Decision Rule Approaches for Pessimistic Bilevel Linear Programs under Moment Ambiguity with Facility Location Applications

Akshit Goyal
Department of Industrial and Systems Engineering, University of Minnesota, goyal080@umn.edu

Yiling Zhang
Department of Industrial and Systems Engineering, University of Minnesota, yiling@umn.edu

Chuan He
Department of Industrial and Systems Engineering, University of Minnesota, he000233@umn.edu

We study a pessimistic stochastic bilevel program in the context of sequential two-player games, where the leader makes a binary here-and-now decision, and the follower responds a continuous wait-and-see decision after observing the leader's action and revelation of uncertainty. Only the information of the mean, covariance, and support is known. We formulate the problem as a distributionally robust (DR) two-stage problem. The pessimistic DR bilevel program is shown to be equivalent to a generic two-stage distributionally robust stochastic (nonlinear) program with both a random objective and random constraints under proper conditions of ambiguity sets. Under continuous distributions, using linear decision rule approaches, we construct upper bounds on the pessimistic DR bilevel program based on (1) 0-1 semidefinite programming (SDP) approximation and (2) an exact 0-1 copositive programming reformulations. When the ambiguity set is restricted to discrete distributions, an exact 0-1 SDP reformulation is developed, and explicit construction of the worst-case distribution is derived. To further improve the computation of the proposed 0-1 SDPs, a cutting-plane framework is developed. Moreover, based on a mixed-integer linear programming approximation, another cutting-plane algorithm is proposed. Extensive numerical studies are conducted to demonstrate the effectiveness of the proposed approaches on a facility location problem.

Key words: Distributionally Robust Optimization, Pessimistic Bilevel Program, Semidefinite Program, Copositive Program, Linear Decision Rules

1. Introduction

Two-stage stochastic bilevel programming often arises in the context of sequential two-player games where players strive to optimize their individual objectives under uncertainty. In the game, the leader makes a decision first to optimize their utility function, and then the follower responds after observing the uncertain parameters and the leader's action. Specifically, the leader first chooses a here-and-now decision $x \in \mathcal{X} \subseteq \{0,1\}^d$, before the revelation of uncertain parameter $\xi \in \mathbb{R}^k$. Next, after the revelation, the follower makes a wait-and-see decision $y \in \mathbb{R}^n$ to minimize their utility function $c(\xi)^\top y$ subject to $Ay \leq b_x(\xi)$ for some $A \in \mathbb{R}^{m\times n}$ and $b_x(\xi) \in \mathbb{R}^m$. We emphasize that the leader's decision affects the follower's feasible region as the right-hand side $b_x(\xi)$ depends on the leader's decision x. By taking the follower's potential choice y into account, the leader minimizes their direct cost $w^\top x$ and an indirect expected cost $\mathbb{E}_F[v(\xi)^\top y]$ via the impact on the follower's feasible region. That is, the leader minimizes $w^\top x + \mathbb{E}_F[v(\xi)^\top y]$ for some $w \in \mathbb{R}^d$ and $v(\xi) \in \mathbb{R}^n$, where $\mathbb{E}_F[\cdot]$ denotes the expectation with respect to distribution F of ξ.

One important concern in solving bilevel optimization problems regards the follower's decision when multiple alternative optimal solutions are present. The leader's decision might have very different values depending on the choice of follower's optimal solutions. Denote the set of alternative optimal solutions for the follower $\Omega(x,\xi) := \arg\min_y \left\{c(\xi)^\top y :\ Ay \leq b_x(\xi)\right\} \subset \mathbb{R}^n$. First, assuming that the follower is cooperative, i.e., the follower chooses the solution in favor of the leader, the *optimistic* stochastic bilevel program is formulated as

$$\min_{x\in\mathcal{X}} w^\top x + \mathbb{E}_F\left[\min_{y\in\Omega(x,\xi)} v(\xi)^\top y\right]. \tag{1}$$

The solution of (1) can be risky for the leader assuming cooperation on the follower. The *pessimistic* formulation considers the worst-case scenario in the follower's optimal solution set $\Omega(x,\xi)$ as follows.

$$\min_{x\in\mathcal{X}} w^\top x + \mathbb{E}_F\left[\max_{y\in\Omega(x,\xi)} v(\xi)^\top y\right]. \tag{2}$$

Both the optimistic and the pessimistic formulations have been recently studied in Yanıkoğlu and Kuhn (2018). Using linear decision rules (LDRs), they propose an upper-bound mixed-integer linear

programming (MILP) approximation of the pessimistic problem and a lower-bound MILP approximation of the optimistic problem. The approximations are robust with respect to distributions matching the first- and second-order (empirical) moments.

However, there may exist estimation errors in the moments when, for example, the historical data is inadequate. As a consequence, the solution obtained from the stochastic bilevel program may yield poor out-of-sample performance (which is demonstrated in Section 7.5). In this paper, we employ a set of plausible probability distributions, termed as $\mathcal{D}$, to take into account the "moment ambiguity". From a robust perspective, we seek for a solution to the bilevel programs that hedges against all probability distributions belonging to the ambiguity set $\mathcal{D}$, as follows.

$$(\mathcal{O}): \quad \min_{x\in\mathcal{X}} w^\top x + \sup_{F\in\mathcal{D}} \mathbb{E}_F \left[\min_{y\in\Omega(x,\xi)} v(\xi)^\top y \right], \tag{3}$$

$$(\mathcal{P}): \quad \min_{x\in\mathcal{X}} w^\top x + \sup_{F\in\mathcal{D}} \mathbb{E}_F \left[\max_{y\in\Omega(x,\xi)} v(\xi)^\top y \right]. \tag{4}$$

We call $(\mathcal{O})$ and $(\mathcal{P})$ the optimistic and pessimistic distributionally robust bilevel programs (DRBPs). They minimize the worst-case expected indirect costs of the follower's decision regarding the ambiguity set $\mathcal{D}$. The two DRBPs can be viewed as special two-stage distributionally robust stochastic programs with $\arg\min$ operators in $\Omega(x,\xi)$ which returns the set of optimizers of the follower's utility minimization programs.

1.1. Assumptions and Ambiguity Sets

In this paper, we make the following assumptions on the DRBPs.

ASSUMPTION 1. *(Square integrability) We assume that $c(\xi), v(\xi) \in \mathcal{L}_n^2(F)$ and $b_x(\xi) \in \mathcal{L}_m^2(F)$ for all $x \in \mathcal{X}$, where $\mathcal{L}_r^2(F)$ denotes the set of r-dimensional square-integrable functions of ξ with respect to its probability distribution F. That is, all Borel-measurable functions g with $\mathbb{E}_F[\|g(\xi)\|^2] < \infty$.*

ASSUMPTION 2. *(Linearity) We assume that $c(\xi) = C\xi + c_0$ for $C \in \mathbb{R}^{n\times k}$, $c_0 \in \mathbb{R}^n$, $v(\xi) = V\xi + v_0$ for $V \in \mathbb{R}^{n\times k}$, $v_0 \in \mathbb{R}^n$, $b_x(\xi) = B_x\xi + b_{x0}$ for $B_x \in \mathbb{R}^{m\times k}$, $b_{x0} \in \mathbb{R}^m$, where $B_x = \sum_{i=1}^d B_i x_i + B_0$ for $B_i \in \mathbb{R}^{m\times k}$, $i = 0, 1, \ldots, d$ and $b_{x0} = \sum_{i=1}^d b_i x_i + b_0$ for $b_i \in \mathbb{R}^m$, $i = 0, 1, \ldots, d$.*

ASSUMPTION 3. *(Relatively complete recourse) We assume that* $\{y \in \mathbb{R}^n \,|Ay \leq b_x(\xi)\}$ *is non-empty and bounded almost surely for every* $x \in \mathcal{X}$ *and for every* ξ *of any* $F \in \mathcal{D}$.

We note that the boundedness assumption is equivalent to either statements: (i) $\{y \in \mathbb{R}^n \mid Ay \leq 0\} = \{0\}$ or (ii) $\{A^\top p \mid p \geq 0,\ p \in \mathbb{R}^m\} = \mathbb{R}^n$.

Suppose that a series of independent data samples $\{\xi^n\}_{n=1}^N$ are drawn from the true probability distribution of ξ. We calculate the sample mean μ_0 and covariance matrix Σ_0 as

$$\mu_0 := \frac{1}{N}\sum_{n=1}^{N} \xi^n \quad \text{and} \quad \Sigma_0 := \frac{1}{N-1}\sum_{n=1}^{N} (\xi^n - \mu_0)(\xi^n - \mu_0)^\top.$$

Two moment-based ambiguity sets are considered and defined as follows.

DEFINITION 1. Moment-based ambiguity set with continuous distributions (Delage and Ye 2010).

$$\mathcal{D}_M\left(\mathcal{S}, \mu_0, \Sigma_0, \gamma_1, \gamma_2\right) := \left\{ F \in \mathcal{M} \,\middle|\, \begin{array}{l} \mathbb{P}_F(\xi \in \mathcal{S}) = 1 \\ \left(\mathbb{E}_F[\xi] - \mu_0\right)^\top \Sigma_0^{-1} \left(\mathbb{E}_F[\xi] - \mu_0\right) \leq \gamma_1 \\ \mathbb{E}_F\left[\left(\xi - \mu_0\right)\left(\xi - \mu_0\right)^\top\right] \preceq \gamma_2 \Sigma_0 \end{array} \right\},$$

where $\gamma_1 \geq 0$, $\gamma_2 \geq 1$, $\mathcal{M}$ is the set of all probability measures, and the support set of ξ is

$$\mathcal{S} := \{\xi \in \mathbb{R}^k \mid W\xi \geq h\},$$

where $W = [w_1, \ldots, w_l]^\top \in \mathbb{R}^{l \times k}$ and $h = [h_1, \ldots, h_l]^\top \in \mathbb{R}^l$.

ASSUMPTION 4. *(Compact support) The support set* $\mathcal{S}$ *is compact.*

DEFINITION 2. Moment-based ambiguity set with discrete distributions. Suppose that the distribution of uncertainty ξ is supported on a finite set $\mathcal{N} := \{\xi^1, \ldots, \xi^N\}$ with probability $p_1, \ldots p_N$. The ambiguity set is given by

$$\mathcal{D}_{\text{dis}}\left(\mathcal{N}, \mu_0, \Sigma_0, \gamma_1, \gamma_2\right) := \left\{ p \in \mathbb{R}_+^N \,\middle|\, \begin{array}{l} \sum_{s=1}^N p_s = 1 \\ \left(\sum_{s=1}^N p_s \xi^s - \mu_0\right)^\top \Sigma_0^{-1} \left(\sum_{s=1}^N p_s \xi^s - \mu_0\right) \leq \gamma_1 \\ \sum_{s=1}^N p_s \left[\left(\xi^s - \mu_0\right)\left(\xi^s - \mu_0\right)^\top\right] \preceq \gamma_2 \Sigma_0 \end{array} \right\},$$

where $\gamma_1 \geq 0$, $\gamma_2 \geq 1$.

The constraints in the sets $\mathcal{D}_M$ and $\mathcal{D}_{\text{dis}}$ ensure that (i) the true first-order moment $\mathbb{E}[\xi]$ lies in an ellipsoid centered at μ_0 and (ii) the true second-order moment lies in a cone bounded above by $\gamma_2\Sigma_0$. Parameters γ_1 and γ_2 control the distance between the true moments and the moment estimates. Theoretically, Delage and Ye (2010)provide guidance on the choice of γ's to guarantee out-of-sample performance. In practice, the parameters can be decided using cross-validation.

1.2. Contributions

In this paper, we focus on the pessimistic DRBP where the follower does not act in favor of the leader. The leader therefore makes decisions against the worst case of the follower's decision. The primary motivation of this work is to take the first step towards modeling and solving distributionally robust pessimistic bilevel programs. Specifically, we show that, under proper conditions of ambiguity sets (not only restricted to the two moment-based ambiguity sets), the pessimistic DRBP is equivalent to a generic two-stage DR stochastic (nonlinear) program (TSDR) with random objective and right-hand side. For moment-based ambiguity sets of continuous distributions, even a two-stage distributionally robust linear program (TSDRLP), a special case of the resulting TSDR, is NP-hard with both random objective and right-hand side (e.g., Bertsimas et al. 2010). Therefore, we approach the problem by linear decision rules (LDRs) to tractably approximate the TSDR. When the ambiguity set is restricted to discrete distributions, an exact reformulation is derived for the resulting TSDR. Our main contributions are summarized as follows.

1. We derive an equivalent TSDR reformulation (Section 3) for the pessimistic DRBP under ambiguity sets satisfying proper conditions. The binary restriction of the leader's decision is not necessarily required for the reformulation. The derivation is based on finding the optimal value function of the follower's problem and establishing the equivalence between the least favorable outcome and the robust outcome with respect to the ambiguity set.
2. We consider the resulting TSDR under moment-based ambiguity sets (Delage and Ye 2010).
   - For the ambiguity set $\mathcal{D}_M$ of continuous distributions (Section 4), we use LDR techniques which leads to a non-convex program. (1) We derive a 0-1 semidefinte programming (SDP)

approximation via the standard convex duality theory. (2) Alternatively, leveraging the modern conic programming techniques, we develop an exact 0-1 copositive program (COP) reformulation, which, although intractable, admits a 0-1 SDP inner approximation. Both 0-1 SDP approximations provide upper bounds to the DRBP under the ambiguity set of continuous distributions.

- A lower bound to the DRBP with $\mathcal{D}_M$ can be provided by solving the DRBP under a constructed ambiguity set of discrete distributions. For the ambiguity set $\mathcal{D}_{\text{dis}}$ of discrete distributions (Section 5), the resulting TSDR is exactly reformulated as a 0-1 SDP and explicit constructions of the worst-case distributions are provided.

3. The resulting 0-1 SDPs still remain computationally challenging. In Section 6, we propose a cutting-plane framework to solve 0-1 SDPs. Moreover, based on the MILP approximation proposed in Yanıkoğlu and Kuhn (2018), another cutting-plane algorithm is proposed for the continuous ambiguity set $\mathcal{D}_M$. We computationally evaluate efficiency and effectiveness of the proposed 0-1 SDP approximations in solving a facility location problem under various settings (Section 7).

2. Related Literature

2.1. Bilevel Programs

The bilevel program can be interpreted as an optimization problem of the leader who searches for the global minimum. The feasible solutions lie among the optimal solutions of the follower's problem, of which the objective function and (or) the feasible region depend on the leader's decision. In the situation where the follower's problem have more than one optimal solution, the follower returns solutions either in favor (*optimistic version*) of or adverse (*pessimistic version*) to the leader's objective. Bilevel program has been widely applied to various problems including revenue management (Côté et al. 2003), supply chain management (Ryu et al. 2004), production planning (Iyer and Grossmann 1998), security (Scaparra and Church 2008), transportation (Migdalas 1995), energy market (Carrión et al. 2009), and many other fields.

The study of deterministic bilevel programs can be dated back to 1934 from the investigation of market equilibria by Stackelberg (Fudenberg and Tirole 1991). However, comprehensive studies began only over the last few decades. Most existing work focuses on optimistic bilevel programs. The reason is that, in the optimistic version, the minimization over the follower's and leader's decisions can be combined into a joint minimization in the leader's objective function. Thus, the three interrelated optimizations (i.e., (1) the leader minimizing over his decision, (2) the leader minimizing over the follower's optimal decisions, and (3) the follower optimizing her decisions) can be transformed into a two-level problem. Moreover, if the follower's problem is convex, the two-level problem can be further reformulated as a single-level program by replacing the constraints, which enforce the solution being optimal for the follower's problem, with the corresponding Karush-Kuhn-Tucker optimality conditions (see e.g., Dempe 2020, Beck and Schmidt 2021). Global optimization algorithms have been developed to solve the resulting single-level (non-convex) problems (e.g., see, Faísca et al. 2007, Tuy et al. 2007). Mitsos et al. (2008) and Tsoukalas et al. (2009) further study the case where the follower's problem is nonconvex. We refer interested readers to Dempe (2002) and Dempe and Zemkoho (2013) for a comprehensive review to deterministic bilevel program research (up to 2002) and more details on the optimality conditions.

Due to inherit nature of successive decision making under uncertainty, many applications can be modeled as stochastic bilevel programs. However, stochastic bilevel programs attract less attention than their deterministic counterparts. Ivanov (2014) consider optimistic stochastic bilevel linear programs with right-hand side uncertainty in the follower's problem using quantile criterion. They develop a two-stage stochastic programming equivalence with equilibrium constraints and a mixed-integer linear programming reformulation under a discrete distribution. More coherent or convex risk measures are considered in Burtscheidt and Claus (2019) and Burtscheidt et al. (2020) for stochastic bilevel linear programs when the right-hand side of the follower's problem is stochastic. They outline Lipschitzian properties, conditions for existence and optimality, and stability results. Zhang and Özaltın (2021) studies stochastic bilevel programs with follower's integer

recourse and stochastic right-hand side using an exact value function-based approach. All the work mentioned above studies optimistic stochastic bilevel programs, which can be viewed as special cases of stochastic mathematical programming with equilibrium constraints (SMPEC) problem (see, e.g., Patriksson and Wynter 1999, Xu 2005, Shapiro 2006, Shapiro and Xu 2008).

Recently, Yanıkoğlu and Kuhn (2018) study pessimistic stochastic bilevel programs with binary leader's decision and continuous follower's decision. They apply decision rules to construct upper and lower bounds on the leader's problem. Tavaslıoğlu et al. (2019) extend the problem to mixed-integer follower's decision and develop a generalized value function-based approach.

2.2. Distributionally Robust Optimization

In contrast to the previous stochastic bilevel programming work requiring full distributional information of uncertainties, we study the DRO variant, which assumes that only limited distributional information is known when making decisions. A set (termed as ambiguity set) of distributions are considered based on the partially known information. Two types of information are widely considered in literature, including the moment information such as mean and covariance (see, e.g., Bertsimas et al. 2010, Delage and Ye 2010, Wiesemann et al. 2014, Zhang et al. 2018), and the distance to a reference distribution, for example, Wasserstein metrics (see, e.g., Gao and Kleywegt 2016, Esfahani and Kuhn 2018, Zhao and Guan 2018).

In this paper, we show that the pessimistic DRBPs admit equivalent TSDR nonlinear reformulations with uncertainty in both the objective and right-hand side. A special case of the TSDR is TSDRLPs, which have attracted many interests since a decade ago. Bertsimas et al. (2010) consider risk averse TSDRLPs when the first- and second-order moments are known. When the uncertainty only appears in the objective function, they obtain a tight SDP reformulation. When the uncertainty is only in the right-hand side, they show that the problem is NP-hard in general. Xie and Ahmed (2018) study TSDRLPs with simple integer round-up recourse using mean and support to form the ambiguity set. They obtain a mixed-integer second-order cone programming reformulation. Recently, Fan and Hanasusanto (2021) consider TSDRLP with random recourse

matrix using piecewise decision rules based on a partitioning scheme and accordingly a conditional ambiguity set is constructed for marginal probabilities on the partitions. Using Wasserstein metrics, the TSDRLPs admit tractable reformulations under various conditions (see, e.g., Hanasusanto and Kuhn 2018, Xie 2019, Wang et al. 2020). For ambiguity sets of distributions supported on a finite set, decomposition methods have been studied for various settings: Love and Bayraksan (2015) based on Φ-divergence ambiguity set, Bansal et al. (2018), Luo and Mehrotra (2021) for problems with binary variables or/and conic constraints. Using L_1-norm ball ambiguity sets, Jiang and Guan (2018) present a sample average approximation algorithm to solve a TSDRLP with binary first-stage decisions.

3. Distributionally Robust Two-stage Stochastic Programming Equivalence

In this section, we will show that the pessimistic DRBPs can be reformulated to equivalent TSDRs with a random objective and right-hand side when the ambiguity set is weakly compact. We first consider a special case when the objective coefficients of the follower's decision in the leader's problem and in the follower's problem coincide, the DRBP admits a TSDRLP reformulation in Section 3.1. Then we focus on the more general case when they are not equal in Section 3.2.

3.1. A Special Case: $c(\xi) = v(\xi)$

When the objective coefficients of the leader's and follower's are equal (or when $v(\xi) = Kc(\xi)$, K is a positive constant), the pessimistic version ($\mathcal{P}$) and the optimistic version ($\mathcal{O}$) coincide as the following TSDRLP with uncertainties in the objective and the right-hand side.

$$\min_{x\in\mathcal{X}} w^\top x + \sup_{F\in\mathcal{D}} \mathbb{E}_F \left[\min_{Ay\le b_x(\xi)} c(\xi)^\top y \right]. \tag{5}$$

3.2. General Case: $c(\xi) \neq v(\xi)$

Now, we consider the more general case when the follower and the leader concern different objective coefficients for the pessimistic version ($\mathcal{P}$). We denote $\Psi(x,\xi) := \max_{y\in\Omega(x,\xi)} v(\xi)^\top y$. Then the pessimistic DRBP ($\mathcal{P}$) is rewritten as $\min_{x\in\mathcal{X}} w^\top x + \sup_{F\in\mathcal{D}} \mathbb{E}_F[\Psi(x,\xi)]$. Recall that the second-stage feasible region $\Omega(x,\xi) := \arg\min_y \left\{ c(\xi)^\top y :\ Ay \le b_x(\xi) \right\} \subset \mathbb{R}^n$ involves an arg min operator.

The most frequently used solution approach for bilevel problems is to reformulate the bilevel problem into a single-level problem by eliminating the arg min operator. To eliminate the arg min operator, for a given distribution $F \in \mathcal{D}$, we first rewrite the second-stage problem in a square-integrable functional space in Lemma 1.

LEMMA 1. *For a given leader's solution x and distribution $F \in \mathcal{D}$, the second-stage expected cost of the pessimistic DRBP ($\mathcal{P}$) with respect to distribution F, i.e., $\mathbb{E}_F[\Psi(x,\xi)]$, is equivalent to the following stochastic program.*

$$Q_F(x) := \max_{y(\xi)\in\mathcal{L}_n^2(F)} \quad \mathbb{E}_F[v(\xi)^\top y(\xi)] \tag{6a}$$

$$s.t. \quad y(\xi) \in \operatorname*{arg\,min}_{y'(\xi)\in\mathcal{L}_n^2(F)} \left\{ c(\xi)^\top y'(\xi) : Ay'(\xi) \le b_x(\xi) \right\}. \tag{6b}$$

See the detailed proof in Section A of the Appendices. Next, following Yanıkoğlu and Kuhn (2018), an optimal value function $\bar{Q}_F(x)$ of the follower's problem is derived in the following Proposition.

PROPOSITION 1. *(Adapted from Proposition 2.1 in Yanıkoğlu and Kuhn (2018)) For a given leader's solution x and a distribution $F \in \mathcal{D}$, the second-stage expected cost $\mathbb{E}_F[\Psi(x,\xi)]$ in pessimistic DRBP ($\mathcal{P}$) is equivalent to*

$$\sup_{y(\xi)\in\mathcal{L}_n^2(F)} \quad \mathbb{E}_F[v(\xi)^\top y(\xi)] \tag{7a}$$

$$s.t. \quad \mathbb{E}_F[c(\xi)^\top y(\xi)] \le \bar{Q}_F(x) \tag{7b}$$

$$Ay(\xi) \le b_x(\xi), \tag{7c}$$

where $\bar{Q}_F(x) := \min_{y(\xi)\in\mathcal{L}_n^2(F)} \left\{ \mathbb{E}_F[c(\xi)^\top y(\xi)] : \; Ay(\xi) \le b_x(\xi) \right\}$.

In constraint (7b), the right-hand side $\bar{Q}_F(x)$ involves a two-stage stochastic program (with only wait-and-see decisions). Lemma 2 incorporates the two-stage stochastic problem into the objective according to the duality of convex stochastic program using risk functions (Ruszczyński and Shapiro 2006). The proof is presented in Section B of the Appendices.

LEMMA 2. *For a given leader's solution x and distribution $F \in \mathcal{D}$, problem* (7)*, or equivalently* $\mathbb{E}_F[\Psi(x,\xi)]$*, is equivalent to the two-stage stochastic program as follows.*

$$\min_{\lambda \geq 0} \mathbb{E}_F \left[\Phi_\lambda(x,\xi)\right], \tag{8}$$

where

$$\Phi_\lambda(x,\xi) := \min_{p \in \mathbb{R}^m, y \in \mathbb{R}^n} \; b_x(\xi)^\top p + c(\xi)^\top y \tag{9a}$$

$$s.t. \qquad A^\top p + \lambda c(\xi) = v(\xi), \; p \geq 0 \tag{9b}$$

$$Ay \leq \lambda b_x(\xi). \tag{9c}$$

REMARK 1. The feasibility of problem (9) is guaranteed by Assumption 3. Specifically, constraint (9b) is feasible for any given ξ as $\left\{A^\top p \colon \; p \geq 0\right\} = \mathbb{R}^n$. For constraint (9c): if $\lambda > 0$, then $\{y \colon \; Ay \leq b_x(\xi)\} \neq \emptyset$; if $\lambda = 0$, $\{y \colon \; Ay \leq 0\} = \{0\}$. $\square$

Combining the minimization problem in Lemma 2 with the outer maximization over the ambiguity set, the worst-case outcome $\sup_{F \in \mathcal{D}} \min_{\lambda \geq 0} \mathbb{E}_F[\Phi_\lambda(x,\xi)]$ is upper bounded by the robust outcome $\min_{\lambda \geq 0} \sup_{F \in \mathcal{D}} \mathbb{E}_F[\Phi_\lambda(x,\xi)]$ with respect to the ambiguity set $\mathcal{D}$. The following theorem further establishes the equivalence between the worst-case outcome and the robust outcome, whose proof is provided in Section C of the Appendices.

THEOREM 1. *Given a leader's solution $x \in \mathcal{X}$ and a convex and weakly compact ambiguity set $\mathcal{D}$ of probability measures on $(\mathcal{S}, \mathcal{F})$ with a compact metric space $\mathcal{S}$ and its Borel σ-algebra, the worst-case second-stage expected cost* $\sup_{F \in \mathcal{D}} \mathbb{E}_F[\Psi(x,\xi)] =$

$$\min_{\lambda \geq 0} \sup_{F \in \mathcal{D}} \mathbb{E}_F[\Phi_\lambda(x,\xi)]. \tag{10}$$

We are now ready to show that the pessimistic DRBP ($\mathcal{P}$) is equivalent to a generic TSDR, which is formalized in Corollary 1.

COROLLARY 1. *Given a convex and weakly compact ambiguity set $\mathcal{D}$ of probability measures on $(\mathcal{S}, \mathcal{F})$ with a compact metric space $\mathcal{S}$ and its Borel σ-algebra, the pessimistic DRBP ($\mathcal{P}$) under $\mathcal{D}$ is equivalent to the following TSDR*

$$\min_{x \in \mathcal{X}, \lambda \geq 0} \left\{ w^\top x + \sup_{F \in \mathcal{D}} \mathbb{E}_F \left[\Phi_\lambda(x,\xi)\right] \right\}. \tag{11}$$

Note that the two-stage distributionally robust problem (11) is nonconvex, inherent from the nonconvexity of bilevel program, due to the bilinear term $b_x(\xi)^\top p$ in the objective function of $\Phi_\lambda(x,\xi)$ and $\lambda b_x(\xi)$ in the constraint of $\Phi_\lambda(x,\xi)$, where $b_x(\xi)$ is an affine function of the leader's decision x and p is a recourse decision.

REMARK 2. The binary constraints in the leader's feasible region $\mathcal{X}$ are not necessarily required in Theorem 1 and Corollary 1. The results can be easily extended (1) to the case where linking constraints of the here-and-now leader's decision x and the follower's optimal wait-and-see decision y by properly assuming boundedness and relatively complete recourse for the second-stage problem, and (2) to the case where the objective coefficient of the follower's problem is an affine function of the leader's decision x. □

REMARK 3. The moment-based ambiguity set $\mathcal{D}_M$, which will be discussed in Section 4, satisfies the weak compactness (Proposition 7, Sun and Xu 2016) required by Theorem 1 and Corollary 1. Other common choices of ambiguity sets are also compact with respect to the weak topology under some proper conditions. For example, a p-Wasserstein ball is weakly compact when the reference distribution has a finite pth moment (Theorem 1, Yue et al. 2021). For an f-divergence (of which examples include Kullback-Leibler divergence, Hellinger divergence, J-divergence, etc.) ball, if the metric space is a compact Polish space, then it is weakly compact (Lemma 3.2, Birghila et al. 2021). We refer interested readers to Sun and Xu (2016) for the compactness results of other types of ambiguity sets constructed through moments and mixture distributions. Furthermore, given that the finite support set $\mathcal{N}$ is compact and $\mathcal{D}_{\text{dis}}$ is closed (see Proposition 7 Sun and Xu 2016), by Prokhorov's theorem, the ambiguity set $\mathcal{D}_{\text{dis}}$ of discrete distributions (further discussed in Section 5) is as well weakly compact (see, e.g., Shapiro and Kleywegt 2002, Prokhorov 1956). □

Corollary 1 extends the results in Yanıkoğlu and Kuhn (2018) from a stochastic bilevel program to a distributionally robust variant. Given that the TSDRLP is a special case of the resulting TSDR (11) in Theorem 1, in the rest of the paper, we will focus on the solution approaches to TSDR (11) with binary leader's decisions under the two types of moment-based ambiguity sets $\mathcal{D}_M$ and $\mathcal{D}_{\text{dis}}$.

4. Decision Rule Approximations under Ambiguity Set of Continuous Distributions

In this section, we consider the pessimistic problem $(\mathcal{P})$ with binary leader's decisions under the moment-based ambiguity set of continuous distributions $\mathcal{D} = \mathcal{D}_M$. As pointed out in Bertsimas et al. (2010), even the TSDRLP (5) is NP-hard even under ambiguity sets that match the moment information. We employ LDRs to conservatively approximate the resulting TSDRs under moment-based ambiguity sets $\mathcal{D}_M$. Specifically, we consider the recourse decisions p and y as affine functions of the uncertainty ξ: $y(\xi) := Y\xi + y_0$ and $p(\xi) := P\xi + p_0$, where $P \in \mathbb{R}^{m\times k},\ p_0 \in \mathbb{R}^m$, $Y \in \mathbb{R}^{n\times k}$, and $y_0 \in \mathbb{R}^n$. In this section, we focus on a conservative approximation of (11) in the following form:

$$\min_{x\in\mathcal{X},\lambda\geq 0,P,p_0,Y,y_0} \left\{ w^\top x + \sup_{F\in\mathcal{D}_M} \mathbb{E}_F[b_x(\xi)^\top p(\xi) + c(\xi)^\top y(\xi)] :\ A^\top p(\xi) + \lambda c(\xi) = v(\xi),\ p(\xi)\geq 0,\ Ay(\xi) \leq \lambda b_x(\xi) \right\}. \tag{12}$$

By deriving the dual of the inner maximization problem (see Lemma 1 of Delage and Ye 2010), we obtain an equivalent formulation:

$$\begin{aligned}
\min_{Q,q,r,t,\lambda,P,Y,p_0,y_0,x\in\mathcal{X}} \quad & w^\top x + r + t && \text{(13a)}\\
\text{s.t.} \quad & r \geq \xi^\top \left(B_x^\top P + C^\top Y - Q\right)\xi + \left(p_0^\top B_x + b_{x0}^\top P + y_0^\top C + c_0^\top Y - q\right)\xi + b_{x0}^\top p_0 + c_0^\top y_0, \\
& \forall \xi \in \mathcal{S} && \text{(13b)}\\
& A(Y\xi + y_0) \leq \lambda(B_x\xi + b_{x0}),\ \forall \xi \in \mathcal{S} && \text{(13c)}\\
& A^\top(P\xi + p_0) + \lambda(C\xi + c_0) = V\xi + v_0,\ P\xi + p_0 \geq 0,\ \forall \xi \in \mathcal{S} && \text{(13d)}\\
& t \geq \left(\gamma_2\Sigma_0 + \mu_0\mu_0^\top\right)\cdot Q + \mu_0^\top q + \sqrt{\gamma_1}\left\|\Sigma_0^{1/2}\left(q + 2Q\mu_0\right)\right\| && \text{(13e)}\\
& Q \succeq 0,\ \lambda \geq 0. && \text{(13f)}
\end{aligned}$$

The problem (13) is a semi-infinite problem with bilinear terms: $B_x^\top P, p_0^\top B_x, b_{x0}^\top P$ in constraint (13b) and $\lambda B_x, \lambda b_{x0}$ in (13c). As the leader's decision x is binary and $B_x = \sum_{i=1}^d B_i x_i + B_0,\ b_{x0} = \sum_{i=1}^d b_i x_i + b_0$ for some $B_i \in \mathbb{R}^{m\times k},\ b_i \in \mathbb{R}^m,\ i = 0, 1, \ldots, d$, one can linearize the bilinear terms using McCormick inequalities (e.g., McCormick 1976). Introduce auxiliary variables $\Gamma_i = B_i^\top P x_i,\ \theta_i = \lambda x_i,\ \omega_i = p_0 x_i,\ \rho_i = P^\top b_i x_i,\ i = 1, \ldots, d$ by the following McCormick inequalities:

$$B_i^\top TW - (1-x_i)M\mathbf{1}_{k\times k} \leq \Gamma_i \leq B_i^\top TW + (1-x_i)M\mathbf{1}_{k\times k},\ -x_i M\mathbf{1}_{k\times k} \leq \Gamma_i \leq x_i M\mathbf{1}_{k\times k},\ i = 1, \ldots, d \tag{14}$$

$$\lambda - M(1-x_i) \leq \theta_i \leq Mx_i,\ 0 \leq \theta_i \leq \lambda,\ i = 1, \ldots, d \tag{15}$$

$$p_0-(1-x_i)M\mathbf{1}_m\le\omega_i\le p_0+(1-x_i)M\mathbf{1}_m,\ -x_iM\mathbf{1}_m\le\omega_i\le x_iM\mathbf{1}_m, i=1,\dots,d \tag{16}$$

$$(TW)^\top b_i-(1-x_i)M\mathbf{1}_k\le\rho_i\le(TW)^\top b_i+(1-x_i)M\mathbf{1}_k\ -x_iM\mathbf{1}_k\le\rho_i\le x_iM\mathbf{1}_k,\ i=1,\dots,d, \tag{17}$$

where $M>0$ is a sufficiently large big-M constant, $\mathbf{1}_r\in\mathbb{R}^r$ is a r-dimensional vector of all ones, and $\mathbf{1}_{k\times k}\in\mathbb{R}^{k\times k}$ is a k-by-k all-ones matrix.

Recall that $\mathcal{S}:=\{\xi\in\mathbb{R}^k\mid W\xi\ge h\}$. The two semi-infinite constraints (13c)-(13d) can be transformed into finite constraints as follows, by applying standard duality techniques for robust optimization.

$$AY+\Lambda W=\sum_{i=1}^d\theta_iB_i+\lambda B_0,\ \Lambda h-Ay_0+\sum_{i=1}^d\theta_ib_i+\lambda b_0\ge0,\ \Lambda\ge0 \tag{18}$$

$$A^\top TW+\lambda C=V,\ A^\top p_0+\lambda c_0=v_0,\ Th+p_0\ge0,\ T\ge0 \tag{19}$$

The remaining semi-infinite constraint (13b), given any fixed r,Q,q,Y,y_0,P,p_0, involves a nonconvex quadratic function of ξ. In Section 4.1, we derive a conservative approximation of (13b) and present a 0-1 SDP approximation of (12). In Section 4.2, we utilize the copositve programming scheme to derive an exact reformulation of (13b) and thus provide an exact 0-1 COP reformulation of (12) which, although intractable, also admits a 0-1 SDP approximation.

4.1. 0-1 Semidefinite Programming Approximation

THEOREM 2. *The following 0-1 SDP is a conservative approximation of Problem* (12).

$$\min_{\substack{Q,q,r,t,Y,\lambda,\tau,\theta_i,\Gamma_i,T,\Lambda,\\ \omega_i,\rho_i,y_0,p_0,x\in\mathcal{X}}} \quad w^\top x+r+t \tag{20a}$$

$$\text{s.t.}\quad \begin{bmatrix} Q-\Big[\frac{1}{2}\Big(\sum_{i=1}^d\Gamma_i+B_0^\top TW+C^\top Y\Big) & \frac{1}{2}\Big(q-\sum_{i=1}^dB_i^\top\omega_i-B_0^\top p_0-\sum_{i=1}^d\rho_i \\ \quad+\frac{1}{2}\Big(\sum_{i=1}^d\Gamma_i+B_0^\top TW+C^\top Y\Big)^\top\Big] & \quad -(TW)^\top b_0-C^\top y_0-Y^\top c_0-W^\top\tau\Big) \\ \frac{1}{2}\Big(q-\sum_{i=1}^dB_i^\top\omega_i-B_0^\top p_0-\sum_{i=1}^d\rho_i & r-\sum_{i=1}^d b_i^\top\omega_i-b_0^\top p_0-c_0^\top y_0+\tau^\top h \\ \quad -(TW)^\top b_0-C^\top y_0-Y^\top c_0-W^\top\tau\Big)^\top & \end{bmatrix}\succeq0 \tag{20b}$$

$$\tau\ge0 \tag{20c}$$

$$(13e)-(13f),\ (14)-(19).$$

Section D in the Appendices presents the detailed proof.

4.2. 0-1 Copositive Programming Exact Reformulation

Constraint (13b) is equivalent to the following constraint involving a nonconvex quadratic maximization problem:

$$r \geq \max_{\xi \in \mathcal{S}} \ \xi^\top \left(B_x^\top P + C^\top Y - Q\right)\xi + \left(p_0^\top B_x + b_{x0}^\top P + y_0^\top C + c_0^\top Y - q\right)\xi + b_{x0}^\top p_0 + c_0^\top y_0. \tag{21}$$

In this section, we reformulate the constraint above using the generalized copositive programming techniques (e.g., Burer and Dong 2012). First, we review the definitions of *copositive cone* and *completely positive cone*.

Definition 3. Let $\mathcal{K} \subset \mathbb{R}^n$ be a closed convex cone. The *copositive cone* with respect to $\mathcal{K}$ is

$$\mathcal{COP}(\mathcal{K}) := \left\{ M \in \mathbb{S}^n : \ z^\top M z \geq 0 \ \forall z \in \mathcal{K} \right\},$$

and its dual cone, the *completely positive cone*, with respect to $\mathcal{K}$, is

$$\mathcal{CP}(\mathcal{K}) := \left\{ X \in \mathbb{S}^n : \ X = \sum_i z^i (z^i)^\top, \ z^i \in \mathcal{K} \right\},$$

where the summation is finite and the cardinality is unspecified.

We define $z := \begin{bmatrix} \xi \\ 1 \end{bmatrix}$. The polyhedral support $\mathcal{S}$ of ξ then induces a polyhedral cone of z, $\hat{\Xi} := \left\{ z \in \mathbb{R}^{k+1} : \ \mathcal{H} z \geq 0 \right\}$, where $\mathcal{H} = [W \mid -h] \in \mathbb{R}^{l \times (k+1)}$. In addition to the compactness assumption 4 on the support $\mathcal{S}$, the polyhedral cone is assumed to satisfy the following condition.

Assumption 5. *$\hat{\Xi}$ is full dimensional. That is, there is no implicit equalities in $\hat{\Xi}$.*

Under Assumption 5, the polyhedral cone $\hat{\Xi}$ is proper (see Lemma 2 of Mittal and Hanasusanto 2021). Now, consider

$$\mathcal{L} := \left\{ \begin{bmatrix} 1 \\ z \end{bmatrix} \begin{bmatrix} 1 \\ z \end{bmatrix}^\top \in \mathbb{S}^{k+2} : \ e_{k+1}^\top z = 1, \ z \in \hat{\Xi} \right\},$$

where $e_{k+1} \in \mathbb{R}^{k+1}$ denotes the vector with one in the $k+1$th coordinate and zeros elsewhere. Then constraint (21) is rewritten as $r \geq v^*$, where

$$v^* := \max \left\{ \mathrm{Tr}(\mathcal{Q} Z) : \ \begin{bmatrix} 1 & z^\top \\ z & Z \end{bmatrix} \in \mathrm{clconv}\,(\mathcal{L}) \right\}. \tag{22}$$

with

$$\mathcal{Q} := \begin{bmatrix} \Big[\frac{1}{2}\Big(\sum_{i=1}^{d}\Gamma_i + B_0^\top TW + C^\top Y\Big) & \frac{1}{2}\Big(\sum_{i=1}^{d} B_i^\top \omega_i + B_0^\top p_0 + \sum_{i=1}^{d}\rho_i \\ + \frac{1}{2}\left(\sum_{i=1}^{d}\Gamma_i + B_0^\top TW + C^\top Y\right)^\top\Big] - Q & \quad +(TW)^\top b_0 + C^\top y_0 + Y^\top c_0 - q\Big) \\ \frac{1}{2}\Big(\sum_{i=1}^{d} B_i^\top \omega_i + B_0^\top p_0 + \sum_{i=1}^{d}\rho_i & b_{x0}^\top p_0 + c_0^\top y_0 \\ \quad +(TW)^\top b_0 + C^\top y_0 + Y^\top c_0 - q\Big)^\top & \end{bmatrix} \in \mathbb{S}^{k+1}.$$

Proposition 2 **(Adapted from Theorem 8.1 of Burer (2012))**. *clconv*$(\mathcal{L}) = \mathcal{R}$, *where*

$$\mathcal{R} := \left\{ \begin{bmatrix} 1 & z^\top \\ z & Z \end{bmatrix} \in \mathcal{CP}(\mathbb{R}_+ \times \hat{\Xi}) : \begin{array}{c} e_{k+1}^\top z = 1, \\ e_{k+1}^\top Z e_{k+1} = 1 \end{array} \right\}.$$

Then the optimal value v^* of (22) equals to that of the following completely positive program (CPP).

$$\max\left\{ \mathrm{Tr}(\mathcal{Q}Z) : \begin{bmatrix} 1 & z^\top \\ z & Z \end{bmatrix} \in \mathcal{R} \right\}. \tag{23}$$

It is possible to further eliminate the last variable z_{k+1}.

Proposition 3. *The CPP* (23) *is equivalent to*

$$v^* = \max\left\{ \mathrm{Tr}\,(\mathcal{Q}Z) :\ Z \in \mathcal{CP}(\hat{\Xi}),\ e_{k+1}^\top Z e_{k+1} = 1 \right\}. \tag{24}$$

The proof is provided in Section E of the Appendices. Besides a more compact feasible region, an additional benefit of (24) over (23) is that (24) has nonempty interior due to Assumption 5 under which the polyhedral cone $\hat{\Xi}$ is a proper cone (see Lemma 4 of Mittal and Hanasusanto 2021). Thus, strong duality holds between (24) and its copositive programming (COP) dual problem:

$$\min\left\{ u :\ u e_{k+1} e_{k+1}^\top - \mathcal{Q} \in \mathcal{COP}(\hat{\Xi}) \right\}.$$

We are now ready to present the 0-1 COP reformulation.

THEOREM 3. *The following 0-1 COP is an exact reformulation of Problem* (12).

$$\min_{\substack{Q,q,r,t,Y,\lambda,\theta_i,\Gamma_i,T,\Lambda,\\ \omega_i,\rho_i,y_0,p_0,x\in\mathcal{X}}} \quad w^\top x + r + t \tag{25a}$$

$$\text{s.t.} \quad re_{k+1}e_{k+1}^\top - \mathcal{Q} \in \mathcal{COP}(\hat{\Xi}) \tag{25b}$$

$$(13e) - (13f),\ (14) - (19).$$

The detailed proof is in Section F of the Appendices. The resulting generalized copositve program is, however, in general intractable (Burer 2012) and one alternative is to replace the copositive cone $\mathcal{COP}(\hat{\Xi})$ with a semidefinite-based inner approximation (Xu and Burer 2018): $\text{IA}(\hat{\Xi}) = \left\{\mathcal{H}^\top U\mathcal{H} \ :\ U \geq 0,\ U \in \mathbb{S}^l\right\} \subset \mathcal{COP}(\hat{\Xi})$.

COROLLARY 2. *A conservative 0-1 SDP approximation of* (12) *is given by*

$$\min_{\substack{Q,q,r,t,Y,\lambda,\theta_i,\Gamma_i,T,\Lambda,\\ \omega_i,\rho_i,y_0,p_0,x\in\mathcal{X}}} \quad w^\top x + r + t \tag{26a}$$

$$\text{s.t.} \quad re_{k+1}e_{k+1}^\top - \mathcal{Q} = \mathcal{H}^\top U\mathcal{H},\ U \geq 0,\ U \in \mathbb{S}^l \tag{26b}$$

$$(13e) - (13f),\ (14) - (19).$$

5. Exact Reformulation under Ambiguity Set of Discrete Distributions

Under an ambiguity set of discrete distributions $\mathcal{D} = \mathcal{D}_{\text{dis}}$, the pessimistic bilevel program $(\mathcal{P})$ admits exact 0-1 SDP reformulation, which is formally stated in Theorem 4 and constitutes a simple corollary of Theorem 2. Thus, the proof is omitted for brevity.

THEOREM 4. *Under the ambiguity set of discrete distributions $\mathcal{D}_{dis}$, the pessimistic bilevel program $(\mathcal{P})$ is equivalent to the 0-1 SDP:*

$$\min_{x\in\mathcal{X},Q,q,r,t,\lambda,p^s,y^s} \quad w^\top x + r + t \tag{27a}$$

$$\text{s.t.} \quad r \geq \sum_{i=1}^{d} (B_i\xi^s + b_i)^\top \omega_i^s + (B_0\xi^s + b_0)^\top p^s + c(\xi^s)^\top y^s - {\xi^s}^\top Q\xi^s - {\xi^s}^\top q,$$

$$A^\top p^s + \lambda c(\xi^s) = v(\xi^s),\ Ay^s \leq \sum_{i=1}^{d} \theta_i(B_i\xi^s + b_i) + \lambda(B_0\xi^s + b_0),\ s = 1,\ldots,N \tag{27b}$$

$$t \geq \left(\gamma_2\Sigma_0 + \mu_0\mu_0^\top\right)\cdot Q + \mu_0^\top q + \sqrt{\gamma_1}\left\|\Sigma_0^{1/2}\left(q + 2Q\mu_0\right)\right\| \tag{27c}$$

$$\omega_i^s \le p^s + (1-x_i)M\mathbf{1}_m,\ \omega_i^s \ge p^s - (1-x_i)M\mathbf{1}_m,\ i=1,\dots,d,\ s=1,\dots,N \quad (27d)$$

$$-x_i M\mathbf{1}_m \le \omega_i^s \le x_i M\mathbf{1}_m,\ i=1,\dots,d,\ s=1,\cdots,N \quad (27e)$$

$$\theta_i \le Mx_i,\ \theta_i \ge \lambda - M(1-x_i),\ 0\le\theta_i\le\lambda,\ i=1,\dots,d \quad (27f)$$

$$Q\succeq 0,\ \lambda\ge 0,\ p^s\ge 0,\ s=1,\dots,N, \quad (27g)$$

where $M>0$ is a sufficiently large big-M constant.

Given a solution of the leader's decision $\hat{x}$, the worst-case distribution can be derived based on the dual problem of the following SDP equivalent of the second-stage problem.

$$\min_{Q,q,r,t,\lambda,p^s,y^s} \quad r+t \quad (28a)$$

$$\text{s.t.}\ \ r \ge b_{\hat{x}}(\xi^s)^\top p^s + c(\xi^s)^\top y^s - {\xi^s}^\top Q\xi^s - {\xi^s}^\top q,\ \forall s=1,\dots,N \quad (28b)$$

$$A^\top p^s + \lambda c(\xi^s) = v(\xi^s),\ p^s\ge 0,\ Ay^s \le \lambda b_{\hat{x}}(\xi^s),\ \forall s=1,\dots,N \quad (28c)$$

$$(27c),\ (27g).$$

THEOREM 5. *Given a leader's decision $\hat{x}$, assume the dual optimal solution associated with constraints* (28b)*, denoted as $\gamma^*\in\mathbb{R}^N$. The worst-case distribution is characterized as $\mathbb{P}\{\xi=\xi^s\}=\gamma^{s*},\ s=1,\dots,N$.*

The proof is given in Section G of the Appendices. For any given continuous ambiguity set $\mathcal{D}_M$, one can generate samples in the support set of $\mathcal{D}_M$ and construct a discrete ambiguity set as an approximation. Naturally, the DRBP model under the discrete ambiguity set provides a lower bound to that under the continuous ambiguity set. Given that the two 0-1 SDP formulations proposed in Section 4 provide upper bounds to the DRBP, together with the lower bounds using discrete ambiguity sets, the approximate gaps of the LDR method can be quantified. Computational studies on the gaps are presented in Section 7.3.

6. Solution Algorithms

In this section, we first propose a cutting-plan framework to efficiently solve the proposed 0-1 SDP formulations. Then another cutting-plane procedure is developed based on the MILP upper-bounding approximation proposed in Yanıkoğlu and Kuhn (2018).

6.1. Cutting-Plane Framework for the Three 0-1 SDP

The three 0-1 SDP formulations: (20) for conservatively approximating the pessimistic problem $(\mathcal{P})$ under the ambiguity set of continuous distributions, (26) for an inner approximation of the exact 0-1 COP reformulation under the ambiguity set of continuous distributions, and (27) for $(\mathcal{P})$ under the ambiguity set of discrete distributions, are computationally challenging as the problem size increases. In this section, we develop a Benders-type cutting-plane algorithm to solve the three 0-1 SDP problems. The cutting-plane algorithm iteratively solves a relaxed master problem and a subproblem to generate optimality cuts (due to the relatively complete recourse assumption 3, no feasibility cuts are needed.) We define variable ν as an underestimator of the worst-case second-stage expected cost $\sup_{F\in\mathcal{D}} \mathbb{E}_F\left[\Psi(x,\xi)\right]$. The relaxed master problem is given by

$$\mathbf{MP}: \quad \min_{x\in\mathcal{X},\nu} \; w^\top x + \nu \tag{29a}$$

$$\text{s.t.} \;\; \nu \geq u_l^\top x + a_l, \; l = 1,\ldots,L, \tag{29b}$$

where the parameter $u_l \in \mathbb{R}^d$ is the coefficient of x and $a_l \in \mathbb{R}$ is the scalar parameter of the valid inequality generated from the l-th subproblem. After obtaining the optimal solution $(\hat{x},\hat{\nu})$, we solve a subproblem. If the optimal value of the subproblem is greater than the underestimator $\hat{\nu}$, an optimality cut is generated and added to the MP. The details of the algorithm are presented in Algorithm 1. The algorithm guarantees finite termination for any given $\epsilon \geq 0$ (Geoffrion 1972). The subproblem solved in Line 5 and the optimality cut in Line (10) are specified in Section H of the Appendices for the three 0-1 SDPs, respectively.

6.2. MILP-based Cutting-Plane Algorithm

When the two moments (μ,Ω) of the worst-case distribution are known, the second-stage cost of the conservative decision-rule based approximation (12) is equivalent to an MILP (Yanıkoğlu and Kuhn 2018).

$$\begin{aligned}\min_{\substack{Y,\lambda,\theta_i,\Gamma_i,T,\Lambda,\\ \omega_i,\rho_i,y_0,p_0,}} \;& \text{Tr}\left[\Omega\left(\sum_{i=1}^d \Gamma_i + B_0^\top TW + C^\top Y\right)\right] + \mu^\top\left(\sum_{i=1}^d B_i^\top\omega_i + B_0^\top p_0 + \sum_{i=1}^d \rho_i + (TW)^\top b_0 + C^\top y_0 + Y^\top c_0\right)\\ &+ \sum_{i=1}^d b_i^\top\omega_i + b_0^\top p_0 + c_0^\top y_0\\ \text{s.t.} \;& (14)-(19).\end{aligned}$$

Algorithm 1 A cutting-plane framework for solving 0-1 SDP formulations ((20), (26), and (27))

Set $LB \leftarrow -\infty,\ UB \leftarrow +\infty$ and $\epsilon > 0$.
for $\ell = 0, 1, 2, \ldots$ **do**
  Solve the relaxed master problem MP. If the problem is infeasible, claim the infeasibility and quit the loop. Otherwise let (x^ℓ, ν^ℓ) be the optimal solution and z^ℓ be the optimal value. (When $\ell = 0$, let $\nu^\ell = -\infty$, and $x^\ell = \arg\min_{x \in \mathcal{X}} \{w^\top x\}$.)
  Update the lower bound: $LB \leftarrow z^\ell$.
  Solve the subproblem SP(x^ℓ). Obtain the optimal solution and optimal value obj^ℓ.
  if $w^\top x^\ell + obj^\ell < UB$ **then**
    Let $x^* \leftarrow x^\ell$ be the incumbent solution and update the upper bound: $UB \leftarrow w^\top x^\ell + obj^\ell$.
  end if
  if $UB - LB > \varepsilon$ **then**
    Add an optimality cut $\nu \geq u_l^\top x + a_l$.
  else
    return x^* as the optimal solution to 0-1 SDP problem.
  end if
end for

It is easy to see that the decision-rule based approximation (12) is upper bounded by

$$\min_{\substack{v,Y,\lambda,\theta_i,\Gamma_i,T,\Lambda,\\ \omega_i,\rho_i,y_0,p_0,x\in\mathcal{X}}} \quad w^\top x + v \tag{31a}$$

$$\text{s.t.} \quad v \geq \mathrm{Tr}\left[\Omega\left(\sum_{i=1}^d \Gamma_i + B_0^\top TW + C^\top Y\right)\right] + \mu^\top\left(\sum_{i=1}^d B_i^\top \omega_i + B_0^\top p_0 + \sum_{i=1}^d \rho_i + (TW)^\top b_0 + C^\top y_0 + Y^\top c_0\right)$$
$$+ \sum_{i=1}^d b_i^\top \omega_i + b_0^\top p_0 + c_0^\top y_0,\ \forall (\mu, \Omega) \in \mathcal{D}_{\text{moment}} \tag{31b}$$
$$\lambda \geq 0,\ (14) - (19),$$

where $\mathcal{D}_{\text{moment}} = \left\{(\mu, \Omega) :\ (\mu - \mu_0)^\top \Sigma_0^{-1} (\mu - \mu_0) \leq \gamma_1,\ \Omega - \mu\mu_0^\top - \mu_0\mu^\top + \mu_0\mu_0^\top \preceq \gamma_2 \Sigma_0,\ \mu\mu^\top \preceq \Omega\right\}$, which contains all the moments of any distributions in the ambiguity set $\mathcal{D}_M$. Although (31b) incorporates an infinite number of constraints, we can relax constraints (31b) and iteratively add them back if needed. Specifically, in each iteration, we obtain an incumbent solution

$(\hat{x}, \hat{v}, \hat{\Gamma}_i, \hat{T}, \hat{Y}, \hat{y}_0, \hat{\rho}_i, \hat{\omega}_i, \hat{p}_0)$ from the relaxed formulation. Then, we solve the following separation problem to decide if any solution violates constraints (31b).

$$\begin{aligned}\max_{(\mu,\Omega)\in\mathcal{D}_{\text{moment}}} \quad & \operatorname{Tr}\left[\Omega\left(\sum_{i=1}^{d}\hat{\Gamma}_i + B_0^\top \hat{T}W + C^\top \hat{Y}\right)\right] + \mu^\top\left(\sum_{i=1}^{d} B_i^\top \hat{\omega}_i + B_0^\top p_0 + \sum_{i=1}^{d}\hat{\rho}_i + (\hat{T}W)^\top b_0 + C^\top \hat{y}_0 + \hat{Y}^\top c_0\right) \\ & + \sum_{i=1}^{d} b_i^\top \hat{\omega}_i + b_0^\top \hat{p}_0 + c_0^\top \hat{y}_0 \end{aligned} \tag{32}$$

If not, then $(\hat{x}, \hat{v}, \hat{\Gamma}_i, \hat{T}, \hat{Y}, \hat{y}_0, \hat{\rho}_i, \hat{\omega}_i, \hat{p}_0)$ is an optimal solution to (31); otherwise, we obtain the optimal solution $(\hat{\mu}, \hat{\Omega})$ to (32) which violates constraints (31b). We add this violated constraint back into the relaxed formulation of (31) to cut off the incumbent solution. Since $\mathcal{X}$ is finite, the iterative cutting-plane algorithm terminates within a finite number of iterations. Note that the proposed cuts can be added in a delayed constraint generation fashion in a branch-and-cut framework.

7. Computational Studies

We evaluate the performance of the proposed 0-1 SDP approximations for the pessimistic DRBP ($\mathcal{P}$) on a facility location problem following Yanıkoğlu and Kuhn (2018). The facility location problem is described in Section 7.1. The experimental setup is described in Section 7.2. The comparison of CPU time and approximate optimality gap of solving the pessimistic DRBP using different formulations are presented in Section 7.3. We then consider two cases : (i) when the uncertainty is only in the constraints ($C = V = 0$), and (ii) when the uncertainty is in both the constraints and objective ($C, V \neq 0$). The solution details of the two cases are presented in Section 7.4 and the out-of-sample performance in Section 7.5. All the computational tests are performed on a 64-bit Windows 10 machine with Intel Core i7-4770 CPU 3.40 GHz and 16 GB memory. All the models have been implemented using YALMIP in MATLAB. Specifically, all the 0-1 SDP formulations are solved using the cutting-plane algorithm proposed in Section 6, of which the master problems are solved using Gurobi 9.1.2 and the subproblems are solved using MOSEK 9.3.

7.1. Facility Location Problem of a Market Entrant

Consider two market companies A and B selling homogeneous products to d demand locations. Among the d locations, a subset of them denoted by vector $l^S \in \{0,1\}^d$ are eligible for accommodating retail stores where $l_i^S = 1$ if a retail store can be accommodated at location i, 0 otherwise. Company A already operates retail stores at a subset of these eligible locations. We define $l^A \in \{0,1\}^d$ for Company A where $l_i^A = 1$ if Company A owns a store at location i, 0 otherwise. Company B wants to enter the market to build at most N_B number of new stores where Company A does not have stores yet. As a leader in the bilevel program, Company B decides where to build stores denoted by a binary vector $x \in \{0,1\}^d$ where $x_i = 1$ if Company B opens a store at location i, 0 otherwise. We assume that at each eligible location, at most one new store can be built either by Company A or by Company B. The feasible region of x is given by $\mathcal{X} = \left\{x \in \{0,1\}^d \,|\, x + l^A \le l^S,\ \|x\|_1 \le N_B\right\}$.

Let ξ_i be the uncertain demand at location i and $\xi = (\xi_i,\ i = 1,\ldots,d)^\top$ be the vector of the demand. We assume that the customers select the nearest store for purchase regardless of the owner. An aggregate customer is considered as the follower whose decision $y = (y_{ij},\ i,j = 1,\ldots,d)^\top \in \mathbb{R}^{d^2}$ denotes the amount of the products which are supplied from location i to location j. Let b_i be the store capacity at location i. Given the leader's decision x and demand ξ, the follower solves a transportation problem as follows.

$$\min_y \sum_{i=1}^{d}\sum_{j=1}^{d} c_{ij}y_{ij} \tag{33a}$$

$$\text{s.t.} \sum_{i=1}^{d} y_{ij} \ge \xi_j,\ j = 1,\ldots,d \tag{33b}$$

$$\sum_{j=1}^{d} y_{ij} \le b_i(x_i + l_i^A),\ i = 1,\ldots,d \tag{33c}$$

$$y_{ij} \ge 0,\ i,j = 1,\ldots,d. \tag{33d}$$

The objective (33a) minimizes the total shipping cost where c_{ij} is the unit transportation cost from location i to location j. Constraints (33b) ensure that all the demands are satisfied and constraints (33c) enforce the store capacity if there is a store at location i. Constraints (33d)

ensure the nonnegativity. Let $Y^*(x,\xi)$ be the set of optimal solutions of the follower's problem (33). We assume that $\max_{\xi\in\Xi}\|\xi\|_1 \le \sum_{i=1}^d b_i l_i^A$ and thus the follower's problem is always feasible (i.e., $Y^*(x,\xi)\neq\emptyset$) for every x and ξ.

For the leader (i.e., Company B), let w_i be the cost of opening a store at location i and v_{ij} be the negative of revenue per unit product sold to location j from location i. Denote vector $w=(w_i,\ i=1,\dots,d)^\top$ and vector $v=(v_{ij},\ i,j=1,\dots,d)^\top$. The pessimistic DRBP is formulated as follows.

$$\inf_{x\in\mathcal{X}}\ \left\{w^\top x+\sup_{F\in\mathcal{D}}\ \mathbb{E}_F\left[\sup_{y\in\mathbb{R}^n} v(\xi)^\top y\right]:\ y\in Y^*(x,\xi)\right\} \tag{34}$$

The model decides how many stores to open and where to build them for Company B under the uncertain demand and the aggregate customer's optimal decisions.

7.2. Computational Setup

We consider first $d=8$ locations of the SGB128 dataset* of North America cities, among which $\|l^S\|_1=5$ are eligible for building a retail store, where $l_i^S=1$ for $i=1,2,3,4,6$, and $l_i^S=0$ for other i's. Company A already operates at location $i=6$, i.e., $\|l^A\|_1=1$, where $l_6^A=1$ with store capacity $b_6=240\cdot(d-\|l^S\|_1)/\|l^A\|_1=720$†. Company B can build at most $N_B=4$ stores among the remaining eligible locations. The store capacities at the eligible locations are $b_i=360$ for $i=1,2,3,4$. The fixed cost of opening a new store is $w_i=305$ and the fixed revenue of selling one product is \$5 at eligible locations for company B, i.e., $v_{ij}=-5$ if $l_i^S=1$ and $l_i^A=0$, otherwise 0. The fixed unit transportation cost c_{ij} is set to be the Euclidean distance between location i and j. We generate N realizations of the demands $\xi_1^n,\dots,\xi_d^n,\ n=1,\dots,N$ sampled using the following procedure. We assume that all the demands are independent, identical and generate ξ_i, at each

*The dataset is available from John Burkardt's website: `https://people.sc.fsu.edu/~jburkardt/datasets/cities/cities.html`

†The capacity is chosen such that the follower's problem is always feasible irrespective of whether Company B opens its stores or not for the relatively complete recourse assumption.

demand location i with $l_i^S = 0$ (not eligible to build stores), following the uniform distribution on $[30, 240]$. There is no demand, $\xi_i = 0$, at eligible location i where $l_i^S = 1$. We solve the 0-1 SDP approximations using the empirical mean μ_0 and covariance Σ_0 calculated from $N = 10$ samples and support set $\mathcal{S} = \{30 \leq \xi_i \leq 240, \text{ for } i \text{ such that } l_i^S = 0\}$. All big-M constants are set to 10^6. Given the optimal solutions obtained by solving different models, we generate $N' = 5000$ data samples for out-of-sample test with details given in Section 7.5.

7.3. CPU Time and Computational Details

To assess the accuracy and runtime of the proposed approximations, we randomly generate instances of the facility location problem with $d \in \{15, 20, 25\}$ locations, $\|l^S\|_1 \in \{5, 10\}$ eligible locations for building stores, $\|l^A\|_1 \in \{2, 3, 5\}$ locations occupied by Company A. The coordinates of the d locations are chosen independently and uniformly from $[0, 1]^2$. The random demand ξ_i at location i follows the uniform distribution on $[50, 150]$. The store capacity is $b_i = 150 \cdot (d - \|l^S\|_1)/\|l^A\|_1$ for i such that $l_i^S = 1$. There are $N_B \in \{2, 3, 5\}$ candidate locations for Company B to establish new stores at a fixed cost of $w_i \in \{750, 1000\}$ if the uncertainty only appears in the constraints ($C = V = 0$). When the uncertainty is in both the objective and the constraints ($C, V \neq 0$), we consider a double cost $w_i \in \{1500, 2000\}$ so that the ratio of the fixed store opening cost (w_i) and the product selling price (v_{ij}) remains compatible. The entries of $C \in \mathbb{R}^{d^2 \times k}$ are independently and randomly generated so that each entry of vector $C\xi$ falls into $[0.09, 0.092]$. The entries of the unit revenue $V \in \mathbb{R}^{d^2 \times k}$ are independently and randomly generated so that each entry is on the interval $[4v_{ij}, 0]$ (note that $v_{ij} = -5$). In particular, we randomly generate 10 instances for each setting as shown in Table 1. For each instance, we solve it using the 0-1 SDP approximation (SDP) proposed in Section 4.1, the inner approximation of the 0-1 COP formulation (IA-COP), and the MILP-based cutting-plane (MILP-Cut) method in Section 6.2, by constructing the ambiguity set $\mathcal{D}_M$ with $N = 10$ samples of the demand. By generating 10 more additional samples of the demand, we solve the same instance under the discrete ambiguity set $\mathcal{D}_{\text{dis}}$ with $N = 20$ samples. Given that the optimal value of the discrete ambiguity set serves as a valid lower bound of the problems under the

continuous ambiguity set, an optimality gap of the LDR approximations can be computed based on the lower bound. We denote the optimal value of SDP (IA-COP, or MILP-Cut) as V_{approx} and that of the discrete ambiguity set as V_{dis}. The optimality gap for SDP (IA-COP, or MILP-Cut) is calculated as

$$Gap = \frac{V_{\text{approx}} - V_{\text{dis}}}{|V_{\text{dis}}|} \times 100\%.$$

All the models are solved with $(\gamma_1, \gamma_2) = (0, 1)$. The results for $(\gamma_1, \gamma_2) = (1, 1)$ are similar and are presented in Section I of the Appendices.

Table 1 reports the 25%, 50% and 75% quantiles of the optimality gaps for the two approximations using linear decision rules. When the uncertainty only appears in the constraints ($C = V = 0$), the three approaches provide the same solutions and the same optimal values. Considering both random objective and constraints ($C, V \neq 0$), SDP and MILP-Cut provide better approximations than IA-COP. Table 2 summarizes the computational performance, across the same test instances as those reported in Table 1, of using the cutting-plane algorithms and the MILP-Cut approach proposed in Section 6. The columns t_{tot}, # It., and Gap report the average total run time, the average number of iterations (of the cutting-plane algorithm), and the average optimality gap, over the 10 instances of each setting. The numbers in bold are either the fastest run times or the smallest gaps among the SDP, the IA-COP, and the MILP-Cut approaches. When $C = V = 0$, the solutions obtained by solving the three approaches are almost the same and thus yield the similar gaps except that the MILP-Cut performs slightly better at for the last three settings. All three approaches have better performance (except for settings 1, 3, 8, and 10) in terms of faster computation time and smaller optimality gaps when $C = V = 0$ compared to $C, V \neq 0$. In contrast, the computational performance of the discrete ambiguity set is similar in both cases. For either case ($C = V = 0$ or $C, V \neq 0$), most of the time, MILP-Cut is faster than the other two approaches and it takes relatively longer time to solve SDP than IA-COP. The optimality gaps of the three approaches are almost the same when $C = V = 0$. Whereas when $C, V \neq 0$, SDP and MILP-Cut yield better gaps than IA-COP. Tables 1 and 2 suggest that when the uncertainty appears only

in the constraints ($C = V = 0$), the decision maker may prefer MILP-Cut for better computational performance. Otherwise, SDP and MILP-Cut are preferred for smaller gaps or IA-COP and MILP-Cut for faster computational time.

Table 1 Quantiles of optimality gaps with $(\gamma_1, \gamma_2) = (0, 1)$

Setting	d	l^S	l^A	N_B	w_i		$C=V=0$			$C,V\neq 0$								
							SDP/IA-COP/MILP-Cut			SDP			IA-COP			MILP-Cut		
					$C=V=0$	$C,V\neq 0$	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q
1	15	5	2	3	750	1500	0.00	11.81	18.00	2.41	12.14	15.12	20.11	25.21	27.32	2.41	12.14	15.12
2	15	5	3	2	750	1500	0.00	18.68	26.65	20.88	24.75	66.21	30.64	35.20	72.50	20.65	24.75	66.21
3	15	5	2	3	1000	2000	7.03	15.29	21.83	3.56	17.73	21.70	26.45	32.42	35.69	3.50	17.73	21.70
4	15	5	3	2	1000	2000	0.00	23.22	33.80	0.00	28.96	36.39	0.00	37.59	43.08	0.00	25.65	30.59
5	20	5	2	3	750	1500	0.00	0.00	0.01	0.83	2.65	9.18	13.83	14.01	14.85	0.65	0.70	0.77
6	20	5	3	2	750	1500	0.00	11.48	45.23	2.02	17.20	46.25	16.30	20.65	54.11	0.81	6.49	46.22
7	20	5	2	3	1000	2000	0.00	0.00	0.01	0.97	3.51	9.94	14.40	15.26	15.78	0.70	0.74	0.78
8	20	5	3	2	1000	2000	0.00	25.37	54.88	3.99	21.38	53.63	19.35	30.79	62.74	1.02	14.42	53.60
9	25	10	5	5	750	1500	0.00	29.96	61.67	1.08	25.46	60.81	23.46	36.00	67.17	1.04	25.43	60.81
10	25	10	5	5	1000	2000	14.77	51.58	70.41	1.75	35.76	74.31	37.24	51.56	82.08	1.69	35.61	74.31

Table 2 Computational comparison of the three approximation approaches with $(\gamma_1, \gamma_2) = (0, 1)$

Setting	$C=V=0$												$C,V\neq 0$											
	SDP			IA-COP			MILP-Cut			Discrete			SDP			IA-COP			MILP-Cut			Discrete		
	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.		t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	
1	5.00	9.0	**10.79**	3.74	8.8	**10.79**	**0.40**	1.0	**10.79**	1.20	8.8		6.00	9.0	**10.62**	4.71	8.6	24.23	**1.75**	3.0	**10.62**	1.32	8.8	
2	2.48	5.0	**26.81**	1.69	4.5	**26.81**	**0.62**	1.7	**26.81**	0.50	4.9		3.15	5.0	38.58	**1.88**	4.5	44.11	2.69	6.8	**38.32**	0.59	4.9	
3	4.83	9.0	**14.99**	3.49	8.9	**14.99**	**0.35**	1.0	**14.99**	0.86	8.9		5.97	9.0	**14.58**	4.42	8.7	30.38	**1.71**	3.0	**14.58**	0.98	8.8	
4	2.64	5.0	**30.65**	1.71	4.5	**30.65**	**0.71**	1.9	**30.65**	0.48	4.9		3.23	5.0	32.55	**1.84**	4.5	38	2.74	6.8	**31.28**	0.60	5.0	
5	8.17	9.0	**1.79**	4.54	9.0	**1.79**	**0.43**	1.0	**1.79**	2.02	8.8		25.19	9.0	5.67	29.09	8.9	15.87	**1.63**	2.2	**2.33**	3.15	8.7	
6	4.47	5.0	**23.36**	2.50	4.6	**23.36**	**0.79**	1.6	**23.26**	1.12	4.6		**9.61**	5.0	26.57	12.67	4.7	33.35	13.77	15.9	**21.75**	1.62	4.7	
7	8.14	9.0	**1.86**	4.48	9.0	**1.86**	**0.44**	1.0	**1.86**	1.85	8.8		22.87	9.0	6.74	26.89	8.9	16.7	**2.03**	2.7	**2.71**	2.77	8.8	
8	4.47	5.0	31.31	2.28	4.4	31.31	**0.77**	1.5	**31.19**	1.09	4.6		**9.64**	5.0	30.85	12.66	4.7	40.76	13.97	16.3	**27.50**	1.41	4.5	
9	82.01	33.0	33.97	34.22	33.0	33.97	**3.97**	1.2	**33.94**	15.71	33.0		478.34	33.0	**31.07**	**55.52**	32.9	42.5	176.59	15.5	33.55	24.47	33.0	
10	83.19	33.0	49.11	33.82	33.0	49.11	**4.18**	1.2	**49.05**	16.65	33.0		458.24	33.0	41.14	**57.02**	32.9	55.3	208.78	19.2	**41.10**	22.79	33.0	

7.4. In-Sample Results

Given that in general the 0-1 SDP approximations in Section 4.1 yield better gaps than IA-COP, we focus on the solution details of the 0-1 SDP approximations. Specifically, Section 7.4.1 shows

the solutions and profits obtained using different values of γ_1 and γ_2. In Section 7.4.2, we further investigate how the profits change with γ's and the support size.

7.4.1. Solution details. Table 3 presents the solution details when the uncertainty is (i) only in the constraints ($C = V = 0$), and (ii) in both the objectives and constraints ($C, V \neq 0$) under various combination of γ_1 and γ_2. The profit is the negative of the objective value (20a) associated with the optimal solution. For the optimal solution, only the first four components of x are reported as they correspond to the four candidate locations for Company B to operate stores.

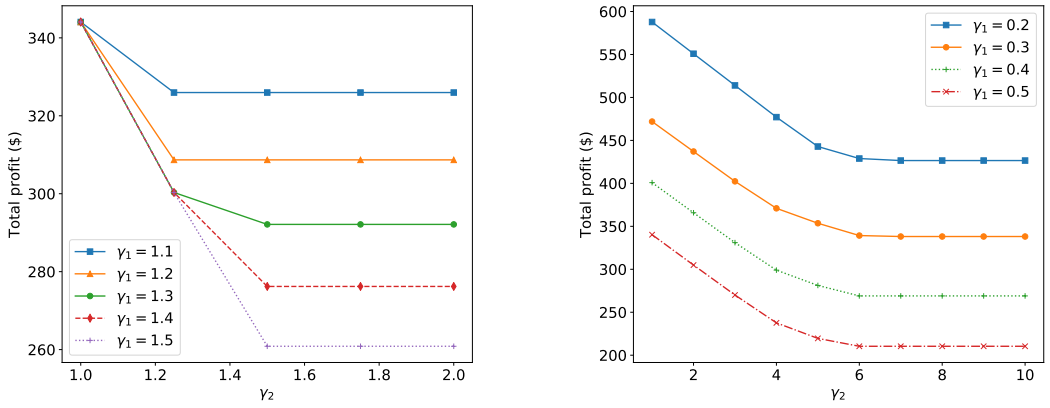
In both cases ($C = V = 0$ and $C, V \neq 0$), fewer stores are decided to open and the profits are non-increasing with larger γ's. The solutions are invariant for a fixed γ_1 with the values of γ_2 considered here. In the next section, we show more results on how the profits change under various values of γ's.

We note that the MILP model proposed in Yanıkoğlu and Kuhn (2018) can be viewed as a DRBP problem under an distributional ambiguity set of distributions requiring to match a given mean and a given covariance. Let the mean and covariance be the same empirical mean and covariance used to solve the 0-1 SDP approximations for the MILP model. We solve for the optimal solution and the corresponding (estimated) profit using the MILP. Specifically, when $C = V = 0$, the solutions and profits are the same as those of the 0-1 SDP with $(\gamma_1, \gamma_2) = (0, 1)$; when $C, V \neq 0$, the solution is the same as that of the 0-1 SDP with $(\gamma_1, \gamma_2) = (0, 1)$, whereas, the profit is 1519.05, higher than that of the 0-1 SDP.

Table 3 Profits and solutions of the 0-1 SDP approximations under various γ_1 and γ_2

$C = V = 0$							$C, V \neq 0$						
γ_1	0		1		1.5		γ_1	0		0.2		0.5	
γ_2	Profit	Sol.	Profit	Sol.	Profit	Sol.	γ_2	Profit	Sol.	Profit	Sol.	Profit	Sol.
1	808.97	(0,0,1,1)	344.07	(0,1,0,0)	344.07	(0,1,0,0)	1	1329.57	(0,0,1,1)	587.80	(0,0,1,1)	340.26	(0,1,0,0)
2	808.97	(0,0,1,1)	344.07	(0,1,0,0)	260.85	(0,1,0,0)	3	1255.79	(0,0,1,1)	513.98	(0,0,1,1)	270.06	(0,1,0,0)

7.4.2. Impact of γ's and support. In Figure 1, each curve corresponds to the profits of a fixed γ_1 and a varying γ_2. The total profits are less when γ's are larger as Company B needs to hedge against more (ambiguous) uncertainties. In both cases (Figures 1a and 1b), for a fixed γ_1, the profits decrease at the beginning and remain unchanged. Comparing across different γ_1, the turning point of γ_2, beyond which the profits become unchanged, are non-decreasing as γ_1 increases. The ambiguity set's parameter γ_2 impacts the profits by implicitly impacting the uncertainties' dispersion via controlling the size of the ambiguity set. When γ_2 is large enough (at and beyond the turning point), the support size becomes the major factor on the uncertainties dispersion.



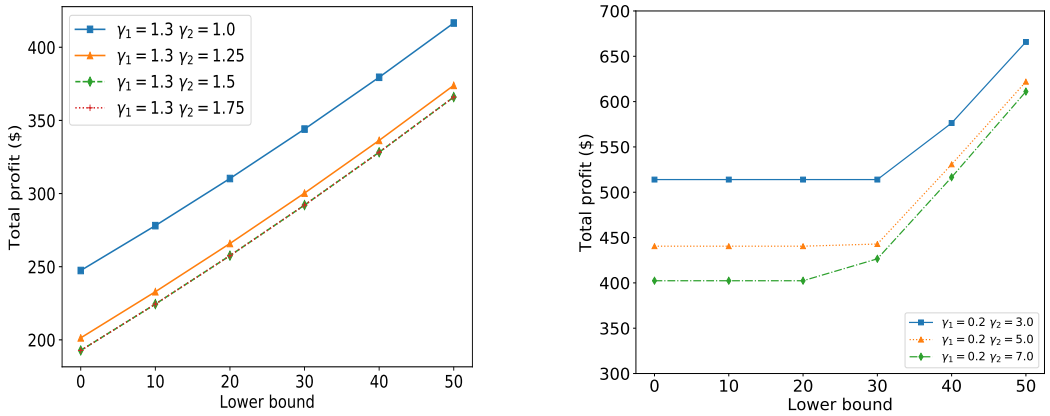
(a) $C = V = 0$ (b) $C, V \neq 0$

Figure 1 Impact of γ_1, γ_2

Next, we investigate how the support size impacts the total profits by varying the lower bound of the support set. In Figure 2, each curve represents the total profits at a combination of γ's using various lower bounds. When $C = V = 0$, larger lower bounds yield higher profits as smaller uncertainties' dispersion is considered. When $C, V \neq 0$, for a fixed (γ_1, γ_2), the profits remain the same when the lower bound is relatively small and the profits start to climb up when continuing improving the lower bound.

7.5. Out-of-Sample Performance

We perform out-of-sample simulations on the optimal solutions obtained by the pessimistic DRBP models under the same distribution used for the in-sample computation, i.e., uniform distribution



(a) $C = V = 0$ (b) $C, V \neq 0$

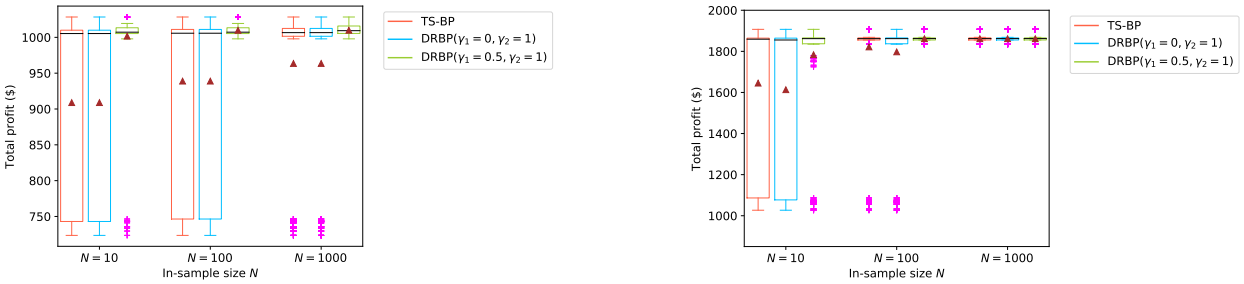
Figure 2 Impact of the support's lower bound

on $[30, 240]$. Specifically, we generate ten out-of sample sets, each containing $N' = 5,000$ i.i.d. out-of-sample data $\xi_1^m, \ldots, \xi_d^m,\ m = 1, \ldots, N'$ of the demand vector ξ following the uniform distribution. In Section J of the Appendices, we perform simulations on the out-of-sample data generated from distributions different than the in-sample uniform distribution to demonstrate the robustness when the distribution is misspecified.

Figure 3 shows the total profits of the pessimistic DRBP models under two γ combinations: $(\gamma_1, \gamma_2) = (0, 1)$ and $(0.5, 1)$. We benchmark our models with the MILP (TS-BP) proposed in Yanıkoğlu and Kuhn (2018). The boxplots are generated by simulating the expected profits using the 100 solutions, each of an independent in-sample set with sizes $N = 10, 100, 1000$, on the ten out-of-sample sets. In both cases $C = V = 0$ and $C, V \neq 0$, the expected profits vary in a wide range for TS-BP and DRBP with $(\gamma_1, \gamma_2) = (0, 1)$. For larger γ's, $(\gamma_1, \gamma_2) = (0.5, 1)$, the expected profits show less variation. As the in-sample size N increases, the variability of the expected profits reduces for all three models.

8. Conclusions

This paper develops a general distributionally robust two-stage programming equivalence for the pessimistic bilevel program under proper conditions of the ambiguity set. Typical choices of the ambiguity sets such as the moment-based and Wasserstein ball ambiguity sets satisfy the conditions



(a) $C = V = 0$ (b) $C, V \neq 0$

Figure 3 Out-of-sample performance: expected profits

under mild assumptions. The resulting two-stage formulation involves both uncertain objective and right-hand side. The paper then focuses on the pessimistic DRBP when the leader chooses a binary decision under moment-based ambiguity sets. Under the ambiguity set of continuous distributions, using LDRs, a 0-1 SDP approximation and an exact 0-1 COP reformulation are provided. When the ambiguity set is restricted to only discrete distributions, an exact 0-1 SDP reformulation is developed and the worst-case distribution is explicitly constructed. To solve the resulting three 0-1 SDPs, we developed a Bender's type cutting-plane framework. Furthermore, another cutting-plane method is proposed based on the MILP approximation developed in Yanıkoğlu and Kuhn (2018). Via numerical studies, we showed the efficiency and effectiveness of the proposed approximations. When the uncertainty only presents in the constraints, the approximations are solved faster and yield smaller gaps compared to the case when the uncertainty is both in the objectives and constraints. The γ parameters and the support size of the ambiguity sets play an important role in in-sample objectives. A reasonable choice of the γ parameters has great impact on the variation of the out-of-sample objectives.

Future research can be extended to distributionally robust bilevel programs with (1) integer or (and) nonlinear follower's problem, (2) continuous leader's decisions, (3) recourse leader's decisions. It would also be interesting to investigate the impact of the choice of ambiguity sets on the bilevel program's performance.

References

Bansal, M., Huang, K.-L., and Mehrotra, S. (2018). Decomposition algorithms for two-stage distributionally robust mixed binary programs. *SIAM Journal on Optimization*, 28(3):2360–2383.

Beck, Y. and Schmidt, M. (2021). A gentle and incomplete introduction to bilevel optimization. *Available at Optimization-Online:* `http://www.optimization-online.org/DB_FILE/2021/06/8450.pdf`.

Bertsimas, D., Doan, X. V., Natarajan, K., and Teo, C.-P. (2010). Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research*, 35(3):580–602.

Birghila, C., Aigner, M., and Engelke, S. (2021). Distributionally robust tail bounds based on Wasserstein distance and f-divergence. *arXiv preprint arXiv:2106.06266*.

Burer, S. (2012). Copositive programming. In *Handbook on semidefinite, conic and polynomial optimization*, pages 201–218. Springer.

Burer, S. and Dong, H. (2012). Representing quadratically constrained quadratic programs as generalized copositive programs. *Operations Research Letters*, 40(3):203–206.

Burtscheidt, J. and Claus, M. (2019). Bilevel optimization under uncertainty. *arXiv preprint arXiv:1907.04663*.

Burtscheidt, J., Claus, M., and Dempe, S. (2020). Risk-averse models in bilevel stochastic linear programming. *SIAM Journal on Optimization*, 30(1):377–406.

Carrión, M., Arroyo, J. M., and Conejo, A. J. (2009). A bilevel stochastic programming approach for retailer futures market trading. *IEEE Transactions on Power Systems*, 24(3):1446–1456.

Côté, J.-P., Marcotte, P., and Savard, G. (2003). A bilevel modelling approach to pricing and fare optimisation in the airline industry. *Journal of Revenue and Pricing Management*, 2(1):23–36.

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

Dempe, S. (2002). *Foundations of bilevel programming*. Springer Science & Business Media.

Dempe, S. (2020). Bilevel optimization: Theory, algorithms, applications and a bibliography. In *Bilevel optimization*, pages 581–672. Springer.

Dempe, S. and Zemkoho, A. B. (2013). The bilevel programming problem: Reformulations, constraint qualifications and optimality conditions. *Mathematical Programming*, 138(1-2):447–473.

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166.

Faísca, N. P., Dua, V., Rustem, B., Saraiva, P. M., and Pistikopoulos, E. N. (2007). Parametric global optimisation for bilevel programming. *Journal of Global Optimization*, 38(4):609–623.

Fan, X. and Hanasusanto, G. A. (2021). A decision rule approach for two-stage data-driven distributionally robust optimization problems with random recourse. *arXiv preprint arXiv:2110.00088*.

Fudenberg, D. and Tirole, J. (1991). Game theory, 1991. *Cambridge, Massachusetts*, 393(12):80.

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.

Geoffrion, A. M. (1972). Generalized benders decomposition. *Journal of optimization theory and applications*, 10(4):237–260.

Hanasusanto, G. A. and Kuhn, D. (2018). Conic programming reformulations of two-stage distributionally robust linear programs over Wasserstein balls. *Operations Research*, 66(3):849–869.

Ivanov, S. V. (2014). Bilevel stochastic linear programming problems with quantile criterion. *Automation and Remote Control*, 75(1):107–118.

Iyer, R. R. and Grossmann, I. E. (1998). A bilevel decomposition algorithm for long-range planning of process networks. *Industrial & Engineering Chemistry Research*, 37(2):474–481.

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

Love, D. and Bayraksan, G. (2015). Phi-divergence constrained ambiguous stochastic programs for data-driven optimization. *Technical report, Department of Integrated Systems Engineering, The Ohio State University, Columbus, Ohio*.

Luo, F. and Mehrotra, S. (2021). A decomposition method for distributionally-robust two-stage stochastic mixed-integer conic programs. *Mathematical Programming*, pages 1–45.

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I-Convex underestimating problems. *Mathematical programming*, 10(1):147–175.

Migdalas, A. (1995). Bilevel programming in traffic planning: Models, methods and challenge. *Journal of Global Optimization*, 7(4):381–405.

Mitsos, A., Lemonidis, P., and Barton, P. I. (2008). Global solution of bilevel programs with a nonconvex inner program. *Journal of Global Optimization*, 42(4):475–513.

Mittal, A. and Hanasusanto, G. A. (2021). Finding minimum volume circumscribing ellipsoids using generalized copositive programming. *Operations Research*.

Patriksson, M. and Wynter, L. (1999). Stochastic mathematical programs with equilibrium constraints. *Operations research letters*, 25(4):159–167.

Pólik, I. and Terlaky, T. (2007). A survey of the S-lemma. *SIAM review*, 49(3):371–418.

Prokhorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214.

Ruszczyński, A. and Shapiro, A. (2003). Optimality and duality in stochastic programming. *Handbooks in Operations Research and Management Science*, 10:65–139.

Ruszczyński, A. and Shapiro, A. (2006). Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452.

Ryu, J.-H., Dua, V., and Pistikopoulos, E. N. (2004). A bilevel programming framework for enterprise-wide process networks under uncertainty. *Computers & Chemical Engineering*, 28(6-7):1121–1129.

Scaparra, M. P. and Church, R. L. (2008). A bilevel mixed-integer program for critical infrastructure protection planning. *Computers & Operations Research*, 35(6):1905–1923.

Shapiro, A. (2006). Stochastic programming with equilibrium constraints. *Journal of Optimization Theory and Applications*, 128(1):221–243.

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

Shapiro, A. and Xu, H. (2008). Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization*, 57(3):395–418.

Sun, H. and Xu, H. (2016). Convergence analysis for distributionally robust optimization and equilibrium problems. *Mathematics of Operations Research*, 41(2):377–401.

Tavaslıoğlu, O., Prokopyev, O. A., and Schaefer, A. J. (2019). Solving stochastic and bilevel mixed-integer programs via a generalized value function. *Operations Research*, 67(6):1659–1677.

Todd, M. J. (2001). Semidefinite optimization. *Acta Numerica*, 10:515–560.

Tsoukalas, A., Rustem, B., and Pistikopoulos, E. N. (2009). A global optimization algorithm for generalized semi-infinite, continuous minimax with coupled constraints and bi-level problems. *Journal of Global Optimization*, 44(2):235–250.

Tuy, H., Migdalas, A., and Hoai-Phuong, N. (2007). A novel approach to bilevel nonlinear programming. *Journal of Global Optimization*, 38(4):527–554.

Wang, Z., You, K., Song, S., and Zhang, Y. (2020). Second-order conic programming approach for Wasserstein distributionally robust two-stage linear programs. *arXiv preprint arXiv:2002.06751*.

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

Xie, W. (2019). Tractable reformulations of distributionally robust two-stage stochastic programs with $\infty-$ Wasserstein distance. *arXiv preprint arXiv:1908.08454*.

Xie, W. and Ahmed, S. (2018). Distributionally robust simple integer recourse. *Computational Management Science*, 15(3-4):351–367.

Xu, G. and Burer, S. (2018). A data-driven distributionally robust bound on the expected optimal value of uncertain mixed 0-1 linear programming. *Computational Management Science*, 15(1):111–134.

Xu, H. (2005). An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium. *Optimization*, 54(1):27–57.

Yanıkoğlu, I. and Kuhn, D. (2018). Decision rule bounds for two-stage stochastic bilevel programs. *SIAM Journal on Optimization*, 28(1):198–222.

Yue, M.-C., Kuhn, D., and Wiesemann, W. (2021). On linear optimization over Wasserstein balls. *Mathematical Programming*, pages 1–16.

Zhang, J. and Özaltın, O. Y. (2021). Bilevel integer programs with stochastic right-hand sides. *INFORMS Journal on Computing*, 33(4):1644–1660.

Zhang, Y., Jiang, R., and Shen, S. (2018). Ambiguous chance-constrained binary programs under mean-covariance information. *SIAM Journal on Optimization*, 28(4):2922–2944.

Zhao, C. and Guan, Y. (2018). Data-driven risk-averse stochastic optimization with Wasserstein metric. *Operations Research Letters*, 46(2):262–267.

Appendix A: Proof of Lemma 1

Proof of Lemma 1: First, we have $\mathbb{E}_F[\Psi(x,\xi)] \geq Q_F(x)$ as the recourse variables $y(\xi)$ are restricted to a smaller class of function map. It remains to show that there exists a $y^*(\xi) \in \underset{y'(\xi)\in\mathcal{L}_n^2(F)}{\arg\min} \left\{c(\xi)^\top y'(\xi) : Ay'(\xi) \leq b_x(\xi)\right\}$ such that $\mathbb{E}_F[v(\xi)^\top y^*(\xi)] \geq \mathbb{E}_F[\Psi(x,\xi)]$.

According to Assumption 3, for a given leader's decision $x \in \mathcal{X}$ and any given $\xi \in \mathcal{S}$, the set of optimal solutions of the follower's problem $\Omega(x,\xi)$ is a polyhedral set spanned over a subset of the extreme points of the follower's feasible region $\mathcal{Y}(\xi) := \{Ay \leq b_x(\xi)\}$. Hence, one optimal solution of $\Psi(x,\xi)$ must be obtained at some extreme point of the feasible region $\mathcal{Y}(\xi)$. There exist finitely many matrices $L_j^y \in \mathbb{R}^{m\times n},\ j = 1,\ldots,J_1$, for every basis of the constraint matrix A, such that each extreme point of $\mathcal{Y}(\xi)$ is represented as $L_j^y b_x(\xi)$. For every $x \in \mathcal{X}$ and $\xi \in \mathcal{S}$, we define the following index set.

$$\mathcal{J}^y(\xi,x) := \left\{j :\ L_j^y b_x(\xi) \in \mathcal{Y}(\xi),\ c(\xi)^\top \left(L_j^y b_x(\xi) - L_{j'}^y b_x(\xi)\right) \leq 0,\ \forall 1 \leq j' \leq J_1 \text{ with } L_{j'}^y b_x(\xi) \in \mathcal{Y}(\xi)\right\},$$

where the set $\mathcal{J}^y(\xi,x)$ consists of all the indices j such that $L_j^y b_x(\xi)$ is an optimal solution in the set $\Omega(x,\xi)$ of optimal solutions to the follower's problem. We then consider a partition of the support set $\mathcal{S}$ of ξ using $\mathcal{J}^y(\xi,x)$.

$$\begin{aligned}\Xi_j^y(x) = \Big\{\xi \in \mathbb{R}^k :\ & v(\xi)^\top \left(L_j^y b_x(\xi) - L_{j'}^y b_x(\xi)\right) \geq 0,\ \forall j' \in \{1,\ldots,j-1\} \cap \mathcal{J}^y(\xi,x),\\ & v(\xi)^\top \left(L_j^y b_x(\xi) - L_{j'}^y b_x(\xi)\right) > 0,\ \forall j' \in \{j+1,\ldots,J_1\} \cap \mathcal{J}^y(\xi,x)\Big\}.\end{aligned}$$

The partition $\Xi_j^y(x)$ consists of all the ξ such that j is the largest index for which $j \in \mathcal{J}^y(\xi,x)$ and $y^*(\xi) := L_j^y b_x(\xi)$ maximizes $v(\xi)^\top y$. Moreover,

$$\begin{aligned}\mathbb{E}_F[\|y^*(\xi)\|^2] &\leq \mathbb{E}_F\left[\max_{1\leq j\leq J_1} \|L_j^y b_x(\xi)\|^2\right] = \mathbb{E}_F\left[\max_{1\leq j\leq J_1} \left\{b_x(\xi)^\top (L_j^y)^\top L_j^y b_x(\xi)\right\}\right]\\ &\leq \max_{1\leq j\leq J_1} \|L_j^y\|^2\, \mathbb{E}_F[\|b_x(\xi)\|^2] < \infty\end{aligned}$$

The first inequality holds since $\|y^*(\xi)\| \leq \max_{1\leq j\leq J_1} \|L_j^y b_x(\xi)\|$ for any given $\xi \in \Xi_j^y(x)$. The second inequality holds because $b_x(\xi)^\top (L_j^y)^\top L_j^y b_x(\xi) \leq \|L_j^y\|^2 \|b_x(\xi)\|^2$ for any ξ, where the norm of a matrix refers to the Frobenius norm, i.e., $\|X\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n X_{ij}^2}$, for $X \in \mathbb{R}^{m\times n}$. Thus, $y^*(\xi) \in \mathcal{L}_n^2(F)$ and is feasible to the maximization problem of $Q_F(x)$. Therefore, $Q_F(x)$ is no less than $\mathbb{E}_F[\Psi(x,\xi)]$ and we conclude the proof. $\square$

Appendix B: Proof of Lemma 2

Proof of Lemma 2: Define

$$\begin{aligned}\varphi_1(x,\lambda) &:= \inf_{p(\xi)\in\mathcal{L}_m^2(F)} \left\{\mathbb{E}_F[b_x(\xi)^\top p(\xi)] :\ A^\top p(\xi) + \lambda c(\xi) = v(\xi),\ p(\xi) \geq 0\right\},\\ \varphi_2(x,\lambda) &:= \inf_{y(\xi)\in\mathcal{L}_n^2(F)} \left\{\mathbb{E}_F[c(\xi)^\top y(\xi)] :\ Ay(\xi) \leq \lambda b_x(\xi)\right\}.\end{aligned}$$

To establish the equivalence between problem (7) and (8), we first will show that (i) problem (7) is equivalent to the following minimization problem over a scaler variable $\lambda \geq 0$, $p(\xi) \in \mathcal{L}_m^2(F)$, and $y(\xi) \in \mathcal{L}_n^2(F)$.

$$\min_{\lambda \geq 0} \ \varphi_1(x, \lambda) + \varphi_2(x, \lambda), \tag{36}$$

Then it remains to (ii) show the equivalence between (36) and (8).

(i) Equivalence between (7) and (36). For this part, the proof is inspired by the proof of Theorem 4.1 in Yanıkoğlu and Kuhn (2018). Dualizing problem (7) yields the following dual problem.

$$\inf_{\lambda, y(\xi) \in \mathcal{L}_n^2(F)} \varphi_1(x, \lambda) + \lambda \mathbb{E}_F[\bar{Q}_F(x)] \tag{37}$$

The strong duality holds for the primal and dual pair (Ruszczyński and Shapiro 2006). Given that $\lambda \geq 0$ and the definition of $\bar{Q}_F(x)$ in Proposition 1, problem (37) is rewritten as

$$\inf_{\lambda, y(\xi) \in \mathcal{L}_n^2(F)} \quad \varphi_1(x, \lambda) + \lambda \mathbb{E}_F[c(\xi)^\top y(\xi)] \tag{38a}$$

$$\text{s.t.} \quad Ay(\xi) \leq \ b_x(\xi). \tag{38b}$$

Problem (38) is nonlinear because of the product of λ and $y(\xi)$ in the objective. We linearize it by letting $y(\xi)$ be $\lambda y(\xi)$ and obtain the reformulation (36). The reformulation is exact. First, given a feasible solution (λ, y, p) to (38), then $(\lambda, \lambda y, p)$ is feasible to (36) with the same objective value. Now, consider a feasible solution (λ, y, p) to (36). If $\lambda > 0$, then $(\lambda, y/\lambda, p)$ is feasible to (38) with the same objective value. If $\lambda = 0$, the constraint of $\varphi_2(x, \lambda)$ becomes $Ay(\xi) \leq 0$ and thus implies that $y(\xi) = 0$ almost surely due to Assumption 3. Hence, $(0, y', p)$, with any $y' \in \mathcal{L}_n^2$ such that $Ay'(\xi) \leq b_x(\xi)$ holds almost surely, is feasible to (38) and achieves the same objective value as (λ, y, p). Therefore, (36) is an exact reformulation of (38) or, equivalently, (7).

(ii) Equivalence between (36) and (8). To establish the equivalence, we need to show the following equivalence.

$$\varphi_1(x, \lambda) = \phi_1(x, \lambda) := \mathbb{E}_F \left[\min_{p \in \mathbb{R}^m} \left\{ b_x(\xi)^\top p \mid A^\top p + \lambda c(\xi) = v(\xi), p \geq 0 \right\} \right], \tag{39}$$

$$\varphi_2(x, \lambda) = \phi_2(x, \lambda) := \mathbb{E}_F \left[\min_{y \in \mathbb{R}^n} \left\{ c(\xi)^\top y \mid Ay \leq \lambda b_x(\xi) \right\} \right]. \tag{40}$$

Note that $\varphi_1(x, \lambda)$ and $\varphi_2(x, \lambda)$ are bounded below as the second-stage expected cost is bounded below with respect to distribution $F \in \mathcal{D}$ by virtue of Assumption 3.

(39): It is clear that the right-hand side $\phi_1(x, \lambda)$ is no greater than $\varphi_1(x, \lambda)$ as $\varphi_1(x, \lambda)$ is more restrictive on the function map $p(\xi)$. It remains to show that $\varphi_1(x, \lambda)$ is no greater than $\phi_1(x, \lambda)$.

We first show that $\phi_1(x,\lambda)$ is bounded below. Denote $u(x,\lambda,\xi) := \min_{p\in\mathbb{R}^m}\{b_x(\xi)^\top p : A^\top p + \lambda c(\xi) = v(\xi),\ p \geq 0\}$. Rewrite $u(x,\lambda,\xi) = \max_{\eta\in\mathbb{R}^n}\left\{(v(\xi)-\lambda c(\xi))^\top \eta :\ A\eta \leq b_x(\xi)\right\}$ with the dual variable η. Due to Assumption 3, $u(x,\lambda,\xi)$ is bounded for every ξ and x.

Next, we establish the equivalence of (39). We form the second partition $\left\{\bar{\Xi}_j^p(x,\lambda)\right\}_{j=1}^{J_2}$ of $\mathcal{S}$ for $x\in\mathcal{X}$ and $\lambda\geq 0$, such that j is the largest index so that $L_j^p d(\lambda,\xi)$ is an optimal solution of $u(x,\lambda,\xi)$.

$$\begin{aligned}\bar{\Xi}_j^p(x,\lambda) := \Big\{\xi\in\mathbb{R}^k :\ & A^\top L_j^p d(\lambda,\xi) = d(\lambda,\xi),\ L_j^p d(\lambda,\xi)\geq 0\\ & b_x(\xi)^\top\left(L_j^p - L_{j'}^p\right) d(\lambda,\xi)\leq 0\ \forall j'=1,\ldots,j-1\ \text{ with } A^\top L_{j'}^p d(\lambda,\xi) = d(\lambda,\xi),\ L_{j'}^p d(\lambda,\xi)\geq 0,\\ & b_x(\xi)^\top\left(L_j^p - L_{j'}^p\right) d(\lambda,\xi)\leq 0\ \forall j'=j+1,\ldots,J_2\ \text{ with } A^\top L_{j'}^p d(\lambda,\xi) = d(\lambda,\xi),\ L_{j'}^p d(\lambda,\xi)\geq 0\Big\}.\end{aligned}$$

Denote $p^*(x,\lambda,\xi) := L_j^p d(\lambda,\xi)$ for $\xi\in\bar{\Xi}(x,\lambda),\ j=1,\ldots,J_2$. We can show that $p^*(x,\lambda,\xi)\in\mathcal{L}_m^2(F)$ and the equivalence between $\phi_1(x,\lambda)$ and $\varphi_1(x,\lambda)$ is established. The proof is similar to that of Lemma 1 and thus is omitted.

(40): According to Assumption 3, $\{y\in\mathbb{R}^n\,|Ay\leq\lambda b_x(\xi)\}$ is a compact polyhedron and thus $\min\limits_{y\in\mathbb{R}^n}\left\{c(\xi)^\top y|Ay\leq\lambda b_x(\xi)\right\}$ is bounded below almost surely for $\xi\in\mathcal{S}$. Then following the similar proof of Lemma 1, we establish the equivalence as desired.

Therefore, the proof is complete. □

Appendix C: Proof of Theorem 1: Interchangeability between the Minimization and Maximization

Proof of Theorem 1: As Lemma 2 implies that $\sup_{F\in\mathcal{D}}\mathbb{E}_F[\Psi(x,\xi)] = \sup_{F\in\mathcal{D}}\min_{\lambda\geq 0}\mathbb{E}_F[\Phi_\lambda(x,\xi)]$, to prove Theorem 1, it suffices to show that

$$\sup_{F\in\mathcal{D}}\min_{\lambda\geq 0}\mathbb{E}_F[\Phi_\lambda(x,\xi)] = \min_{\lambda\geq 0}\sup_{F\in\mathcal{D}}\mathbb{E}_F[\Phi_\lambda(x,\xi)].$$

Consider a weakly compact ambiguity set $\mathcal{D}$ of probability measures on $(\mathcal{S},\mathcal{F})$ and a convex neighborhood $N_\delta := \{\lambda :\ \lambda\geq -\delta\}$ of the feasible region of λ, where $\delta>0$. According to the minimax analysis of stochastic problems (see Theorem 2.1 of Shapiro and Kleywegt (2002), Theorem 49 of Ruszczyński and Shapiro (2003)), we need to prove the following statements to show the interchangeability of the maximization and minimization.

1. For $\lambda\in N_\delta$, $\sup_{F\in\mathcal{D}}\mathbb{E}_F[\Phi_\lambda(x,\xi)]<+\infty$.
2. For every $\xi\in\mathcal{S}$, function $\Phi_\lambda(x,\xi)$ is convex for $\lambda\in N_\delta$.
3. Given an optimization solution $\bar{\lambda}$ of (10), for every λ in a neighborhood of $\bar{\lambda}$, the function $\Phi_\lambda(x,\xi)$ is bounded and upper semicontinuous on $\xi\in\mathcal{S}$ and function $\Phi_{\bar{\lambda}}(x,\xi)$ is bounded and continuous on $\xi\in\mathcal{S}$.

Statement 1. Denote $\Phi^1_\lambda(x,\xi) := \min_y \{c(\xi)^\top y:\ Ay \le \lambda b_x(\xi)\}$ and $\Phi^2_\lambda(x,\xi) := \min_p \{b_x(\xi)^\top p:\ A^\top p + \lambda c(\xi) = v(\xi),\ p \ge 0\}$. To show that $\sup_{F\in\mathcal{D}} \mathbb{E}_F[\Phi_\lambda(x,\xi)] < +\infty$, it suffices to show that $\Phi_\lambda(x,\xi) = \Phi^1_\lambda(x,\xi) + \Phi^2_\lambda(x,\xi)$. Due to Assumption 3, $\Phi^1_\lambda(x,\xi)$ is bounded for every ξ and x. It remains to prove that $\Phi^2_\lambda(x,\xi)$ is bounded. Rewrite $\Phi^2_\lambda(x,\xi) = \max_\eta \{(v(\xi) - \lambda c(\xi))^\top \eta:\ A\eta \le b_x(\xi)\}$ with the dual variable η. Due to Assumption 3, $\Phi^2_\lambda(x,\xi)$ is bounded for every ξ and x. Thus $\sup_{F\in\mathcal{D}} \mathbb{E}_F[\Phi_\lambda(x,\xi)] < +\infty$.

Statement 2. Given any $\xi \in \mathcal{S}$, we derive the dual problem of $\Phi_\lambda(x,\xi)$.

$$\begin{aligned}\Phi_\lambda(x,\xi) = \max_{\varrho\in\mathbb{R}^m,\eta\in\mathbb{R}^n}\ & v(\xi)^\top\eta - \lambda c(\xi)^\top\eta - \lambda b_x(\xi)^\top\varrho \\ \text{s.t. }\ & A^\top\varrho + c(\xi) = 0,\ A\eta \le b_x(\xi),\ \varrho \ge 0,\end{aligned}$$

where $\eta \in \mathbb{R}^n$ and $\varrho \in \mathbb{R}^m$ are the dual variables. As $\Phi_\lambda(x,\xi)$ is bounded below, the optimal value $\Phi_\lambda(x,\xi)$ can be view as a pointwise maximum of (finitely many) linear functions. Hence $\Phi_\lambda(x,\xi)$ is a convex function of $\lambda \in N_\delta$.

Statement 3. As Statement 1 shows that $\Phi_\lambda(x,\xi)$ is bounded below for any $\xi \in \mathcal{S}$, it remains to show that, for any $\lambda \in N_\delta$, $\Phi_\lambda(x,\xi)$ is Lipschitz continuous on $\xi \in \mathcal{S}$. Given a $\lambda \ge 0$, consider any $\xi_1 \ne \xi_2$, $\xi_1, \xi_2 \in \mathcal{S}$. Without loss of generality, assume that $\Phi_\lambda(x,\xi_1) \le \Phi_\lambda(x,\xi_2)$. Let (y^*, p^*) denote the optimal solution of $\Phi_\lambda(x,\xi_1)$. We then rewrite $\Phi_\lambda(x,\xi_2)$ as

$$\begin{aligned}\min_{\Delta y\in\mathbb{R}^n,\Delta p\in\mathbb{R}^m}\ & b(\xi_2)^\top(p^* + \Delta p) + c(\xi_2)^\top(y^* + \Delta y) && \text{(41a)}\\ \text{s.t. }\ & A^\top(p^* + \Delta p) + \lambda c(\xi_2) = v(\xi_2),\ p^* + \Delta p \ge 0 && \text{(41b)}\\ & A(y^* + \Delta y) \le \lambda b_x(\xi_2). && \text{(41c)}\end{aligned}$$

Note that (y^*, p^*) is feasible to $\Phi_\lambda(x,\xi_1)$ and thus $A^\top p^* + \lambda c(\xi_1) = v(\xi_1),\ p^* \ge 0,\ Ay^* \le \lambda b_x(\xi_1)$. Define

$$\phi^y_\lambda(x,\xi_1,\xi_2) := \min_{\Delta y\in\mathbb{R}^n} \{c(\xi_2)^\top \Delta y:\ A\Delta y \le \lambda(b_x(\xi_2) - b_x(\xi_1))\}, \quad (42)$$

$$\phi^p_\lambda(x,\xi_1,\xi_2) := \min_{\Delta p\in\mathbb{R}^m} \{b(\xi_2)^\top \Delta p:\ \Delta p \ge 0,\ A^\top\Delta p = (v(\xi_2) - v(\xi_1)) - \lambda(c(\xi_2) - c(\xi_1))\}. \quad (43)$$

Then $\Phi_\lambda(x,\xi_2) \le b(\xi_2)^\top p^* + c(\xi_2)^\top y^* + \phi^y_\lambda(x,\xi_1,\xi_2) + \phi^p_\lambda(x,\xi_1,\xi_2)$ because of two reasons: (1) the constraint in (42) is more restrictive than (41c) as $\lambda(b_x(\xi_2) - b_x(\xi_1)) \le \lambda b_x(\xi_2) - Ay^*$, and (2) the constraint $\Delta p \ge 0$ of (43)is more restrictive than $\Delta p \ge -p^*$ in (41b).

Now, consider

$$\begin{aligned}|\Phi_\lambda(x,\xi_2) - \Phi_\lambda(x,\xi_1)| &= \Phi_\lambda(x,\xi_2) - \Phi_\lambda(x,\xi_1) && \text{(44a)}\\ &\le |\,(b(\xi_2) - b(\xi_1))^\top p^*| + |\,(c(\xi_2) - c(\xi_1))^\top y^*| + |\phi^y_\lambda(x,\xi_1,\xi_2)| + |\phi^p_\lambda(x,\xi_1,\xi_2)|. && \text{(44b)}\end{aligned}$$

We then show that each term in (44b) is bounded above by a product of a constant and $\|\xi_2 - \xi_1\|$ for Lipschitz continuity. Recall the linearity assumption 2 where $c(\xi) = C\xi + c_0$ for $C \in \mathbb{R}^{n\times k}$, $c_0 \in \mathbb{R}^n$, $v(\xi) = V\xi + v_0$ for $V \in \mathbb{R}^{n\times k}$, $v_0 \in \mathbb{R}^n$, $b_x(\xi) = B_x\xi + b_{x0}$ for $B_x \in \mathbb{R}^{m\times k}$, $b_{x0} \in \mathbb{R}^m$.

1. $|\left(b(\xi_2)-b(\xi_1)\right)^\top p^*|$: Similarly as in the proof of Lemma 2, there are finitely many matrices $L_j^p \in \mathbb{R}^{n\times m},\ j=1,\ldots,J_2$, one for every basis of the constraint matrix $A^\top$, so that $L_j^p\left(v(\xi_1)-\lambda c(\xi_1)\right)$ represents a basic feasible solution in $\left\{p\in\mathbb{R}^m:\ A^\top p+\lambda c(\xi_1)=v(\xi_1),\ p\geq 0\right\}$. Let $L_{j^*}^p$ be the matrix associated with p^*. We then have
$$|\left(b(\xi_2)-b(\xi_1)\right)^\top p^*| \leq \|B_x(\xi_2-\xi_1)\|\|L_{j^*}^p\left(v(\xi_1)-\lambda c(\xi_1)\right)\| \leq L_1\|\xi_2-\xi_1\|, \tag{45}$$
where $L_1 := \|B_x\|\max_{1\leq j\leq J_2}\max_{\xi\in\mathcal{S}}\|L_j^p\left(v(\xi)-\lambda c(\xi)\right)\|$. The maximum is attainable due to the compactness of $\mathcal{S}$.
2. $|\left(c(\xi_2)-c(\xi_1)\right)^\top y^*|$: Similarly as in the proof of Lemma 2, there are finitely many matrices $L_j^y\in\mathbb{R}^{m\times n},\ j=1,\ldots,J_1$, one for each basis of the constraint matrix A such that $\lambda L_j^y b_x(\xi_1)$ is a basic feasible solution in $\{y\in\mathbb{R}^n:\ Ay\leq\lambda b_x(\xi_1)\}$. Let $L_{j^*}^y$ be the matrix associated with y^*. So
$$|\left(c(\xi_2)-c(\xi_1)\right)^\top y^*| \leq \|C(\xi_2-\xi_1)\|\|\lambda L_{j^*}^y b_x(\xi_1)\| \leq L_2\|\xi_2-\xi_1\|, \tag{46}$$
where $L_2 := |\lambda|\|C\|\max_{1\leq j\leq J_1}\max_{\xi\in\mathcal{S}}\|L_j^y b_x(\xi)\|$ and the maximum is attainable due to the compactness of $\mathcal{S}$.
3. $|\phi_\lambda^y(x,\xi_1,\xi_2)|$: Following a similar argument as for the term $|\left(c(\xi_2)-c(\xi_1)\right)^\top y^*|$, let $L_{\bar{j}}^y$ be the matrix associated with an optimal Δy. Then
$$|\phi_\lambda^y(x,\xi_1,\xi_2)| \leq \|c(\xi_2)\|\|\lambda L_{\bar{j}}^y B_x(\xi_2-\xi_1)\| \leq L_3\|\xi_2-\xi_1\|, \tag{47}$$
where $L_3 := |\lambda|\max_{\xi\in\mathcal{S}}\|c(\xi)\|\max_{1\leq j\leq J_1}\|L_j^y B_x\|$.
4. $|\phi_\lambda^p(x,\xi_1,\xi_2)|$: Let $L_{\bar{j}}^p$ be the matrix associated with an optimal Δp. We have
$$|\phi_\lambda^p(x,\xi_1,\xi_2)| \leq \|b(\xi_2)\|\|L_{\bar{j}}^p\left(V-\lambda C\right)(\xi_2-\xi_1)\| \leq L_4\|\xi_2-\xi_1\|, \tag{48}$$
where $L_4 := \max_{\xi\in\mathcal{S}}\|b(\xi)\|\max_{1\leq j\leq J_2}\|L_j^p\left(V-\lambda C\right)\|_{HS}$.

Combining (44)-(48), we obtain $|\Phi_\lambda(x,\xi_2)-\Phi_\lambda(x,\xi_1)|\leq L\|\xi_2-\xi_1\|$, where $L := L_1+L_2+L_3+L_4$.

Therefore, the proof is completed. □

Appendix D: Proof of Theorem 2: 0-1 SDP Approximation for $\mathcal{D}_M$

Proof of Theorem 2: The constraint (13b) is conservatively approximated by requiring the right-hand side to be a convex function. Denote
$$\tilde{\mathcal{Q}} := \begin{bmatrix} Q-\frac{1}{2}(B_x^\top P+C^\top Y)-\frac{1}{2}\left(B_x^\top P+C^\top Y\right)^\top & \frac{1}{2}(q-B_x^\top p_0-P^\top b_{x0}-C^\top y_0-Y^\top c_0-W^\top\tau) \\ \frac{1}{2}(q-B_x^\top p_0-P^\top b_{x0}-C^\top y_0-Y^\top c_0-W^\top\tau)^\top & r-b_{x0}^\top p_0-c_0^\top y_0+\tau^\top h \end{bmatrix} \in \mathbb{S}^{k+1},$$
where $\mathbb{S}^n$ denotes the space of $n\times n$ symmetric matrices. Applying the S-lemma (e.g., Pólik and Terlaky 2007), given a nonnegative variable $\tau\geq 0$, the constraint (13b) is implied by the following SDP constraint: $\tilde{\mathcal{Q}}\succeq 0$, which is further equivalent to constraints (20b)-(20c) by replacing the bilinear terms with the auxiliary variables Γ_i,ω_i,ρ_i. We complete the proof. □

Appendix E: Proof of Proposition 3

Proof of Proposition 3:

1. A feasible solution of (23) can be expressed as

$$\begin{bmatrix} 1 & (z^*)^\top \\ z^* & Z^* \end{bmatrix} = \sum_i \begin{bmatrix} \nu_i^{*2} & \nu_i^* \boldsymbol{\beta}_i^{*\top} \\ \nu_i^* \boldsymbol{\beta}_i^* & \boldsymbol{\beta}_i^* \boldsymbol{\beta}_i^{*\top} \end{bmatrix} = \begin{bmatrix} \sum_i \nu_i^{*2} & \sum_i \nu_i^* \boldsymbol{\beta}_i^{*\top} \\ \sum_i \nu_i^* \boldsymbol{\beta}_i^* & \sum_i \boldsymbol{\beta}_i^* \boldsymbol{\beta}_i^{*\top} \end{bmatrix}, \quad \sum_i \nu_i^* \boldsymbol{\beta}_{i,k+1}^* = 1 \text{ and } \sum_i \boldsymbol{\beta}_{i,k+1}^{*2} = 1,$$

   where $\boldsymbol{\beta}_i^* \in \mathbb{R}^{k+1},\ \forall i$ and $\boldsymbol{\nu}_i^* \in \mathbb{R},\ \forall i$. Let $\hat{Z} = \sum_i \boldsymbol{\beta}_i^* \boldsymbol{\beta}_i^{*\top}$. It is easy to verify the feasibility of Z^* to (24).

2. A feasible solution of (24) can be expressed as $\bar{Z} = \begin{bmatrix} \Psi^* & \xi^{*\top} \\ \xi^* & 1 \end{bmatrix} = \sum_i \boldsymbol{\zeta}_i^* \boldsymbol{\zeta}_i^{*\top}$, where $\boldsymbol{\zeta}_i^* \in \mathbb{R}^{k+1},\ \forall i$.

   Let $z' = \sum_i \zeta_{i,k+1}^* \zeta_i^*$ and $Z' = \sum_i \zeta_i^* \zeta_i^{*\top}$ and $\begin{bmatrix} 1 & (z')^\top \\ z' & Z' \end{bmatrix}$ is feasible to (23).

The proof completes. □

Appendix F: Proof of Theorem 3

Proof of Theorem 3: Constraint (21) is equivalent to

$$r \geq \min \left\{ u :\ u e_{k+1} e_{k+1}^\top - \mathcal{Q} \in \mathcal{COP}(\hat{\Xi}) \right\}. \tag{49}$$

Given that Problem (12) is a minimization problem and that constraint (49) is equivalent to (13b), for an optimal solution to (12), $r = u$. The proof concludes. □

Appendix G: Proof of Theorem 5: Worst-case Distribution for $\mathcal{D}_{\text{dis}}$

Proof of Theorem 5: Given a leader's solution $\hat{x}$, the dual of (28) is derived as follows.

$$\begin{aligned}
\mathbf{SP}_{\text{dis}}(\hat{x}):\quad \max_{\sigma,\mu,\chi,\gamma}\ & \textstyle\sum_{s=1}^N \chi^{s\top} v(\xi^s) && \text{(50a)}\\
\text{s.t.}\quad & \gamma_2 \Sigma_0 + \mu_0 \mu_0^\top - \sum_{s=1}^N \gamma^s \xi^s \xi^{s\top} + \sqrt{\gamma_1} \left(\mu_0 \mu^\top \Sigma_0^{1/2} + \Sigma_0^{1/2} \mu \mu_0^\top \right) \succeq 0 && \text{(50b)}\\
& \sum_{s=1}^N \gamma^s = 1, \quad \sqrt{\gamma_1} \Sigma_0^{1/2} \mu + \mu_0 - \sum_{s=1}^N \gamma^s \xi^s = 0, \quad \|\mu\|_2 \leq 1 && \text{(50c)}\\
& -A\chi^s + \gamma^s b_{\hat{x}}(\xi^s) \geq 0,\ s = 1, \ldots, N && \text{(50d)}\\
& \gamma^s c(\xi^s) + A^\top \sigma^s = 0,\ s = 1, \ldots, N && \text{(50e)}\\
& \sum_{s=1}^N \left(\chi^{s\top} c(\xi^s) + \sigma^{s\top} b_{\hat{x}}(\xi^s) \right) \leq 0 && \text{(50f)}\\
& \gamma^s \geq 0,\ \sigma^s \geq 0,\ s = 1, \ldots, N. && \text{(50g)}
\end{aligned}$$

Denote the optimal solution $(\sigma^*, \mu^*, \chi^*, \gamma^*)$. We first show that the optimal dual solution γ^* satisfy the three conditions required by the ambiguity set $\mathcal{D}_{\text{dis}}$, and then show that the expected value: $\min_{\lambda \geq 0} \mathbb{E}\left[\Phi_\lambda(\hat{x}, \xi)\right]$ with respect to the distribution $\{\mathbb{P}(\xi = \xi^s) = \gamma^{s*}\}_{s=1,\ldots,N}$ equals to the worst-case expectation.

1. $\sum_{s=1}^N \gamma^{s*} = 1,\ \gamma^{s*} \geq 0,\ s = 1, \ldots, N$: Because γ^* is feasible to (50).
2. $\left(\sum_{s=1}^N \gamma^{s*}\xi^s - \mu_0\right)^\top \Sigma_0^{-1} \left(\sum_{s=1}^N \gamma^{s*}\xi^s - \mu_0\right) \leq \gamma_1$:

$$\left(\sum_{s=1}^N \gamma^{s*}\xi^s - \mu_0\right)^\top \Sigma_0^{-1} \left(\sum_{s=1}^N \gamma^{s*}\xi^s - \mu_0\right) = \gamma_1 {\mu^*}^\top \Sigma_0^{1/2}\Sigma_0^{-1}\Sigma_0^{1/2}\mu^* = \gamma_1\|\mu^*\|^2 \leq \gamma_1.$$

   The first equality is due to the first constraint in (50c): $\sum_{s=1}^N \gamma^{s*}\xi^s - \mu_0 = \sqrt{\gamma_1}\Sigma_0^{1/2}\mu^*$. The last inequality is because of the last constraint in (50c).
3. $\sum_{s=1}^N \gamma^{s*}\left[(\xi^s - \mu_0)(\xi^s - \mu_0)^\top\right] \preceq \gamma_2\Sigma_0$:

$$\begin{aligned}\sum_{s=1}^N \gamma^{s*}(\xi^s - \mu_0)(\xi^s - \mu_0)^\top &= \sum_{s=1}^N \gamma^{s*}\xi^s{\xi^s}^\top - \mu_0\sum_{s=1}^N \gamma^{s*}{\xi^s}^\top - \sum_{s=1}^N \gamma^{s*}\xi^s\mu_0^\top + \mu_0\mu_0^\top \\ &= \sum_{s=1}^N \gamma^{s*}\xi^s{\xi^s}^\top - \mu_0(\sqrt{\gamma_1}\Sigma_0^{1/2}\mu^* + \mu_0)^\top - (\sqrt{\gamma_1}\Sigma_0^{1/2}\mu^* + \mu_0)\mu_0^\top + \mu_0\mu_0^\top \preceq \gamma_2\Sigma_0.\end{aligned}$$

   The second equality holds because of the first constraint in (50c). The inequality is according to constraint (50b).

Thus, the distribution $\{\mathbb{P}(\xi = \xi^s) = \gamma^{s*}\}_{s=1,\ldots,N} \in \mathcal{D}_{\text{dist}}$ and the expected value with respect to γ^*:

$$\min_{\lambda\geq 0} \mathbb{E}_{\gamma^*}[\Phi_\lambda(\hat{x}, \xi)] := \min_{\lambda\geq 0} \sum_{s=1}^N \gamma^{s*}\Phi_\lambda(\hat{x}, \xi^s) \leq \sup_{\mathcal{F}\in\mathcal{D}_{\text{dis}}} \min_{\lambda\geq 0} \mathbb{E}_F[\Phi_\lambda(\hat{x}, \xi)].$$

It remains to show that the expected value $\min_{\lambda\geq 0} \mathbb{E}_{\gamma^*}[\Phi_\lambda(\hat{x}, \xi)] \geq \sup_{\mathcal{F}\in\mathcal{D}_{\text{dis}}} \min_{\lambda\geq 0} \mathbb{E}_F[\Phi_\lambda(\hat{x}, \xi)]$, where the equality is according to Theorem 1. To see this, denote $\tilde{\lambda}, \tilde{p}^s, \tilde{y}^s,\ s = 1, \ldots, N$ the optimal solution to

$$\begin{aligned}\min_{\lambda\geq 0} \mathbb{E}_{\gamma^*}[\Phi_\lambda(\hat{x}, \xi^s)] = \min_{\lambda\geq 0, p, y}\ & \sum_{s=1}^N \gamma^{s*}\left(b_{\hat{x}}(\xi^s)^\top p^s + c(\xi^s)^\top y^s\right) \\ \text{s.t.}\ & A^\top p^s + \lambda c(\xi^s) = v(\xi^s),\ s = 1, \ldots, N \\ & Ay^s \leq \lambda b_{\hat{x}}(\xi^s),\ p^s \geq 0,\ s = 1, \ldots, N.\end{aligned}$$

We then have $\min_{\lambda\geq 0} \mathbb{E}_{\gamma^*}[\Phi_\lambda(\hat{x}, \xi)] =$

$$\begin{aligned}\sum_{s=1}^N \gamma^{s*}\left(b_{\hat{x}}(\xi^s)^\top \tilde{p}^s + c(\xi^s)^\top \tilde{y}^s\right) &\geq \sum_{s=1}^N \left({\chi^{s*}}^\top A^\top \tilde{p}^s - {\sigma^{s*}}^\top A\tilde{y}^s\right) \\ &\geq \sum_{s=1}^N \left[{\chi^{s*}}^\top \left(v(\xi^s) - \tilde{\lambda}c(\xi^s)\right)\tilde{p}^s - \tilde{\lambda}{\sigma^{s*}}^\top b_{\hat{x}}(\xi^s)\right] \\ &\geq \sum_{s=1}^N {\chi^{*s}}^\top v(\xi^s) = \sup_{\mathcal{F}\in\mathcal{D}_{\text{dis}}} \min_{\lambda\geq 0} \mathbb{E}_F[\Phi_\lambda(\hat{x}, \xi)].\end{aligned}$$

The first inequality holds because of constraints (50d)-(50e) and the nonnegativity of $\tilde{p}^s$. The second inequality holds due to constraints (51)-(51) and the nonnegativity of σ^{s*}. The third inequality

holds because of (50f) and $\lambda \geq 0$. The equality holds since strong duality holds between (28) and (50) (Todd 2001).

It follows that $\min_{\lambda\geq 0} \mathbb{E}_{\gamma^*}[\Phi_\lambda(\hat{x},\xi)] = \sup_{\mathcal{F}\in\mathcal{D}_{\rm dis}} \min_{\lambda\geq 0} \mathbb{E}_F[\Phi_\lambda(\hat{x},\xi)]$ and $\{\mathbb{P}(\xi = \xi^s) = \gamma^{s*}\}_{s=1,\ldots,N} \in \mathcal{D}_{\rm dist}$ characterize the worst-case distribution. □

Appendix H: Subproblems for Cutting-Plane Algorithms

In the cutting-plane algorithm, at each iteration, we solve the relaxed master problem MP (29) and obtain the optimal solution $(\hat{x},\hat{\nu})$. Then the leader's decision $\hat{x}$ is plugged into a subproblem, which admits attractable SDP formulations. In this section, we present the three subproblems for (1) the 0-1 SDP formulation (20) obtained by conservatively approximating the the nonconvex quadratic constraint (13b), for (2) the 0-1 SDP approximation (26) of the 0-1 COP exact reformulation of constraint (13b), and for the exact 0-1 SDP reformulation (27) under the discrete ambiguity set.

H.1. Subproblems and Optimality Cuts for 0-1 SDPs

H.1.1. Subproblem and optimality cut for (20) The worst-case second-stage problem is approximated using the following SDP formulation:

$$\begin{aligned}
\min\ & r+t && \text{(51a)}\\
\text{s.t.}\ & AY+\Lambda W=\lambda B_x,\ \Lambda h - Ay_0 + \lambda b_{x0}\geq 0,\ \Lambda\geq 0 && \text{(51b)}\\
& \tilde{\mathcal{Q}}\succeq 0 && \text{(51c)}\\
& (13e)-(13f),\ (19),\ (20c).
\end{aligned}$$

The subproblem is the dual problem of (51).

$$\begin{aligned}
\mathbf{SP}_{\rm SDP}(\hat{x}):\ \max_{U,G,H,\zeta,\sigma,\eta,\mu,\chi}\ & {\rm Tr}(VG)+\chi^\top v_0 && \text{(52a)}\\
\text{s.t.}\ & WUB_{\hat{x}}^\top + W\eta b_{\hat{x}0}^\top - WGA^\top - h\zeta^\top \geq 0,\ \zeta\geq 0 && \text{(52b)}\\
& {\rm Tr}(B_{\hat{x}}H)\geq {\rm Tr}(CG)+\chi^\top c_0+\sigma^\top b_{\hat{x}0} && \text{(52c)}\\
& B_{\hat{x}}\eta + b_{\hat{x}0} = A\chi+\zeta && \text{(52d)}\\
& -U+\gamma_2\Sigma_0+\mu_0\mu_0^\top+\sqrt{\gamma_1}(\mu_0\mu^\top\Sigma_0^{1/2}+\Sigma_0^{1/2}\mu\mu_0^\top)\succeq 0 && \text{(52e)}\\
& \eta=\mu_0+\sqrt{\gamma_1}\Sigma_0^{1/2}\mu,\ W\eta\geq h,\ \|\mu\|_2\leq 1 && \text{(52f)}\\
& \begin{bmatrix} U & \eta\\ \eta^\top & 1\end{bmatrix}\succeq 0 && \text{(52g)}\\
& WH+h\sigma^\top\leq 0,\ \sigma\geq 0 && \text{(52h)}\\
& UC^\top+\eta c_0^\top = HA && \text{(52i)}\\
& A^\top\sigma + C\eta + c_0 = 0. && \text{(52j)}
\end{aligned}$$

Note that the subproblem does not involve the big-M constants as in the 0-1 SDP formulation (20) in Section 4.1. Let $(\widehat{U},\widehat{G},\widehat{H},\widehat{\zeta},\widehat{\sigma},\widehat{\eta},\widehat{\mu},\widehat{\chi})$ be the optimal solution to the subproblem SP($\hat{x}$). When the optimal value is great than $\hat{\nu}$, we generate a valid cut into MP in the following form.

PROPOSITION 4. *Let $[\cdot]^+$ ($[\cdot]^-$) denote the element-wise positive (negative) part of a matrix and M be a sufficiently large big-M constant. The inequality*

$$\begin{aligned}
&\nu \geq \mathrm{Tr}(V\widehat{G}) + \widehat{\chi}^\top v_0 \\
&-M\textstyle\sum_{i=1}^d \Big\{\hat{x}_i \Big(\mathrm{Tr}[\mathbf{1}_{k\times k}([\widehat{U}]^+ + [\widehat{U}]^-)] + [\mathrm{Tr}(B_i\widehat{H}) - b_i^\top\widehat{\sigma}]^+ + ([B_i\widehat{\eta}+b_i]^+ + [B_i\widehat{\eta}+b_i]^-)^\top\mathbf{1}_m + ([\widehat{\eta}]^+ + [\widehat{\eta}]^-)^\top\mathbf{1}_k\Big)(1-x_i) \\
&+(1-\hat{x}_i)\Big(\mathrm{Tr}[\mathbf{1}_{k\times k}([\widehat{U}]^+ + [\widehat{U}]^-)] + [\mathrm{Tr}(B_i\widehat{H}) - b_i^\top\widehat{\sigma}]^- + ([B_i\widehat{\eta}+b_i]^+ + [B_i\widehat{\eta}+b_i]^-)^\top\mathbf{1}_m + ([\widehat{\eta}]^+ + [\widehat{\eta}]^-)^\top\mathbf{1}_k\Big)x_i\Big\}
\end{aligned} \tag{53}$$

is a specific optimality cut in the form of $\nu \geq u_l^\top x + a_l$ to the master problem to solve 0-1 SDP (20).

Proof of Proposition 4: To see that the inequality (53) is a valid cut, we first present an SDP formulation fo the worst-case second-stage problem with big-M constants.

$$\min\{r+t:\ (20b)-(20c),\ (13e)-(13f),\ (14)-(19)\}. \tag{54}$$

The SDP (54) is equivalent to (51) with binary $x=\hat{x}$. So their dual problems are equivalent as well. Specifically, the dual of (54) is

$$\begin{aligned}
\max_{\substack{U,G,H,\zeta,\sigma,\eta,\mu,\Pi_i^j,\pi_i^j,\\ \chi,\phi_i^j,\psi_i^j}}\quad & \mathrm{Tr}(VG) + \chi^\top v_0 - M\sum_{i=1}^d \Big(\big[\mathrm{Tr}[\mathbf{1}_{k\times k}(\Pi_i^1+\Pi_i^2)] + \pi_i^2 + (\phi_i^1+\phi_i^2)^\top\mathbf{1}_m + (\psi_i^1+\psi_i^2)^\top\mathbf{1}_k\big](1-\hat{x}_i) \\
& + \big[\mathbf{1}_{k\times k}\,\mathrm{Tr}(\Pi_i^3+\Pi_i^4) + \pi_i^1 + (\phi_i^3+\phi_i^4)^\top\mathbf{1}_m + (\psi_i^3+\psi_i^4)^\top\mathbf{1}_k\big]\hat{x}_i\Big) && \text{(55a)} \\
& \sum_{i=1}^d W(\Pi_i^2-\Pi_i^1)B_i^\top + \sum_{i=1}^d W(\psi_i^2-\psi_i^1)b_i^\top + WUB_0^\top + W\eta b_0^\top - WGA^\top - h\zeta^\top \geq 0, \\
& \zeta \geq 0 && \text{(55b)} \\
& U + \Pi_i^1 - \Pi_i^2 - \Pi_i^3 + \Pi_i^4 = 0,\ i=1,\ldots,d && \text{(55c)} \\
& B_0\eta + b_0 + \sum_{i=1}^d(\phi_i^2-\phi_i^1) = A\chi + \zeta && \text{(55d)} \\
& -\mathrm{Tr}(CG) - \chi^\top c_0 + \mathrm{Tr}(B_0H) - \sigma^\top b_0 + \sum_{i=1}^d(\pi_i^2-\pi_i^4) \geq 0 && \text{(55e)} \\
& \mathrm{Tr}(B_iH) - b_i^\top\sigma + \pi_i^1 - \pi_i^2 - \pi_i^3 + \pi_i^4 = 0,\ i=1,\ldots,d && \text{(55f)} \\
& B_i\eta + b_i + \phi_i^1 - \phi_i^2 - \phi_i^3 + \phi_i^4 = 0,\ i=1,\ldots,d && \text{(55g)} \\
& \eta + \psi_i^1 - \psi_i^2 - \psi_i^3 + \psi_i^4 = 0,\ i=1,\ldots,d && \text{(55h)}
\end{aligned}$$

$$\pi_i^j \geq 0,\ \Pi_i^j \geq 0,\ \phi_i^j \geq 0,\ \psi_i^j \geq 0,\ i=1,\ldots,d,\ j=1,\ldots,4 \tag{55i}$$

(52e) – (52j).

Given an optimal solution $(\overline{U},\overline{G},\overline{H},\overline{\zeta},\overline{\sigma},\overline{\eta},\overline{\mu},\overline{\Pi}_i^j,\overline{\pi}_i^j,\overline{\chi},\overline{\phi}_i^j,\overline{\psi}_i^j)$ to (55), if the optimal value of (55) is greater than $\hat{\nu}$, following strong duality (Todd 2001), an optimality cut $\nu \geq u_l^\top x + a_l$ to the master problem MP is specified as

$$\begin{aligned}\nu \ \geq\ & \mathrm{Tr}(V\overline{G}) + \overline{\chi}^\top v_0 - M\sum_{i=1}^{d}\Big(\big[\,\mathrm{Tr}[\mathbf{1}_{k\times k}(\overline{\Pi}_i^1+\overline{\Pi}_i^2)]+\overline{\pi}_i^2+(\overline{\phi}_i^1+\overline{\phi}_i^2)^\top\mathbf{1}_m+(\overline{\psi}_i^1+\overline{\psi}_i^2)^\top\mathbf{1}_k\big](1-x_i)\\ &+\big[\,\mathrm{Tr}[\mathbf{1}_{k\times k}(\overline{\Pi}_i^3+\overline{\Pi}_i^4)]+\overline{\pi}_i^1+(\overline{\phi}_i^3+\overline{\phi}_i^4)^\top\mathbf{1}_m+(\overline{\psi}_i^3+\overline{\psi}_i^4)^\top\mathbf{1}_k\big]x_i\Big).\end{aligned} \tag{56}$$

To show that (53) is a valid cut, it remains to prove that the optimal $(\Pi_i^j,\pi_i^j,\phi_i^j,\psi_i^j),\ i=1,\ldots,d,\ j=1,\ldots,4$, of (55) can be expressed using $(\widehat{U},\widehat{H},\widehat{\eta},\widehat{\sigma})$ which are optimal to SP($\hat{x}$). To this end, we present the following complementary slackness conditions and feasibility conditions associated with $(\Pi_i^j,\pi_i^j,\phi_i^j,\psi_i^j),\ i=1,\ldots,d,\ j=1,\ldots,4$.

$$\Pi_i^1(B_i^\top TW+(1-\hat{x}_i)M\mathbf{1}_{k\times k}-\Gamma_i)=0,\ \Pi_i^2(B_i^\top TW-(1-\hat{x}_i)M\mathbf{1}_{k\times k}-\Gamma_i)=0 \tag{57a}$$

$$\Pi_i^3(\hat{x}_iM\mathbf{1}_{k\times k}+\Gamma_i)=0,\ \Pi_i^4(\hat{x}_iM\mathbf{1}_{k\times k}-\Gamma_i)=0 \tag{57b}$$

$$\pi_i^1(\theta_i-M\hat{x}_i)=0,\ \pi_i^2(\theta_i-\lambda+M(1-\hat{x}_i))=0 \tag{57c}$$

$$\pi_i^3\theta_i=0,\ \pi_i^4(\theta_i-\lambda)=0,\ i=1,\ldots,d \tag{57d}$$

$$(p_0+(1-\hat{x}_i)M\mathbf{1}_m-\omega_i)^\top\phi_i^1=0,\ (p_0-(1-\hat{x}_i)M\mathbf{1}_m-\omega_i)^\top\phi_i^2=0 \tag{57e}$$

$$(\hat{x}_iM\mathbf{1}_m+\omega_i)^\top\phi_i^3=0,\ (\hat{x}_iM\mathbf{1}_m-\omega_i)^\top\phi_i^4=0 \tag{57f}$$

$$\big((TW)^\top b_i+(1-\hat{x}_i)M\mathbf{1}_k-\rho_i\big)^\top\psi_i^1=0,\ \big((TW)^\top b_i-(1-\hat{x}_i)M\mathbf{1}_k-\rho_i\big)^\top\psi_i^2=0 \tag{57g}$$

$$(\hat{x}_iM\mathbf{1}_k+\rho_i)^\top\psi_i^3=0,\ (\hat{x}_iM\mathbf{1}_k-\rho_i)^\top\psi_i^4=0 \tag{57h}$$

(55b) – (55i).

Denote $(\widehat{Q},\widehat{q},\widehat{r},\widehat{t},\widehat{Y},\widehat{\lambda},\widehat{\tau},\widehat{T},\widehat{\Lambda},\widehat{\Gamma}_i,\widehat{\theta}_i,\widehat{\omega}_i,\widehat{\rho}_i,\widehat{p}_0,\widehat{y}_0)$ be one optimal solution to (54). Let

$$\widehat{\Pi}_i^1=\hat{x}_i[\widehat{U}]^-,\ \widehat{\Pi}_i^2=\hat{x}_i[\widehat{U}]^+,\ \widehat{\Pi}_i^3=(1-\hat{x}_i)[\widehat{U}]^+,\ \widehat{\Pi}_i^4=(1-\hat{x}_i)[\widehat{U}]^-. \tag{58}$$

$$\widehat{\pi}_i^1=(1-\hat{x}_i)[\mathrm{Tr}(B_i\widehat{H})-b_i^\top\widehat{\sigma}]^-,\ \widehat{\pi}_i^2=\hat{x}_i[\mathrm{Tr}(B_i\widehat{H})-b_i^\top\widehat{\sigma}]^+,\ \widehat{\pi}_i^3=(1-\hat{x}_i)[\mathrm{Tr}(B_i\widehat{H})-b_i^\top\widehat{\sigma}]^+, \tag{59}$$

$$\widehat{\pi}_i^4 = \hat{x}_i[\mathrm{Tr}(B_i\widehat{H}) - b_i^\top \widehat{\sigma}]^-, \tag{60}$$

$$\widehat{\phi}_i^1 = \hat{x}_i[B_i\widehat{\eta} + b_i]^-,\ \widehat{\phi}_i^2 = \hat{x}_i[B_i\widehat{\eta} + b_i]^+,\ \widehat{\phi}_i^3 = (1-\hat{x}_i)[B_i\widehat{\eta} + b_i]^+,\ \widehat{\phi}_i^4 = (1-\hat{x}_i)[B_i\widehat{\eta} + b_i]^-, \tag{61}$$

$$\widehat{\psi}_i^1 = \hat{x}_i[\widehat{\eta}]^-,\ \widehat{\psi}_i^2 = \hat{x}_i[\widehat{\eta}]^+,\ \widehat{\psi}_i^3 = (1-\hat{x}_i)[\widehat{\eta}]^+,\ \widehat{\psi}_i^4 = (1-\hat{x}_i)[\widehat{\eta}]^-. \tag{62}$$

It is easy to verify that (58)-(62) satisfy (57) and thus are optimal to (55). Substituting (58)-(62) in (56), the optimality cut (56) is equivalent to (53). □

H.1.2. Subproblem and optimality cut for (26) Denote

$$\hat{\mathcal{Q}} := \begin{bmatrix} \frac{1}{2}(B_x^\top P + C^\top Y) + \frac{1}{2}(P^\top B_x + Y^\top C) - Q & \frac{1}{2}(B_x^\top p_0 + P^\top b_{x0} + C^\top y_0 + Y^\top c_0 - q) \\ \frac{1}{2}(B_x^\top p_0 + P^\top b_{x0} + C^\top y_0 + Y^\top c_0 - q)^\top & b_{x0}^\top p_0 + c_0^\top y_0 \end{bmatrix} \in \mathbb{S}^{(k+1)}.$$

The worst-case second-stage problem is approximated using the following SDP formulation:

$$\begin{aligned}
\min\ & r + t && \text{(63a)}\\
\text{s.t.}\ & re_{k+1}e_{k+1}^\top - \hat{\mathcal{Q}} = \mathcal{H}^\top U \mathcal{H},\ U \geq 0,\ U \in \mathbb{S}^l && \text{(63b)}\\
& (13e) - (13f),\ (19),\ (51b).
\end{aligned}$$

The subproblem is the dual of (63):

$$\begin{aligned}
\mathbf{SP}_{\mathrm{COP}}(\hat{x}):\ \max_{U,G,H,E,\zeta,\sigma,\eta,\mu,\chi}\ & \mathrm{Tr}(VG) + \chi^\top v_0 && \text{(64a)}\\
\text{s.t.}\ & -\mathrm{Tr}(CG) - \chi^\top c_0 + \mathrm{Tr}(B_{\hat{x}}H) - \sigma^\top b_{\hat{x}0} \geq 0 && \text{(64b)}\\
& -WH - h\sigma^\top \geq 0 && \text{(64c)}\\
& -WGA^\top - h\zeta^\top + \frac{1}{2}W(U+U^\top)B_{\hat{x}}^\top + W\eta b_{\hat{x}0}^\top \geq 0 && \text{(64d)}\\
& -\zeta - A\chi + B_{\hat{x}}\eta + b_{\hat{x}0} = 0 && \text{(64e)}\\
& -HA + \frac{1}{2}(U+U^\top)C^\top + \eta c_0^\top = 0,\ A^\top\sigma + C\eta + c_0 = 0 && \text{(64f)}\\
& -\frac{1}{2}(U+U^\top) + \gamma_2\Sigma_0 + \mu_0\mu_0^\top + \sqrt{\gamma_1}(\mu_0\mu^\top\Sigma_0^{1/2} + \Sigma_0^{1/2}\mu\mu_0^\top) \succeq 0 && \text{(64g)}\\
& \mu_0 + \sqrt{\gamma_1}\Sigma_0^{1/2}\mu - \eta = 0 && \text{(64h)}\\
& \frac{1}{2}W(U+U^\top)W^\top - h\eta^\top W^\top - W\eta h^\top + hh^\top - \frac{1}{2}(E+E^\top) = 0 && \text{(64i)}\\
& \zeta \geq 0,\ \sigma \geq 0,\ \|\mu\|_2 \leq 1,\ E \geq 0. && \text{(64j)}
\end{aligned}$$

Given an optimal solution $(U^*, G^*, H^*, E^*, \zeta^*, \sigma^*, \eta^*, \mu^*, \chi^*)$ to problem (64), if the optimal value of (64) is great than $\hat{\nu}$, we generate the following optimality cut into the relaxed master problem MP.

PROPOSITION 5. *The inequality*

$$\begin{aligned}\nu \geq \operatorname{Tr}(VG^*) + \chi^{*\top} v_0 - M\sum_{i=1}^{d}(1-\hat{x}_i)\Bigg\{ &\left[\operatorname{Tr}(B_i H^*) - \sigma^{*\top} b_i\right]^- + \frac{1}{2}\operatorname{Tr}\left[\mathbf{1}_{k\times k}\left(\left[U^* + U^{*\top}\right]^+ + \left[U^* + U^{*\top}\right]^-\right)\right] \\ &+ \mathbf{1}_m^\top\left([B_i\eta^* + b_i]^+ + [B_i\eta^* + b_i]^-\right) + \mathbf{1}_k^\top\left([\eta^*]^+ + [\eta^*]^-\right)\Bigg\} x_i - M\sum_{i=1}^{d}\hat{x}_i\Bigg\{\left[\operatorname{Tr}(B_i H^*) - \sigma^{*\top} b_i\right]^+ \\ &+ \frac{1}{2}\operatorname{Tr}\left[\mathbf{1}_{k\times k}\left(\left[U^* + U^{*\top}\right]^+ + \left[U^* + U^{*\top}\right]^-\right)\right] + \mathbf{1}_m^\top\left([B_i\eta^* + b_i]^+ + [B_i\eta^* + b_i]^-\right) + \mathbf{1}_k^\top\left([\eta^*]^+ + [\eta^*]^-\right)\Bigg\}(1-x_i)\end{aligned} \tag{65}$$

is a specific optimality cut in the form of $\nu \geq u_l^\top x + a_l$ *to the master problem to solve 0-1 SDP* (26).

The proof is similar to that of Proposition 4 and the details are omitted for brevity.

H.2. Subproblem and optimality cut for (27)

The subproblem is $\mathrm{SP}_{\mathrm{dis}}(\hat{x})$ presented in Section G of the Appendices. Given an optimal solution $(\hat{\sigma}, \hat{\mu}, \hat{\chi}, \hat{\gamma})$, the optimality cut is specified in Proposition 6.

PROPOSITION 6. *The inequality*

$$\begin{aligned}\nu \geq \sum_{s=1}^{N}\hat{\chi^s}^\top v(\xi^s) - M\sum_{i=1}^{d}\Bigg\{\hat{x}_i\left[\sum_{s=1}^{N}([\hat{\gamma^s}(B_i\xi^s + b_i)]^- + [\hat{\gamma^s}(B_i\xi^s + b_i)]^+)^\top\mathbf{1}_m + [-\sum_{s=1}^{N}\hat{\sigma^s}^\top(B_i\xi^s + b_i)]^+\right](1-x_i) \\ + (1-\hat{x}_i)\left[\sum_{s=1}^{N}([\hat{\gamma^s}(B_i\xi^s + b_i)]^+ + [\hat{\gamma^s}(B_i\xi^s + b_i)]^-)^\top\mathbf{1}_m + [-\sum_{s=1}^{N}\hat{\sigma^s}^\top(B_i\xi^s + b_i)]^-\right]x_i\Bigg\}\end{aligned} \tag{66}$$

is a specific optimality cut in the form of $\nu \geq u_l^\top x + a_l$ *to the master problem to solve 0-1 SDP* (27).

The proof is similar to that of Proposition 4 and thus is omitted for brevity.

Appendix I: Computational Performance with $(\gamma_1, \gamma_2) = (1, 1)$

Table 4 reports the 25%, 50% and 75% quantiles of the optimality gap for the three approximation approaches using linear decision rules when $(\gamma_1, \gamma_2) = (1, 1)$. Table 5 summarizes the computational performance, across the same test instances as those reported in Table 4.

Appendix J: Out-of-Sample Performance on Misspecified Distributions

Denote μ and σ^2 the mean and variance associated with the in-sample uniform distribution over $[30, 240]$. We generate two out-sample sets with $N' = 5,000$ using the following two types of misspecified distributions, respectively.

Table 4 Quantiles of optimality gaps with $(\gamma_1, \gamma_2) = (1, 1)$

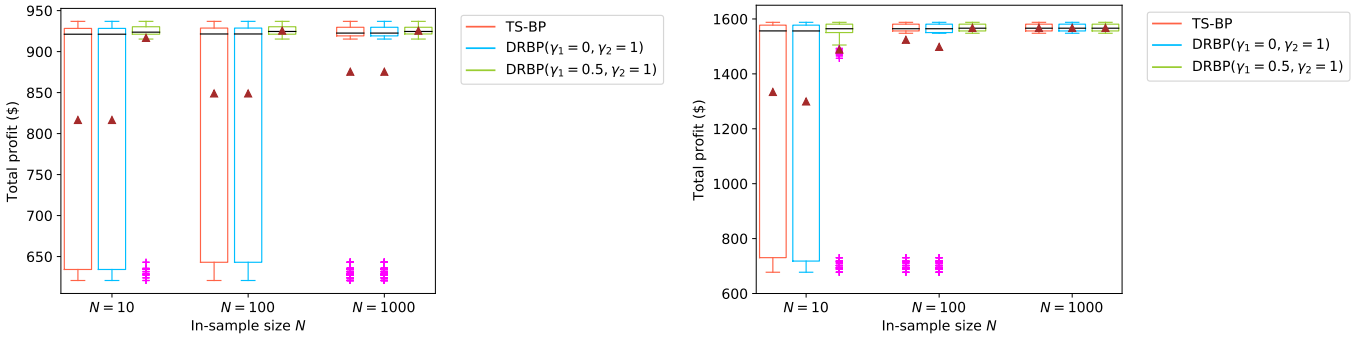
Setting	$C = V = 0$									$C, V \neq 0$								
	SDP			IA-COP			MILP-Cut			SDP			IA-COP			MILP-Cut		
	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q	25%-Q	50%-Q	75%-Q
1	24.46	28.12	31.10	24.46	28.12	31.10	24.46	28.12	31.10	27.80	32.10	36.21	41.84	42.98	46.06	27.80	32.10	36.21
2	0.00	34.48	41.15	0.00	34.48	41.15	0.00	34.48	41.15	0.00	42.43	45.24	0.00	51.31	54.10	0.00	42.43	45.24
3	29.88	32.71	38.77	29.88	32.71	38.77	29.88	32.71	38.77	36.66	38.01	43.29	48.57	49.64	52.10	36.66	38.01	43.29
4	0.00	43.60	52.54	0.00	43.60	52.54	0.00	43.60	52.54	0.00	49.68	52.96	0.00	60.08	64.57	0.00	49.68	52.89
5	11.53	12.80	15.50	11.53	12.80	15.50	11.53	12.80	15.50	14.91	17.25	19.48	26.06	27.93	30.08	14.91	16.47	18.25
6	17.69	26.46	56.50	17.69	26.36	56.50	17.69	26.36	56.50	19.88	29.05	59.46	31.23	39.79	64.81	19.45	26.53	55.59
7	13.07	13.89	16.98	13.07	13.89	16.98	13.07	13.89	16.98	16.06	18.83	21.38	28.35	30.85	32.41	16.06	18.03	19.16
8	22.24	41.57	69.07	22.24	41.57	69.07	22.24	41.57	69.07	25.89	36.96	69.43	39.38	52.23	78.39	23.76	36.94	69.43
9	27.58	43.66	75.87	27.57	43.65	75.87	27.57	43.65	75.87	27.68	42.23	75.77	44.20	53.79	83.59	27.59	42.18	75.76
10	53.61	68.75	84.44	53.61	68.72	84.44	53.61	68.72	84.44	45.94	56.06	90.66	62.18	77.44	95.75	45.84	55.91	90.64

Table 5 Computational comparison of the three approximation approaches with $(\gamma_1, \gamma_2) = (1, 1)$

Setting	$C = V = 0$											$C, V \neq 0$										
	SDP			IA-COP			MILP-Cut			Discrete		SDP			IA-COP			MILP-Cut			Discrete	
	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.	Gap (%)	t_{tot} (s)	# It.
1	**3.87**	9.0	**27.65**	4.67	8.8	**27.65**	3.99	6.5	27.66	1.36	8.8	8.29	9.0	**31.96**	10.45	8.7	43.59	**4.86**	6.6	**31.69**	1.58	8.9
2	2.12	5.0	**36.64**	2.08	4.5	**36.64**	**1.78**	5.7	**36.64**	0.75	5.0	4.10	5.0	**39.94**	**3.75**	4.5	44.93	10.48	22.0	**39.94**	0.85	5.0
3	**3.73**	9.0	**36.47**	4.65	8.9	**36.47**	9.05	12.7	**36.47**	1.39	8.9	8.19	9.0	**39.32**	10.76	8.9	51.98	**5.83**	8.7	**39.32**	1.57	8.9
4	2.09	5.0	**43.07**	2.14	4.6	**43.07**	**1.26**	4.6	**43.07**	0.76	5.0	4.13	5.0	44.99	**3.85**	4.6	50.48	11.01	22.7	**44.98**	0.85	5.0
5	9.01	9.0	13.5	8.30	9.0	**13.43**	**1.00**	2.1	**13.43**	3.37	9.0	45.97	9.0	17.53	91.07	8.9	28.54	**1.88**	2.2	**16.44**	3.71	8.9
6	5.09	5.0	38.17	**3.73**	4.7	37.91	15.62	19.8	**37.43**	1.73	4.6	**19.72**	5.0	40.63	33.24	4.7	48.14	49.99	37.0	**37.48**	2.20	4.7
7	9.11	9.0	**14.55**	8.08	9.0	**14.55**	**1.09**	2.2	**14.55**	3.30	9.0	44.10	9.0	18.91	85.55	8.9	30.52	**1.79**	2.2	**17.66**	3.63	8.9
8	5.00	5.0	47.78	**3.60**	4.7	47.78	7.79	14.4	**47.43**	1.75	4.6	19.94	5.0	48.33	31.21	4.7	57.96	**17.60**	16.6	**46.01**	2.18	4.7
9	83.78	33.0	49.95	**42.82**	33.0	49.95	46.72	11.6	49.95	22.81	33.0	387.88	33.0	48.63	**80.62**	33.0	58.98	163.48	20.1	**48.60**	26.47	33.0
10	83.79	33.0	66.45	**43.08**	33.0	**66.44**	57.82	13.2	**66.44**	22.90	33.0	390.89	33.0	60.19	**82.09**	33.0	72.27	160.40	17.1	**60.15**	27.58	33.0

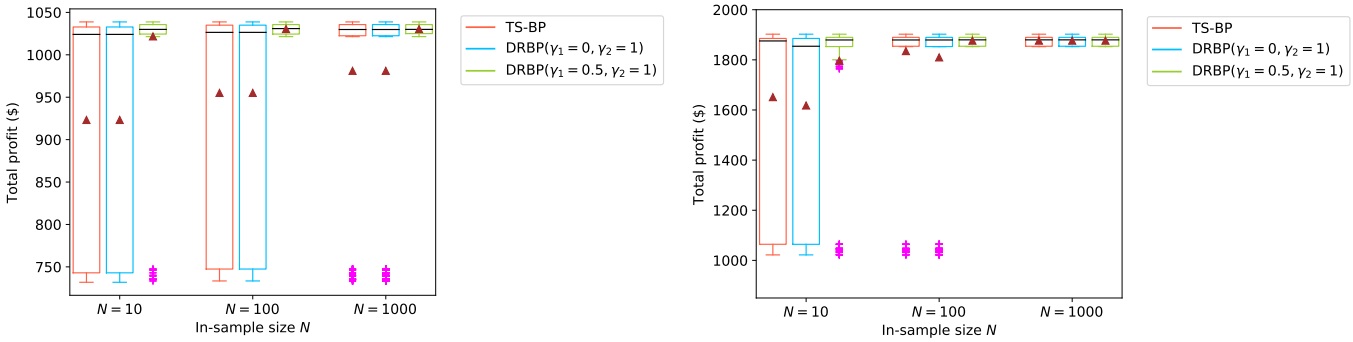
- Misspecified moment information: uniform distribution with mean $\lambda^\mu \mu = 124$ and variance $\lambda^\sigma \sigma^2 = 2940$ (i.e., uniform distribution on [30, 218]), where μ and σ^2 are the mean and the variance of the in-sample uniform distribution.
- Misspecified distribution type: truncated normal distributions[‡] $\mathcal{N}(\mu, \sigma^2)$ over the interval $[30, 240]$.

[‡]See `https://en.wikipedia.org/wiki/Truncated_normal_distribution`.



(a) $C = V = 0$ (b) $C, V \neq 0$

Figure 4 Out-of-sample performance using misspecified uniform distributions: expected profits



(a) $C = V = 0$ (b) $C, V \neq 0$

Figure 5 Out-of-sample performance using misspecified normal distributions: expected profits