^2) \leq \sigma$, for some vector, $(\boldsymbol{f},h) \in \mathbb{R}^{I_1+1}$.
- **Upper bound on p-ordered deviation:** $\mathbb{E}_{\mathbb{P}}((|\boldsymbol{f}'\boldsymbol{z}+h|)^p) \leq \sigma$, for some vector, $(\boldsymbol{f},h) \in \mathbb{R}^{I_1+1}$ and some rational $p \geq 1$.
- **Upper bound on semi-variance:** $\mathbb{E}_{\mathbb{P}}(((\boldsymbol{f}'\boldsymbol{z}+h)^+)^2) \leq \sigma$, for some vector, $(\boldsymbol{f},h) \in \mathbb{R}^{I_1+1}$.
- **Approximate upper bound on entropy:** $\mathbb{E}_{\mathbb{P}}(\exp(\boldsymbol{f}'\boldsymbol{z})) \leq \sigma$, for some vector, $\boldsymbol{f} \in \mathbb{R}^{I_1}$. We refer readers to Ben-Tal and Nemirovski (2001a) for the approximate SOC representation.
- **Upper bound on convex piecewise linear function:** $\mathbb{E}_{\mathbb{P}}(\max_{p\in[P]}\{\boldsymbol{f}_p'\boldsymbol{z}+h_p\}) \leq \sigma$, for some vectors, $(\boldsymbol{f}_p,h_p) \in \mathbb{R}^{I_1+1}$, $p \in [P]$.

Although the ambiguity set of Wiesemann et al. (2014) is more general, the corresponding robust counterpart may become a semidefinite optimization problem (SDP), which, despite being polynomially time solvable, the current solvers for SDPs are still not as scalable and stable as those for solving linear and second order conic optimization problems. In this regard, one important class of ambiguity set that could be modeled in the ambiguity set of Wiesemann et al. (2014) but not ours is the conic representable cross moment ambiguity set as follows,

$$\mathbb{F}_{CM} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1}\right) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \in \mathbb{R}^{I_1} \\ \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}((\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})') \preceq \boldsymbol{\Sigma} \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1 \end{array} \right\}.$$

Observe that the semidefinite constraint

$$\mathbb{E}_{\mathbb{P}}((\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})') \preceq \boldsymbol{\Sigma}$$

is equivalent to the following semi-infinite quadratic constraints

$$\mathbb{E}_{\mathbb{P}}((\boldsymbol{f}'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2) \leq \boldsymbol{f}'\boldsymbol{\Sigma}\boldsymbol{f} \qquad \forall \boldsymbol{f} \in \mathbb{R}^{I_1}. \tag{5}$$

Hence, as a conservative approximation of the cross moment ambiguity set, we propose the partial cross moment ambiguity set, which is an SOC ambiguity set as follows:

$$\mathbb{F}_{PCM} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1}\right) \left| \begin{array}{ll} \tilde{\boldsymbol{z}} \in \mathbb{R}^{I_1} & \\ \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{P}}((\boldsymbol{f}_k'(\tilde{\boldsymbol{z}}-\boldsymbol{\mu}))^2) \leq \boldsymbol{f}_k'\boldsymbol{\Sigma}\boldsymbol{f}_k & \forall k \in [K] \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1 & \end{array} \right. \right\},$$

for some choice of parameters $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_K \in \mathbb{R}^{I_1}$. Observe that the approximation never deteriorates with addition of new vectors, $\boldsymbol{f}_k, k > K$. However, in general, it is difficult to determine the "right choice" of these vectors. Nevertheless, in our applications to inventory control and appointment scheduling problems, we will demonstrate how the partial cross moment ambiguity set can yield tractable models and provide far less conservative solutions than those obtained from the marginal moment ambiguity set, an ambiguity set that does not consider cross moment information.

As in Wiesemann et al. (2014), we also define the *lifted ambiguity set*, $\mathbb{G}$ that encompasses the primary random variable $\tilde{\boldsymbol{z}}$ and the lifted or auxiliary random variable $\tilde{\boldsymbol{u}}$ as follows:

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right) \left| \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \\ \mathbb{E}_{\mathbb{Q}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma} \\ \mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right) = 1 \end{array} \right. \right\}, \tag{6}$$

where $\bar{\mathcal{W}}$ is the *lifted support set* defined as

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \mid \boldsymbol{z} \in \mathcal{W}, \boldsymbol{g}(\boldsymbol{z}) \leq \boldsymbol{u} \right\}, \tag{7}$$

with $\boldsymbol{g}(\boldsymbol{z}) = (g_1(\boldsymbol{z}), \ldots, g_{I_2}(\boldsymbol{z}))$. Observe that the lifted ambiguity set has only linear expectation constraints and that the corresponding lifted support set is SOC representable.

Proposition 1. *The ambiguity set, $\mathbb{F}$ is equivalent to the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}$, for all $\mathbb{Q} \in \mathbb{G}$.*

Proof. The proof is similar to Wiesemann et al. (2014). We first show that $\prod_{\tilde{\boldsymbol{z}}} \mathbb{G} \subseteq \mathbb{F}$, where $\prod_{\tilde{\boldsymbol{z}}} \mathbb{G}$ represents the set of marginal distributions of $\tilde{\boldsymbol{z}}$ under any $\mathbb{Q} \in \mathbb{G}$. Indeed, for any $\mathbb{Q} \in \mathbb{G}$, and $\mathbb{P}$ is the marginal distribution of $\tilde{\boldsymbol{z}}$ under all $\mathbb{Q}$, then we have $\mathbb{E}_{\mathbb{P}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \mathbb{E}_{\mathbb{Q}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu}$. Moreover, since $\mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right) = 1$, we have $\mathbb{Q}\left(\tilde{\boldsymbol{z}} \in \mathcal{W}\right) = 1$ and $\mathbb{Q}(\boldsymbol{g}(\tilde{\boldsymbol{z}}) \leq \tilde{\boldsymbol{u}}) = 1$. Hence, $\mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1$ and

$$\mathbb{E}_{\mathbb{P}}(\boldsymbol{g}(\tilde{\boldsymbol{z}})) = \mathbb{E}_{\mathbb{Q}}\left(\boldsymbol{g}(\tilde{\boldsymbol{z}})\right) \leq \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma}.$$

Conversely, suppose $\mathbb{P} \in \mathbb{F}$, we observe that $\mathbb{P}\left((\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}})) \in \bar{\mathcal{W}}\right) = 1$. Hence, we can then construct a probability distribution $\mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right)$ associated with the random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2}$ so that

$$(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) = (\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}})) \;\; \mathbb{P}\text{-a.s.}$$

Observe that

$$\mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) = \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{g}(\tilde{\boldsymbol{z}})\right) \leq \boldsymbol{\sigma}$$

and

$$\mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right) = 1.$$

Hence, $\mathbb{F} \subseteq \prod_{\tilde{\boldsymbol{z}}} \mathbb{G}$. □

Although the ambiguity sets, $\mathbb{F}$ and $\mathbb{G}$ are essentially the same, as it will become clearer in the subsequent sections, the auxiliary random variable of the lifted ambiguity set, $\mathbb{G}$ has important ramification for improving the approximation of the adaptive distributionally robust optimization problem.

Reformulation as a classical robust optimization problem

For simplicity and clarity of the exposition, we will focus on deriving the exact reformulation of $\beta(\boldsymbol{x})$ for a relatively complete recourse problem. We can easily incorporate the result to Problem (3) to obtain the corresponding *here-and-now* decision, $\boldsymbol{x} \in X$.

THEOREM 1. *Suppose the second stage problem (1) has relatively complete recourse, then Problem (2) is equivalent to the following robust optimization problem,*

$$\begin{aligned}
\beta(\boldsymbol{x}) = \min\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq Q(\boldsymbol{x}, \boldsymbol{z}) && \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\
& \boldsymbol{t} \geq \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} \\
= \min\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \boldsymbol{p}_i'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) && \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}\ \forall i \in [P] \\
& \boldsymbol{t} \geq \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2},
\end{aligned} \tag{8}$$

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_P$ are the extreme points of the polyhedron, $\mathcal{P}$ given by

$$\mathcal{P} = \left\{\boldsymbol{p} \in \mathbb{R}^M_+ : \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d}\right\}.$$

Proof. From Proposition 1, we have equivalently

$$\beta(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(Q(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right) = \sup_{\mathbb{Q} \in \mathbb{G}} \mathbb{E}_{\mathbb{Q}}\left(Q(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right).$$

By weak duality of Isii (1962), we have

$$\begin{aligned}\beta(\boldsymbol{x}) \leq \beta_1(\boldsymbol{x}) = \inf\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq Q(\boldsymbol{x},\boldsymbol{z})\ \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2},\end{aligned}$$

where $r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2}$ are the dual variables corresponding to the expectation constraints of $\mathbb{G}$.

Under the assumption of relatively complete recourse, $Q(\boldsymbol{x},\boldsymbol{z})$ is finite and by strong duality of linear optimization, we have equivalently

$$Q(\boldsymbol{x},\boldsymbol{z}) = \max_{\boldsymbol{p} \in \mathcal{P}} \boldsymbol{p}'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) = \max_{p \in [P]} \left\{ \boldsymbol{p}_p'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \right\}.$$

Therefore,

$$\begin{aligned}\beta_1(\boldsymbol{x}) = \inf\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.}\ & r \geq \sup_{(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}}} \left\{ \left(\begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \right)' \boldsymbol{z} - \boldsymbol{t}'\boldsymbol{u} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) \right\} \forall p \in [P] \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2}.\end{aligned} \tag{9}$$

To further obtain explicit formulations, we express the lifted support set, which is SOC representable, as follows:

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z},\boldsymbol{u}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \mid \exists \boldsymbol{v} \in \mathbb{R}^{I_3}\ :\ \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{E}\boldsymbol{v} \preceq_{\mathcal{K}} \boldsymbol{h} \right\}, \tag{10}$$

with $\boldsymbol{C} \in \mathbb{R}^{L_2 \times I_1}$, $\boldsymbol{D} \in \mathbb{R}^{L_2 \times I_2}$, $\boldsymbol{E} \in \mathbb{R}^{L_2 \times I_3}$, $\boldsymbol{h} \in \mathbb{R}^{L_2}$ and $\mathcal{K} \subseteq \mathbb{R}^{L_2}$ is a Cartesian product of second-order cones. By weak conic duality (see, for instance, Ben-Tal and Nemirovski (2001a)), we have for all $p \in [P]$,

$$\begin{aligned}& \sup_{(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}}} \left\{ \left(\begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \right)' \boldsymbol{z} - \boldsymbol{t}'\boldsymbol{u} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) \right\} \\ \leq\ & \\ & \quad \inf\ \boldsymbol{\pi}_p{}'\boldsymbol{h} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) \\ & \quad \text{s.t.}\ \boldsymbol{C}'\boldsymbol{\pi}_p = \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \\ & \qquad\ \ \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} \\ & \qquad\ \ \boldsymbol{E}'\boldsymbol{\pi}_p = \boldsymbol{0} \\ & \qquad\ \ \boldsymbol{\pi}_p \succeq_{\mathcal{K}} \boldsymbol{0} \\ & \qquad\ \ \boldsymbol{\pi}_p \in \mathbb{R}^{L_2},\end{aligned}$$

where $\boldsymbol{\pi}_p \in \mathbb{R}^{L_2}, \forall p \in [P]$ are the dual variables associated with the conic constants in $\bar{\mathcal{W}}$. Note that since the Cartesian product of second-order cones are self-dual, we have $\mathcal{K}^* = \mathcal{K}$,

Hence, using standard robust counterpart techniques, we substitute the dual formulations in Problem (9) to yield the following compact conic optimization problem

$$
\begin{aligned}
\beta_2(\boldsymbol{x}) = \inf\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t. } & r \geq \boldsymbol{\pi}_p{}'\boldsymbol{h} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) && \forall p \in [P] \\
& \boldsymbol{C}'\boldsymbol{\pi}_p = \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} && \forall p \in [P] \\
& \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} && \forall p \in [P] \\
& \boldsymbol{E}'\boldsymbol{\pi}_p = 0 && \forall p \in [P] \\
& \boldsymbol{\pi}_p \succeq_{\mathcal{K}} 0 && \forall p \in [P] \\
& \boldsymbol{t} \geq \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} \\
& \boldsymbol{\pi}_p \in \mathbb{R}^{L_2} && \forall p \in [P].
\end{aligned}
\tag{11}
$$

Note that replacing the dual formulation in (9) leads to a restriction to the inf problem, hence $\beta_1(\boldsymbol{x}) \leq \beta_2(\boldsymbol{x})$. Observe that $\beta(\boldsymbol{x}) \leq \beta_1(\boldsymbol{x}) \leq \beta_2(\boldsymbol{x})$, and our goal is to establish strong duality by showing $\beta_2(\boldsymbol{x}) \leq \beta(\boldsymbol{x})$. Then we will next approach Problem (11) by taking the dual, which is

$$
\begin{aligned}
\beta_3(\boldsymbol{x}) = \sup\ & \sum_{p\in[P]} \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x})\alpha_p + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \bar{\boldsymbol{z}}_p \right) \\
\text{s.t. } & \sum_{p\in[P]} \alpha_p = 1 \\
& \alpha_p \geq 0 && \forall p \in [P] \\
& \sum_{p\in[P]} \boldsymbol{G}\bar{\boldsymbol{z}}_p = \boldsymbol{\mu} \\
& \sum_{p\in[P]} \bar{\boldsymbol{u}}_p \leq \boldsymbol{\sigma} \\
& \boldsymbol{C}\bar{\boldsymbol{z}}_p + \boldsymbol{D}\bar{\boldsymbol{u}}_p + \boldsymbol{E}\bar{\boldsymbol{v}}_p \preceq_{\mathcal{K}} \alpha_p\boldsymbol{h} && \forall p \in [P] \\
& \alpha_p \in \mathbb{R}, \bar{\boldsymbol{z}}_p \in \mathbb{R}^{I_1}, && \forall p \in [P] \\
& \bar{\boldsymbol{u}}_p \in \mathbb{R}^{I_2}, \bar{\boldsymbol{v}}_p \in \mathbb{R}^{I_3} && \forall p \in [P],
\end{aligned}
\tag{12}
$$

where $\alpha_p, \bar{\boldsymbol{z}}, \bar{\boldsymbol{u}}_p, \bar{\boldsymbol{v}}_p, \forall p \in [P]$ are the dual variables associated with the specified constraints respectively. Under the SOC ambiguity set, there exists $\boldsymbol{z}^\dagger$ in the relative interior of $\mathcal{W}$ such that $\boldsymbol{G}\boldsymbol{z}^\dagger = \boldsymbol{\mu}$ and $\boldsymbol{g}(\boldsymbol{z}^\dagger) < \boldsymbol{\sigma}$. Hence, there also exists $\boldsymbol{u}^\dagger \in \mathbb{R}^{I_2}$ and $\boldsymbol{v}^\dagger \in \mathbb{R}^{I_3}$ such that

$$
\begin{aligned}
& \boldsymbol{u}^\dagger < \boldsymbol{\sigma} \\
& \boldsymbol{C}\boldsymbol{z}^\dagger + \boldsymbol{D}\boldsymbol{u}^\dagger + \boldsymbol{E}\boldsymbol{v}^\dagger \prec_{\mathcal{K}} \boldsymbol{h}.
\end{aligned}
$$

Therefore, we can construct a strictly feasible solution

$$
\alpha_p = \frac{1}{P}, \bar{\boldsymbol{z}}_p = \frac{\boldsymbol{z}^\dagger}{P}, \bar{\boldsymbol{u}}_p = \frac{\boldsymbol{u}^\dagger}{P}, \bar{\boldsymbol{v}}_p = \frac{\boldsymbol{v}^\dagger}{P},
$$

for all $p \in [P]$. Hence, since Problem (12) is strictly feasible and, as we will subsequently show, is also bounded from above, strong duality holds and $\beta_2(\boldsymbol{x}) = \beta_3(\boldsymbol{x})$. Moreover, there exists a sequence of strictly feasible or interior solutions

$$\left\{(\alpha_p^k, \bar{\boldsymbol{z}}_p^k, \bar{\boldsymbol{u}}_p^k, \bar{\boldsymbol{v}}_p^k)_{p\in[P]}\right\}_{k\geq 0}$$

such that

$$\lim_{k\to\infty} \sum_{p\in[P]} \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x})\alpha_p^k + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \bar{\boldsymbol{z}}_p^k \right) = \beta_3(\boldsymbol{x}).$$

Observe that for all k, $\alpha_p^k > 0$, $\sum_{p\in[P]} \alpha_p^k = 1$ and we can construct a sequence of discrete probability distributions $\{\mathbb{Q}_k \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2})\}_{k\geq 0}$ on random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}$ such that

$$\mathbb{Q}_k\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) = \left(\frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k}, \frac{\bar{\boldsymbol{u}}_p^k}{\alpha_p^k}, \frac{\bar{\boldsymbol{v}}_p^k}{\alpha_p^*} \right) \right) = \alpha_p^k \qquad \forall p \in [P].$$

Observe that,

$$\mathbb{E}_{\mathbb{Q}_k}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{Q}_k}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma}, \mathbb{Q}_k((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}) = 1,$$

and hence $\mathbb{Q}_k \in \mathbb{G}$ for all k. Moreover,

$$\begin{aligned}
\beta_3(\boldsymbol{x}) &= \lim_{k\to\infty} \sum_{p\in[P]} \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x})\alpha_p^k + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \bar{\boldsymbol{z}}_p^k \right) \\
&= \lim_{k\to\infty} \sum_{p\in[P]} \alpha_p^k \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k} \right) \\
&\leq \lim_{k\to\infty} \sum_{p\in[P]} \alpha_p^k \left(\max_{q\in[P]} \left\{ \boldsymbol{p}_q'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) + \begin{bmatrix} \boldsymbol{p}_q'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_q'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \frac{\bar{\boldsymbol{z}}_q^k}{\alpha_q^k} \right\} \right) \\
&= \lim_{k\to\infty} \mathbb{E}_{\mathbb{Q}_k} \left(\max_{q\in[P]} \left\{ \boldsymbol{p}_q'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) + \begin{bmatrix} \boldsymbol{p}_q'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_q'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \tilde{\boldsymbol{z}} \right\} \right) \\
&\leq \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(Q(\boldsymbol{x}, \boldsymbol{z})) \\
&= \beta(\boldsymbol{x}).
\end{aligned}$$

Hence, $\beta(\boldsymbol{x}) \leq \beta_1(\boldsymbol{x}) \leq \beta_2(\boldsymbol{x}) = \beta_3(\boldsymbol{x}) \leq \beta(\boldsymbol{x})$. $\square$

Remarks: Note that similar results have been derived in Bertsimas and Popescu (2005), Shapiro and Kleywegt (2002), among others, by applying some sufficient conditions to attain strong duality. The assumption of relatively complete recourse ensures that the dual of Problem (1) is feasible and finite so that it suffices to focus on the extreme points of the polyhedron $\mathcal{P}$.

Theorem 1 shows that we can transform the two stage adaptive distributionally robust optimization problem to a well-known classical robust optimization problem with SOC representable uncertainty set, $\bar{\mathcal{W}}$. Hence, if $\bar{\mathcal{W}}$ is a polyhedron, then Problem (8) will become a linear optimization problem (LP). This is the case, for instance, when the ambiguity set has a hyper-rectangular support set and expectation constraints over absolute deviations. On the other hand, if the ambiguity set has expectation constraints over variances, the adaptive distributionally robust optimization problem will become an SOCP. Nevertheless, we note that Problem (8) is generally an intractable problem unless the number of extreme points of $\mathcal{P}$ is small. Hence, in the next section, we will show how we can tractably compute an upper bound of $\beta(\boldsymbol{x})$ using linear decision rule (LDR) approximation techniques.

3. Linear decision rule (LDR) approximation

Observe that we can express Problem (2) as a minimization problem over a measurable decision function map, $\boldsymbol{y} \in \mathcal{R}^{I_1,N_2}$ as follows:

$$\begin{aligned} \beta(\boldsymbol{x}) = \min \ & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\ \text{s.t.} \ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y} \in \mathcal{R}^{I_1,N_2}. \end{aligned} \tag{13}$$

However, problem (13) is generally an intractable optimization problem because $\boldsymbol{y}$ is a function map instead of a finite vector of decision variables. Nevertheless, we can obtain an upper bound of the problem by restricting $\boldsymbol{y}$ to a smaller class of function map or so called decision rule, which can be characterized by a polynomial number of decision variables. In a regular LDR approximation, the decision rule is restricted to one that is affinely dependent on the primary random variable $\tilde{\boldsymbol{z}}$.

Under the regular LDR approximation, we obtain an upper bound of $\beta(\boldsymbol{x})$ by solving the following problem,

$$\begin{aligned} \beta_L(\boldsymbol{x}) = \min \ & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\ \text{s.t.} \ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y} \in \mathcal{L}^{N_2}, \end{aligned} \tag{14}$$

where

$$\mathcal{L}^N = \left\{ \boldsymbol{y} \in \mathcal{R}^{I_1,N} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y}^0, \boldsymbol{y}_i^1, \ i \in [I_1] : \\ \boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \sum_{i\in[I_1]} \boldsymbol{y}_i^1 z_i \end{array} \right\}.$$

Unfortunately, despite its simplicity, the LDR approximation may lead to infeasibility even if the problem has complete recourse (see, for instance, Chen et al. 2008, Goh and Sim 2009). As an illustration, we consider the following complete recourse problem,

$$\begin{aligned} \beta = \min \ & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(y(\tilde{z})) \\ \text{s.t.} \ & y(z) \geq z \qquad \forall z \in \mathbb{R} \\ & y(z) \geq -z \qquad \forall z \in \mathbb{R} \\ & y \in \mathcal{R}^{1,1} \end{aligned} \tag{15}$$

where

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) \ \middle| \ \mathbb{E}_{\mathbb{P}}(|\tilde{z}|) \leq 1 \right\}.$$

Clearly, $y(z) = |z|$ is the optimal decision rule that yields $\beta = 1$. However, under the regular LDR approximation, we require $y(z) = y_0 + y_1 z$ for some $y_0, y_1 \in \mathbb{R}$, which will be infeasible in the following set of semi-infinite constraints,

$$\begin{aligned} y_0 + y_1 z \geq z \quad \forall z \in \mathbb{R} \\ y_0 + y_1 z \geq -z \ \forall z \in \mathbb{R}. \end{aligned} \tag{16}$$

Incorporating auxiliary variables in the LDR

We propose a powerful extension by of the LDR approximation by simply including the auxiliary random variable $\tilde{\boldsymbol{u}}$ associated with the lifted ambiguity set. Specifically, under this LDR approximation, we solve the following problem:

$$\begin{aligned} \beta_E(\boldsymbol{x}) = \min \ & \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})) \\ \text{s.t.} \ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) \qquad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{y} \in \bar{\mathcal{L}}^{N_2}, \end{aligned} \tag{17}$$

where

$$\bar{\mathcal{L}}^N = \left\{ \boldsymbol{y} \in \mathcal{R}^{I_1+I_2,N} \ \middle| \ \begin{array}{l} \exists \boldsymbol{y}^0, \boldsymbol{y}_i^1, \boldsymbol{y}_j^2 \in \mathbb{R}^N, \forall i \in [I_1], j \in [I_2]: \\ \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) = \boldsymbol{y}^0 + \sum_{i\in[I_1]} \boldsymbol{y}_i^1 z_i + \sum_{j\in[I_2]} \boldsymbol{y}_j^2 u_j \end{array} \right\}.$$

THEOREM 2.

$$\beta(\boldsymbol{x}) \leq \beta_E(\boldsymbol{x}) \leq \beta_L(\boldsymbol{x})$$

Proof. Given the flexibility of the new LDR approximation, it follows trivially that $\beta_E(\boldsymbol{x}) \leq \beta_L(\boldsymbol{x})$. Let $\boldsymbol{y}^*$ be an optimal decision rule of Problem (17) and we define the decision rule $\boldsymbol{y}^\dagger \in \mathcal{R}^{I_1,N_2}$ such that

$$\boldsymbol{y}^\dagger(\boldsymbol{z}) = \boldsymbol{y}^*(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z})).$$

Observe that $\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}})$ is feasible in Problem (13) and hence,

$$\beta(\boldsymbol{x}) \leq \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}})) = \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}}, \boldsymbol{g}(\tilde{\boldsymbol{z}}))) \leq \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}^*(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})) = \beta_E(\boldsymbol{x}). \quad \square$$

Proposition 2. *Problem (17) is equivalent to the affinely adjustable robust optimization problem of Ben-Tal et al. (2004) as follows,*

$$\begin{array}{rll} \beta_E(\boldsymbol{x}) = \min & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} & \\ \text{s.t.} & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) & \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) & \forall (\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} & \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} & \\ & \boldsymbol{y} \in \bar{\mathcal{L}}^{N_2}. & \end{array} \tag{18}$$

Proof. We omit the proof as it follows straightforwardly from Theorem 1. □

As a consequence of Proposition 2, we can transform an adaptive distributionally robust optimization problem under the LDR approximation to a classical affinely adjustable robust optimization problem. Hence, software packages for handling adjustable robust optimization problems such as AIMMS or JuMPeR can also be used to address distributional ambiguity as well.

A major improvement of the LDR approximation that incorporates the auxiliary random variables is the ability to resolve the issue of infeasibility in complete recourse problems.

Theorem 3. *Suppose Problem (13) has complete recourse and the objective is bounded from below. For any ambiguity set, $\mathbb{F}$ such that*

$$\mathbb{E}_{\mathbb{P}}(|\tilde{z}_i|) < \infty \qquad \forall \mathbb{P} \in \mathbb{F}$$

there exists a lifted ambiguity set, $\mathbb{G}$ whose corresponding LDR is feasible in Problem (17).

Proof. Let $\boldsymbol{\sigma} > \boldsymbol{0}$ be such that

$$\mathbb{E}_{\mathbb{P}}(|\tilde{z}_i|) \leq \sigma_i \qquad \forall \mathbb{P} \in \mathbb{F}.$$

Consider the following lifted ambiguity set $\mathbb{F}$ is

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_1}) \;\middle|\; \begin{array}{l} \prod_{\tilde{\boldsymbol{z}}} \mathbb{Q} \in \mathbb{F} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma} \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}) = 1 \end{array} \right\},$$

in which the lifted support set is $\bar{\mathcal{W}} = \{(\boldsymbol{z},\boldsymbol{u}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_1} \mid \boldsymbol{u} \geq \boldsymbol{z}, \boldsymbol{u} \geq -\boldsymbol{z}\}$. The decision rule under the new LDR approximation is given by

$$\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) = \boldsymbol{y}^0 + \sum_{i \in [I_1]} \boldsymbol{y}_i^1 z_i + \sum_{i \in [I_1]} \boldsymbol{y}_i^2 u_i.$$

To show that Problem (17) is feasible and finite under the new LDR, we consider the following relaxed lifted ambiguity set

$$\bar{\mathbb{G}} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_1}) \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \leq \boldsymbol{\sigma} \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}) = 1 \end{array} \right\},$$

and the following problem,

$$\begin{aligned}\bar{\beta}_E(\boldsymbol{x}) = \min\ & \sup_{\mathbb{Q}\in\bar{\mathbb{G}}} \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}},\tilde{\boldsymbol{u}})) \\ \text{s.t.}\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{y} \in \bar{\mathcal{L}}^{N_2}.\end{aligned} \tag{19}$$

Since $\mathbb{G} \subseteq \bar{\mathbb{G}}$, we have $\beta_E(\boldsymbol{x}) \leq \bar{\beta}_E(\boldsymbol{x})$. Consequently, it suffices to show that there exists a solution $\boldsymbol{y} \in \bar{\mathcal{L}}^{N_2}$ that is feasible in the following problem,

$$\begin{aligned}\min\ & r + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.}\ & r + \boldsymbol{t}'\boldsymbol{u} \geq \boldsymbol{d}'\boldsymbol{y}^0 + \sum_{i\in[I_1]} \boldsymbol{d}'\boldsymbol{y}_i^2 u_i & \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{B}\boldsymbol{y}^0 + \sum_{i\in[I_1]} \boldsymbol{B}\boldsymbol{y}_i^1 z_i + \sum_{i\in[I_1]} \boldsymbol{B}\boldsymbol{y}_i^2 u_i \geq \boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x} + \sum_{i\in I_1}(\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x})z_i & \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} \\ & \boldsymbol{y}^0, \boldsymbol{y}_i^1, \boldsymbol{y}_i^2 \in \mathbb{R}^{N_2} \quad i \in [I_1].\end{aligned} \tag{20}$$

Since $\boldsymbol{B}$ is a matrix that satisfies the condition of complete recourse, there exist $\bar{\boldsymbol{y}}^0, \bar{\boldsymbol{y}}_i^2$ $i \in [I_1]$, such that

$$\boldsymbol{B}\bar{\boldsymbol{y}}^0 \geq \boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x},$$

$$\boldsymbol{B}\bar{\boldsymbol{y}}_i^2 \geq (\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x}), \quad \boldsymbol{B}\bar{\boldsymbol{y}}_i^2 \geq -(\boldsymbol{b}^i - \boldsymbol{A}^i\boldsymbol{x}) \quad \forall i \in [I_1].$$

Observe that given any $a \in \mathbb{R}$, $\boldsymbol{b} \in \mathbb{R}^{I_1}$, $x \geq a$, and $y_i \geq |b_i|$, $i \in [I_1]$, we have

$$x + \boldsymbol{y}'\boldsymbol{u} \geq a + \boldsymbol{b}'\boldsymbol{z} \qquad \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}}.$$

Hence, a feasible solution for Problem (20) would be

$$\begin{aligned} r &= \boldsymbol{d}'\bar{\boldsymbol{y}}^0 \\ t_i &= \max\{\boldsymbol{d}'\bar{\boldsymbol{y}}_i^2, 0\} & \forall i \in [I_1] \\ \boldsymbol{y}^0 &= \bar{\boldsymbol{y}}^0 \\ \boldsymbol{y}_i^1 &= \boldsymbol{0} & \forall i \in [I_1] \\ \boldsymbol{y}_i^2 &= \bar{\boldsymbol{y}}_i^2 & \forall i \in [I_1].\end{aligned}$$

□

Remarks: Note that the class of ambiguity set depicted in Theorem 3 encompasses any random variable with finite deviation, i.e., $\mathbb{E}_{\mathbb{P}}(|\tilde{z}_i - \mu_i|^{p_i}) < \infty$ for some $\mu_i \in \mathbb{R}$, $p_i \geq 1$, $i \in [I_1]$, since

$$\mathbb{E}_{\mathbb{P}}(|\tilde{z}_i|) \leq \mathbb{E}_{\mathbb{P}}(|\tilde{z}_i - \mu_i|) + |\mu_i| \leq (\mathbb{E}_{\mathbb{P}}(|\tilde{z}_i - \mu_i|^{p_i}))^{1/p_i} + |\mu_i| < \infty.$$

More interestingly, we show in the following result that the new LDR approximation can attain the optimal objective values for a class of adaptive distributionally robust linear optimization problems.

THEOREM 4. *Suppose Problem (13) is a complete recourse problem with only one second stage decision variable, i.e.* $N_2 = 1$*, then*

$$\beta(\boldsymbol{x}) = \beta_E(\boldsymbol{x}).$$

Proof. Since $\boldsymbol{B}$ is a matrix that satisfies the condition of complete recourse, for $N_2 = 1$, the matrix $\boldsymbol{B} \in \mathbb{R}^{M\times 1}$ must satisfy either $\boldsymbol{B} > \boldsymbol{0}$ or $\boldsymbol{B} < \boldsymbol{0}$. Observe that the problem

$$\begin{aligned} Q(\boldsymbol{x},\boldsymbol{z}) = \min\ & dy \\ \text{s.t.}\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}y \geq \boldsymbol{b}(\boldsymbol{z})\ \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & y \in \mathbb{R}, \end{aligned}$$

is unbounded below whenever $d\boldsymbol{B} < \boldsymbol{0}$. Since the second stage decision variable y is unconstrained, for the problem to be finite, we can assume without loss of generality that $\boldsymbol{B} > 0$ and $d \geq 0$. In which case,

$$Q(\boldsymbol{x},\boldsymbol{z}) = d \max_{i\in[M]} \left\{ \frac{[\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_i}{[\boldsymbol{B}]_i} \right\}.$$

Observe that terms in $Q(\boldsymbol{x},\boldsymbol{z})$ corresponds to the extreme points of polyhedron

$$\mathcal{P} = \left\{ \boldsymbol{p} \in \mathbb{R}^M_+ : \boldsymbol{B}'\boldsymbol{p} = d \right\}.$$

Hence, applying Theorem 1, we have

$$\begin{aligned} \beta(\boldsymbol{x}) = \min\ & d(r + \boldsymbol{s}' + \boldsymbol{t}'\boldsymbol{\sigma}) \\ \text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \tfrac{[\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_i}{[\boldsymbol{B}]_i}\ \ \forall i \in [M], \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2}. \end{aligned} \tag{21}$$

The solution derived under the LDR approximation would be

$$\begin{aligned} \beta_E(\boldsymbol{x}) = \min\ & r + \boldsymbol{s}' + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq dy(\boldsymbol{z},\boldsymbol{u})\ \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}y(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} \\ & y \in \bar{\mathcal{L}}^1 \end{aligned}$$

or equivalently

$$\begin{aligned} \beta_E(\boldsymbol{x}) = \min\ & r + \boldsymbol{s}' + \boldsymbol{t}'\boldsymbol{\sigma} \\ \text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq d(y^0 + \boldsymbol{y}^{1\prime}\boldsymbol{z} + \boldsymbol{y}^{2\prime}\boldsymbol{u})\ \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & (y^0 + \boldsymbol{y}^{1\prime}\boldsymbol{z} + \boldsymbol{y}^{2\prime}\boldsymbol{u}) \geq \tfrac{[\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}]_i}{[\boldsymbol{B}]_i} \qquad \forall i \in [M], \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} \\ & y^0 \in \mathbb{R}, \boldsymbol{y}^1 \in \mathbb{R}^{I_1}, \boldsymbol{y}^2 \in \mathbb{R}^{I_2}. \end{aligned} \tag{22}$$

Let $(r^\dagger, \boldsymbol{s}^\dagger, \boldsymbol{t}^\dagger)$ be a feasible solution of Problem (21). We can construct a feasible solution $(r, \boldsymbol{s}, \boldsymbol{t}, y^0, \boldsymbol{y}^1, \boldsymbol{y}^2)$ to Problem (22) by letting

$$y^0 = r^\dagger, \boldsymbol{y}^1 = \boldsymbol{G}'\boldsymbol{s}^\dagger, \boldsymbol{y}^2 = \boldsymbol{t}^\dagger, r = dr^\dagger, \boldsymbol{s} = d\boldsymbol{s}^\dagger, \boldsymbol{t} = d\boldsymbol{t}^\dagger,$$

which yields the same objective as Problem (21). Hence, $\beta(\boldsymbol{x}) \le \beta_E(\boldsymbol{x}) \le \beta(\boldsymbol{x})$. □

Remarks: Note that for complete recourse problem with $N_2 = 1$, Problem (8) becomes a tractable problem since the number of extreme points of the polyhedron $\mathcal{P}$ becomes M. Nevertheless, notwithstanding the simplicity, we are not aware of other types of decision rules that would yield tight results for this instance.

A natural question is whether we could extend the results of Theorem 3 and 4 to the case of relatively complete recourse. However, this is not the case as depicted in the following negative result even for the case of $N_2 = 1$.

Proposition 3. *There exists a relatively complete recourse problem with $N_2 = 1$ for which Problem (17) is infeasible under any LDR that incorporates both the primary and auxiliary random variables associated with the lifted ambiguity set.*

Proof. Consider the following problem

$$\begin{array}{lll} \min & 0 & \\ \text{s.t.} & y(\boldsymbol{z}) \ge z_1 - z_2 & \forall \boldsymbol{z} \in \mathcal{W} \\ & y(\boldsymbol{z}) \ge z_2 - z_1 & \forall \boldsymbol{z} \in \mathcal{W} \\ & y(\boldsymbol{z}) \le z_1 + z_2 + 2 & \forall \boldsymbol{z} \in \mathcal{W} \\ & y(\boldsymbol{z}) \le -z_1 - z_2 + 2 & \forall \boldsymbol{z} \in \mathcal{W} \\ & y \in \mathcal{R}^{2,1} & \end{array} \tag{23}$$

where

$$\mathcal{W} = \left\{ \boldsymbol{z} \in \mathbb{R}^2 \;\middle|\; \|\boldsymbol{z}\|_\infty \le 1 \right\}.$$

We verify that Problem (23) is one with relatively complete recourse by showing that $y(\boldsymbol{z}) = |z_1 - z_2|$ is a feasible solution. Indeed, $y(\boldsymbol{z})$ is feasible in the first two sets of constraints of Problem (23). Moreover, for all $\|\boldsymbol{z}\|_\infty \le 1$,

$$\begin{aligned} & |z_1 - z_2| + |z_1 + z_2| \\ = & \max\{z_1 - z_2 + z_1 + z_2, -z_1 + z_2 + z_1 + z_2, z_1 - z_2 - z_1 - z_2, -z_1 + z_2 - z_1 - z_2\} \\ = & \max\{2z_1, 2z_2, -2z_2, -2z_1\} \le 2. \end{aligned}$$

Hence, for all $\boldsymbol{z} \in \mathcal{W}$, $y(\boldsymbol{z}) \le -|z_1 + z_2| + 2$ and it is therefore feasible in the last two sets of constraints of Problem (23). We also note that there does not exists a feasible LDR in which y is affinely dependent on $\boldsymbol{z}$, i.e. $y(\boldsymbol{z}) = y_0 + y_1 z_1 + y_2 z_2$ for some $y_0, y_1, y_2 \in \mathbb{R}$. Indeed, when substituting the extreme points of $\mathcal{W}$ in Problem (23), we will obtain the following set of infeasible equations,

$$\begin{aligned} z_1 = z_2 = 1 &\Rightarrow y_0 + y_1 + y_2 = 0 \\ z_1 = z_2 = -1 &\Rightarrow y_0 - y_1 - y_2 = 0 \\ z_1 = -z_2 = 1 &\Rightarrow y_0 + y_1 - y_2 = 2 \\ z_1 = -z_2 = -1 &\Rightarrow y_0 - y_1 + y_2 = 2. \end{aligned}$$

For any lifted ambiguity set, the corresponding lifted support set $\bar{\mathcal{W}}$ is defined in (6), where the parameter $\boldsymbol{u}$ is associated with the auxiliary random variable. Incorporating the auxiliary random variable in the LDR, we have

$$y(\boldsymbol{z}, \boldsymbol{u}) = y_0 + \boldsymbol{y}^1 \boldsymbol{z} + \boldsymbol{y}^2 \boldsymbol{u}.$$

Since $\boldsymbol{u}$ is unbounded from above in the lifted ambiguity set $\bar{\mathcal{W}}$, the constraints

$$y(\boldsymbol{z}, \boldsymbol{u}) \geq z_1 - z_2 \qquad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$$

imply $\boldsymbol{y}^2 \geq \boldsymbol{0}$. Similarly, the constraints

$$y(\boldsymbol{z}, \boldsymbol{u}) \leq z_1 + z_2 + 2 \qquad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$$

imply $\boldsymbol{y}^2 \leq \boldsymbol{0}$. Hence, the LDR is reduced one that is only affinely dependent on $\boldsymbol{z}$, which, as we have shown, is infeasible in Problem (23).

Quite surprisingly, by simply extending the LDR approximation to include the auxiliary random variable of the lifted ambiguity set, we are able to attain the optimum objective values for the class of complete recourse problems described in Theorem 4. In contrast, the proposed deflected LDR (DLDR) approximations of Chen et al. (2008), Goh and Sim (2009), albeit more complex, do not necessarily recover the optimum objective values for this class of complete recourse problems. In Appendix A, we will further demonstrate that the new LDR approximation can indeed improve over the more sophisticated DLDR approximations.

As a useful consequence of Theorem 4, the worst-case expectation of a convex piecewise linear function,

$$\mathbb{E}_{\mathbb{P}}\left(\max_{i\in[K]}\{\boldsymbol{a}_i'(\tilde{\boldsymbol{z}})\boldsymbol{x} + b_i(\tilde{\boldsymbol{z}})\}\right)$$

can also be expressed as

$$\begin{aligned} \min \ & \sup_{\mathbb{Q}\in\mathbb{G}} \mathbb{E}_{\mathbb{Q}}(y(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})) \\ \text{s.t.} \ & y(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{a}_k(\boldsymbol{z})'\boldsymbol{x} + b_k(\boldsymbol{z}) \qquad \forall k \in [K], (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}, \\ & \boldsymbol{y} \in \bar{\mathcal{L}}^1, \end{aligned}$$

which can be modeled directly using algebraic modeling software package. In fact, this technique can be applied straightforwardly to obtain exact solutions in adaptive distributionally robust optimization problems found in recent applications such as Meng et al. (2015) and Qi (2015). We will use the case study of medical appointment scheduling to show how we could easily apply our results to study various types of ambiguity sets.

Generalization to multistage problems

Another important feature the LDR approximation is the ability to easily enforce non-anticipativity in multistage decision making. For given subsets $\mathcal{S}^i \subseteq [I_1]$, $i \in [N_2]$ that reflect the information dependency of the adaptive decisions, y_i, we consider the generalization of Problem (13) as follows:

$$\begin{aligned} \gamma(\boldsymbol{x}) = \min \sup_{\mathbb{P}\in\mathbb{F}} & \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) \\ \text{s.t. } & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{b}(\boldsymbol{z}) \quad && \forall \boldsymbol{z} \in \mathcal{W} \\ & y_i \in \mathcal{R}^{I_1}(\mathcal{S}^i) && \forall i \in [N_2], \end{aligned} \tag{24}$$

where we define the space of restricted measurable functions as

$$\mathcal{R}^I(\mathcal{S}) = \left\{ y \in \mathcal{R}^{I,1} \;\middle|\; y(\boldsymbol{v}) = y(\boldsymbol{w}) \;\forall \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^I : v_j = w_j, \forall j \in \mathcal{S} \right\}.$$

Problem (24) solves for the optimal decision rule $\boldsymbol{y} \in \mathcal{R}^{I_1,N_2}$ that minimizes the worst-case expected objective taking into account of the information dependency requirement. Under the new LDR approximation, we have

$$\begin{aligned} \gamma_E(\boldsymbol{x}) = \min \sup_{\mathbb{Q}\in\mathbb{G}} & \mathbb{E}_{\mathbb{Q}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}})) \\ \text{s.t. } & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u})) \geq \boldsymbol{b}(\boldsymbol{z}) \quad && \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\ & y_i \in \bar{\mathcal{L}}(\mathcal{S}^i, \mathcal{T}^i) && \forall i \in [N_2], \end{aligned} \tag{25}$$

where

$$\bar{\mathcal{L}}(\mathcal{S}, \mathcal{T}) = \left\{ y \in \mathcal{R}^{I_1+I_2,1} \;\middle|\; \begin{array}{l} \exists y^0, y_i^1, y_j^2 \in \mathbb{R}, \forall i \in \mathcal{S}, j \in \mathcal{T}: \\ y(\boldsymbol{z}, \boldsymbol{u}) = y^0 + \sum_{i\in\mathcal{S}} y_i^1 z_i + \sum_{j\in\mathcal{T}} y_j^2 u_j \end{array} \right\}$$

and the subsets $\mathcal{T}^i \subseteq [I_2]$, $i \in [N_2]$ are consistent with the information restriction imposed by $\mathcal{S}^i \subseteq [I_1]$. In Section 5, we will illustrate how we can use this model to formulate a multi-period inventory control problem.

Enhancements of LDR approximations

As further enhance to the LDR approximation, we can incorporate the auxiliary variables associated with the support set as proposed in Chen et al. (2008), Chen and Zhang (2009), Goh and Sim (2010). However, these approaches do not provide systematic ways of refining the approximations towards optimality. More recently, Zheng et al. (2016) demonstrate how an adaptive robust or distributionally robust optimization problem can be transformed to a static robust optimization problems via Fourier-Motzkin elimination (FME). For instance, without imposing complete recourse, if $N_2 = 1$, we can eliminate the only recourse variable via FME, and solve the static robust optimization problem to optimality in polynomial time. Although this approach would generally create exponential number of constraints, to keep the model tractable, we can perform partial FME and apply our LDR approximation to improve the solutions. Hence, this generic reformulation technique enhances our LDR approximation and enables us to solve adaptive distributionally robust optimization problems to the level of optimality within the limits of the available computational resources.

On interpreting decision rule as policy and the issue of time consistency

For a given $\boldsymbol{x} \in X$, let $\boldsymbol{y}^*$ be the optimal decision rule of Problem (13), and $\boldsymbol{y}^*_E$ be the optimal decision rule of Problem (17). Consider a policy based on the decision rule, $\boldsymbol{y}^\dagger \in \mathcal{R}^{I_1,N_2}$ such that

$$\boldsymbol{y}^\dagger(\boldsymbol{z}) = \boldsymbol{y}^*_E(\boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z})).$$

Observe that $\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}})$ is feasible in Problem (13) and from the proof of Theorem 2, it follows that

$$\beta(\boldsymbol{x}) \leq \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}})) \leq \beta_E(\boldsymbol{x}).$$

Suppose $\beta(\boldsymbol{x}) = \beta_E(\boldsymbol{x})$, which is the case for complete recourse problems and $N_2 = 1$, there is a tendency to infer the optimality of $\boldsymbol{y}^\dagger$ vis-à-vis $\boldsymbol{y}^*$ so that

$$\boldsymbol{d}'\boldsymbol{y}^\dagger(\boldsymbol{z}) = \boldsymbol{d}'\boldsymbol{y}^*(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{W}.$$

However, this is not the case and we will demonstrate this fallacy in the following simple example. Consider the following complete recourse problem,

$$\begin{aligned} \beta = \min\ & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(y(\tilde{z})) \\ \text{s.t.}\ & y(z) \geq z && \forall z \in \mathbb{R} \\ & y(z) \geq -z && \forall z \in \mathbb{R} \\ & y \in \mathcal{R}^{1,1}, \end{aligned} \tag{26}$$

where

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) \;\middle|\; \mathbb{E}_{\mathbb{P}}(\tilde{z}) = 0, \mathbb{E}_{\mathbb{P}}(\tilde{z}^2) \leq 1 \right\}.$$

Clearly, $y^*(z) = |z|$ is the optimal solution and it is also the optimal objective value for all $z \in \mathbb{R}$. However, under the LDR approximation, we obtain $y^\dagger(z) = \frac{1+z^2}{2}$, which is essentially greater than the optimum policy $y^*(z)$ except at $z = 1$ and $z = -1$. Incidentally, the worst-case distribution $\mathbb{P} \in \mathbb{F}$ corresponds to the two point distributions with $\mathbb{P}(\tilde{z} = 1) = \mathbb{P}(\tilde{z} = -1) = 1/2$, which explains why their worse-case expectations coincide. This is similar to Delage and Iancu (2015) observation that the worst-case policy generated by decision rule can be inefficient and such degeneracy is common in robust multistage decision models.

Another issue with using the optimal decision rule as a policy is the potential violation of time consistency. In dynamic decision making, time inconsistency arrises when an optimal policy perceived in one time period may not be recognized as optimal in another. Delage and Iancu (2015), Xin et al. (2015) show that in addressing multiperiod robust or distributionally robust optimization problems, time consistency may be affected by how the ambiguity sets are being updated dynamically. While time consistency is a desirable feature in rational decision making, policies that may violate time consistency have also been justified in the literature (see, for instance, Basak and Chabakauri 2010, Kydland and Prescott 1977, Richardson 1989, Bajeux-Besnainou and Portait 1998).

Consequently, when solving the adaptive distributionally robust optimization problem, we caution against using the optimal decision rule as a policy. In many practical applications of dynamic decision making, it suffices to implement the *here-and-now* decision without having to commit to a policy that dictates how the solution might change as uncertainty unfolds. For a two stage problem, the second stage decision should be determined by solving a linear optimization problem after the uncertainty is resolved. In addressing a multistage decision problem, we advocate using the LDR approximation to obtain the *here-and-now* decision, $\boldsymbol{x} \in X$, which accounts for how decisions might adapt as uncertainty unfolds over the stages. As we proceed to the next stage, we should adopt the folding horizon approach and solve for new *here-and-now* decision using the latest available information as inputs to another adaptive distributionally robust optimization problem.

Software packages

As a proof of concept, we have develop the software package, ROC (`http://www.meilinzhang.com/software`) to provide an intuitive environment for formulating and solving our proposed adaptive distributionally robust optimization models. ROC is developed in the C++ programming language, which is fast, highly portable and well suited for deployment in decision support systems. A typical algebraic modeling package provides the standardized format for declaration of decision variables, transcription of constraints and objective functions, and interface with external solvers. ROC has additional features including declaration of uncertain parameters and linear decision rules, transcriptions of ambiguity sets and automatic reformulation of standard and distributionally robust counterparts using the techniques described in this paper. Interestingly, XProg (`http://xprog.weebly.com`), a new MATLAB based algebraic modeling package that implements our proposed framework has independently emerged. The design of

XProg is similar to ROC. Since MATLAB platform is a more user friendly environment, XProg can be used for rapid prototyping of models, while ROC would be better suited for deployment of the solutions. The examples in our numerical studies below can easily be implemented in both ROC and XProg.

4. An application in medical appointment scheduling

For the first application, we consider a medical appointment scheduling problem where patients arrive at their stipulated schedule and may have to wait in a queue to be served by a physician. The patients' consultation times are uncertain and their arrival schedules are determined at the first stage, which can influence the waiting times of the patients and the overtime of the physician.

To formulate the problem, we consider N patients arriving in sequence with their indices $j \in [N]$ and the uncertain consultation times are denoted by $\tilde{z}_j,\ j \in [N]$. We let the first stage decision variable, x_j to represent the inter-arrival time between patient j to the adjacent patient $j+1$ for $j \in [N-1]$ and x_N to denote the time between the arrival of the last patient and the scheduled completion time for the physician before overtime commences. The first patient will be scheduled to arrive at the starting time of zero and subsequent patients i, $i \in [N], i \geq 2$ will be scheduled to arrive at $\sum_{j \in [i-1]} x_j$. Consequently, the feasible region of $\boldsymbol{x}$ is given by

$$X = \left\{ \boldsymbol{x} \in \mathbb{R}^N_+ \;\middle|\; \sum_{i \in [N]} x_i \leq T \right\},$$

where T is the scheduled completion time for the physician before overtime commences.

A common decision criterion in the medical appointment schedule is to minimize the expected total cost of patients waiting and physician overtime, where the cost of a patient waiting is normalized to one per unit delay and the physician's overtime cost is γ per unit delay. For a given arrival schedule $\boldsymbol{x} \in X$, and a realization of consultation times $\boldsymbol{z} \in \mathbb{R}^N_+$, the total cost can be determined by solving the following linear optimization problem

$$\begin{aligned} Q(\boldsymbol{x}, \boldsymbol{z}) = \min\ & \sum_{i \in [N]} y_i + \gamma y_{N+1} \\ \text{s.t.}\ & y_i - y_{i-1} + x_{i-1} \geq z_{i-1}\ \forall i \in \{2, \ldots, N+1\} \\ & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned} \tag{27}$$

where the y_i denotes the waiting time of patient i, $i \in [N]$ and y_{N+1} represents the overtime of the physician. Note that the appointment scheduling problem is one that has complete recourse. From Theorem 1, we can compute the worst-case expectation over an ambiguity set $\mathbb{F}$,

$$\min_{\boldsymbol{x} \in X} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(Q(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right),$$

by enumerating all the extreme points of the corresponding dual feasible set,

$$\mathcal{P} = \left\{ \boldsymbol{p} \in \mathbb{R}^N_+ \;\middle|\; \begin{array}{l} p_i - p_{i-1} \geq -1 \ \forall i \in \{2, \ldots, N\} \\ p_N \leq \gamma \end{array} \right\}.$$

However, given the exponentially large number of extreme points of $\mathcal{P}$, it would be generally intractable to obtain exact solutions.

Kong et al. (2013) is first to propose a distributional robust optimization for the appointment scheduling problem. They consider a cross moment ambiguity set that characterizes the distributions of all nonnegative random variables with some specified mean values, $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ as follows:

$$\bar{\mathbb{F}}_{CM} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N) \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}\left((\boldsymbol{z} - \boldsymbol{\mu})(\boldsymbol{z} - \boldsymbol{\mu})'\right) = \boldsymbol{\Sigma} \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathbb{R}^N_+) = 1 \end{array} \right\}. \tag{28}$$

As the problem is intractable, they formulate a SDP relaxation that solves the problem approximately.

To obtain a tractable formulation, Mak et al. (2014) ignore information on covariance and consider a marginal moment ambiguity set as follows:

$$\bar{\mathbb{F}}_{MM} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{P}}((\tilde{z}_i - \mu_i)^2) = \sigma_i^2 & \forall i \in [N] \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathbb{R}^N_+) = 1 & \end{array} \right\}, \tag{29}$$

where $\sigma_i^2, i \in [N]$ is the variance of $\tilde{z}_i$. Surprsingly, Mak et al. (2014) show that the model has a hidden tractable reformulation, which they have cleverly exploited to obtain exact solutions. Observe that due to equality constraints on variances, $\bar{\mathbb{F}}_{MM}$ is not a SOC ambiguity set. Nevertheless, by relaxing the equality constraints to inequality constraints, we will obtain the following SOC ambiguity set:

$$\mathbb{F}_{MM} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{P}}((\tilde{z}_i - \mu_i))^2 \leq \sigma_i^2 & \forall i \in [N] \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathbb{R}^N_+) = 1 & \end{array} \right\}. \tag{30}$$

Correspondingly, the lifted ambiguity set, $\mathbb{G}_{MM}$ is

$$\mathbb{G}_{MM} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^N) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}(\tilde{u}_i) \leq \sigma_i^2 & \forall i \in [N] \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}_{PCM}) = 1 & \end{array} \right\},$$

where

$$\bar{\mathcal{W}}_{MM} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^N \times \mathbb{R}^N \;\middle|\; \begin{array}{l} \boldsymbol{z} \geq \boldsymbol{0} \\ \sqrt{(z_i - \mu_i)^2 + \left(\frac{u_i - 1}{2}\right)^2} \leq \frac{u_i + 1}{2} \ \forall i \in [N] \end{array} \right\}.$$

In the proof of Mak et al. (2014) Theorem 1, the authors show that the unconstrained dual variables that are associated with the second moment equality constraints must be positive at optimality. Incidentally, when the problem is replaced by the relaxed ambiguity set, these variables would have nonnegative constraints. Hence, optimizing over the relaxed ambiguity set, $\mathbb{F}_{MM}$ would yield the same set of optimal solutions as those obtained via the ambiguity set, $\bar{\mathbb{F}}_{MM}$.

However, since the marginal moment ambiguity set is incapable of incorporating covariance, despite the ease of computing the optimal solution, it may lead to conservative solutions. As a compromise, we propose the following partial cross moment SOC ambiguity set,

$$\mathbb{F}_{PCM} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^N) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{P}}((\tilde{z}_i - \mu_i)^2) \le \sigma_i^2 & \forall i \in [N] \\ \mathbb{E}_{\mathbb{P}}((\mathbf{1}'(\tilde{\boldsymbol{z}} - \boldsymbol{\mu}))^2) \le \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} & \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathbb{R}^N_+) = 1 & \end{array} \right\}. \tag{31}$$

The corresponding lifted ambiguity set, $\mathbb{G}_{PCM}$ is

$$\mathbb{G}_{PCM} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^N \times \mathbb{R}^{N+1}) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} & \\ \mathbb{E}_{\mathbb{Q}}(\tilde{u}_i) \le \sigma_i^2 & \forall i \in [N] \\ \mathbb{E}_{\mathbb{Q}}(\tilde{u}_{N+1}) \le \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} & \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}_{PCM}) = 1 & \end{array} \right\},$$

where

$$\bar{\mathcal{W}}_{PCM} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^N \times \mathbb{R}^{N+1} \;\middle|\; \begin{array}{ll} \boldsymbol{z} \ge \mathbf{0} & \\ (z_i - \mu_i)^2 \le u_i & \forall i \in [N] \\ \left(\sum_{i \in [N]} (z_i - \mu_i) \right)^2 \le u_{N+1} & \end{array} \right\} \tag{32}$$

or equivalently as a second order cone representable feasible set,

$$\bar{\mathcal{W}}_{PCM} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^N \times \mathbb{R}^{N+1} \;\middle|\; \begin{array}{ll} \boldsymbol{z} \ge \mathbf{0} & \\ \sqrt{(z_i - \mu_i)^2 + \left(\frac{u_i - 1}{2}\right)^2} \le \frac{u_i + 1}{2} & \forall i \in [N] \\ \sqrt{\left(\sum_{i \in [N]} (z_i - \mu_i) \right)^2 + \left(\frac{u_{N+1} - 1}{2} \right)^2} \le \frac{u_{N+1} + 1}{2} & \end{array} \right\}.$$

For these SOC ambiguity sets, we can obtain approximate solutions to the appointment scheduling problem based on the new LDR approximation as follows,

$$\begin{array}{lll} \min \sup_{\mathbb{Q} \in \mathbb{G}} \mathbb{E}_{\mathbb{Q}} \left(\sum_{i \in [N]} y_i(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) + \gamma y_{N+1}(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \right) & & \\ \text{s.t.} \;\; y_i(\boldsymbol{z}, \boldsymbol{u}) - y_{i-1}(\boldsymbol{z}, \boldsymbol{u}) + x_{i-1} \ge z_{i-1} & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}, \forall i \in \{2, \dots, N+1\} & \\ \quad\;\; \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \ge \mathbf{0} & \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} & \\ \quad\;\; \boldsymbol{x} \in X & & \\ \quad\;\; \boldsymbol{y} \in \bar{\mathcal{L}}^{N+1}. & & \end{array} \tag{33}$$

In our numerical study, we investigate the performance of appointment scheduling problem among the following approaches,

- **(Regular LDR)**: Solutions based on the regular LDR approximation.
- **(Exact MM)**: Exact solutions (Mak et al. 2014) under the marginal moment ambiguity set, $\bar{\mathbb{F}}_{MM}$.
- **(Approx MM)**: Solutions based on the new LDR approximation under the relaxed marginal moment ambiguity set, $\mathbb{F}_{MM}$.
- **(Approx PCM)**: Solutions based on the new LDR approximation under the partial cross moment ambiguity set, $\mathbb{F}_{PCM}$.
- **(Approx CM)**: Solutions based on Kong et al. (2013) approximation under the cross moment ambiguity set, $\bar{\mathbb{F}}_{CM}$. We implement the model in YALMIP and use SDPT3 (Tutuncu et al. 2003) as the underlying SDP solver.

The numerical settings of our computational experiments are similar to Mak et al. (2014). We first consider $N = 8$ patients. The unit overtime cost is $\gamma = 2$. For each patient $i \in [N]$, we randomly select μ_i based on uniform distribution over $[30, 60]$ and $\sigma_i = \mu_i \cdot \epsilon$ where ϵ is randomly selected based on uniform distribution over $[0, 0.3]$. The covariance matrix is given by

$$[\mathbf{\Sigma}]_{ij} = \begin{cases} \alpha\sigma_i\sigma_j & \text{if } i \neq j \\ \sigma_j^2 & \text{otherwise.} \end{cases}$$

where $\alpha \in [0, 1]$ is the correlation coefficient between any two different random consultation times. The evaluation period, T depends on the instance parameters as follows,

$$T = \sum_{i=1}^{N} \mu_i + 0.5\sqrt{\sum_{i=1}^{N} {\sigma_i}^2}.$$

For each approach and $\alpha \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$, we obtain the objective values of seven randomly generated instances. We report the results in Table 1. Observe that Regular LDR performs extremely poorly. Indeed, as noted in Model (14), the regular LDR approximation is unable to incorporate most of the information of the ambiguity set other than the mean, $\boldsymbol{\mu}$ and the nonnegative support set, which leads to the ultra-conservative result. We note that Exact MM improves marginally over Approx MM, and Approx PCM improves over Approx MM. For uncorrelated or mildly correlated random variables, i.e., $\alpha \in \{0, 0.25\}$, Approx PCM would yield a lower objective then. Under perfect correlation, i.e., $\alpha = 1$, the objective values of Approx PCM and Approx MM coincide and they are marginally higher than the objective values of Exact MM. Hence, while there are benefits from the new LDR approximation, it does not replicate the optimal solution of Mak et al. (2014).

α	Instance Index	Objective Value				
		Regular LDR	Exact MM	Approx MM	Approx PCM	Approx CM
0.00	1	1987.66	22.12	22.67	20.71	19.27
	2	2089.22	46.32	47.21	43.06	33.52
	3	1987.66	37.03	37.79	34.51	27.40
	4	2089.22	66.32	66.88	61.00	49.28
	5	2009.91	89.47	91.62	83.59	72.25
	6	2100.27	121.33	123.65	112.33	*108.52
	7	1979.56	219.35	222.36	202.51	*190.01
0.25	1	1987.66	22.12	22.67	21.76	20.17
	2	2089.22	46.32	47.21	45.27	41.38
	3	1987.66	37.03	37.79	36.27	32.08
	4	2089.22	66.32	66.88	64.13	*67.45
	5	2009.91	89.47	91.62	87.80	*106.29
	6	2100.27	121.33	123.65	118.27	*171.25
	7	1979.56	219.35	222.36	213.09	*498.92
0.50	1	1987.66	22.12	22.67	22.39	*21.44
	2	2089.22	46.32	47.21	46.59	*48.25
	3	1987.66	37.03	37.79	37.31	*36.26
	4	2089.22	66.32	66.88	66.01	*84.77
	5	2009.91	89.47	91.62	90.43	*138.14
	6	2100.27	121.33	123.65	121.90	*229.40
	7	1979.56	219.35	222.36	219.43	*710.98
0.75	1	1987.66	22.12	22.67	22.66	*21.91
	2	2089.22	46.32	47.21	47.17	*53.96
	3	1987.66	37.03	37.79	37.76	*39.89
	4	2089.22	66.32	66.88	66.82	*100.54
	5	2009.91	89.47	91.62	91.55	*166.80
	6	2100.27	121.33	123.65	123.53	*282.34
	7	1979.56	219.35	222.36	222.18	*899.37
1.00	1	1987.66	22.12	22.67	22.67	*20.71
	2	2089.22	46.32	47.21	47.21	*57.34
	3	1987.66	37.03	37.79	37.79	*37.89
	4	2089.22	66.32	66.88	66.88	*108.85
	5	2009.91	89.47	91.62	91.62	*184.71
	6	2100.27	121.33	123.65	123.65	*330.60
	7	1979.56	219.35	222.36	222.36	*888.06

Table 1 Objective values under different instances with varying correlation coefficient α.

For Approx CM, unlike the previous approaches, we are unable to obtain its optimally verified solutions for all the instances. The suboptimal instances are marked by asterisks within Table 1. Among the instances that the solutions of Approx CM could be optimally verified, we observe that the corresponding objective values attained are the lowest values among the approaches. However, when Approx CM is suboptimal, its objective value could be significantly larger than the marginal moment approaches, especially when the correlation is high.

In Table 2, we increase the number of jobs from $N = 8$ to $N = 20$ and report the objective values and computational times for the different approaches. In this numerical study, we set $\alpha =$

0 and $\epsilon = 0.15$. Observe that it takes significantly longer time to solve Approx CM and more seriously, its solution may not necessarily be optimally verified by the SDP solver. Moreover, the suboptimal objectives of Approx CM deteriorates significantly as the size of the problem increases. In contrast, the Approx PCM can be computed instantaneously and reliably and its solution consistently improves over Exact MM and Approx MM for the case when $\alpha = 0$. Hence, this underscores the importance of having a stable optimization format and reaffirm our restriction to SOC ambiguity set.

Similar to Mak et al. (2014), we also compare the performance of the approaches in out-of-sample study using truncated normal and log normal probability distributions. We assume that the underlying random variables are independently distributed and the parameters of the distributions correspond to the respective ambiguity sets for which the distributionally robust solutions are obtained. Upon obtaining the solutions from the various methods, we compare the performance and present the results in Table 3. These values are evaluated via Monte Carlo simulation with $10,000$ trials under each specific distribution. Interestingly, despite the differences in objective values attained, the out-of-sample study alludes to the closeness of results between Exact MM and Approx MM as well as between Approx PCM and Approx CM. Since the random variables are independent, and hence uncorrelated, as we have expected, incorporating covariance or partial covariance information in the ambiguity would lead to improvement in the out-of-sample performances.

5. An application in multi-period inventory control

For the second application, we illustrate how the adaptive distributionally robust linear optimization framework can model a multistage decision problem. We consider a finite horizon, T period single product inventory control problem. Demands are filled from on-hand inventory and unfilled demands are fully backlogged. At the beginning of period t, a quantity of $x_t \in [0, \bar{x}_t]$ is ordered, which will arrive immediately to replenish the stock before the demand is realized. The unit ordering cost is c_t, excess inventory will incur a per-unit holding cost of h_t, while backlogged demand will be penalized with a per-unit underage cost of b_t. At the last period $t = T$, lost sales could be accounted for via the backorder cost. We use y_t to indicate the net inventory level at the beginning of period t. The initial net inventory level of the system is $y_1 = 0$.

As in Graves (1999) and See and Sim (2009), we model the demand process is an integrated moving average (IMA) process of order $(0, 1, 1)$ as follows:

$$\begin{aligned} d_t(\tilde{\boldsymbol{z}}) &= \tilde{z}_t + \alpha\tilde{z}_{t-1} + \alpha\tilde{z}_{t-2} + \cdots + \alpha\tilde{z}_1 + \mu \\ &= d_{t-1}(\tilde{\boldsymbol{z}}) - (1-\alpha)\tilde{z}_{t-1} + \tilde{z}_t, \end{aligned}$$

N	Models	Objective Values	Optimality Status	Computational Time (sec)
8	Exact MM	112.98	optimal	<1
	Approx MM	113.37	optimal	<1
	Approx PCM	103.53	optimal	<1
	Approx CM	96.06	optimal	5
9	Exact MM	142.36	optimal	<1
	Approx MM	145.22	optimal	<1
	Approx PCM	131.53	optimal	<1
	Approx CM	123.94	not optimal	7
10	Exact MM	156.36	optimal	<1
	Approx MM	158.42	optimal	<1
	Approx PCM	142.19	optimal	<1
	Approx CM	135.71	not optimal	11
11	Exact MM	199.98	optimal	<1
	Approx MM	203.32	optimal	<1
	Approx PCM	182.35	optimal	<1
	Approx CM	176.74	not optimal	16
12	Exact MM	216.74	optimal	<1
	Approx MM	219.65	optimal	<1
	Approx PCM	196.64	optimal	<1
	Approx CM	215.11	not optimal	36
13	Exact MM	252.45	optimal	<1
	Approx MM	260.58	optimal	<1
	Approx PCM	231.87	optimal	<1
	Approx CM	256.63	not optimal	58
14	Exact MM	278.75	optimal	<1
	Approx MM	288.48	optimal	<1
	Approx PCM	255.07	optimal	<1
	Approx CM	432.80	not optimal	43
15	Exact MM	308.65	optimal	<1
	Approx MM	316.09	optimal	<1
	Approx PCM	278.82	optimal	<1
	Approx CM	1065.87	not optimal	68
16	Exact MM	368.62	optimal	<1
	Approx MM	374.71	optimal	<1
	Approx PCM	329.23	optimal	<1
	Approx CM	1739.64	not optimal	93
17	Exact MM	372.45	optimal	<1
	Approx MM	382.84	optimal	<1
	Approx PCM	336.69	optimal	<1
	Approx CM	9931.75	not optimal	561
18	Exact MM	364.84	optimal	<1
	Approx MM	380.50	optimal	<1
	Approx PCM	333.67	optimal	<1
	Approx CM	10109.55	not optimal	689
19	Exact MM	438.97	optimal	<1
	Approx MM	455.86	optimal	<1
	Approx PCM	389.28	optimal	<1
	Approx CM	17551.99	not optimal	903
20	Exact MM	520.24	optimal	<1
	Approx MM	538.52	optimal	<1
	Approx PCM	459.48	optimal	<1
	Approx CM	36696.95	not optimal	1328

Table 2 Objective values and CPU time under different instances with increasing jobs, N.

Instance Index	Models	Objective Value	Truncated Normal	Log Normal
1	Regular LDR	1987.66	1306.37(0.58)	1306.36(0.61)
	Exact MM	22.12	11.12(0.25)	11.07(0.3)
	Approx MM	22.67	11.08(0.30)	11.16(0.31)
	Approx PCM	20.71	10.99(0.26)	10.10(0.27)
	Approx CM	19.27	10.68(0.25)	10.07(0.26)
2	Regular LDR	2089.22	1328.41(1.20)	1328.40(1.17)
	Exact MM	46.32	22.17(0.50)	22.12(0.51)
	Approx MM	47.21	22.27(0.51)	22.48(0.52)
	Approx PCM	43.06	22.09(0.49)	20.19(0.51)
	Approx CM	33.52	22.12(0.51)	19.86(0.52)
3	Regular LDR	1987.66	1307.89(0.96)	1307.88(1.01)
	Exact MM	37.03	18.75(0.41)	18.25(0.42)
	Approx MM	37.79	18.47(0.40)	18.65(0.41)
	Approx PCM	34.51	16.72(0.41)	16.90(0.42)
	Approx CM	27.40	16.67(0.40)	16.76(0.41)
4	Regular LDR	2089.22	1329.58(1.70)	1329.54(1.68)
	Exact MM	66.32	31.13(0.76)	31.36(0.78)
	Approx MM	66.88	31.55(0.73)	31.96(0.76)
	Approx PCM	61.00	28.32(0.74)	28.70(0.78)
	Approx CM	49.28	28.20(0.76)	28.60(0.78)
5	Regular LDR	2009.91	1323.02(2.52)	1323.23(2.56)
	Exact MM	89.47	42.08(0.96)	43.93(1.03)
	Approx MM	91.62	43.10(0.93)	44.09(0.99)
	Approx PCM	83.59	39.32(0.97)	40.26(1.05)
	Approx CM	72.25	38.94(0.98)	39.15(1.03)
6	Regular LDR	2100.27	1357.65(3.03)	1357.59(3.15)
	Exact MM	121.33	55.69(1.23)	56.15(1.28)
	Approx MM	123.65	56.45(1.19)	57.75(1.25)
	Approx PCM	112.33	50.66(1.18)	51.87(1.32)
	Approx CM	*108.52	48.13(1.21)	51.062(1.31)
7	Regular LDR	1979.56	1271.30(5.87)	1272.27(6.13)
	Exact MM	219.35	103.36(2.51)	107.89(3.05)
	Approx MM	222.36	105.61(2.49)	109.97(2.98)
	Approx PCM	202.51	96.49(2.52)	100.54(3.11)
	Approx CM	*190.01	95.86(2.52)	99.16(3.14)

Table 3 Results of Monte Carlo simulation under truncated normal and log normal distributions. Asterisks indicate suboptimal solutions and standard errors are reported in parenthesis .

for $t \in [T]$, where $\alpha \in [0,1]$ and the uncertain factors, $\tilde{z}_t$ are zero means and uncorrelated random variables. Hence, with $\alpha = 0$, the demands are uncorrelated. As α grows, the variances and correlation of the demands also increase.

The objective of the problem is to minimize the worst-case expected total cost over the entire horizon as follows,

$$
\begin{array}{lll}
\min \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\sum_{t=1}^{T}(c_t x_t(\tilde{\boldsymbol{z}}_{t-1}) + v_t(\tilde{\boldsymbol{z}}_t))\right) & & \\
\text{s.t. } y_{t+1}(\boldsymbol{z}_t) = y_t(\boldsymbol{z}_{t-1}) + x_t(\boldsymbol{z}_{t-1}) - d_t(\boldsymbol{z}_t) & \forall \boldsymbol{z}\in\mathcal{W}, t\in[T] & \\
\quad v_t(\boldsymbol{z}_t) \geq h_t y_{t+1}(\boldsymbol{z}_t) & \forall \boldsymbol{z}\in\mathcal{W}, t\in[T] & (34)\\
\quad v_t(\boldsymbol{z}_t) \geq -b_t y_{t+1}(\boldsymbol{z}_t) & \forall \boldsymbol{z}\in\mathcal{W}, t\in[T] & \\
\quad 0 \leq x_t(\boldsymbol{z}_{t-1}) \leq \bar{x}_t & \forall \boldsymbol{z}\in\mathcal{W}, t\in[T] & \\
\quad x_t \in \mathcal{R}^{t-1,1}, y_{t+1}\in\mathcal{R}^{t,1}, v_t\in\mathcal{R}^{t,1} & \forall t\in[T]. &
\end{array}
$$

We consider the following partial cross moment ambiguity set,

$$
\mathbb{F}_{PCM} = \left\{ \mathbb{P}\in\mathcal{P}_0(\mathbb{R}^T) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{z}}) = \boldsymbol{0} & \\ \mathbb{E}_{\mathbb{P}}\left(\left(\sum_{r=s}^{t}\tilde{z}_r\right)^2\right) \leq \phi_{st}^2 & \forall s\leq t, s,t\in[T] \\ \mathbb{P}(\tilde{\boldsymbol{z}}\in\mathbb{R}^T_+) = 1 & \end{array} \right\}, \tag{35}
$$

where the lifted ambiguity set, $\mathbb{G}$ is

$$
\mathbb{G}_{PCM} = \left\{ \mathbb{Q}\in\mathcal{P}_0(\mathbb{R}^T\times\mathbb{R}^{(T+1)T/2}) \;\middle|\; \begin{array}{ll} \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{z}}) = \boldsymbol{0} & \\ \mathbb{E}_{\mathbb{Q}}(\tilde{u}_{st}) \leq \phi_{st}^2 & \forall s\leq t, s,t\in[T] \\ \mathbb{Q}(\tilde{\boldsymbol{z}}\in\bar{\mathcal{W}}_{PCM3}) = 1 & \end{array} \right\}, \tag{36}
$$

$$
\bar{\mathcal{W}}_{PCM} = \left\{ (\boldsymbol{z},\boldsymbol{u})\in\mathbb{R}^T\times\mathbb{R}^{(T+1)T/2} \;\middle|\; \begin{array}{ll} \tilde{\boldsymbol{z}}\in\mathbb{R}^T_+ & \\ u_{st} \geq \left(\sum_{r=s}^{t} z_r\right)^2 & \forall s\leq t, s,t\in[T] \end{array} \right\}.
$$

This partial cross moment incorporates the variances of the sum of factors leading to the time period t. We let $\tilde{\boldsymbol{u}}_t = (\tilde{u}_{rs})_{1\leq r\leq s\leq t}$, $t\in[T]$, which, together with $\tilde{\boldsymbol{z}}_t$ are associated with the information available at the end of period t. Consequently, we use ROC to formulate the problem via the new LDR approximation and solve it using CPLEX.

Following the numerical study of See and Sim (2009), we set the parameters $\bar{x}_t = 260$, $c_t = 0.1$, $h_t = 0.02$, for all $t\in[T]$, $b_t = 0.2$, for all $t\in[T-1]$ and $b_T = 2$. We assume that random factors $\tilde{z}_t$ are uncorrected random variables in $[-\bar{z},\bar{z}]$ with standard deviations bounded below by $\frac{1}{\sqrt{3}}\bar{z}$. In characterizing the partial cross moment ambiguity set, we have $\mu_t = \mu$, $t\in[T]$ and

$$
\phi_{st}^2 = \frac{(t-s+1)}{3}\bar{z}^2 \qquad \forall s\leq t, s,t\in[T].
$$

Observe that iid uniformly distributed random variables in $[-\bar{z},\bar{z}]$ would be a feasible distribution in the ambiguity set and we use this to obtain a lower bound to the inventory control problem. Specifically, we investigate the performance of the multi-period inventory control problem among different approaches as follows,

• **(Lower Bound)**: A lower bound obtained by using iid uniformly distributed random factors and solving the dynamic inventory control problem to optimality (for the dynamic programming implementation, see See and Sim 2009).

• **(Approx MM)**: Solutions based on the new LDR approximation under the marginal moment ambiguity set (i.e. known mean values, upper bound on variances and nonnegative support).

• **(Approx PCM)**: Solutions based on the new LDR approximation under the partial cross moment ambiguity set, $\mathbb{F}_{PCM}$.

In Table 4, we report the objective values attained for the different approaches under various parameters. As in the previous computational study, we observe that by incorporating partial cross moment information, we can significantly improve the objectives of the adaptive distributionally robust optimization problems. Moreover, the objectives of Approx PCM are reasonably close to the lower bounds. It has well been known that despite the gaps from the lower bounds, numerical experiments on robust inventory control problems have demonstrated that the actual objectives attained in out-of-sample analysis are often closer to the optimal values than what the model objectives have reflected (see, for instance, Bertsimas and Thiele 2006, See and Sim 2009). Moreover, the benefit of distributional robustness arrises when there is disparity between the actual demand distribution and the demand distribution in which the optimal policy is derived. In such cases, the robust solution could perform significantly better than the misspecified optimum policy (see, numerical experiments of Bertsimas and Thiele 2006).

It is important to note that a realistic inventory control problem may be confounded by issues such as, multi-products, storage constraints, transshipment decisions, among others, which are often too complex to analyze and solve using conventional approaches such as dynamic programming and stochastic optimization. Notwithstanding the intractability of these problems, we can still model these problems as adaptive distributionally robust linear optimization problems and solve them to optimality in a deterministic framework using commercially available solver packages to obtain stable and consistent solutions.

6. Future work

In our numerical studies, we show the benefits of the partial cross moments ambiguity set. However, the choice of such ambiguity set appears ad hoc and it begs an interesting question as to how we can systematically adapt and improve the partial cross moments ambiguity set. Chen et al. (2016) have recently proposed a new class of infinitely constrained ambiguity sets where the number of expectation constraints could be infinite. To solve the problem, they consider a relaxed ambiguity set with finite number of expectation constraints, as in the case of the partial cross moments ambiguity set. More interestingly, for static robust optimization problems, the “violating” expectation

T	μ	$\bar{z}$	α	Lower Bound	Approx MM	Approx PCM
5	200	20	0	108.0	167.3	115.7
10	200	10	0	206.0	272.5	214.9
20	240	6	0	486.0	583.2	499.2
30	240	4	0	725.0	838.6	740.8
5	200	20	0.25	108.0	181.0	124.8
10	200	10	0.25	206.0	303.7	232.8
20	240	6	0.25	487.0	684.7	543.6
30	240	4	0.25	725.0	1028.8	811.2
5	200	20	0.50	109.0	195.1	133.6
10	200	10	0.50	207.0	335.2	250.7
20	240	6	0.50	496.0	795.1	588.4
30	240	4	0.50	732.0	1232.2	882.9

Table 4 Objective values of the various models under different instances.

constraint can be identified and added to the relaxed ambiguity set to improve the solution. While the approach works for static distributionally robust optimization problems, the extension to adaptive problems has not been studied. We believe this is an important extension of this framework that will help us model and solve adaptive distributiuoanlly robust optimization problems for a larger variety of ambiguity sets.

A. Improvement over deflected LDR (DLDR) approximation

Chen et al. (2008), Goh and Sim (2010) propose a class of of piecewise LDR approximation known as deflected LDR (DLDR) approximation, which can also circumvent the issues of infeasibility in complete recourse problems. The approach requires to solve a set of subproblems given by

$$\begin{aligned} f_i^* = \min\ & \boldsymbol{d}'\boldsymbol{y} \\ \text{s.t.}\ & \boldsymbol{B}\boldsymbol{y} = \boldsymbol{q} \\ & \boldsymbol{q} \geq \boldsymbol{e}_i \\ & \boldsymbol{y} \in \mathbb{R}^{N_2}, \boldsymbol{q} \in \mathbb{R}^M, \end{aligned} \tag{37}$$

for all $i \in [M]$, which are not necessarily feasible optimization problems. Let $\mathcal{M} \subseteq [M]$ denote the subset of indices in which their corresponding subproblems are feasible, i.e., $\mathcal{M} = \{i \in [M] \mid f_i^* < \infty\}$, and $\bar{\mathcal{M}} = [M] \setminus \mathcal{M}$. Correspondingly, let $(\bar{\boldsymbol{y}}_i, \bar{\boldsymbol{q}}_i)$ be the optimal solution of Problem (37) for all $i \in \mathcal{M}$. Here, $f_i^* = \boldsymbol{d}'\bar{\boldsymbol{y}}_i \geq 0$, $i \in \mathcal{M}$ is assumed or otherwise, $Q(\boldsymbol{x}, \boldsymbol{z})$ would be unbounded from below. The solution to deflected linear decision is obtained by solving the following optimization problem,

$$\begin{aligned} \beta_D(\boldsymbol{x}) = \min\ & \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}(\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})) + \sum_{i\in\mathcal{M}} f_i^* \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}((-q_i(\tilde{\boldsymbol{z}}))^+) \\ \text{s.t.}\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{b}(\boldsymbol{z}) + \boldsymbol{q}(\boldsymbol{z}) && \forall \boldsymbol{z} \in \mathcal{W} \\ & q_i(\boldsymbol{z}) \geq 0 && \forall i \in \bar{\mathcal{M}}, \forall \boldsymbol{z} \in \mathcal{W} \\ & \boldsymbol{y} \in \mathcal{L}^{N_2} \\ & \boldsymbol{q} \in \mathcal{L}^M. \end{aligned} \tag{38}$$

Suppose $(\boldsymbol{y}^*, \boldsymbol{q}^*)$ is the optimal solution of Problem (38), the corresponding decision rule under DLDR approximation is given by

$$\boldsymbol{y}_D(\boldsymbol{z}) = \boldsymbol{y}^*(\boldsymbol{z}) + \sum_{i\in\mathcal{M}} \bar{\boldsymbol{y}}_i((-q_i^*(\boldsymbol{z}))^+).$$

Chen et al. (2008), Goh and Sim (2010) show that $\boldsymbol{y}_D(\tilde{\boldsymbol{z}})$ is a feasible solution of Problem (13). Moreover,

$$\beta(\boldsymbol{x}) \leq \sup_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(\boldsymbol{d}'\boldsymbol{y}_D(\tilde{\boldsymbol{z}})\right) \leq \beta_D(\boldsymbol{x}) \leq \beta_L(\boldsymbol{x}).$$

Our next result shows that the new LDR approximation may improve the bound provided by the DLDR approximation.

THEOREM 5. *The bound obtained via the new LDR approximation is no larger then the bound obtained via DLDR approximation, i.e.,*

$$\beta_E(\boldsymbol{x}) \leq \beta_D(\boldsymbol{x}).$$

Moreover, there exists instances such that $\beta_E(\boldsymbol{x}) < \beta_D(\boldsymbol{x})$.

Proof. Following similar exposition of Theorem 1, we have the equivalent form of $\beta_D(\boldsymbol{x})$ as follows:

$$\begin{array}{rlr}\beta_D(\boldsymbol{x}) = \min & r_0 + \boldsymbol{s}_0'\boldsymbol{\mu} + \boldsymbol{t}_0'\boldsymbol{\sigma} + \sum_{i\in\mathcal{M}} f_i^*(r_i + \boldsymbol{s}_i'\boldsymbol{\mu} + \boldsymbol{t}_i'\boldsymbol{\sigma}) & \\ \text{s.t.} & r_0 + \boldsymbol{s}_0'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_0'\boldsymbol{u} \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z}) & \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & r_i + \boldsymbol{s}_i'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i'\boldsymbol{u} \geq -q_i(\boldsymbol{z}) & \forall i \in \mathcal{M}, \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & r_i + \boldsymbol{s}_i'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i'\boldsymbol{u} \geq 0 & \forall i \in \mathcal{M}, \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} & \forall i \in \{0\}\cup\mathcal{M} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{b}(\boldsymbol{z}) + \boldsymbol{q}(\boldsymbol{z}) & \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & q_i(\boldsymbol{z}) \geq 0 & \forall i \in \bar{\mathcal{M}} \\ & r_i \in \mathbb{R}, \boldsymbol{s}_i \in \mathbb{R}^{L_1}, \boldsymbol{t}_i \in \mathbb{R}^{I_2} & \forall i \in \{0\}\cup\mathcal{M} \\ & \boldsymbol{y} \in \mathcal{L}^{N_2} & \\ & \boldsymbol{q} \in \mathcal{L}^{M}. & \end{array} \tag{39}$$

Similarly, we have the equivalent form of $\beta_E(\boldsymbol{x})$ as follows:

$$\begin{array}{rlr}\beta_E(\boldsymbol{x}) = \min & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} & \\ \text{s.t.} & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) & \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) \geq \boldsymbol{b}(\boldsymbol{z}) & \forall(\boldsymbol{z},\boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \boldsymbol{t} \geq \boldsymbol{0} & \\ & r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{I_2} & \\ & \boldsymbol{y} \in \bar{\mathcal{L}}^{N_2}. & \end{array} \tag{40}$$

Let $\boldsymbol{y}^\dagger$, $\boldsymbol{q}^\dagger$, $r_i^\dagger, \boldsymbol{s}_i^\dagger, \boldsymbol{t}_i^\dagger$, $i \in \{0\}\cup\mathcal{M}$ be a feasible solution of Problem (39). We will show that there exists a corresponding feasible solution for Problem (40) with the same objective value. Let

$$\begin{aligned} r &= r_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}\bar{\boldsymbol{y}}_i r_i^\dagger \\ \boldsymbol{s} &= \boldsymbol{s}_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{s}_i^\dagger \\ \boldsymbol{t} &= \boldsymbol{t}_0^\dagger + \sum_{i\in\mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{t}_i^\dagger, \\ \boldsymbol{y}(\boldsymbol{z},\boldsymbol{u}) &= \boldsymbol{y}^\dagger(\boldsymbol{z}) + \sum_{i\in\mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right)\bar{\boldsymbol{y}}_i. \end{aligned}$$

Observe that the objective value of Problem (40) becomes

$$\begin{aligned} r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} &= r_0^\dagger + \boldsymbol{s}_0^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_0^{\dagger\prime}\boldsymbol{\sigma} + \sum_{i\in\mathcal{M}} (r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{\sigma})\boldsymbol{d}'\bar{\boldsymbol{y}}_i \\ &= r_0^\dagger + \boldsymbol{s}_0^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_0^{\dagger\prime}\boldsymbol{\sigma} + \sum_{i\in\mathcal{M}} f_i^*(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}\boldsymbol{\mu} + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{\sigma}). \end{aligned}$$

We next check the feasibility of the solution in Problem (40). Note that $\boldsymbol{t} \geq \boldsymbol{0}$ and for all $(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$,

$$
\begin{aligned}
r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} &= r_0^\dagger + \sum_{i \in \mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i r_i^\dagger + \left(\boldsymbol{s}_0^\dagger + \sum_{i \in \mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{s}_i^\dagger\right)'(\boldsymbol{G}\boldsymbol{z}) + \left(\boldsymbol{t}_0^\dagger + \sum_{i \in \mathcal{M}} \boldsymbol{d}'\bar{\boldsymbol{y}}_i \boldsymbol{t}_i^\dagger\right)' \boldsymbol{u} \\
&= r_0^\dagger + \boldsymbol{s}_0^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_0^{\dagger\prime}\boldsymbol{u} + \sum_{i \in \mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \boldsymbol{d}'\bar{\boldsymbol{y}}_i \\
&\geq \boldsymbol{d}'\boldsymbol{y}^\dagger(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \boldsymbol{d}'\bar{\boldsymbol{y}}_i \\
&= \boldsymbol{d}'\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}),
\end{aligned}
$$

where the inequality follows from the first robust counterpart constraint in Problem (39). We now show the feasibility of second robust robust counterpart constraint in Problem (40). Observe that for all $(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$,

$$
\begin{aligned}
\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) &= \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}^\dagger(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \boldsymbol{B}\bar{\boldsymbol{y}}_i \\
&= \boldsymbol{b}(\boldsymbol{z}) + \boldsymbol{q}^\dagger(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \bar{\boldsymbol{q}}_i \\
&= \boldsymbol{b}(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} q_i^\dagger(\boldsymbol{z})\boldsymbol{e}_i + \sum_{j \in \bar{\mathcal{M}}} q_j^\dagger(\boldsymbol{z})\boldsymbol{e}_j + \sum_{i \in \mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \bar{\boldsymbol{q}}_i \\
&\geq \boldsymbol{b}(\boldsymbol{z}) + \sum_{i \in \mathcal{M}} q_i^\dagger(\boldsymbol{z})\boldsymbol{e}_i + \sum_{j \in \bar{\mathcal{M}}} q_j^\dagger(\boldsymbol{z})\boldsymbol{e}_j + \sum_{i \in \mathcal{M}} \left(r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \boldsymbol{e}_i \\
&= \boldsymbol{b}(\boldsymbol{z}) + \sum_{j \in \bar{\mathcal{M}}} q_j^\dagger(\boldsymbol{z})\boldsymbol{e}_j + \sum_{i \in \mathcal{M}} \left(q_i^\dagger(\boldsymbol{z}) + r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u}\right) \boldsymbol{e}_i \\
&\geq \boldsymbol{b}(\boldsymbol{z}).
\end{aligned}
$$

The first inequality holds because $\bar{\boldsymbol{q}}_i \geq \boldsymbol{e}_i$ and $r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u} \geq 0$ for all $i \in \mathcal{M}$, $(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$. The second inequality is due to $r_i^\dagger + \boldsymbol{s}_i^{\dagger\prime}(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}_i^{\dagger\prime}\boldsymbol{u} \geq -q_i^\dagger(\boldsymbol{z})$ for all $i \in \mathcal{M}$, $(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$ and $q_i^\dagger(\boldsymbol{z}) \geq 0$ for all $i \in \bar{\mathcal{M}}$, $(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}$.

To show that in general $\beta_E(\boldsymbol{x}) \neq \beta_D(\boldsymbol{x})$, we consider the following complete recourse problem,

$$
\begin{aligned}
\beta^* = \min\ & \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}\left(y(\tilde{\boldsymbol{z}})\right) \\
\text{s.t.}\ & y(\boldsymbol{z}) \geq z_i \qquad \forall \boldsymbol{z} \in \mathbb{R}^3, i \in \{1,2,3\} \\
& y \in \mathcal{R}^{1,1}.
\end{aligned}
\tag{41}
$$

where

$$
\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^3) \;\middle|\; \begin{array}{l} \mathbb{E}_{\mathbb{P}}(\boldsymbol{z}) = \boldsymbol{0} \\ \mathbb{E}_{\mathbb{P}}\left(\sum_{k=1}^3 \tilde{z}_k^2\right) \leq 1 \end{array} \right\}.
$$

From Theorem 4, we can obtain the optimal objective using the new LDR approach, which is $\beta_E = \beta^* = 0.8165$. On the other hand, under the DLDR approximation, the objective value is $\beta_D = 1.154$, which is significantly higher than β_E. $\square$

B. Ambiguity set of Wiesemann et al. (2014)

In this section, we consider a more general ambiguity set based on Wiesemann et al. (2014) and derive similar results presented in the paper.

DEFINITION 4. A WKS ambiguity set, $\mathbb{F}$ is an ambiguity set of probability distributions that can be expressed as

$$\mathbb{F} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_1}\right) : \begin{array}{l} \tilde{\boldsymbol{z}} \in \mathbb{R}^{I_1} \\ \mathbb{E}_{\mathbb{P}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{P}}(\boldsymbol{g}(\tilde{\boldsymbol{z}})) \preceq_{\mathcal{K}_0} \boldsymbol{\sigma} \\ \mathbb{P}(\tilde{\boldsymbol{z}} \in \mathcal{W}) = 1 \end{array} \right\} \tag{42}$$

with $\boldsymbol{G} \in \mathbb{R}^{L_1 \times I_1}$, $\boldsymbol{\mu} \in \mathbb{R}^{L_1}$, $\boldsymbol{\sigma} \in \mathbb{R}^{L_2}$, $\boldsymbol{g} \in \mathcal{R}^{I_1, L_2}$ and proper cone $\mathcal{K}_0 \subseteq \mathbb{R}^{L_2}$, where the support set $\mathcal{W}$ and the $\mathcal{K}_0$-epigraph of the function $\boldsymbol{g}$,

$$\mathcal{G} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} : \boldsymbol{g}(\boldsymbol{z}) \preceq_{\mathcal{K}_0} \boldsymbol{u} \right\}$$

are both conic representable.

The corresponding lifted ambiguity set, $\mathbb{G}$ is as follows:

$$\mathbb{G} = \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2}\right) : \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \\ \mathbb{E}_{\mathbb{Q}}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu} \\ \mathbb{E}_{\mathbb{Q}}(\tilde{\boldsymbol{u}}) \preceq_{K_0} \boldsymbol{\sigma} \\ \mathbb{Q}\left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}\right) = 1 \end{array} \right\}, \tag{43}$$

where $\bar{\mathcal{W}}$ is the joint support set of the actual and auxiliary random variables defined as

$$\bar{\mathcal{W}} = \{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathcal{G} : \boldsymbol{z} \in \mathcal{W} \}. \tag{44}$$

To obtain explicit formulations of our distributional robust optimization models, we express joint support set as

$$\bar{\mathcal{W}} = \left\{ (\boldsymbol{z}, \boldsymbol{u}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} : \exists \boldsymbol{v} \in \mathbb{R}^{I_3}, \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} + \boldsymbol{E}\boldsymbol{v} \preceq_{\mathcal{K}} \boldsymbol{h} \right\}, \tag{45}$$

with $\boldsymbol{C} \in \mathbb{R}^{L_3 \times I_1}$, $\boldsymbol{D} \in \mathbb{R}^{L_3 \times I_2}$, $\boldsymbol{E} \in \mathbb{R}^{L_3 \times I_3}$, $\boldsymbol{h} \in \mathbb{R}^{L_3}$ and $\mathcal{K} \subseteq \mathbb{R}^{L_3}$ being a proper cone. We also impose the following Slater's like conditions:

ASSUMPTION 2. *There exists* $(\boldsymbol{z}^\dagger, \boldsymbol{u}^\dagger, \boldsymbol{v}^\dagger) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}$ *such that*

$$\begin{array}{l} \boldsymbol{G}\boldsymbol{z}^\dagger = \boldsymbol{\mu} \\ \boldsymbol{u}^\dagger \prec_{\mathcal{K}_0} \boldsymbol{\sigma} \\ \boldsymbol{C}\boldsymbol{z}^\dagger + \boldsymbol{D}\boldsymbol{u}^\dagger + \boldsymbol{E}\boldsymbol{v}^\dagger \prec_{\mathcal{K}} \boldsymbol{h}. \end{array}$$

THEOREM 6. *Suppose the second stage problem (1) has relatively complete recourse, then under the WKS ambiguity set, Problem (2) is equivalent to the following problem,*

$$\begin{array}{rl} \beta(\boldsymbol{x}) = & \inf\ r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\ & \text{s.t. } r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq \max_{\boldsymbol{p} \in \mathcal{P}} \boldsymbol{p}'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \ \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\ & \quad \boldsymbol{t} \succeq_{\mathcal{K}_0^*} \boldsymbol{0} \\ & \quad r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{L_2}, \end{array} \tag{46}$$

where $\mathcal{P}$ is the dual feasible set of the Problem (1) given by

$$\mathcal{P} = \left\{ \boldsymbol{p} \in \mathbb{R}^M_+ : \boldsymbol{B}'\boldsymbol{p} = \boldsymbol{d} \right\}.$$

In particular, its explicit formulation is given by

$$\begin{aligned}
\beta(\boldsymbol{x}) = \min\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.}\ & r \geq \boldsymbol{\pi}_i{}'\boldsymbol{h} + \boldsymbol{p}_i{}'\boldsymbol{b}^0 - \boldsymbol{p}_i{}'\boldsymbol{A}^0\boldsymbol{x} && \forall i \in [P] \\
& \boldsymbol{C}'\boldsymbol{\pi}_i = \begin{bmatrix} \boldsymbol{p}_i'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_i'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} && \forall i \in [P] \\
& \boldsymbol{D}'\boldsymbol{\pi}_i = -\boldsymbol{t} && \forall i \in [P] \\
& \boldsymbol{E}'\boldsymbol{\pi}_i = \boldsymbol{0} && \forall i \in [P] \\
& \boldsymbol{\pi}_i \succeq_{\mathcal{K}^*} \boldsymbol{0} && \forall i \in [P] \\
& \boldsymbol{t} \succeq_{\mathcal{K}_0^*} \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{L_2} \\
& \boldsymbol{\pi}_i \in \mathbb{R}^{L_3} && \forall i \in [P].
\end{aligned} \tag{47}$$

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_P$ are the extreme points of $\mathcal{P}$.

Proof. Observe that

$$\beta(\boldsymbol{x}) = \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(Q(\boldsymbol{x}, \boldsymbol{z}) \right) = \sup_{\mathbb{Q} \in \mathbb{G}} \mathbb{E}_{\mathbb{P}} \left(Q(\boldsymbol{x}, \boldsymbol{z}) \right).$$

By weak duality of Isii (1962), we have

$$\begin{aligned}
\beta(\boldsymbol{x}) \leq \beta_1(\boldsymbol{x}) = \inf\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.}\ & r + \boldsymbol{s}'(\boldsymbol{G}\boldsymbol{z}) + \boldsymbol{t}'\boldsymbol{u} \geq Q(\boldsymbol{x}, \boldsymbol{z})\ \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}} \\
& \boldsymbol{t} \succeq_{\mathcal{K}_0^*} \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{L_2},
\end{aligned}$$

where $r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{L_2}$ are the dual variables corresponding to the expectation constraints of $\mathbb{G}$. Since $Q(\boldsymbol{x}, \boldsymbol{z})$ is finite, by strong duality of linear optimization, we have equivalently

$$Q(\boldsymbol{x}, \boldsymbol{z}) = \max_{\boldsymbol{p} \in \mathcal{P}} \boldsymbol{p}'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) = \max_{p \in [P]} \left\{ \boldsymbol{p}_p'(\boldsymbol{b}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}) \right\}.$$

Therefore,

$$\begin{aligned}
\beta_1(\boldsymbol{x}) = \inf\ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t.}\ & r \geq \sup_{(\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{W}}} \left\{ \left(\begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \right)' \boldsymbol{z} - \boldsymbol{t}'\boldsymbol{u} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) \right\}\ \forall p \in [P] \\
& \boldsymbol{t} \succeq_{\mathcal{K}_0^*} \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{L_2}.
\end{aligned} \tag{48}$$

By weak conic duality (see, for instance, Ben-Tal and Nemirovski (2001a)), we have for all $p \in [P]$,

$$\begin{aligned}
&\sup_{(\boldsymbol{z},\boldsymbol{u})\in\bar{\mathcal{W}}} \left\{ \left(\begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \right)' \boldsymbol{z} - \boldsymbol{t}'\boldsymbol{u} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) \right\} \\
\le \quad & \inf \ \boldsymbol{\pi}_p{}'\boldsymbol{h} + \boldsymbol{p}_i'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) \\
& \text{s.t. } \boldsymbol{C}'\boldsymbol{\pi}_p = \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} \\
& \quad\ \ \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} \\
& \quad\ \ \boldsymbol{E}'\boldsymbol{\pi}_p = \boldsymbol{0} \\
& \quad\ \ \boldsymbol{\pi}_p \succeq_{\mathcal{K}^*} \boldsymbol{0} \\
& \quad\ \ \boldsymbol{\pi}_p \in \mathbb{R}^{L_3},
\end{aligned}$$

where $\boldsymbol{\pi}_p \in \mathbb{R}^{L_3}, \forall p \in [P]$ are the dual variables associated with the conic constants in $\bar{\mathcal{W}}$. Hence, using standard robust counterpart techniques, we substitute the dual formulations in Problem (9) to yield the following compact conic optimization problem

$$\begin{aligned}
\beta_2(\boldsymbol{x}) = \inf \ & r + \boldsymbol{s}'\boldsymbol{\mu} + \boldsymbol{t}'\boldsymbol{\sigma} \\
\text{s.t. } & r \ge \boldsymbol{\pi}_p{}'\boldsymbol{h} + \boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x}) && \forall p \in [P] \\
& \boldsymbol{C}'\boldsymbol{\pi}_p = \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix} - \boldsymbol{G}'\boldsymbol{s} && \forall p \in [P] \\
& \boldsymbol{D}'\boldsymbol{\pi}_p = -\boldsymbol{t} && \forall p \in [P] \\
& \boldsymbol{E}'\boldsymbol{\pi}_p = 0 && \forall p \in [P] \\
& \boldsymbol{\pi}_p \succeq_{\mathcal{K}^*} 0 && \forall p \in [P] \\
& \boldsymbol{t} \succeq_{\mathcal{K}_0^*} \boldsymbol{0} \\
& r \in \mathbb{R}, \boldsymbol{s} \in \mathbb{R}^{L_1}, \boldsymbol{t} \in \mathbb{R}^{L_2} \\
& \boldsymbol{\pi}_p \in \mathbb{R}^{L_3} && \forall p \in [P].
\end{aligned} \tag{49}$$

Observe that $\beta(\boldsymbol{x}) \le \beta_1(\boldsymbol{x}) \le \beta_2(\boldsymbol{x})$, and our goal is to establish strong duality by showing $\beta_2(\boldsymbol{x}) \le \beta(\boldsymbol{x})$. Then we will next approach Problem (11) by taking the dual, which is

$$\begin{aligned}
\beta_3(\boldsymbol{x}) = \sup \ & \sum_{p\in[P]} \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0\boldsymbol{x})\alpha_p + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1\boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1}\boldsymbol{x}) \end{bmatrix}' \bar{\boldsymbol{z}}_p \right) \\
\text{s.t. } & \sum_{p\in[P]} \alpha_p = 1 \\
& \alpha_p \ge 0 && \forall p \in [P] \\
& \sum_{p\in[P]} \boldsymbol{G}\bar{\boldsymbol{z}}_p = \boldsymbol{\mu} \\
& \sum_{p\in[P]} \bar{\boldsymbol{u}}_p \preceq_{\mathcal{K}_0} \boldsymbol{\sigma} \\
& \boldsymbol{C}\bar{\boldsymbol{z}}_p + \boldsymbol{D}\bar{\boldsymbol{u}}_p + \boldsymbol{E}\bar{\boldsymbol{v}}_p \preceq_{\mathcal{K}} \alpha_p\boldsymbol{h} && \forall p \in [P] \\
& \alpha_p \in \mathbb{R}, \bar{\boldsymbol{z}}_p \in \mathbb{R}^{I_1}, && \forall p \in [P] \\
& \bar{\boldsymbol{u}}_p \in \mathbb{R}^{I_2}, \bar{\boldsymbol{v}}_p \in \mathbb{R}^{I_3} && \forall p \in [P],
\end{aligned} \tag{50}$$

where $\alpha_p, \bar{\boldsymbol{z}}, \bar{\boldsymbol{u}}_p, \bar{\boldsymbol{v}}_p, \forall p \in [P]$ are the dual variables associated with the specified constraints respectively. Since $(\boldsymbol{z}^\dagger, \boldsymbol{u}^\dagger, \boldsymbol{v}^\dagger) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}$ satisfy the conditions in Assumption 2, we can construct a strictly feasible solution

$$\alpha_p = \frac{1}{P}, \bar{\boldsymbol{z}}_p = \frac{\boldsymbol{z}^\dagger}{P}, \bar{\boldsymbol{u}}_p = \frac{\boldsymbol{u}^\dagger}{P}, \bar{\boldsymbol{v}}_p = \frac{\boldsymbol{v}^\dagger}{P},$$

for all $p \in [P]$. Hence, since Problem (12) is strictly feasible and, as we will subsequently show, is also bounded from above, strong duality holds and $\beta_2(\boldsymbol{x}) = \beta_3(\boldsymbol{x})$. Moreover, there exists a sequence of strictly feasible or interior solutions

$$\left\{(\alpha_p^k, \bar{\boldsymbol{z}}_p^k, \bar{\boldsymbol{u}}_p^k, \bar{\boldsymbol{v}}_p^k)_{p \in [P]}\right\}_{k \geq 0}$$

such that

$$\lim_{k \to \infty} \sum_{p \in [P]} \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0 \boldsymbol{x}) \alpha_p^k + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1 \boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1} \boldsymbol{x}) \end{bmatrix}' \bar{\boldsymbol{z}}_p^k \right) = \beta_3(\boldsymbol{x}).$$

Observe that for all k, $\alpha_p^k > 0$, $\sum_{p \in [P]} \alpha_p^k = 1$ and we can construct a sequence of discrete probability distributions $\{\mathbb{Q}_k \in \mathcal{P}_0(\mathbb{R}^{I_1} \times \mathbb{R}^{I_2})\}_{k \geq 0}$ on random variable $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) \in \mathbb{R}^{I_1} \times \mathbb{R}^{I_2} \times \mathbb{R}^{I_3}$ such that

$$\mathbb{Q}_k \left((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}}) = \left(\frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k}, \frac{\bar{\boldsymbol{u}}_p^k}{\alpha_p^k}, \frac{\bar{\boldsymbol{v}}_p^k}{\alpha_p^*} \right) \right) = \alpha_p^k \qquad \forall p \in [P].$$

Observe that,

$$\mathbb{E}_{\mathbb{Q}_k}(\boldsymbol{G}\tilde{\boldsymbol{z}}) = \boldsymbol{\mu}, \mathbb{E}_{\mathbb{Q}_k}(\tilde{\boldsymbol{u}}) \preceq_{\mathcal{K}_0} \boldsymbol{\sigma}, \mathbb{Q}_k((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{W}}) = 1,$$

and hence $\mathbb{Q}_k \in \mathbb{G}$ for all k. Moreover,

$$\begin{aligned}
\beta_3(\boldsymbol{x}) &= \lim_{k \to \infty} \sum_{p \in [P]} \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0 \boldsymbol{x}) \alpha_p^k + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1 \boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1} \boldsymbol{x}) \end{bmatrix}' \bar{\boldsymbol{z}}_p^k \right) \\
&= \lim_{k \to \infty} \sum_{p \in [P]} \alpha_p^k \left(\boldsymbol{p}_p'(\boldsymbol{b}^0 - \boldsymbol{A}^0 \boldsymbol{x}) + \begin{bmatrix} \boldsymbol{p}_p'(\boldsymbol{b}^1 - \boldsymbol{A}^1 \boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_p'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1} \boldsymbol{x}) \end{bmatrix}' \frac{\bar{\boldsymbol{z}}_p^k}{\alpha_p^k} \right) \\
&\leq \lim_{k \to \infty} \sum_{p \in [P]} \alpha_p^k \left(\max_{q \in [P]} \left\{ \boldsymbol{p}_q'(\boldsymbol{b}^0 - \boldsymbol{A}^0 \boldsymbol{x}) + \begin{bmatrix} \boldsymbol{p}_q'(\boldsymbol{b}^1 - \boldsymbol{A}^1 \boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_q'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1} \boldsymbol{x}) \end{bmatrix}' \frac{\bar{\boldsymbol{z}}_q^k}{\alpha_q^k} \right\} \right) \\
&= \lim_{k \to \infty} \mathbb{E}_{\mathbb{Q}_k} \left(\max_{q \in [P]} \left\{ \boldsymbol{p}_q'(\boldsymbol{b}^0 - \boldsymbol{A}^0 \boldsymbol{x}) + \begin{bmatrix} \boldsymbol{p}_q'(\boldsymbol{b}^1 - \boldsymbol{A}^1 \boldsymbol{x}) \\ \vdots \\ \boldsymbol{p}_q'(\boldsymbol{b}^{I_1} - \boldsymbol{A}^{I_1} \boldsymbol{x}) \end{bmatrix}' \tilde{\boldsymbol{z}} \right\} \right) \\
&\leq \sup_{\mathbb{Q} \in \mathbb{G}} \mathbb{E}_{\mathbb{Q}}(Q(\boldsymbol{x}, \boldsymbol{z})) \\
&= \beta(\boldsymbol{x}).
\end{aligned}$$

Hence, $\beta(\boldsymbol{x}) \leq \beta_1(\boldsymbol{x}) \leq \beta_2(\boldsymbol{x}) = \beta_3(\boldsymbol{x}) \leq \beta(\boldsymbol{x})$, and strong duality holds. Since $\beta(\boldsymbol{x})$ is finite, Problem (12) is bounded from above and hence, Problem (11) also solvable. □

References

Alizadeh, F., D. Goldfarb (2003) Second-order cone programming. *Math. Programming* 95(1):2–51.

Bajeux-Besnainou, Isabelle, Roland Portait (1998) Dynamic asset allocation in a mean-variance framework. *Management Science* 44.11-part-2 (1998): S79-S95.

Basak, Suleyman, Georgy Chabakauri (2010) Dynamic mean-variance asset allocation. *Review of financial Studies* 23(8): 2970–3016.

Ben-Tal, A., A. Nemirovski (1998) Robust convex optimization. *Math. Oper. Res.* 23(4):769–805.

Ben-Tal, A., A. Nemirovski (1999) Robust solutions of uncertain linear programs. *Oper. Res. Lett* 25(1, Auguest):1–13.

Ben-Tal, A., A. Nemirovski (2000) Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming Ser. A* 88(3):411–424.

Ben-Tal, A., A. Nemirovski (2001a) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* SIAM.

Ben-Tal, A., A. Nemirovski (2001b) On polyhedral approximations of the second-order cone *Mathematics of Operations Research* 26:193–205

Ben-Tal, A., A. Goryashko, E.Guslitzer, A. Nemirovski (2004) Ajustable robust solutions of uncertain linear programs. *Math. Programming* 99:351-376.

Bertsimas, D., D. A. Iancu, P. A. Parrilo (2010) Optimality of Affine Policies in Multistage Robust Optimization. *Mathematics of Operations Research* 35(2):363–394.

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

Bertsimas, D., Brown, D. B., Caramanis, C. (2011). Theory and applications of robust optimization. SIAM review, 53(3), 464-501.

Bertsimas, D., Popescu, I. (2005) Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* 15(3):780-804.

Bertsimas, D., M. Sim (2004) The price of robustness. *Oper. Res.* 52(1):35–53.

Bertsimas, D., A. Thiele (2006) A robust optimization approach to inventory theory. *Oper. Res.* 54(1):150–168.

Birge, J. R., F. Louveaux (1997) Introduction to Stochastic Programming. Springer, New York.

Breton, M., S. El Hachem (1995) Algorithms for the solution of stochastic dynamic minimax problems. *Comput. Optim. Appl.* 4:317–345.

Chen, W., M. Sim (2009) Goal-driven optimization. *Operations Research* 57(2)(2):342–357.

Chen, Z., M. Sim., H. Xu (2016) Distributionally Robust Optimization with Infinitely Constrained Ambiguity Sets *Optimization online*

Chen, X., Y. Zhang (2009) Uncertain linear programs: Extended affinely adjustable robust counterparts. *Oper. Res.* 57(6):1469–1482.

Chen, X., M. Sim, P. Sun (2007) A robust optimization perspective on stochastic programming. *Operations Research*, 55(6):1058–1071.

Chen, X., M. Sim, P. Sun, J. Zhang (2008) A linear decision-based approximation approach to stochastic programming. *Oper. Res.* 56(2):344–357.

Delage, E., Dan Iancu (2015) Robust multi-stage decision making. *INFORMS Tutorials in Operations Research* (2015): 20–46.

Delage, E., Y. Ye (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):596–612.

Dupacova, J.(1987) The minimax approach to stochastic programming and an illustrative application. *Stochastics* 20(1):73–88.

Dyer, M., L. Stougie (2006) Computational complexity of stochastic programming problems. *Math. Programming Ser. A* 106:423–432.

Ellsberg, D. (1961) Risk, ambiguity and the Savage axioms. *Quarterly Journal of Economics*, 75(4), pp. 643-669.

Garstka, S. J., R. J.-B. Wets (1974) On decision rules in stochastic programming. *Math. Programming* 7(1):117-143.

Ghaoui, El. L., H. Lebret (1997) Robust solutions to least-squares problems with uncertain data. *SIAM J.Matrix Anal. Appl.*18(4):1035–1064.

Ghaoui, El. L., F. Oustry, H. Lebret (1998) Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* 9:33–53.

Gilboa, I., D. Schmeidler (1989) Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2):141–153.

Goh, J., M. Sim (2009) Robust optimization made easy with ROME. *Oper. Res.* 59(4):973–985.

Goh, J., M. Sim (2010) Distributionally robust optimization and its tractable approximations. *Oper. Res.* 58(4):902–917.

Graves, S. C. (1999) A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* 1(1):50–61.

Hsu M., M. Bhatt, R. Adolphs, D. Tranel, C.F. Camerer. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310 1680–1683.

Isii, K. (1962) On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics* 14(1):185–197.

Knight, F. H. (1921) Risk, uncertainty and profit. *Hart, Schaffner and Marx.*

Kong, Q., Lee, C. Y., Teo, C. P., Zheng, Z. (2013). Scheduling arrivals to a stochastic service delivery system using copositive cones. *Oper. Res.* 61(3): 711726.

Kuhn, D., W. Wiesemann, A. Georghiou (2011) Primal and dual linear decision rules in stochastic and robust optimization. *Math. Programming* 130(1):177–209.

Kydland, Finn E., Edward C. Prescott (1977) Rules rather than discretion: The inconsistency of optimal plans. *The journal of political Economy*: 473-491.

Lobo, M., L. Vandenberghe, S. Boyd, H. Lebret (1998) Applications of second-order cone programming. *Linear Algebra and its Applications* 284(1-3):193–228.

Löberg, J. (2012) Automatic robust convex programming. *Optimization methods and software* 27(1): 115–129.

Mak, H. Y., Rong, Y., Zhang, J. (2014) Appointment scheduling with limited distributional information. *Management Science* 61(2): 316-334.

F. Meng, J. Qi, M. Zhang, J. Ang, S. Chu, M. Sim (2015) A robust optimization model for managing elective admission in a public hospital. *Operations Research* 63(6):1452–1467.

Popescu, I. (2007) Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* 55(4):98–112.

J. Qi (2015) Mitigating delays and unfairness in appointment systems. *Forthcoming in Management Science.*

Richardson, Henry R (1989) A minimum variance result in continuous trading portfolio optimization. *Management Science* 35(9): 1045–1055.

Ruszczyński, A., A. Shapiro (2003) Stochastic Programming. *Handbooks in Operations Research and Management Science* 10. Elsevier Science, Amsterdam.

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

Scarf, H. (1958) A min-max solution of an inventory problem. K. Arrow, ed. *Studies in the Mathematical Theory of Inventory and Production.* Stanford University Press, Stanford, CA, 201–209.

See, C.-T., M. Sim (2009) Robust approximation of multiperiod inventory management. *Oper. Res.* 58(3):583–594.

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

Shapiro, A., A. Nemirovski (2005) On complexity of stochastic programming problems. V. Jeyakumar, A. Rubinov, eds. *Continuous Optimization.* Springer, New York, 111–146.

Tutuncu, R.H., K.C. Toh, and M.J. Todd (2003) Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming Ser. B* 95:189–217.

Wiesemann, W., D. Kuhn, M. Sim (2014) Distributionally Robust Convex Optimization. *Operations Research* 62(6): 1358–1376.

Xin, L., DA. Goldberg, and A. Shapiro (2015) Time (in)consistency of multistage distributionally robust inventory models with moment constraints. https://arxiv.org/abs/1304.3074.

Xu, H., S. Mannor (2012) Distributionally robust Markov decision processes. *Mathematics of Operations Research* 37(2):288–300.

Žáčková, J . (1966) On minimax solution of stochastic linear programming problems. *Časopis pro Pěstování Matematiky*, 91:423–430.

Zheng, J., M. Sim, den Hertog, D. (2016) Adjustable Robust Optimization via Fourier-Motzkin Elimination *Optimization online*