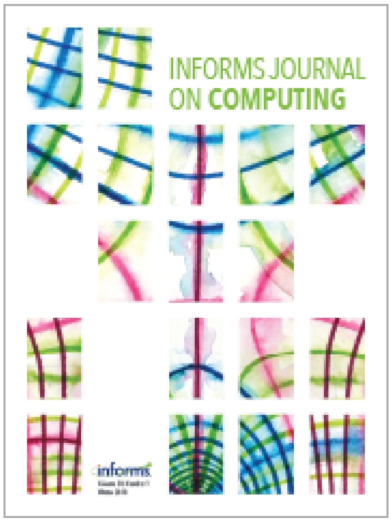


INFORMS Journal on Computing

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Decision Rule Approach for Two-Stage Data-Driven Distributionally Robust Optimization Problems with Random Recourse

Xiangyi Fan, Grani A. Hanasusanto

INFORMS JOURNAL ON COMPUTING
Articles in Advance, pp. 1–17
https://pubsonline.informs.org/journal/ijoc
ISSN 1091-9856 (print), ISSN 1526-5528 (online)

A Decision Rule Approach for Two-Stage Data-Driven Distributionally Robust Optimization Problems with Random Recourse

Xiangyi Fan,[a,*] **Grani A. Hanasusanto**[b]

[a] Graduate Program in Operations Research and Industrial Engineering, The University of Texas at Austin, Austin, Texas 78705; [b] Department of Industrial and Enterprise Systems Engineering, University of Illinois Urbana-Champaign, Urbana, Illinois 61820

*Corresponding author

Contact: fxy@utexas.edu, https://orcid.org/0000-0003-4553-2687 (XF); gah@illinois.edu, https://orcid.org/0000-0003-4900-2958 (GAH)

Received: November 1, 2021
Revised: September 29, 2022; May 30, 2023; October 4, 2023
Accepted: October 13, 2023
Published Online in Articles in Advance: November 29, 2023

https://doi.org/10.1287/ijoc.2021.0306

Abstract. We study two-stage stochastic optimization problems with random recourse, where the coefficients of the adaptive decisions involve uncertain parameters. To deal with the infinite-dimensional recourse decisions, we propose a scalable approximation scheme via piecewise linear and piecewise quadratic decision rules. We develop a data-driven distributionally robust framework with two layers of robustness to address distributional uncertainty. We also establish out-of-sample performance guarantees for the proposed scheme. Applying known ideas, the resulting optimization problem can be reformulated as an exact copositive program that admits semidefinite programming approximations. We design an iterative decomposition algorithm, which converges under some regularity conditions, to reduce the runtime needed to solve this program. Through numerical examples for various known operations management applications, we demonstrate that our method produces significantly better solutions than the traditional sample-average approximation scheme especially when the data are limited. For the problem instances for which only the recourse cost coefficients are random, our method exhibits slightly inferior out-of-sample performance but shorter runtimes compared with a competing approach.

History: Accepted by Nicola Secomandi, Area Editor for Stochastic Models & Reinforcement Learning.
Funding: This work was supported by the National Science Foundation [Grants 2342505 and 2343869].
Supplemental Material: The software that supports the findings of this study is available within the paper and its Supplemental Information (https://pubsonline.informs.org/doi/suppl/10.1287/ijoc.2021.0306) as well as from the IJOC GitHub software repository (https://github.com/INFORMSJoC/2021.0306). The complete IJOC Software and Data Repository is available at https://informsjoc.github.io/.

Keywords: distributionally robust optimization • random recourse • piecewise decision rules • copositive programming • decomposition algorithm

1. Introduction

Two-stage decision-making under uncertainty considers settings where recourse actions can be taken once the realizations of the uncertain parameters are revealed. The setting is prevalent across many problems in operations management, transportation, and finance (Birge and Louveaux 2011). The classical two-stage stochastic programming approach for these problems assumes that the uncertain parameters are random with complete knowledge of the underlying probability distribution. Unfortunately, precise distributional information is rarely available in practice. Optimizing in view of the discrete empirical distribution based on historical samples often yields inferior solutions that perform poorly in out-of-sample tests (Van Parys et al. 2021). To mitigate these overfitting effects, recent interest has grown in using the distributionally robust optimization (DRO) methodology (Bertsimas et al. 2022, Cheramin et al. 2022, Delage and Saif 2022, Li et al. 2022). In DRO, one constructs an *ambiguity set* of different plausible distributions consistent with the available information. Optimal decisions are then obtained in view of the worst-case probability distribution taken from the ambiguity set. Hence, the DRO model yields decisions that can safely anticipate adverse outcomes and exhibit superior performance in out-of-sample tests (Delage and Ye 2010, Wiesemann et al. 2014). Despite this promising observation, two-stage DRO problems are generically intractable because they optimize over functions describing the recourse policies (Bertsimas et al. 2010).

Current solution schemes for two-stage DRO problems focus on problems with fixed recourse matrices, that is, the coefficients of the adaptive decisions in the constraints are fixed parameters. However, many optimization problems under uncertainty in finance (Rocha and Kuhn 2012), energy systems (Martins da Silva Rocha 2013), and inventory control (Bertsimas and Georghiou 2018) have *random* (or nonfixed) recourse. In this paper, we consider the popular decision rule approximation scheme, which restricts the recourse variables to simple functions (Garstka and Wets 1974, Georghiou et al. 2019). The simplest approximation scheme is obtained using linear decision rules (LDR) (Ben-Tal et al. 2005, Atamtürk and Zhang 2007, Chen et al. 2007), which use affine functions of the uncertain parameters. Decision rules are attractive because they usually lead to tractable approximations for problems with fixed recourse. However, when the problem has random recourse, even the simplest LDR are intractable (Guslitser 2002, Ben-Tal et al. 2004).

Compared with LDR, more advanced decision rules may improve the approximation quality but are more computationally demanding. One can use the polynomial or piecewise affine functions of the uncertain parameters, respectively, denoted as polynomial decision rules (Bampou and Kuhn 2011, Bertsimas et al. 2011) and piecewise linear decision rules (PLDR) (Bertsimas and Georghiou 2015, Ben-Tal et al. 2020). Quadratic decision rules (QDR) are a class of polynomial decision rules with degree 2 that provide a reasonable tradeoff between solution quality and computational cost. PLDR also provide a tighter approximation than basic LDR. Although highly effective, these methods can only cope with two-stage DRO problems with fixed recourse.

Two-stage DRO problems have been widely studied in the literature using different ambiguity sets. Bertsimas et al. (2010) study two-stage DRO problems with a non-data-driven ambiguity set based on first- and second-order moments. Bertsimas et al. (2018) propose a modified sample-average approximation (SAA) to approximate a two-stage DRO model in which the ambiguity set is a confidence region of a goodness-of-fit hypothesis test. Jiang and Guan (2018) derive an equivalent reformulation of two-stage DRO problems with an L^1-norm ambiguity set and uses SAA as an approximation. Bayraksan and Love (2015) study two-stage DRO problems using the ϕ-divergence (e.g., Kullback-Leibler divergence and χ^2-distance) ambiguity sets, which lead to tractability. Unfortunately, these papers do not provide any out-of-sample guarantees.

Although data-driven DRO problems with both random recourse costs and random recourse matrices have not been studied in the literature, several solution schemes for problems with random recourse have been devised in other settings. In robust optimization, Xu and Hanasusanto (2023) derive exact copositive programming reformulations for LDR approximations; Postek and Hertog (2016) construct adjustable approximation policy for recourse variables by iteratively splitting the uncertainty set. However, the resulting decision can be too conservative because it does not take into account the distributional information embedded in the historical data. At the other extreme, Kuhn et al. (2011) apply LDR for multistage stochastic optimization, an approach that requires complete distributional knowledge of the random parameters. Hanasusanto and Kuhn (2018) derive copositive programming approximations for DRO problems with random recourse costs only in the objective function.

In this paper, we combine the enhanced PLDR and piecewise quadratic decision rules (PQDR) to tackle two-stage linear DRO problems with random recourse. To the best of our knowledge, we do so for the first time in the case of random recourse cost and matrices. Our work uses a partitioning scheme based on Voronoi diagrams (Aurenhammer 1991). The Voronoi diagrams have previously been used for an iterative finite adaptability approach to solving multistage robust optimization problems (Bertsimas and Dunning 2016). In a similar vein, Chen et al. (2020) introduce the event-wise affine adaptive solution scheme for two-stage DRO with K-means clustering ambiguity sets. Here, the K-means clustering is used to construct Voronoi partitions by using perpendicular bisectors of cluster centroids.

The traditional reformulation technique based on standard convex duality theory (Boyd et al. 2004) is not applicable for the decision rule approximations because of the random recourse setting. We leverage the modern conic programming machinery (Burer 2012) to derive a concise copositive program (COP) that is intractable but admits high-quality semidefinite programming approximations. The reformulation method we use relies on the approach developed by Xu and Hanasusanto (2023) for robust optimization.

We construct a tailored ambiguity set with two layers of robustness: (1) an ambiguity set of conditional distributions given that the random parameters fall within a partition and (2) an uncertainty set for the marginal probabilities that the random parameters realize in different partitions. Given the particular decision rules that we use, the conditional ambiguity set for each partition can be aptly defined through the conditional second moments of the random parameters. We then use the χ^2-distance (Ben-Tal et al. 2013) to construct the uncertainty set for the marginal probabilities. Thus, we design the ambiguity set by combining the concentration inequalities for second moments (Shawe-Taylor and Cristianini 2003) and χ^2 statistics (Laurent and Massart 2000), which has not been considered in the context of DRO models. The ambiguity set enables us to derive a theoretical out-of-sample performance guarantee that does not suffer from the curse of dimensionality, whereas the finite-sample guarantees for the ambiguity set based on Wasserstein metric (Hanasusanto and Kuhn 2018) are known to suffer from the curse of dimensionality (Esfahani and Kuhn 2018). A recent result by Gao (2022) shows that under some Lipschitz continuity assumptions on the loss function, the curse of dimensionality can be eliminated. However, the assumptions are not satisfied in the general two-stage settings in which the loss function is not finite whenever the second-stage problem is infeasible, such as the problems without complete recourse.

To reduce the computational effort of solving the conic program, we design a Benders-type decomposition algorithm (Benders 1962) that exploits the structure of the partitioned ambiguity set. The decomposition algorithm separates the complexity of finding the optimal first-stage decision, which constitutes a tractable second-order cone program, and the second-stage policy, which comprises several concise copositive programs. Each second-stage subproblem is of a much smaller size compared with the original copositive program and can be solved independently within each partition. We prove that, under some regularity conditions, the algorithm converges in a finite number of iterations.

To demonstrate the performance of our methodology in terms of solution quality and computational effort, we conduct an extensive computational investigation on network inventory allocation, newsvendor, medical scheduling, and facility location problems. The numerical results show that our proposed approach achieves significantly better solution quality than SAA, especially with limited data. Even in the particular case where the random recourse costs appear only in the objective function, as in the network inventory allocation problem, we observe that our method distinctly outperforms the benchmark solution scheme proposed by Hanasusanto and Kuhn (2018) in terms of computational requirement, although it incurs a small out-of-sample performance loss compared with this approach. Additionally, our computational study demonstrates the benefit of using the decomposition algorithm to reduce runtimes as the number of partitions increases.

In Section 2, we introduce the risk-averse two-stage linear DRO problem. In Section 3, we use piecewise decision rules (PDR) to obtain an exact copositive programming reformulation and then derive its out-of-sample performance guarantee. Section 4 develops a decomposition algorithm to solve the reformulation. Section 5 reports our numerical results for the network inventory allocation and newsvendor problems. We conclude and delineate possible avenues for future research in Section 6. Online Appendix A includes all the proofs. Online Appendix B presents the semidefinite approximations of both the copositive cones and completely positive cones. We discuss our numerical study for the medical scheduling and facility location problems in Online Appendix C.

2. Risk-Averse Two-Stage DRO Model

We study the risk-averse two-stage linear distributionally robust optimization problem (Rahimian and Mehrotra 2019) using the conditional value-at-risk (CVaR) (Rockafellar and Uryasev 2000) measure. In this adaptive optimization problem, a decision maker first selects a here-and-now decision $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^{N_1}$, which incurs the immediate cost $\boldsymbol{c}^\top \boldsymbol{x}$. After the realization of the uncertain parameter vector $\boldsymbol{\xi} \in \Xi$, the wait-and-see decision $\boldsymbol{y}(\boldsymbol{\xi}) \in \mathbb{R}^{N_2}$ that minimizes the second-stage cost $(\boldsymbol{D}\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})$ is taken. The decision maker seeks a decision $\boldsymbol{x}$ and a policy $\boldsymbol{y}(\cdot)$ that perform the best in view of the worst-case CVaR at level $\delta \in (0,1]$. The two-stage problem is formally written as

$$\inf_{x \in \mathcal{X}} \quad \boldsymbol{c}^\top \boldsymbol{x} + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}\text{-CVaR}_\delta[Z(\boldsymbol{x}, \boldsymbol{\xi})], \tag{1}$$

where $\mathcal{P}$ is an ambiguity set containing plausible distributions of the uncertain parameter vector $\boldsymbol{\xi} \in \mathbb{R}^{S+1}$ with a known support set Ξ. In this work, we will construct the ambiguity set in a data-driven manner using historical samples $\{\hat{\boldsymbol{\xi}}_i\}_{i \in [N]}$ of the random vector $\boldsymbol{\xi}$. The inf-sup formulation in (1) means that the first-stage decision $\boldsymbol{x}$ is chosen in view the most adverse distribution $\mathbb{P} \in \mathcal{P}$ that maximizes the CVaR of the second-stage recourse function $Z(\boldsymbol{x}, \boldsymbol{\xi})$. When $\delta = 1$, the worst-case CVaR reduces to the worst-case expectation.

The recourse function $Z(\boldsymbol{x}, \boldsymbol{\xi})$ in (1) corresponds to the optimal value of the linear program:

$$\begin{aligned} &\inf\ (\boldsymbol{D}\boldsymbol{\xi})^\top \boldsymbol{y} \\ &\text{s.t. } \boldsymbol{y} \in \mathbb{R}^{N_2} \\ &\qquad \boldsymbol{T}_\ell(\boldsymbol{x})^\top \boldsymbol{\xi} \le (\boldsymbol{W}_\ell \boldsymbol{\xi})^\top \boldsymbol{y} \quad \forall \ell \in [L], \end{aligned} \tag{2}$$

where $\boldsymbol{D} \in \mathbb{R}^{N_2 \times (S+1)}$ and $\boldsymbol{W}_\ell \in \mathbb{R}^{N_2 \times (S+1)}$, $\ell \in [L] := \{1, \dots, L\}$ are coefficient matrices. The matrices $\boldsymbol{T}_\ell(\boldsymbol{x}) \in \mathbb{R}^{S+1}$, $\ell \in [L]$, are assumed to be affine functions in $\boldsymbol{x}$. Two-stage linear DRO problems are NP-hard even with fixed recourse (Bertsimas et al. 2010). The structure of our second-stage Problem (2) has random recourse, where the adaptive decision $\boldsymbol{y}(\boldsymbol{\xi})$ is multiplied with the uncertain parameters $\boldsymbol{\xi}$ in both the objective function and the constraints. The structure introduces significant challenges in addressing the problem. In this paper, we employ the decision rule approach that enables an exact reformulation with tractable approximations. To this end, we first reformulate Problem (1) as a distributionally robust semi-infinite linear program involving a worst-case expectation, as stated in the following proposition.

Proposition 1. *The risk-averse two-stage DRO problem* (1) *can be reformulated as*

$$\begin{array}{ll} \inf & \boldsymbol{c}^\top \boldsymbol{x} + \theta + \dfrac{1}{\delta} \sup\limits_{\mathbb{P} \in \mathcal{P}} \ \mathbb{E}_{\mathbb{P}}[\tau(\boldsymbol{\xi})] \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{X},\ \theta \in \mathbb{R},\ \boldsymbol{y} : \mathbb{R}^{S+1} \to \mathbb{R}^{N_2},\ \tau : \mathbb{R}^{S+1} \to \mathbb{R} \\ & \left. \begin{array}{l} \tau(\boldsymbol{\xi}) \geq 0 \\ \tau(\boldsymbol{\xi}) \geq (\boldsymbol{D}\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi}) - \theta \\ \boldsymbol{T}_\ell(\boldsymbol{x})^\top \boldsymbol{\xi} \leq (\boldsymbol{W}_\ell \boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi}) \quad \forall \ell \in [L] \end{array} \right\} \forall \boldsymbol{\xi} \in \Xi, \end{array} \tag{3}$$

where the second-stage decision variables $\boldsymbol{y}$ *and* τ *are measurable mappings from* $\mathbb{R}^{S+1}$ *to* $\mathbb{R}^{N_2}$ *and from* $\mathbb{R}^{S+1}$ *to* $\mathbb{R}$, *respectively.*

Reformulated Problem (3) contains random recourse matrices in the constraints even if the original two-stage DRO Problem (1) has random recourse costs only in the objective. In the following sections, we focus on addressing the equivalent Problem (3).

We now explain the structure of the support set. The uncertain parameter vector $\boldsymbol{\xi}$ belongs to a support set Ξ defined as a slice of a convex cone $\mathcal{K} \in \mathbb{R}^S \times \mathbb{R}_+$, given by

$$\Xi := \left\{ \boldsymbol{\xi} := \begin{bmatrix} \boldsymbol{\zeta} \\ \nu \end{bmatrix} \in \mathcal{K} : \nu = 1 \right\}. \tag{4}$$

The restriction that the last component of $\boldsymbol{\xi}$ is one enables us to simplify any affine function of the primitive parameter vector $\boldsymbol{\zeta}$ as a linear function of $\boldsymbol{\xi}$. Similarly, it can also represent any quadratic function in a homogenized form. The support set Ξ is assumed to satisfy the following mild condition.

Assumption 1. *The support set* Ξ *is nonempty, compact, convex, and full-dimensional.*

The support set Ξ in (4) can model widely used support sets. For instance, we can define a polytope by setting

$$\mathcal{K} := \left\{ \boldsymbol{\xi} := \begin{bmatrix} \boldsymbol{\zeta} \\ \nu \end{bmatrix} \in \mathbb{R}^S \times \mathbb{R}_+ : \boldsymbol{P}\boldsymbol{\zeta} \geq \boldsymbol{t}\nu \right\},$$

with $\boldsymbol{P} \in \mathbb{R}^{S_p \times S}$, $\boldsymbol{t} \in \mathbb{R}^{S_p}$. In addition, we can model an ellipsoid or two-norm ball by setting

$$\mathcal{K} := \left\{ \boldsymbol{\xi} := \begin{bmatrix} \boldsymbol{\zeta} \\ \nu \end{bmatrix} \in \mathbb{R}^S \times \mathbb{R}_+ : \|\mathbf{R}\boldsymbol{\zeta}\|_2 \leq q\nu \right\},$$

with $\boldsymbol{R} \in \mathbb{R}^{S_r \times S}$, $q \in \mathbb{R}$.

3. Decision Rule Approach

In this section, we develop an approximation scheme for the two-stage linear DRO problems. In Section 3.1, we approximate the problem using a combination of PLDR and PQDR. We then design the ambiguity set and derive an alternative formulation of the objective function in Section 3.2. In Section 3.3, we illustrate that, with appropriate choices of ambiguity set parameters, we can establish a theoretical out-of-sample guarantee for the solution of the PDR problem.

3.1. Decision Rule Framework

We adopt the combination of PLDR and PQDR to conservatively approximate the two-stage Problem (3) by restricting the adaptive decisions $\boldsymbol{y}(\boldsymbol{\xi})$ and $\tau(\boldsymbol{\xi})$ to, respectively, piecewise affine functions and piecewise quadratic functions of the uncertain parameters. To this end, we partition the support set Ξ into K subsets $\Xi_1, \ldots, \Xi_K$, and we optimize basic linear or quadratic decision rules in each partition separately. Specifically, the PLDR scheme for the decision variable $\boldsymbol{y}(\boldsymbol{\xi})$ is given by

$$\boldsymbol{y}(\boldsymbol{\xi}) = \boldsymbol{Y}_k \boldsymbol{\xi} \quad \forall \boldsymbol{\xi} \in \Xi_k \ \ \forall k \in [K],$$

where $\boldsymbol{Y}_k \in \mathbb{R}^{N_2 \times (S+1)}$ is the coefficient matrix of the linear decision rule for partition k.

We observe that at optimality the second-stage epigraphical variable $\tau(\boldsymbol{\xi})$ in Problem (3) coincides with $\max\{(\boldsymbol{D}\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi}) - \theta, 0\}$. After replacing $\boldsymbol{y}(\boldsymbol{\xi})$ with a piecewise linear decision rule, the term $(\boldsymbol{D}\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi})$ constitutes a piecewise quadratic function in $\boldsymbol{\xi}$, and the semi-infinite inequality constraints in (3) are satisfied if $\tau(\boldsymbol{\xi})$ exhibits a piecewise quadratic form in $\boldsymbol{\xi}$. Thus, we express $\tau(\boldsymbol{\xi})$ as a piecewise quadratic decision rule

$$\tau(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} \quad \forall \boldsymbol{\xi} \in \Xi_k \ \forall k \in [K],$$

where $\boldsymbol{Q}_k \in \mathbb{R}^{(S+1)\times(S+1)}$ is the coefficient matrix of the quadratic decision rule for partition k.

Using the combination of PLDR and PQDR and applying the law of total expectation yield the conservative PDR problem

$$\begin{aligned} J^{\text{PDR}} := \inf\ & \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta} \sup_{\mathbb{P}\in\mathcal{P}} \sum_{k\in[K]} \mathbb{P}(\boldsymbol{\xi} \in \Xi_k) \mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} \mid \boldsymbol{\xi} \in \Xi_k] \\ \text{s.t.}\ & \boldsymbol{x} \in \mathcal{X},\ \theta \in \mathbb{R},\ \boldsymbol{Y}_k \in \mathbb{R}^{N_2 \times (S+1)},\ \boldsymbol{Q}_k \in \mathbb{R}^{(S+1)\times(S+1)} \\ & \mathcal{T}_\ell(\boldsymbol{x})^\top \boldsymbol{\xi} \le (\mathcal{W}_\ell \boldsymbol{\xi})^\top \boldsymbol{Y}_k \boldsymbol{\xi} + \lambda_\ell \boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} + \kappa_\ell \theta \quad \forall \boldsymbol{\xi} \in \Xi_k \ \forall k \in [K] \ \forall \ell \in [L+2], \end{aligned} \tag{5}$$

where

$$\begin{aligned} &\mathcal{T}_\ell(\boldsymbol{x}) = \boldsymbol{T}_\ell(\boldsymbol{x}), && \mathcal{W}_\ell = \boldsymbol{W}_\ell, && \lambda_\ell = 0, && \kappa_\ell = 0 && \forall \ell \in [L], \\ &\mathcal{T}_\ell(\boldsymbol{x}) = [\boldsymbol{0}], && \mathcal{W}_\ell = [\boldsymbol{0}], && \lambda_\ell = 1, && \kappa_\ell = 0 && \ell = L+1, \\ &\mathcal{T}_\ell(\boldsymbol{x}) = [\boldsymbol{0}], && \mathcal{W}_\ell = -\boldsymbol{D}, && \lambda_\ell = 1, && \kappa_\ell = 1 && \ell = L+2. \end{aligned} \tag{6}$$

In this work, we construct the partitioning of the support set Ξ corresponding to Voronoi regions. Starting with a set of constructor points $\{\boldsymbol{\xi}'_k\}_{k\in[K]}$, we define the region Ξ_k as the set of all points in Ξ whose Euclidean distance is closer to $\boldsymbol{\xi}'_k$ than any other constructor points. That is, for the k^{th} partition we have

$$\begin{aligned} \Xi_k &= \{\boldsymbol{\xi} \in \Xi : \|\boldsymbol{\xi} - \boldsymbol{\xi}'_k\|_2 \le \|\boldsymbol{\xi} - \boldsymbol{\xi}'_i\|_2 \ \forall i \in [K] : i \ne k\} \\ &= \{\boldsymbol{\xi} \in \Xi : 2(\boldsymbol{\xi}'_i - \boldsymbol{\xi}'_k)^\top \boldsymbol{\xi} \le \boldsymbol{\xi}'^\top_i \boldsymbol{\xi}'_i - \boldsymbol{\xi}'^\top_k \boldsymbol{\xi}'_k \ \forall i \in [K] : i \ne k\} \\ &= \{\boldsymbol{\xi} \in \mathcal{K}_k : \mathbf{e}_{S+1}^\top \boldsymbol{\xi} = 1\}, \end{aligned}$$

where

$$\mathcal{K}_k = \{\boldsymbol{\xi} \in \mathcal{K} : 2(\boldsymbol{\xi}'_i - \boldsymbol{\xi}'_k)^\top \boldsymbol{\xi} \le \boldsymbol{\xi}'^\top_i \boldsymbol{\xi}'_i - \boldsymbol{\xi}'^\top_k \boldsymbol{\xi}'_k \ \forall i \in [K] : i \ne k\} \tag{7}$$

is a convex cone generated by the k^{th} region. Here, $\mathbf{e}_i$ denotes the i^{th} standard basis vector. The constructor points $\{\boldsymbol{\xi}'_k\}_{k\in[K]}$ are selected independently of the samples $\{\hat{\boldsymbol{\xi}}_i\}_{i\in[N]}$. One possibility is to sample the constructor points from an independent uniform distribution over the support set Ξ. Alternatively, one can split the available N historical data points into two parts: samples to build the ambiguity set and samples for the constructor points.

3.2. Design of the Ambiguity Set

The objective function of Problem (5) contains two sets of terms: the partition probabilities $\mathbb{P}(\boldsymbol{\xi} \in \Xi_k)$, $k \in [K]$ and the conditional expectations $\mathbb{E}_{\mathbb{P}}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} \mid \boldsymbol{\xi} \in \Xi_k]$, $k \in [K]$. Their values depend on the distribution of the uncertain parameter $\boldsymbol{\xi}$. One could naïvely define the ambiguity set $\mathcal{P}$ to contain only the empirical distribution $\hat{\mathbb{P}} = \frac{1}{N}\sum_{i\in[N]} \delta_{\hat{\boldsymbol{\xi}}_i}$, where $\delta_{\hat{\boldsymbol{\xi}}_i}$ denotes the Dirac delta measure that places a unit mass at $\hat{\boldsymbol{\xi}}_i$. This choice of ambiguity set reduces the DRO problem into the classical SAA for stochastic programming. The approximation, however, tends to result in poor decisions that are overfitted to the observed data points (Van Parys et al. 2021). To mitigate the overfitting effects, we design a tailored data-driven ambiguity set with two layers of robustness on the partition probabilities and the conditional expectations, as follows.

Definition 1 (Ambiguity Set). The ambiguity set is given by

$$\mathcal{P} := \left\{ \mathbb{P} \in \mathcal{P}(\Xi) : \mathbb{P} = \sum_{k\in[K]} p_k \mathbb{P}_k \ \text{ with } \ \boldsymbol{p} \in \Delta \text{ and } \ \mathbb{P}_k \in \mathcal{P}_k \quad \forall k \in [K] \right\}. \tag{8}$$

Here, the uncertainty set Δ for the partition probability vector $\boldsymbol{p}$ is defined in terms of the χ^2-distance as

$$\Delta := \left\{ \boldsymbol{p} \in \mathbb{R}^K_+ : \mathbf{e}^\top \boldsymbol{p} = 1, \sum_{k=1}^K (p_k - \hat{p}_k)^2 / p_k \le \gamma \right\}, \tag{9}$$

whereas the ambiguity set $\mathcal{P}_k$ for the conditional distribution $\mathbb{P}_k$ is defined as

$$\mathcal{P}_k := \{\mathbb{P}_k \in \mathcal{P}(\Xi_k) : \| \mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}\boldsymbol{\xi}^\top] - \mathbb{E}_{\hat{\mathbb{P}}_k}[\boldsymbol{\xi}\boldsymbol{\xi}^\top] \|_F \le \epsilon_k\}. \tag{10}$$

In the definition, we denote by $\mathcal{P}(\Xi) := \{\mu \in \mathcal{M}_+ : \int_\Xi \mu(d\xi) = 1\}$ the set of all probability measures supported on Ξ, where $\mathcal{M}_+$ is the set of nonnegative Borel measures. The vector of all ones is denoted by $\mathbf{e}$. The empirical partition probabilities are defined as $\hat{p}_k = \frac{1}{N}\sum_{i\in[N]} \mathbb{1}_{\Xi_k}(\hat{\boldsymbol{\xi}}_i)$, $k \in [K]$, where the indicator function $\mathbb{1}_\Xi : \mathbb{R}^{S+1} \to \{0,1\}$ of a set $\Xi \subseteq \mathbb{R}^{S+1}$ is defined through $\mathbb{1}_\Xi(\boldsymbol{\xi}) = 1$ if $\boldsymbol{\xi} \in \Xi$ and $\mathbb{1}_\Xi(\boldsymbol{\xi}) = 0$ if $\boldsymbol{\xi} \notin \Xi$. The set $\mathcal{I}_k = \{i \in [N] : \hat{\boldsymbol{\xi}}_i \in \Xi_k\}$ contains the indices i for which the corresponding sample point $\hat{\boldsymbol{\xi}}_i$ falls in the partition Ξ_k. Accordingly, $\hat{\mathbb{P}}_k = \frac{1}{|\mathcal{I}_k|}\sum_{i\in\mathcal{I}_k} \delta_{\hat{\boldsymbol{\xi}}_i}$ denotes the empirical conditional distribution over that partition. Here, $\|\cdot\|_F$ denotes the Frobenius norm.

Any distribution $\mathbb{P}$ in the ambiguity set $\mathcal{P}$ is defined as a mixture of the conditional distributions $\mathbb{P}_1, \ldots, \mathbb{P}_K$ with weights $p_1, \ldots, p_K$. The weights are assumed to belong to the uncertainty set Δ. The set is described by a single parameter γ and contains all probability vectors $\boldsymbol{p}$ whose χ^2-distance to the empirical marginal probability vector $\hat{\boldsymbol{p}}$ is less than or equal to γ. The χ^2-distance belongs to the class of ϕ-divergences (Pardo 2018), which includes the KL divergence, the Hellinger distance, and the Burg entropy-based divergence. We use the χ^2-distance as it leads to a tractable reformulation and provides an attractive statistical guarantee. Each conditional distribution $\mathbb{P}_k$ belonging to the ambiguity set $\mathcal{P}_k$ is defined as the set of all probability distributions, whose second-moment matrix $\mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]$ is within a distance ϵ_k from the empirical second-moment matrix $\mathbb{E}_{\hat{\mathbb{P}}_k}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]$ with respect to the Frobenius norm. By (4), the matrix $\mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]$ incorporates both the first- and second-order moments of the random parameter $\boldsymbol{\zeta}$.

In view of the ambiguity set $\mathcal{P}$ defined in (8), the objective function of Problem (5) can be rewritten as

$$\begin{aligned}
&\boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}\sup_{\mathbb{P}\in\mathcal{P}} \sum_{k\in[K]} \mathbb{P}(\boldsymbol{\xi} \in \Xi_k)\mathbb{E}_\mathbb{P}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} \mid \boldsymbol{\xi} \in \Xi_k] \\
&= \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}\sup_{\boldsymbol{p}\in\Delta} \sup_{\mathbb{P}_k\in\mathcal{P}_k \forall k} \sum_{k\in[K]} p_k \mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi}] \\
&= \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}\sup_{\boldsymbol{p}\in\Delta} \sum_{k\in[K]} p_k \sup_{\mathbb{P}_k\in\mathcal{P}_k} \mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi}].
\end{aligned} \tag{11}$$

The second equality holds because each innermost maximization problem optimizes the conditional distribution $\mathbb{P}_k$ separately. In Proposition 2, we first derive the reformulation of the kth inner worst-case expectation over the conditional ambiguity set $\mathcal{P}_k$ defined in (10). Then, in Proposition 3, we state the reformulation of the outer worst-case expectation problem over the uncertainty set Δ.

Proposition 2. *Using the conditional ambiguity set $\mathcal{P}_k$ defined in* (10), *the kth worst-case expectation in* (11) *can be reformulated as the semi-infinite program*

$$\begin{aligned}
\sup_{\mathbb{P}_k\in\mathcal{P}_k} \mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi}] = \inf\ & \alpha_k + \mathrm{tr}(\boldsymbol{Q}_k\hat{\boldsymbol{\Omega}}_k) + \mathrm{tr}(\boldsymbol{B}_k\hat{\boldsymbol{\Omega}}_k) + \epsilon_k\|\boldsymbol{Q}_k^\top + \boldsymbol{B}_k\|_F \\
\text{s.t.}\ & \alpha_k \in \mathbb{R},\ \boldsymbol{B}_k \in \mathbb{S}^{S+1} \\
& \alpha_k + \boldsymbol{\xi}^\top \boldsymbol{B}_k \boldsymbol{\xi} \ge 0 \quad \forall \boldsymbol{\xi} \in \Xi_k,
\end{aligned} \tag{12}$$

where $\hat{\boldsymbol{\Omega}}_k = \sum_{i\in\mathcal{I}_k} \hat{\boldsymbol{\xi}}_i \hat{\boldsymbol{\xi}}_i^\top / |\mathcal{I}_k|$ is the empirical second-order moment matrix.

Proposition 3 (Ben-Tal et al. 2013, theorem 4.1). *Let the uncertainty set Δ be defined in* (9). *For any $\boldsymbol{\varphi} \in \mathbb{R}^K$, the worst-case expectation problem $\max_{\boldsymbol{p}\in\Delta} \boldsymbol{\varphi}^\top \boldsymbol{p}$ is equivalent to the second-order cone program (SOCP)*

$$\begin{aligned}
\max\ & \boldsymbol{\varphi}^\top \boldsymbol{p} \\
\text{s.t.}\ & \boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}_+^K,\ \mathbf{e}^\top \boldsymbol{p} = 1,\ \mathbf{e}^\top \boldsymbol{q} \le \gamma \\
& \sqrt{(p_k - \hat{p}_k)^2 + \frac{1}{4}p_k^2 + q_k^2} \le \frac{1}{2}p_k + q_k \quad \forall k \in [K],
\end{aligned} \tag{13}$$

that is, the optimal value and the optimal solution of the expectation problem can be computed by solving Problem (13). *The* *optimal value can also be obtained by solving the following SOCP dual to (13):*

$$\begin{aligned} \min\ & \gamma\omega - \eta - 2\hat{\boldsymbol{p}}^\top \boldsymbol{r} + 2\omega\hat{\boldsymbol{p}}^\top \mathbf{e} \\ \text{s.t.}\ & \omega \in \mathbb{R}_+,\ \eta \in \mathbb{R},\ \boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^K \\ & \varphi_k \le s_k,\ s_k + \eta \le \omega,\ \sqrt{4r_k^2 + (s_k+\eta)^2} \le 2\omega - s_k - \eta \quad \forall k \in [K]. \end{aligned}$$

Combining Propositions 2 and 3 and identifying φ_k with the optimal value of the kth inner worst-case expectation $\sup_{\mathbb{P}_k \in \mathcal{P}_k} \mathbb{E}_{\mathbb{P}_k}[\boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi}]$, we obtain the reformulation of the objective function (11). Incorporating the reformulation of the objective function, we find that the PDR problem (5) is equivalent to the semi-infinite program

$$\begin{aligned} J^{\text{PDR}} = \inf\ & \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}(\gamma\omega - \eta - 2\hat{\boldsymbol{p}}^\top \boldsymbol{r} + 2\omega\hat{\boldsymbol{p}}^\top \mathbf{e}) \\ \text{s.t.}\ & \boldsymbol{x} \in \mathcal{X},\ \theta, \eta \in \mathbb{R},\ \omega \in \mathbb{R}_+,\ \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{\alpha} \in \mathbb{R}^K, \\ & \left.\begin{aligned} & \boldsymbol{Y}_k \in \mathbb{R}^{N_2 \times (S+1)},\ \boldsymbol{Q}_k \in \mathbb{R}^{(S+1)\times(S+1)},\ \boldsymbol{B}_k \in \mathbb{S}^{S+1} \\ & s_k + \eta \le \omega,\ \sqrt{4r_k^2 + (s_k+\eta)^2} \le 2\omega - s_k - \eta \\ & \alpha_k + \operatorname{tr}(\boldsymbol{Q}_k \hat{\boldsymbol{\Omega}}_k) + \operatorname{tr}(\boldsymbol{B}_k \hat{\boldsymbol{\Omega}}_k) + \epsilon_k \|\boldsymbol{Q}_k^\top + \boldsymbol{B}_k\|_F \le s_k \\ & \alpha_k + \boldsymbol{\xi}^\top \boldsymbol{B}_k \boldsymbol{\xi} \ge 0 \quad \forall \boldsymbol{\xi} \in \Xi_k \\ & \mathcal{T}_\ell(\boldsymbol{x})^\top \boldsymbol{\xi} \le (\mathcal{W}_\ell \boldsymbol{\xi})^\top \boldsymbol{Y}_k \boldsymbol{\xi} + \lambda_\ell \boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} + \kappa_\ell \theta \quad \forall \ell \in [L+2]\ \forall \boldsymbol{\xi} \in \Xi_k \end{aligned}\right\} \forall k \in [K]. \end{aligned} \tag{14}$$

3.3. Out-of-Sample Performance Guarantee

In the following, we establish the theoretical performance guarantee for the solution of the PDR problem (5). The result uses Lemmas 1 and 2 in Online Appendix A.3, which provide high confidence bounds on the errors of the empirical estimates $\mathbb{E}_{\hat{\mathbb{P}}_k}[\boldsymbol{\xi}\boldsymbol{\xi}^\top]$ and $\hat{\boldsymbol{p}}$. Based on the confidence bounds, we can ensure the objective value of the PDR problem (14) constitutes an upper bound on the out-of-sample CVaR with a large probability, as follows.

Theorem 1 (Out-of-Sample Guarantee). *Let the ambiguity set $\mathcal{P}$ in Problem (5) be defined in (8). Setting the robustness parameters to $\epsilon_k = R_k^2(2 + \sqrt{2\ln K/\rho_1})/\sqrt{|\mathcal{I}_k|}$, $k \in [K]$, and $\gamma = \left(K - 1 + 2\sqrt{-(K-1)\ln \rho_2} - 2\ln \rho_2\right)/N$, we can ascertain that*

$$\text{Prob}\left(J^{\text{PDR}} \ge \boldsymbol{c}^\top \hat{\boldsymbol{x}} + \mathbb{P}^*\text{-CVaR}_\delta[Z(\hat{\boldsymbol{x}}, \boldsymbol{\xi})]\right) \ge 1 - \rho_1 - (1 + o(1))\rho_2,$$

where $\hat{\boldsymbol{x}}$ is an optimal solution of Problem (5), $\mathbb{P}^*$ is the true underlying distribution, and $\rho_1, \rho_2 \in (0,1)$ are the prescribed tolerance levels that control the tightness of the bound.

Setting $\rho = \rho_1 + (1 + o(1))\rho_2$), the theorem implies that with judicious choices for the robustness parameters, the optimal value J^{PDR} of PDR Problem (14) provides a $1-\rho$ confidence bound on the out-of-sample performance of the data-driven solution. The theorem also provides guidance for choosing the values of the robustness parameters in practice. Specifically, we can determine the values of the robustness parameters $\{\epsilon_k\}_{k\in[K]}$ and γ by approximating $(1+o(1))$ with its asymptotic value of one. Then, for a desired tolerance level ρ, one can set $\epsilon_k = R_k^2(2 + \sqrt{2\ln 2K/\rho})/\sqrt{|\mathcal{I}_k|}$, $k \in [K]$, and $\gamma = \left(K - 1 + 2\sqrt{-(K-1)\ln \rho/2} - 2\ln \rho/2\right)/N$ to ensure that the guarantee $J^{\text{PDR}} \ge \boldsymbol{c}^\top \hat{\boldsymbol{x}} + \mathbb{P}^*\text{-CVaR}_\delta[Z(\hat{\boldsymbol{x}}, \boldsymbol{\xi})]$ holds with an approximate confidence level $1-\rho$. Theorem 1 shows that γ decays at rate $1/N$, which is faster than ϵ_k with rate $1/\sqrt{N}$. Thus, as the size of in-sample data N increases, we give more trust to the empirical partition probabilities than the estimated conditional distributions.

4. Algorithms

In this section, we reformulate semi-infinite Problem (14) as a convex conic program and develop a decomposition algorithm to accelerate the solution process. In Section 4.1, we use the copositive technique to derive an equivalent copositive program that admits tractable approximations in semidefinite programming. In Section 4.2, we develop an iterative Benders-type decomposition algorithm that allows the subproblems to be solved independently to further reduce the runtime.

4.1. Copositive Reformulations

Problem (14) has infinitely many constraints parametrized by the realizations of $\boldsymbol{\xi}$ in the partitions Ξ_k, $k \in [K]$. For any fixed decision $(\boldsymbol{x}, \theta, \boldsymbol{\alpha}, \boldsymbol{Y}_k, \boldsymbol{Q}_k, \boldsymbol{B}_k)$, the constraints are equivalent to the following constraints involving nonconvex quadratic minimization problems:

$$0 \le \inf_{\boldsymbol{\xi}\in\Xi_k} \boldsymbol{\xi}^\top \mathcal{W}_\ell^\top \boldsymbol{Y}_k \boldsymbol{\xi} + \lambda_\ell \boldsymbol{\xi}^\top \boldsymbol{Q}_k \boldsymbol{\xi} - \mathcal{T}_\ell(\boldsymbol{x})^\top \boldsymbol{\xi} + \kappa_\ell \theta \quad \forall \ell \in [L+2] \;\; \forall k \in [K], \tag{15a}$$

$$0 \le \inf_{\boldsymbol{\xi}\in\Xi_k} \boldsymbol{\xi}^\top \boldsymbol{B}_k \boldsymbol{\xi} + \alpha_k \quad \forall k \in [K]. \tag{15b}$$

To convexify these problems, we use the copositive programming scheme to derive equivalent reformulations of (15). The reformulation technique is based on the one developed by Xu and Hanasusanto (2023) in the robust optimization setting.

In this paper, we use the generalized copositive cone $\mathcal{C}(\mathcal{K}) := \{\boldsymbol{X} \in \mathbb{S}^N : \boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{v} \ge 0 \;\; \forall \boldsymbol{v} \in \mathcal{K}\}$ and its dual, that is, the generalized completely positive cone $\mathcal{C}^*(\mathcal{K}) := \{\boldsymbol{X} \in \mathbb{S}^N : \boldsymbol{X} = \sum_{i\in N} \boldsymbol{x}^i (\boldsymbol{x}^i)^\top, \boldsymbol{x}^i \in \mathcal{K}\}$ over a closed and convex cone $\mathcal{K} \in \mathbb{R}^N$. Here, we refer to linear optimization over $\mathcal{C}(\mathcal{K})$ and $\mathcal{C}^*(\mathcal{K})$ as (generalized) copositive programming and (generalized) completely positive programming, respectively. In addition, to simplify the notation, we define the matrices $\boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k)$, $\ell \in [L+2]$, $k \in [K]$ that are affine in their arguments as follows:

$$\boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k) := \frac{1}{2}\Big(\mathcal{W}_\ell^\top \boldsymbol{Y}_k + \boldsymbol{Y}_k^\top \mathcal{W}_\ell + \lambda_\ell \boldsymbol{Q}_k + \lambda_\ell \boldsymbol{Q}_k^\top - \mathcal{T}_\ell(\boldsymbol{x}) \mathbf{e}_{S+1}^\top - \mathbf{e}_{S+1} \mathcal{T}_\ell^\top(\boldsymbol{x})\Big). \tag{16}$$

Based on the technical lemmas in Online Appendix A.4, the following proposition presents copositive reformulations of the constraints in (15).

Proposition 4. *The constraints in* (15a) *are satisfied if and only if there exist* $\pi_\ell^k \in \mathbb{R}$, $\ell \in [L+2]$, $k \in [K]$, *such that*

$$\boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k) - \pi_\ell^k \mathbf{e}_{S+1} \mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}_k),\; \pi_\ell^k + \kappa_\ell \theta \ge 0 \quad \forall \ell \in [L+2] \;\; \forall k \in [K]. \tag{17}$$

Similarly, the constraints in (15b) *are satisfied if and only if there exists a vector* $\boldsymbol{\rho} \in \mathbb{R}^K$ *such that*

$$\boldsymbol{B}_k - \rho_k \mathbf{e}_{S+1} \mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}_k),\; \rho_k + \alpha_k \ge 0 \quad \forall k \in [K]. \tag{18}$$

Proposition 4 enables us to reformulate the semi-infinite program (14) as a COP, as follows.

Theorem 2. *PDR Problem* (5) *is equivalent to the following polynomial-size copositive program:*

$$\begin{aligned}
J^{\mathrm{PDR}} = \inf\; & \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}(\gamma\omega - \eta - 2\hat{\boldsymbol{p}}^\top \boldsymbol{r} + 2\omega \hat{\boldsymbol{p}}^\top \mathbf{e}) \\
\text{s.t.}\; & \boldsymbol{x} \in \mathcal{X},\; \theta, \eta \in \mathbb{R},\; \omega \in \mathbb{R}_+,\; \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{\alpha} \in \mathbb{R}^K \\
& \boldsymbol{\pi}_k \in \mathbb{R}^L,\; \boldsymbol{Y}_k \in \mathbb{R}^{N_2 \times (S+1)},\; \boldsymbol{Q}_k \in \mathbb{R}^{(S+1)\times(S+1)},\; \boldsymbol{B}_k \in \mathbb{S}^{S+1} \qquad \forall k \in [K] \\
& \left.\begin{array}{l}
s_k + \eta \le \omega,\; \sqrt{4r_k^2 + (s_k + \eta)^2} \le 2\omega - s_k - \eta \\
\alpha_k + \mathrm{tr}(\boldsymbol{Q}_k \hat{\boldsymbol{\Omega}}_k) + \mathrm{tr}(\boldsymbol{B}_k \hat{\boldsymbol{\Omega}}_k) + \epsilon_k \|\boldsymbol{Q}_k^\top + \boldsymbol{B}_k\|_F \le s_k \\
\pi_\ell^k + \kappa_\ell \theta \ge 0 \quad \forall \ell \in [L+2] \\
\boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k) - \pi_\ell^k \mathbf{e}_{S+1}\mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}_k) \quad \forall \ell \in [L+2] \\
\boldsymbol{B}_k + \alpha_k \mathbf{e}_{S+1}\mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}_k)
\end{array}\right\} \forall k \in [K],
\end{aligned} \tag{19}$$

where the affine functions $\boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k)$, $\ell \in [L+2]$, $k \in [K]$, *are defined in* (16).

The size of Problem (19) is independent of the number of historical samples N. However, copositive programs are generically intractable (Burer 2012). To develop a tractable solution scheme for the problem, we replace each copositive cone $\mathcal{C}(\mathcal{K}_k)$ with the semidefinite-representable inner approximation developed in Online Appendix B. In this way, we obtain a conservative solution by solving a tractable semidefinite program with $O(KL)$ semidefinite constraints. The time complexity of solving the semidefinite program using the interior point algorithm is $O(K^3L^3(KSL + KS^2 + KN_2S + N_1)^{\frac{1}{2}} + K^2L^2(KSL + KS^2 + KN_2S + N_1)^{\frac{5}{2}} + (KSL + KS^2 + KN_2S + N_1)^{\frac{7}{2}})$ (Ben-Tal and Nemirovski 2001).

4.2. Decomposition Algorithm

The copositive program (19) with $O(KL)$ copositive constraints introduces additional computational challenges when we increase the number of partitions to improve the approximation quality. Fortunately, the structure of Problem (19) allows us to solve the subproblems corresponding to different partitions. Exploiting this decomposition structure, we now develop an iterative Benders-type algorithm to solve the PDR problem (19). The main idea of the algorithm is to divide the original problem into a master problem and K subproblems. The subproblems can be solved independently given a fixed first-stage decision returned by the master problem. They yield optimality cuts and feasibility cuts that are then passed to the master problem as additional constraints. The algorithm proceeds iteratively until a termination criterion is attained and returns an η-optimal solution for a given tolerance level $\eta > 0$.

To design the decomposition algorithm for Problem (19), we consider the following equivalent reformulation:

$$\begin{aligned} J^{\mathrm{PDR}} = \inf\ & \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}(\gamma\omega - \eta - 2\hat{\boldsymbol{p}}^\top \boldsymbol{r} + 2\omega \hat{\boldsymbol{p}}^\top \mathbf{e}) \\ \text{s.t.}\ & \boldsymbol{x} \in \mathcal{X},\ \theta, \eta \in \mathbb{R},\ \omega \in \mathbb{R}_+,\ \boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^K \\ & \left.\begin{array}{l} s_k + \eta \le \omega,\ \sqrt{4r_k^2 + (s_k+\eta)^2} \le 2\omega - s_k - \eta \\ Z_k^{\mathrm{PDR}}(\boldsymbol{x}, \theta) \le s_k \end{array}\right\} \forall k \in [K]. \end{aligned} \tag{20}$$

Here, for fixed $(\boldsymbol{x}, \theta)$, the residual problem corresponding to the k^{th} partition is given by

$$\begin{aligned} Z_k^{\mathrm{PDR}}(\boldsymbol{x}, \theta) := \inf\ & \alpha_k + \mathrm{tr}(\boldsymbol{Q}_k \hat{\boldsymbol{\Omega}}_k) + \mathrm{tr}(\boldsymbol{B}_k \hat{\boldsymbol{\Omega}}_k) + \epsilon_k \|\boldsymbol{Q}_k^\top + \boldsymbol{B}_k\|_F \\ \text{s.t.}\ & \boldsymbol{Y}_k \in \mathbb{R}^{N_2 \times (S+1)},\ \boldsymbol{Q}_k \in \mathbb{R}^{(S+1)\times(S+1)},\ \boldsymbol{B}_k \in \mathbb{S}^{S+1},\ \alpha_k \in \mathbb{R},\ \pi_k \in \mathbb{R}^{L+2} \\ & \pi_\ell^k + \kappa_\ell \theta \ge 0 \quad \forall \ell \in [L+2] \\ & \boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k) - \pi_\ell^k \mathbf{e}_{S+1}\mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}_k) \quad \forall \ell \in [L+2] \\ & \boldsymbol{B}_k + \alpha_k \mathbf{e}_{S+1}\mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}_k), \end{aligned} \tag{21}$$

where the affine functions $\boldsymbol{\Delta}_\ell^k(\boldsymbol{x}, \boldsymbol{Y}_k, \boldsymbol{Q}_k), \ell \in [L+2], k \in [K]$, are defined in (16). Because Problems (19) and (20) are equivalent, we can obtain the optimal solution of Problem (19) by solving the COP (20), which involves K smaller residual problems (21), using a Benders-type algorithm.

Before describing the algorithm, we introduce the following terminology that will be used to obtain its theoretical foundation.

Definition 2 (Complete Recourse Under Linear Decision Rules). We say that the two-stage DRO Problem (1) has complete recourse under linear decision rules if there exists $\boldsymbol{Y} \in \mathbb{R}^{N_2 \times (S+1)}$ such that $(\boldsymbol{W}_\ell \boldsymbol{\xi})^\top \boldsymbol{Y}\boldsymbol{\xi} > 0$ for all $\boldsymbol{\xi} \in \Xi$ and $\ell \in [L]$.

This condition implies that the second-stage Problem (2) is always feasible under linear decision rules; that is, there exists $\boldsymbol{Y} \in \mathbb{R}^{N_2 \times (S+1)}$ such that $\boldsymbol{y} = \boldsymbol{Y}\boldsymbol{\xi}$ is feasible to (2) for every $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$. The existence of feasible LDR also implies the existence of feasible PDR. By the equivalence between Problems (5) and (20), we conclude that Subproblem (21) is feasible for any fixed $\boldsymbol{x} \in \mathcal{X}$ and $\theta \in \mathbb{R}$, that is, $Z_k^{\mathrm{PDR}}(\boldsymbol{x}, \theta) < \infty$, $k \in [K]$. The following lemma provides an equivalent condition for checking complete recourse under linear decision rules.

Lemma 1. *The complete recourse under linear decision rules is satisfied if and only if there exist* $\boldsymbol{Y} \in \mathbb{R}^{N_2 \times (S+1)}$ *and* $\beta_\ell \in \mathbb{R}_{++}$, $\ell \in [L]$, *such that*

$$\frac{1}{2}(\boldsymbol{Y}^\top \boldsymbol{W}_\ell + \boldsymbol{W}_\ell^\top \boldsymbol{Y}) - \beta_\ell \mathbf{e}_{S+1}\mathbf{e}_{S+1}^\top \in \mathcal{C}(\mathcal{K}) \quad \forall \ell \in [L]. \tag{22}$$

Condition (22) can be sufficiently checked by replacing the copositive cone with its inner semidefinite approximations described in Online Appendix B. We also impose the following condition for the theoretical convergence guarantee of the algorithm.

Definition 3 (Sufficiently Expensive Recourse). The two-stage DRO problem (1) has sufficiently expensive recourse if for any fixed $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, the dual of the recourse problem (2) is feasible.

The sufficiently expensive recourse condition guarantees that for any fixed $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, the residual problem (21) is not unbounded, that is, $Z_k^{\mathrm{PDR}}(\boldsymbol{x}, \theta) > -\infty$, for every $k \in [K]$.

We now describe the framework of the decomposition algorithm. For fixed first-stage decision variables $\hat{\boldsymbol{x}}$ and $\hat{\theta}$, the kth subproblem solved by the algorithm corresponds to the dual of the COP (21), as follows:

$$\begin{aligned} Z_k^{\mathrm{PDR}^*}(\hat{\boldsymbol{x}},\hat{\theta}) := \sup\ & \frac{1}{2}\sum_{\ell} \operatorname{tr}\Big(\boldsymbol{H}_\ell^k(\mathbf{e}_{S+1}\mathcal{T}_\ell(\hat{\boldsymbol{x}})^\top + \mathcal{T}_\ell(\hat{\boldsymbol{x}})\mathbf{e}_{S+1}^\top)\Big) - \sum_\ell \kappa_\ell \hat{\theta}\mathbf{e}_{S+1}^\top \boldsymbol{H}_\ell^k \mathbf{e}_{S+1} \\ \text{s.t.}\ & \boldsymbol{G}_k \in \mathbb{R}^{(S+1)\times(S+1)},\ \boldsymbol{O}_k \in \mathbb{S}^{S+1},\ \boldsymbol{H}_\ell^k \in \mathbb{S}^{S+1} \quad \forall \ell \in [L+2] \\ & \|\boldsymbol{G}_k\|_F \le \epsilon_k,\ \sum_\ell \mathcal{W}_\ell \boldsymbol{H}_\ell^k = \mathbf{0} \\ & \mathbf{e}_{S+1}^\top \boldsymbol{O}_k \mathbf{e}_{S+1} = 1,\ \mathbf{e}_{S+1}^\top \boldsymbol{H}_\ell^k \mathbf{e}_{S+1} \ge 0 \quad \forall \ell \in [L+2] \\ & \hat{\boldsymbol{\Omega}}_k + \boldsymbol{G}_k - \boldsymbol{O}_k = \mathbf{0},\ \hat{\boldsymbol{\Omega}}_k + \boldsymbol{G}_k - \sum_\ell \lambda_\ell \boldsymbol{H}_\ell^k = \mathbf{0} \\ & \boldsymbol{O}_k \in \mathcal{C}^*(\mathcal{K}_k),\ \boldsymbol{H}_\ell^k \in \mathcal{C}^*(\mathcal{K}_k) \quad \forall \ell \in [L+2]. \end{aligned} \tag{23}$$

The following proposition states that the primal and dual problems are equivalent.

Proposition 5. *Suppose the complete recourse under linear decision rule assumption is satisfied. Then, strong duality holds between the copositive program* (21) *and the completely positive program* (23), *that is,* $Z_k^{\mathrm{PDR}}(\hat{\boldsymbol{x}},\hat{\theta}) = Z_k^{\mathrm{PDR}^*}(\hat{\boldsymbol{x}},\hat{\theta})$.

At every iteration, the first-stage decision variables $\hat{\boldsymbol{x}}$ and $\hat{\theta}$ are solved via the *master problem* given by

$$\begin{aligned} \underline{J}^{\mathrm{PDR}} := \inf\ & \boldsymbol{c}^\top \boldsymbol{x} + \theta + \frac{1}{\delta}(\gamma\omega - \eta - 2\hat{\boldsymbol{p}}^\top \boldsymbol{r} + 2\omega\hat{\boldsymbol{p}}^\top \mathbf{e}) \\ \text{s.t.}\ & \boldsymbol{x} \in \mathcal{X},\ \theta,\eta \in \mathbb{R},\ \omega \in \mathbb{R}_+,\ \boldsymbol{r},\boldsymbol{s} \in \mathbb{R}^K \\ & s_k + \eta \le \omega,\ \sqrt{4r_k^2 + (s_k+\eta)^2} \le 2\omega - s_k - \eta \quad \forall k \in [K] \end{aligned}$$

$$\begin{aligned} & \frac{1}{2}\sum_\ell \operatorname{tr}\Big(\boldsymbol{H}_\ell^k(\mathbf{e}_{S+1}\mathcal{T}_\ell(\boldsymbol{x})^\top + \mathcal{T}_\ell(\boldsymbol{x})\mathbf{e}_{S+1}^\top)\Big) \\ & -\sum_\ell \kappa_\ell \theta \mathbf{e}_{S+1}^\top \boldsymbol{H}_\ell^k \mathbf{e}_{S+1} \le s_k \quad \forall (\boldsymbol{H}_1^k,\dots,\boldsymbol{H}_L^k) \in \mathcal{V}^k \quad \forall k \in [K], \end{aligned} \tag{24a}$$

$$\begin{aligned} & \frac{1}{2}\sum_\ell \operatorname{tr}\Big(\boldsymbol{H}_\ell^k(\mathbf{e}_{S+1}\mathcal{T}_\ell(\boldsymbol{x})^\top + \mathcal{T}_\ell(\boldsymbol{x})\mathbf{e}_{S+1}^\top)\Big) \\ & -\sum_\ell \kappa_\ell \theta \mathbf{e}_{S+1}^\top \boldsymbol{H}_\ell^k \mathbf{e}_{S+1} \le 0 \quad \forall (\boldsymbol{H}_1^k,\dots,\boldsymbol{H}_L^k) \in \mathcal{W}^k \quad \forall k \in [K], \end{aligned} \tag{24b}$$

which involves tractable linear and second-order cone constraints only.

The constraint systems in (24a) and (24b) serve as the *optimality cuts* and the *feasibility cuts*, respectively. The left-hand sides of the constraints correspond to the objective function of Subproblem (23). To generate an *optimality cut* when the subproblem is feasible but not optimal ($\hat{s}_k \le Z_k^{\mathrm{PDR}^*}(\hat{\boldsymbol{x}},\hat{\theta}) < \infty$), we add the solution $(\boldsymbol{H}_\ell^k)_{\ell\in[L+2]}$ to the set $\mathcal{V}^k$ by solving Subproblem (23). Conversely, to generate a *feasibility cut* in the master problem when Subproblem (23) is unbounded ($Z_k^{\mathrm{PDR}^*}(\hat{\boldsymbol{x}},\hat{\theta}) = \infty$), we add the solution $(\boldsymbol{H}_\ell^k)_{\ell\in[L+2]}$ to the set $\mathcal{W}^k$ by solving

$$\begin{aligned} (\boldsymbol{H}_\ell^k)_{\ell\in[L+2]} \in \operatorname{argmax}\ & \frac{1}{2}\sum_\ell (\operatorname{tr}(\boldsymbol{H}_\ell^k(\mathbf{e}_{S+1}\mathcal{T}_\ell(\hat{\boldsymbol{x}})^\top + \mathcal{T}_\ell(\hat{\boldsymbol{x}})\mathbf{e}_{S+1}^\top))) - \sum_\ell \kappa_\ell \hat{\theta}\mathbf{e}_{S+1}^\top \boldsymbol{H}_\ell^k \mathbf{e}_{S+1} \\ \text{s.t.}\ & \boldsymbol{G}_k \in \mathbb{R}^{(S+1)\times(S+1)},\ \boldsymbol{O}_k \in \mathbb{S}^{S+1},\ \boldsymbol{H}_\ell^k \in \mathbb{S}^{S+1} \quad \forall \ell \in [L+2] \\ & \|\boldsymbol{G}_k\|_F \le \epsilon_k,\ \sum_\ell \mathcal{W}_\ell \boldsymbol{H}_\ell^k = \mathbf{0} \\ & \mathbf{e}_{S+1}^\top \boldsymbol{O}_k \mathbf{e}_{S+1} = 0,\ \mathbf{e}_{S+1}^\top \boldsymbol{H}_\ell^k \mathbf{e}_{S+1} \ge 0 \quad \forall \ell \in [L+2] \\ & \boldsymbol{G}_k - \boldsymbol{O}_k = \mathbf{0},\ \boldsymbol{G}_k - \sum_\ell \lambda_\ell \boldsymbol{H}_\ell^k = \mathbf{0} \\ & \boldsymbol{O}_k \in \mathcal{C}^*(\mathcal{K}_k),\ \boldsymbol{H}_\ell^k \in \mathcal{C}^*(\mathcal{K}_k) \quad \forall \ell \in [L+2] \\ & -\mathbf{e}\mathbf{e}^\top \le \boldsymbol{G}_k, \boldsymbol{O}_k, \boldsymbol{H}_\ell^k \le \mathbf{e}\mathbf{e}^\top \quad \forall \ell \in [L+2]. \end{aligned} \tag{25}$$

Here, $\mathbf{ee}^\top$ denotes the matrix of ones. Problem (25) finds the direction in which Subproblem (23) becomes unbounded. The decomposition algorithm consecutively adds the optimal solution $(\boldsymbol{H}_\ell^k)_{\ell\in[L+2]}$ obtained from solving Problems (23) and (25) to the sets $\mathcal{V}^k$ and $\mathcal{W}^k$ at each iteration.

Based on the previous description, we provide the decomposition procedure in Algorithm 1.

Algorithm 1 (Decomposition Algorithm for Two-Stage DRO)

1: **Input:** Parameters of Problem (5), tolerance level $\eta \geq 0$

2: **Output:** Solution $(\boldsymbol{x}^*, \theta^*)$ to Problem (5) within $100\eta\%$ of optimal

3: INITIALIZE $\overline{J}^{\text{PDR}} = \infty$, $\mathcal{V}^k = \mathcal{W}^k = \emptyset \quad \forall k \in [K]$

4: Step 1 Solve the master Problem (24) and obtain the optimal value $\underline{J}^{\text{PDR}}$. If $\overline{J}^{\text{PDR}} - \underline{J}^{\text{PDR}} \leq \eta \min(|\overline{J}^{\text{PDR}}|, |\underline{J}^{\text{PDR}}|)$, stop.

5: Step 2 For $k \in [K]$, given fixed $(\hat{\boldsymbol{x}}, \hat{\theta})$, solve the Subproblem (23) and get the optimal value $Z_k^{\text{PDR}^*}(\hat{\boldsymbol{x}}, \hat{\theta})$. Solve Problem (20) with input $Z_k^{\text{PDR}^*}(\hat{\boldsymbol{x}}, \hat{\theta})$ and get the optimal value $\hat{J}^{\text{PDR}}$. If $\hat{J}^{\text{PDR}} < \overline{J}^{\text{PDR}}$, set $\overline{J}^{\text{PDR}} = \hat{J}^{\text{PDR}}$, and $(\boldsymbol{x}^*, \theta^*) = (\hat{\boldsymbol{x}}, \hat{\theta})$.

6: Step 3 For $k \in [K]$, if $Z_k^{\text{PDR}^*}(\hat{\boldsymbol{x}}, \hat{\theta}) = \infty$, add the feasibility cut by solving Problem (25) and updating $\mathcal{W}_k = \mathcal{W}_k \cup (\boldsymbol{H}_1^k, \ldots, \boldsymbol{H}_L^k)$; else if $Z_k^{\text{PDR}^*}(\hat{\boldsymbol{x}}, \hat{\theta}) < \infty$, add the optimality cut by solving Problem (23) and updating $\mathcal{V}_k = \mathcal{V}_k \cup (\boldsymbol{H}_1^k, \ldots, \boldsymbol{H}_L^k)$.

The algorithm terminates in finitely many iterations, as stated in the next theorem.

Theorem 3 (Finite η-Convergence). *Assume $\mathcal{X}$ is a nonempty compact convex set or a finite discrete set. If the two-stage DRO problem (1) satisfies both the complete recourse under linear decision rules and sufficiently expensive recourse assumptions, then Algorithm 1 converges in a finite number of steps for any given tolerance level η.*

For the special case where the first-stage decision vector $\boldsymbol{x}$ is integer-valued, the benefit of the decomposition algorithm is substantial because it breaks up the complexity of the original problem, which constitutes a mixed-integer COP. Indeed, the master problem is now a more tractable mixed-integer second-order cone program, for which off-the-shelf solvers are applicable. Furthermore, as the feasible set of the first-stage decision variables is finite, the decomposition algorithm can solve the problem to optimality in a finite number of iterations (Geoffrion 1972, theorem 2.4).

As in the development of the inner semidefinite approximations for the copositive programs, we derive the corresponding semidefinite approximations in Online Appendix B for the completely positive programs (23) and (25) in the decomposition algorithm. After the approximation, each semidefinite subproblem is solvable in $O(L^3(SL+S^2+N_2S)^{\frac{1}{2}} + L^2(SL+S^2+N_2S)^{\frac{5}{2}} + (SL+S^2+N_2S)^{\frac{7}{2}})$ time, whereas the time complexity of solving the second-order conic master problem is $O((N_1+K)^3K^{\frac{1}{2}}n^{\frac{1}{2}} + (N_1+K)K^{\frac{3}{2}}n^{\frac{3}{2}})$ at iteration n (Ben-Tal and Nemirovski 2001). In Online Appendix B, we provide further detailed comparisons of different semidefinite approximations.

5. Computational Study

In this section, we demonstrate the effectiveness of our copositive programming approach to two-stage DRO problems on two applications. In Section 5.1, we compare various data-driven methods and describe the platform on which we run the optimization models. Section 5.2 deals with a network inventory allocation problem that uses expectation as the risk measure and has random recourse costs in the objective function. The multi-item newsvendor problem in Section 5.3 uses CVaR as a risk measure, where its expectation reformulation involves random recourse matrices in the constraints. Two additional applications are discussed in the online appendix: the medical scheduling problem in Online Appendix C.1 and the facility location problem in Online Appendix C.2.

5.1. Setup

We compare different methods by evaluating their out-of-sample performance and computation time. The various data-driven methods include the following:

- **C_0 SDP**: The semidefinite approximation using the cone $\mathcal{IA}_0(\mathcal{K})$ defined in (B.1) in Online Appendix B for the copositive reformulation (19), where the ambiguity set is defined in (8).
- **C_1 SDP**: The semidefinite approximation using the cone $\mathcal{IA}_1(\mathcal{K})$ defined in (B.3) in Online Appendix B for the copositive reformulation (19), where the ambiguity set is defined in (8).
- **Wass SDP**: The semidefinite approximation for the copositive reformulation of two-stage DRO with the Wasserstein ambiguity set, proposed by Hanasusanto and Kuhn (2018). We only apply this method in the first

application because this scheme is only applicable for solving two-stage DRO problems with random recourse costs in the objective.

- **SAA**: The sample-average approximation.

All optimization problems are solved using MOSEK 9.2.28 via the YALMIP interface (a toolbox in MATLAB; Löfberg 2004) on a 10-core, 2.8-GHz Windows PC with 32 GB of RAM. All implementation codes are available in the Github repository (Fan and Hanasusanto 2023).

5.2. Network Inventory Allocation

5.2.1. Problem Description. We consider a two-stage capacitated network inventory problem (Bertsimas et al. 2022) with M locations where one should determine the stock allocations x_i, $i \in [M]$, before knowing the realizations of demands u_i, $i \in [M]$, and per-unit transportation costs v_{ij} from location i to j, $i, j \in [M]$. The unit cost of buying stock in advance at location i is c_i. The demand u_i in location i can be satisfied by existing stock x_i and by transporting y_{ji} units from location j. The stock capacity at each location is T units. The network inventory allocation problem minimizes the worst-case storage and transportation cost given by

$$\begin{aligned} \min \quad & \boldsymbol{c}^\top \boldsymbol{x} + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[Z(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v})] \\ \text{s.t.} \ & \boldsymbol{x} \in \mathbb{R}^M, \ 0 \le x_i \le T \quad \forall i \in [M], \end{aligned}$$

where the second-stage cost is defined as

$$\begin{aligned} Z(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) := \min & \sum_{i \in [M]} \sum_{j \in [M]} v_{ij} y_{ij} \\ \text{s.t.} \quad & \boldsymbol{y} \in \mathbb{R}_+^{M \times M} \\ & x_i + \sum_{j \in [M]} y_{ji} - \sum_{j \in [M]} y_{ij} \ge u_i \quad \forall i \in [M]. \end{aligned}$$

In this problem, we assume the only information provided is the historical observations of the random parameters, $(\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{v}}_1), \ldots, (\hat{\boldsymbol{u}}_N, \hat{\boldsymbol{v}}_N)$, and the description of the support set $\Xi = \{(\boldsymbol{u}, \boldsymbol{v}) \in (\mathbb{R}^M, \mathbb{R}^{M \times M}) : \underline{u}_l \le u_i \le \overline{u}_u, \underline{v}_l \le v_{ij} \le \overline{v}_u, i, j \in [M]\}$.

5.2.2. Experimental Setting. Because the problem has random recourse costs in the objective function and fixed recourse matrices in the constraints, we apply the PLDR introduced in Section 3.1 to the second-stage decision variable $\boldsymbol{y}(\boldsymbol{u}, \boldsymbol{v})$. Throughout our computational study, we use all N samples to construct the ambiguity set for the DRO model. The constructor points are drawn in two ways: (1) from the empirical distribution that assigns equal mass $1/N$ to all the historical samples and (2) from an independent uniform distribution on the support. In the first case, we do not adopt the sample-splitting scheme to ensure the independence between the samples used for the ambiguity set and the constructor points. Instead, we make full use of all available data to build both the ambiguity set and the Voronoi diagram and construct more balanced partitions such that the numbers of samples in different partitions are evenly distributed. This is done to achieve the best out-of-sample performance. In the numerical study, we approximate the copositive problems as semidefinite programs using the cones $\mathcal{IA}_0(\mathcal{K})$ and $\mathcal{IA}_1(\mathcal{K})$. We denote the problems as $\mathbf{C_0}$ **SDP** and $\mathbf{C_1}$ **SDP** when the constructor points are drawn from the empirical distribution; and as $\mathbf{C}_0'$ **SDP** and $\mathbf{C}_1'$ **SDP** when the constructor points are drawn from the uniform distribution. We compare our SDP solution schemes with **Wass SDP** and **SAA**. Furthermore, we apply Algorithm 1 and show the advantages of using the decomposition algorithm in accelerating the solution process. To calculate the runtime of the decomposition algorithm, we choose the maximum among the runtimes of solving the K subproblems and add that to the total time of running the algorithm.

We evaluate the out-of-sample performance of the previous solution schemes over 100 instances by randomly generating $N = 10, 20, 40, 80, 160$ independent training samples, which are taken to be the historical data, and 50,000 independent testing samples. We set the number of partitions K equal to the size of the historical data set, N. For each instance, we construct a network of moderate size $M = 5$ with inventory costs $\boldsymbol{c} = [40, 50, 60, 70, 80]$. We set $T = 80$, $\underline{u}_l = 20$, $\overline{u}_u = 40$, $\underline{v}_l = 40$, $\overline{v}_u = 50$. We assume the true distributions of the demands $\boldsymbol{u}$ are truncated lognormal, which is a conditional distribution that results from restricting the domain of a lognormal distribution to the support Ξ, with means $\boldsymbol{\mu}_1 = [3, 3, 3.5, 3.5, 3.5]$ and standard deviations $\boldsymbol{\sigma}_1 = 0.2\mathbf{e}$. The true distributions of the per-unit transportation costs $\boldsymbol{v}$ are also assumed to be truncated lognormal with means $\boldsymbol{\mu}_2 = 3\mathbf{e}$ and standard deviations $\boldsymbol{\sigma}_2 = 0.1\mathbf{e}$. Based on the theoretical decay rate in Theorem 1, the robustness parameters are set to $\epsilon_k = \frac{R_k^2}{\sqrt{\mathcal{I}_k}} \epsilon_k'$ and $\gamma = \frac{K}{N} \gamma'$, where the values of ϵ_k', $k \in [K]$, and γ' are determined using a twofold cross-validation

Table 1. Average Out-of-Sample Cost for the Network Inventory Problem of Different Approaches

N	C_0 SDP	C_1 SDP	C_0' SDP	C_1' SDP	Wass SDP
10	11,999	11,627	12,000	11,774	11,516
20	12,000	11,622	12,000	11,289	—
40	12,000	—	12,000	—	—
80	12,000	—	12,000	—	—
160	12,000	—	12,000	—	—

procedure so that they give the best performance on the validation set (see the work by Hastie et al. (2009) for a review of the procedure). Numerical results are averaged over 100 random instances.

5.2.3. Analysis of the Results. We calculate the average proportion of realizations where the first-stage decision x is feasible in the out-of-sample tests. The numerical results show that, as expected, all the SDP solution schemes are guaranteed to produce feasible first-stage decisions. Conversely, SAA exhibits poor feasibility performance. It produces no feasible solution on these instances when the number of in-sample data N is less than 80. As such, we no longer consider SAA's out-of-sample performance and runtime in subsequent comparisons for this application.

Table 1 compares the out-of-sample performance of different approaches. In all applications, the time limit is set to 900 seconds. The emdash indicates that the method cannot solve the problem within the specified time limit. In addition, because the objective value obtained from the decomposition method is within the optimality tolerance η of that associated with directly solving the corresponding semidefinite program, we no longer report its value. Under this setting, the optimal value is approximately 10,408, which is computed by solving **SAA** with a large enough sample size, for example, $N = 50,000$. We observe that the out-of-sample costs of **$\mathbf{C_1}$ SDP** and **$\mathbf{C_1'}$ SDP** decrease as the number of in-sample points increases. The semidefinite approximations that use the inner cones $\mathcal{IA}_0(\mathcal{K})$ perform slightly worse than the ones based on $\mathcal{IA}_1(\mathcal{K})$. Also, we can see that it is better to choose the constructor points according to the empirical distribution because the partitions are more balanced. The results in Table 1 further show that, although **Wass SDP** can be solved within the imposed time limit only when the number of samples is 10, in this case, it outperforms the other methods, slightly for **$\mathbf{C_1}$ SDP**, in terms of out-of-sample cost.

With regard to computational costs, **$\mathbf{C_0}$ SDP**, **$\mathbf{C_1}$ SDP**, **$\mathbf{C_0'}$ SDP,** and **$\mathbf{C_1'}$ SDP** have notably lower runtime than **Wass SDP** as shown in Table 2. One possible reason is that **Wass SDP** solves a large SDP problem with $O(N)$ semidefinite constraints involving matrices of size $O((S+N_2+L)^2)$. In comparison, the other semidefinite methods solve an SDP problem with $O(KL)$ semidefinite constraints involving smaller matrices of size $O(S^2)$. However, the SDP solution schemes become less time efficient as the number of partitions increases. Thus, it would be advantageous to use Algorithm 1, which solves the subproblems independently. We further observe that **$\mathbf{C_1}$ SDP** takes much longer time to solve and their runtime exceeds the time limit, that is, 900 seconds with more than 20 in-sample points, but they deliver similar out-of-sample performance to **$\mathbf{C_0}$ SDP**. Therefore, we conclude **$\mathbf{C_0}$ SDP** is a better choice when balancing optimality and scalability. Hence, we only apply the decomposition algorithm to solve **$\mathbf{C_0}$ SDP**, denoted as **Benders $\mathbf{C_0}$**. We set the tolerance level to $\eta = 0.05$ when solving the decomposition algorithm, which indicates that the objective costs of **Benders $\mathbf{C_0}$** are within 5% gap from those of **$\mathbf{C_0}$ SDP**. Numerically this algorithm significantly reduces the runtime compared with the original SDP schemes, particularly when the number of in-sample points is large.

5.3. Multi-Item Newsvendor

5.3.1. Problem Description. In this application, we explore an inventory management problem (Hanasusanto et al. 2015) where we want to decide the order quantities x_i, $i \in [M]$, of M products, before observing the random demands ξ_i, $i \in [M]$, and the random stockout costs s_i, $i \in [M]$. There is a limit B on the total order quantity. We incur a per-unit stockout cost s_i when the order quantity x_i is less than the demand ξ_i and a holding cost g_i if the order quantity x_i is more than the demand ξ_i. The second-stage variables $y_{1,i} := \max\{x_i - \xi_i, 0\}$ and $y_{2,i} := \max\{\xi_i - x_i, 0\}$ correspond to the excess amount and the shortfall amount, respectively. Our goal is to minimize

Table 2. Average Runtime (Seconds) for the Network Inventory Problem of Different Approaches

N	C_0 SDP	C_1 SDP	C_0' SDP	C_1' SDP	Benders C_0	Wass SDP
10	3.728	16.468	2.810	23.321	10.085	737.800
20	9.743	76.215	7.607	104.61	6.178	—
40	28.097	—	14.901	—	7.276	—
80	97.241	—	87.645	—	10.584	—
160	351.984	—	696.5	—	16.524	—

the worst-case CVaR of the total stockout and holding costs given by

$$\begin{aligned} &\min \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{P}\text{-CVaR}_\epsilon[Z(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{s})] \\ &\text{s.t. } \boldsymbol{x}\in\mathbb{R}_+^M,\ \mathbf{e}^\top\boldsymbol{x}\le B, \end{aligned}$$

where the second-stage cost is defined as

$$\begin{aligned} Z(\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{s}) := \inf\ & \boldsymbol{g}^\top\boldsymbol{y}_1 + \boldsymbol{s}^\top\boldsymbol{y}_2 \\ \text{s.t. } & \boldsymbol{y}_1\in\mathbb{R}_+^M,\ \boldsymbol{y}_2\in\mathbb{R}_+^M \\ & \boldsymbol{y}_1 \ge \boldsymbol{x}-\boldsymbol{\xi},\ \boldsymbol{y}_2\ge\boldsymbol{\xi}-\boldsymbol{x}. \end{aligned}$$

According to the definition of CVaR, it can be shown that this problem is equivalent to

$$\begin{aligned} &\inf\ \theta + \frac{1}{\delta}\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_\mathbb{P}[\tau(\boldsymbol{\xi},\boldsymbol{s})] \\ &\text{s.t. } \theta\in\mathbb{R},\ \boldsymbol{x}\in\mathbb{R}_+^M,\ \boldsymbol{y}_1:\Xi\to\mathbb{R}^M,\ \boldsymbol{y}_2:\Xi\to\mathbb{R}^M,\ \tau:\Xi\to\mathbb{R} \\ &\quad \mathbf{e}^\top\boldsymbol{x}\le B \\ &\quad \left.\begin{array}{l} \tau(\boldsymbol{\xi},\boldsymbol{s})\ge 0,\ \tau(\boldsymbol{\xi},\boldsymbol{s})\ge \boldsymbol{g}^\top\boldsymbol{y}_1(\boldsymbol{\xi},\boldsymbol{s})+\boldsymbol{s}^\top\boldsymbol{y}_2(\boldsymbol{\xi},\boldsymbol{s})-\theta \\ \boldsymbol{y}_1(\boldsymbol{\xi},\boldsymbol{s})\ge\mathbf{0},\ \boldsymbol{y}_1(\boldsymbol{\xi},\boldsymbol{s})\ge\boldsymbol{x}-\boldsymbol{\xi} \\ \boldsymbol{y}_2(\boldsymbol{\xi},\boldsymbol{s})\ge\mathbf{0},\ \boldsymbol{y}_2(\boldsymbol{\xi},\boldsymbol{s})\ge\boldsymbol{\xi}-\boldsymbol{x} \end{array}\right\}\ \forall(\boldsymbol{\xi},\boldsymbol{s})\in\Xi. \end{aligned} \tag{26}$$

We assume the only information provided is the historical data $(\hat{\boldsymbol{\xi}}_1,\hat{s}_1),\dots,(\hat{\boldsymbol{\xi}}_N,\hat{s}_N)$, and the description of the support set $\Xi=\{(\boldsymbol{\xi},\boldsymbol{s})\in(\mathbb{R}^M,\mathbb{R}^M):\underline{\xi}_l\le\xi_i\le\overline{\xi}_u,\ \underline{s}_l\le s_i\le\overline{s}_u,\ i\in[M]\}$.

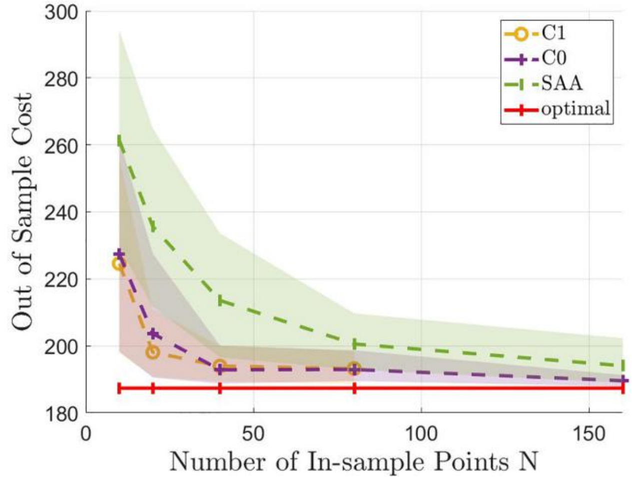
5.3.2. Experimental Setting. Reformulated Problem (26) has random recourse matrices in the constraints. After applying PLDR to the second-stage decision variables $\boldsymbol{y}_1(\boldsymbol{\xi},\boldsymbol{s})$ and $\boldsymbol{y}_2(\boldsymbol{\xi},\boldsymbol{s})$, the right-hand side of the third inequality contains a quadratic term of random parameters $(\boldsymbol{\xi},\boldsymbol{s})$. Thus, PQDR are applied to the second-stage decision variable $\tau(\boldsymbol{\xi},\boldsymbol{s})$ on the left-hand side. After applying the decision rules, we separately use the inner approximations $\mathcal{IA}_0(\mathcal{K})$ and $\mathcal{IA}_1(\mathcal{K})$ proposed in Online Appendix B and solve the corresponding semidefinite programs denoted as $\mathbf{C_0}$ **SDP** and $\mathbf{C_1}$ **SDP**. We further take advantage of the decomposition algorithm to reduce the runtime of the semidefinite programs.

To assess the performance of $\mathbf{C_0}$ **SDP**, $\mathbf{C_1}$ **SDP**, and **SAA**, we compute the out-of-sample performance of a realistic size multi-item newsvendor problem with $M = 5$ products. The training data sets are of sizes $N = 10, 20, 40, 80, 160$, whereas the testing data sets contain 50,000 independent samples. Here we select the constructor points according to the empirical distribution generated from the historical (i.e., training) data. We evaluate the out-of-sample performance by averaging the results obtained on 100 instances. For each instance, we set $B = 30$, $\underline{\xi}_l = 0$, $\overline{\xi}_u = 10$, $\underline{s}_l = 0$, $\overline{s}_u = 50$ and the holding cost vector $\boldsymbol{g} = [5, 6, 7, 8, 9]$. The parameter δ related to the risk attitude of CVaR is fixed to 0.1. The true distributions of the demands $\boldsymbol{\xi}$ are assumed to be truncated lognormal with means $\boldsymbol{\mu}_1 = \mathbf{e}$ and standard deviations $\boldsymbol{\sigma}_1 = \mathbf{e}$, and the true distributions of the stockout costs $\boldsymbol{s}$ are assumed to be truncated lognormal with means $\boldsymbol{\mu}_2 = 3\mathbf{e}$ and standard deviations $\boldsymbol{\sigma}_2 = 2\mathbf{e}$. We determine the values of the robustness parameters γ and ϵ_k, $k\in[K]$ in the ambiguity set using a twofold cross-validation procedure. The tolerance level η of the decomposition algorithm is set to 0.05.

5.3.3. Analysis of the Results. A comparison of the out-of-sample performance of $\mathbf{C_0}$ **SDP**, $\mathbf{C_1}$ **SDP**, and **SAA** is shown in Figure 1. In Figure 1, we observe that the out-of-sample costs of all three methods decrease with the training data size and approach the optimal cost. Both $\mathbf{C_0}$ **SDP** and $\mathbf{C_1}$ **SDP** achieve better out-of-sample performance with approximately 15% improvement compared with **SAA** when the in-sample data are limited (e.g., $n = 10$), whereas the optimality gaps of both the semidefinite approaches are about 20%. When the data size is less limited (e.g., $n = 160$), the gaps between the semidefinite approaches and **SAA** shrink to 3% as the solutions of **SAA** asymptotically converge to the optimal one. Meanwhile, the optimality gaps of both SDP solution schemes become fairly small at around 1%. Comparing the two SDP schemes, we notice that $\mathbf{C_1}$ **SDP** has a marginal advantage over $\mathbf{C_0}$ **SDP** in terms of out-of-sample performance, which indicates $\mathcal{IA}_1(\mathcal{K})$ provides a slightly better approximation than $\mathcal{IA}_0(\mathcal{K})$.

Table 3 shows the runtimes of all three previous methods, with the addition of **Benders** $\mathbf{C_0}$, which solves $\mathbf{C_0}$ **SDP** using the decomposition algorithm. The runtime of $\mathbf{C_1}$ **SDP** exceeds the time limit with 160 in-sample points. Our results confirm the benefit of applying the Benders decomposition algorithm, where the advantage grows with the number of in-sample data. Table 4 shows that the algorithm reaches a near-optimal solution in a small number of iterations and each subproblem is solved within one second.

Figure 1. (Color online) Out-of-Sample Cost for the Newsvendor Problem as a Function of the Number of In-Sample Data Points with a Fixed Number of Partitions $K = N/2$



Note. The dashed lines represent the average values, and the shaded areas depict the 10%–90% quantile range over 100 instances.

Table 3. Average Runtime (Seconds) for the Newsvendor Problem of Different Approaches with Fixed Number of Partitions $K = N/2$

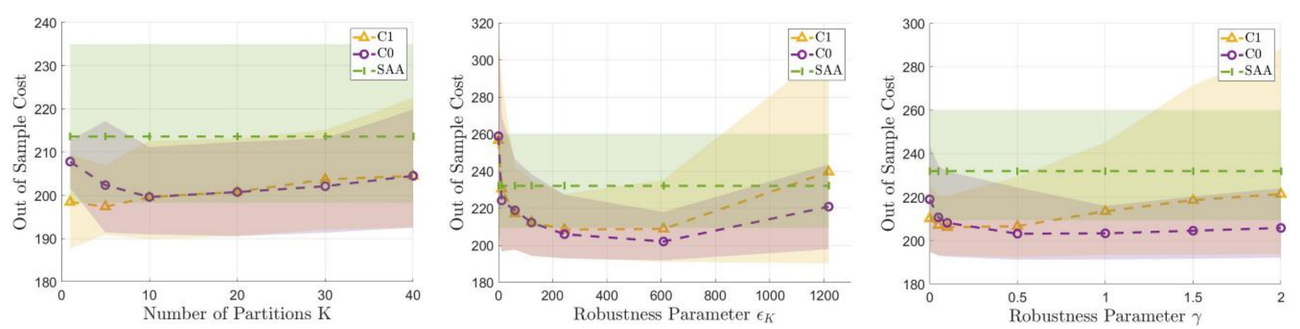
N	C_1 SDP	C_0 SDP	Benders C_0	SAA
10	0.572	0.181	0.992	0.003
20	1.965	0.370	1.001	0.004
40	11.313	0.926	1.032	0.006
80	66.086	2.819	1.074	0.013
160	—	6.614	1.664	0.022

Table 4. Computational Performance of the Decomposition Algorithm for the Newsvendor Problem

	Benders C_0		
N	Runtime	Maximum sub. runtime	No. of iterations
10	0.992	0.081	12
20	1.001	0.099	10
40	1.032	0.127	8
80	1.074	0.177	6
160	1.664	0.274	6

Note. All runtimes are in seconds.

Figure 2. (Color online) Out-of-Sample Cost for the Newsvendor Problem as a Function of the Values of Parameters K, ϵ_K, and γ



Notes. (Left) Out-of-sample cost for the newsvendor problem as a function of the number of partitions K with fixed $n = 40$ and parameters $\epsilon_k = 300\ \forall k \in [K]$, $\gamma = 0.5$. (Middle) Out-of-sample cost as a function of the value of parameter ϵ_K with fixed $n = 20$, $K = 5$, and parameter $\gamma = 0.5$. (Right) Out-of-sample cost as a function of the value of parameter γ with fixed n = 20, $K = 5$, and parameters $\epsilon_k = 300\ \forall k \in [K]$.

When using PDR, the number of partitions K affects the out-of-sample performance. In Figure 2 (left), we observe that the out-of-sample performance reaches its minimum at the midpoint. This could be attributed to the fact that increasing the number of partitions enhances the approximation power of PDR, yet the moment estimation within each partition becomes less accurate as a smaller number of samples fall into each partition. Consequently, the optimal number of partitions strikes a balance. In Figure 2 (middle), we investigate the impact of varying the robustness parameter ϵ_k on the performance of the considered methods. Here, the best value of ϵ_K is 600. This observation is consistent with the fact that a sufficiently small parameter value leads to overfitting, whereas a value that is too large yields overly conservative solutions. Either one of these situations gives rise to poor out-of-sample performance. Figure 2 (right) illustrates the impact of different values of the robustness parameter γ on the out-of-sample performance. We observe that, although the optimal value of γ is achieved at point 0.5, when we select the constructor points based on the empirical distribution, that is, set $\gamma = 0$, the performance is only slightly inferior to the optimal one, whereas the computational effort is reduced.

6. Conclusions

Efficient solution schemes for two-stage DRO problems with random recourse have not yet been developed due to their high level of intractability. In this paper, we leveraged the decision rule approach to alleviate the computational challenge of solving these problems and proposed a novel ambiguity set with two layers of robustness based on the decision rule structure. The resulting distributionally robust model admits an attractive out-of-sample performance guarantee that provides valuable guidance in choosing the robustness parameters in applications. We reformulated the model as a convex copositive program, which is amenable to tractable semidefinite programming approximations. We further developed a decomposition algorithm that can significantly speed up the solution time. This method enables us to efficiently solve the more difficult problem instances with integer first-stage decisions. Through a variety of numerical studies, we demonstrated that our solution method exhibits near-optimal out-of-sample performance with reasonable computational effort, even under limited sample sizes.

We see at least three opportunities for further investigation. First, it would be interesting to explore other partitioning schemes than the Voronoi regions for the piecewise decision rules, which may lead to solutions with different optimality and scalability properties. Second, the decomposition algorithm can naturally be implemented in a parallel computing environment because the subproblems can be solved independently. Investigating the computational implications of this approach could be an interesting avenue of exploration. Third, it would be beneficial to extend the proposed approach to address DRO problems with more than two stages, which can attain the theoretical performance guarantee while maintaining similar scalability in its solution algorithm.

Acknowledgments

The authors thank the area editor, associate editor, and referees for invaluable feedback and insightful comments and Jonathan Bard and Soroosh Shafieezadeh-Abadeh for valuable discussions and constructive suggestions that led to substantial improvements of the paper.

References

Atamtürk A, Zhang M (2007) Two-stage robust network flow and design under demand uncertainty. *Oper. Res.* 55:662–673.

Aurenhammer F (1991) Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Comput. Survey* 23(3):345–405.

Bampou D, Kuhn D (2011) Scenario-free stochastic programming with polynomial decision rules. *Proc. 50th IEEE Conf. on Decision and Control and Eur. Control Conf.* (IEEE, Piscataway, NJ), 7806–7812.

Bayraksan G, Love DK (2015) Data-driven stochastic programming using phi-divergences. *TutORials in Operations Research* (INFORMS, Catonsville, MD), 1–19.

Ben-Tal A, Nemirovski A (2001) *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications* (Society for Industrial and Applied Mathematics, Philadelphia).

Ben-Tal A, El Housni O, Goyal V (2020) A tractable approach for designing piecewise affine policies in two-stage adjustable robust optimization. *Math. Programming* 182(1):57–102.

Ben-Tal A, Golany B, Nemirovski A, Vial JP (2005) Retailer-supplier flexible commitments contracts: A robust optimization approach. *Manufacturing Service Oper. Management* 7(3):248–271.

Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. *Math. Programming* 99(2):351–376.

Ben-Tal A, Den Hertog D, De Waegenaere A, Melenberg B, Rennen G (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Sci.* 59(2):341–357.

Benders JF (1962) Partitioning procedures for solving mixed-variables programming problems. *Numerical Math.* 4(1):238–252.

Bertsimas D, Dunning I (2016) Multistage robust mixed-integer optimization with adaptive partitions. *Oper. Res.* 64(4):980–998.

Bertsimas D, Georghiou A (2015) Design of near optimal decision rules in multistage adaptive mixed-integer optimization. *Oper. Res.* 63:610–627.

Bertsimas D, Georghiou A (2018) Binary decision rules for multistage adaptive mixed-integer optimization. *Math. Programming* 167(2):395–433.

Bertsimas D, Gupta V, Kallus N (2018) Robust sample average approximation. *Math. Programming* 171(1):217–282.

Bertsimas D, Iancu DA, Parrilo PA (2011) A hierarchy of near-optimal policies for multistage adaptive optimization. *IEEE Trans. Automated Control* 56(12):2809–2824.

Bertsimas D, Shtern S, Sturt B (2022) Two-stage sample robust optimization. *Oper. Res.* 70(1):624–640.

Bertsimas D, Doan XV, Natarajan K, Teo CP (2010) Models for minimax stochastic linear optimization problems with risk aversion. *Math. Oper. Res.* 35(3):580–602.

Birge JR, Louveaux F (2011) *Introduction to Stochastic Programming* (Springer Science & Business Media, Boston).

Boyd S, Boyd SP, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).

Burer S (2012) Copositive programming. Anjos MF, Lasserre JB, eds. *Handbook on Semidefinite, Conic and Polynomial Optimization.* International Series in Operations Research & Management Science (Springer, Berlin), 201–218.

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

Chen Z, Sim M, Xiong P (2020) Robust stochastic optimization made easy with rsome. *Management Sci.* 66(8):3329–3339.

Cheramin M, Cheng J, Jiang R, Pan K (2022) Computationally efficient approximations for distributionally robust optimization under moment and Wasserstein ambiguity. *INFORMS J. Comput.* 34(3):1768–1794.

Delage E, Saif A (2022) The value of randomized solutions in mixed-integer distributionally robust optimization problems. *INFORMS J. Comput.* 34(1):333–353.

Delage E, Ye Y (2010) Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.* 58(3):595–612.

Esfahani PM, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Math. Programming* 171(1):115–166.

Fan X, Hanasusanto GA (2023) Github repository: A decision rule approach for two-stage data-driven distributionally robust optimization problems with random recourse. https://dx.doi.org/10.1287/ijoc.2021.0306.cd, https://github.com/INFORMSJoC/2021.0306.

Gao R (2022) Finite-sample guarantees for Wasserstein distributionally robust optimization: Breaking the curse of dimensionality. *Oper. Res.*, ePub ahead of print July 20, https://doi.org/10.1287/opre.2022.2326.

Garstka SJ, Wets RJB (1974) On decision rules in stochastic programming. *Math. Programming* 7:117–143.

Geoffrion AM (1972) Generalized Benders decomposition. *J. Optim. Theory Appl.* 10(4):237–260.

Georghiou A, Kuhn D, Wiesemann W (2019) The decision rule approach to optimization under uncertainty: Methodology and applications. *Comput. Management Sci.* 16(4):545–576.

Guslitser E (2002) Uncertainty-immunized solutions in linear programming. PhD thesis, Technion–Israel Institute of Technology, Haifa, Israel.

Hanasusanto GA, Kuhn D (2018) Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Oper. Res.* 66:849–869.

Hanasusanto GA, Kuhn D, Wallace SW, Zymler S (2015) Distributionally robust multi-item newsvendor problems with multimodal demand distributions. *Math. Programming* 152(1):1–32.

Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2 (Springer, Berlin).

Jiang R, Guan Y (2018) Risk-averse two-stage stochastic program with distributional ambiguity. *Oper. Res.* 66(5):1390–1405.

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

Laurent B, Massart P (2000) Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* 28(5):1302–1338.

Li Y, Li X, Shu J, Song M, Zhang K (2022) A general model and efficient algorithms for reliable facility location problem under uncertain disruptions. *INFORMS J. Comput.* 34(1):407–426.

Löfberg J (2004) Yalmip: A toolbox for modeling and optimization in matlab. *Proc. IEEE Internat. Conf. on Robotics and Automation* (IEEE, Piscataway, NJ), 284–289.

Martins da Silva Rocha PC (2013) Medium-term planning in deregulated energy markets with decision rules. PhD thesis, Imperial College London, London.

Pardo L (2018) *Statistical Inference Based on Divergence Measures* (CRC Press, Boca Raton, FL).

Postek K, Hertog DD (2016) Multistage adjustable robust mixed-integer optimization via iterative splitting of the uncertainty set. *INFORMS J. Comput.* 28(3):553–574.

Rahimian H, Mehrotra S (2019) Distributionally robust optimization: A review. Preprint, submitted August 13, https://arxiv.org/abs/1908.05659.

Rocha P, Kuhn D (2012) Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules. *Eur. J. Oper. Res.* 216(2):397–408.

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *J. Risk* 2:21–42.

Shawe-Taylor J, Cristianini N (2003) Estimating the moments of a random vector with applications. *Proc. GRETSI 2003 Conf.*, 47–52.

Van Parys BP, Esfahani PM, Kuhn D (2021) From data to decisions: Distributionally robust optimization is optimal. *Management Sci.* 67(6):3387–3402.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Oper. Res.* 62(6):1358–1376.

Xu G, Hanasusanto GA (2023) Improved decision rule approximations for multistage robust optimization via copositive programming. *Oper. Res.*, ePub ahead of print September 21, https://doi.org/10.1287/opre.2018.0505.