Integer Programming Approaches for Distributionally Robust Chance Constraints with Adjustable Risks

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

We study distributionally robust chance constrained programs (DRCCPs) with individual chance constraints and random right-hand sides. The DRCCPs treat the risk tolerances associated with the distributionally robust chance constraints (DRCCs) as decision variables to trade off between the system cost and risk of violations by penalizing the risk tolerances in the objective function. We consider two types of Wasserstein ambiguity sets: one with finite support and one with a continuum of realizations. By exploring the hidden discrete structures, we develop mixed integer programming reformulations under the two types of ambiguity sets to determine the optimal risk tolerance for the chance constraint. Valid inequalities are derived to strengthen the formulations. We test instances with transportation problems of diverse sizes and a demand response management problem.

Key words: Distributionally robust optimization, Chance-constrained programming, Wasserstein metric, Mixed-integer programming, Reliability, Adjustable risk

History:

1. Introduction

In many planning and operational problems, chance constraints are often used for ensuring the quality of service (QoS) or system reliability. For example, chance constraints can be used to restrict the risk of under-utilizing renewable energy in power systems (e.g., Ma et al. 2019, Zhang and Dong 2022), to constrain the risk of loss in portfolio optimization (e.g., Lejeune and Shen 2016), and to impose the probability of satisfying demand in humanitarian relief networks (e.g., Elçi et al. 2018). In particular, with a predetermined risk tolerance $\alpha \in [0, 1]$, a generic chance constraint is formulated in the following form.

$$\mathbb{P}_f(T(\xi)x \geq q(\xi)) \geq 1 - \alpha,$$

where $x \in \mathbb{R}^d$ and the probability of violating the constraint $T(\xi)x \leq q(\xi)$ is no more than α with a random vector $\xi \in \mathbb{R}^l$ following distribution distribution f. The technology matrix is obtained using a function $T : \mathbb{R}^l \mapsto \mathbb{R}^{m \times d}$ and the right-hand side is a function $q : \mathbb{R}^l \mapsto \mathbb{R}^m$.

When an accurate estimate of the underlying distribution f is not accessible, *distributionally robust optimization* (DRO) provides tools to accommodate incomplete distributional information. Instead of assuming a known underlying distribution, DRO considers a prescribed set $\mathcal{D}$ of probability distributions, termed as an *ambiguity set*. The distributionally robust variant of chance constraint (1) is as follows.

$$\inf_{f \in \mathcal{D}} \mathbb{P}_f(T(\xi)x \geq q(\xi)) \geq 1 - \alpha.$$

In the distributionally robust chance constraint (DRCC), α represents the worst-case probability of violating constraints $T(\xi)x \geq q(\xi)$ with respect to the ambiguity set $\mathcal{D}$.

In many system planning and operational problems, a higher value of the probability $1-\alpha$ can lead to potentially better customer satisfaction and/or a lower probability of unfavorable events. However, a too-large $1-\alpha$ may lead to problem infeasibility and requires additional resources and operational costs (e.g., Ma et al. 2019). To find a proper balance between the cost and reliability objectives, alternatively, in this paper, we consider DRCC problems with an adjustable risk, where the risk tolerance α is treated as a variable.

1.1. Problem Formulation

In particular, with a variable risk tolerance α, we consider

$$z_0 := \min_{x \in \mathcal{X}, \alpha} \quad c^\top x + g(\alpha) \tag{1a}$$

$$\text{s.t.} \quad \inf_{f \in \mathcal{D}} \mathbb{P}_f(Tx \geq \xi) \geq 1 - \alpha \tag{1b}$$

$$\alpha \in [0, \overline{\alpha}], \tag{1c}$$

where the technology matrix $T \in \mathbb{R}^{m\times d}$ and a random right-hand side (RHS) vector $\xi \in \mathbb{R}^m$. The risk tolerance α is upper bounded by a parameter $\overline{\alpha} < 1$. The parameter $\bar{\alpha}$ is predetermined and can be viewed as the most risk of unfavorable events that the decision maker is willing to take. The objective trades off between the system cost $c^\top x$ with $c \in \mathbb{R}^d$ and the (penalty) cost of allowed violation risk $g(\alpha): [0,\bar{\alpha}] \to \mathbb{R}_0^+$. The risk cost function $g(\alpha)$ is assumed monotonically increasing in α. In this paper, we can assume a linear risk cost function $g(\alpha) = p\alpha$, which, however, is not required in the proposed models and methods. In the following, we focus on individual chance constraints, i.e., $m = 1$ with the random RHS ξ following a univariate distribution.

Another motivation for focusing on (individual) chance constraints with adjustable risks is that they can provide approximation schemes for joint chance constraints. In general, joint chance constraints are significantly harder than individual chance constraints. The Bonferroni approximation replaces a joint chance constraint with m individual chance constraints and requires the sum of individual-chance-constraint risk tolerances to be upper bounded by the risk tolerance of the joint chance constraint. Optimizing those risk tolerances of individual chance constraints potentially leads to better approximations (see, e.g., Prékopa 2003).

Particularly, in the risk-adjustable DRCC (1b), we consider a Wasserstein ambiguity set $\mathcal{D}$ constructed as follows. Given a series of N historical data samples $\{\xi^n\}_{n=1}^N$ drawn from $\mathbb{R}$, the empirical distribution is constructed as $\mathbb{P}_0(\tilde{\xi} = \xi^n) = 1/N,\ n = 1, \ldots, N$. For a positive radius $\epsilon > 0$, the Wasserstein ambiguity set defines a ball around a reference distribution (e.g., the empirical distribution) in the space of probability distributions as follows:

$$\mathcal{D} := \left\{ f : \ \mathbb{P}_f(\tilde{\xi} \in \mathbb{R}) = 1,\ W(\mathbb{P}_f, \mathbb{P}_0) \le \epsilon \right\}. \tag{2}$$

The Wasserstein distance is defined as

$$W(\mathbb{P}_f, \mathbb{P}_0) := \inf_{\mathbb{Q}\sim(\mathbb{P}_1,\mathbb{P}_2)} \mathbb{E}_{\mathbb{Q}}\left[\|\tilde{\xi}_1 - \tilde{\xi}_2\|_p\right], \tag{3}$$

where $\tilde{\xi}_1$ and $\tilde{\xi}_2$ are random variables following distribution $\mathbb{P}_1$ and $\mathbb{P}_2$, $\mathbb{Q}\sim(\mathbb{P}_1,\mathbb{P}_2)$ denotes a joint distribution of $\tilde{\xi}_1$ and $\tilde{\xi}_2$ with marginals $\mathbb{P}_1$ and $\mathbb{P}_2$, and $\|\cdot\|_p$ denotes the p-norm.

Throughout this paper, we focus on the risk-adjustable DRCC model (1) with two uncertainty types:

A1 *Finite distribution:* The random vector ξ has a finite support. The mass probability of each atom is unknown and allowed to vary.

A2 *Continuous distribution:* The random variable ξ has a continuum (infinite number) of realizations.

In the rest of the paper, without loss of generality, we assume that the samples are in a non-increasing order: $\xi^1 \geq \xi^2 \geq \cdots \geq \xi^N$. To exclude trivial special cases, throughout the rest of the paper, we assume that $\epsilon > 0$ and $\alpha \in (0,1)$.

The remainder of the paper is organized as follows. Section 1.2 reviews the prior work related to risk-adjustable chance constraints and DRCCs. Section 2 presents preliminary results regarding individual DRCC with RHS uncertainties. Section 3 utilizes hidden discrete structures to derive mixed integer programming reformulations along with valid inequalities to strengthen the mixed integer programming reformulations under the two distributional assumptions A1 and A2. Section 4 demonstrates the computational efficacy of the proposed approaches for solving a transportation problem with diverse problem sizes and a demand response management problem. Finally, we draw conclusions in Section 5.

1.2. Literature review

The idea of using variable risk tolerances can be dated back to Evers (1967), where they trade off between the cost of charge materials and the probability of not meeting specifications in metal melting furnace operations. It later has wide applications including facility sizing (Rengarajan and Morton 2009), flexible ramping capacity (Wang et al. 2018), power dispatch (Qiu et al. 2016, Ma et al. 2019), portfolio optimization (Lejeune and Shen 2016), humanitarian relief network design (Elçi et al. 2018), and inventory control problem (Rengarajan et al. 2013).

Rengarajan and Morton (2009), Rengarajan et al. (2013) perform Pareto analyses to seek an efficient frontier for a trade-off between the total investment cost and the probability of disruptions that cause undesirable events. In particular, they require to solve a series of chance-constrained programs for a large number of risk-level α choices. Unlike Rengarajan and Morton (2009), Rengarajan et al. (2013), another stream of research treats the risk tolerance α as a decision variable and develops nonparametric approaches to trade off the cost and reliability. With only right-hand side uncertainty, Shen (2014) develops a mixed integer linear programming (MILP) reformulation for individual chance constraints with only RHS uncertainty ($m = 1$ in the chance constraint (1)) under discrete distributions. Along the same line, Elçi et al. (2018) propose an alternative MILP reformulation for the same setting using knapsack inequalities. In the context of joint chance constraints ($m > 1$), Lejeune and Shen (2016) use Boolean modeling framework to develop exact reformulations for the case with RHS uncertainty and inner approximations for the case with left-hand side uncertainty. All these studies assume known underlying (discrete) probability distributions. In recent work, Zhang and Dong (2022) focus on the distributionally robust variants of the risk-adjustable chance constraints under ambiguity sets with moment constraints and Wasserstein metrics, respectively. They consider an individual DRCC with RHS uncertainty. For the moment-based ambiguity set, they develop two second-order cone programs (SOCPs) with α in different ranges; for the Wasserstein ambiguity set, they propose an exact MILP reformulation. In particular, their results for the Wasserstein ambiguity set require decision variables to be all pure binary to facilitate the linearization of bilinear terms. Without assuming pure binary decisions, in this paper, we will develop integer approaches for solving the DRCCs with adjustable risks under Wasserstein metrics.

Main Contributions: The main contributions of the paper are three-fold. First, by exploiting the (hidden) discrete structures of the individual DRCC with random RHS, we provide tractable mixed-integer reformulations for risk-adjustable DRCCs using the Wasserstein ambiguity set. Specifically, a MILP reformulation is proposed under the finite distribution assumption, and a mixed-integer second-order cone programming (MISOCP) reformulation is derived under the continuous distribution assumption. Second, we strengthen the proposed mixed-integer reformulations by deriving valid inequalities by exploring the mixing set structure of the MILP reformulation and submodularity in the MISOCP reformulation. Third, extensive numerical studies are conducted to demonstrate the computational efficacy of the proposed solution approaches.

2. Preliminary Results for DRCC with a Known Risk Tolerance α

Proposition 1 **(Adapted from Theorem 2 in Chen et al. (2022))**. *For a given risk tolerance α, the DRCC* (1b) *is equivalent to*

$$\frac{1}{N}\sum_{n=1}^{\alpha N}(Tx-\xi^n)^+ \geq \epsilon, \tag{4}$$

where $(a)^+ = \max\{a, 0\}$ *and the summation on the left-hand side is a partial sum for fractional* αN*:* $\sum_{n=1}^{\alpha N} k_n = \sum_{n=1}^{\lfloor \alpha N \rfloor} k_n + (\alpha N - \lfloor \alpha N \rfloor) k_{\lfloor \alpha N \rfloor + 1}$.

The equivalent formulation (4) has a *water-filling* interpretation as illustrated in Figure 1. The height of patch n is given by ξ^n and the width is given by $1/N$. The region with a width of α is flooded to a level t_α which uses a total amount of water equal to ϵ. Then the reformulation of DR chance constraint (4) is equivalent to a linear inequality $Tx \geq t_\alpha$. The water level t_α represents the worst-case value-at-risk (VaR):

$$t_\alpha := \inf_v \left\{ v : \inf_{f \in \mathcal{D}} \mathbb{P}(v \geq \xi) \geq 1 - \alpha \right\} = \min_v \left\{ v : \frac{1}{N} \sum_{n=1}^{\alpha N} (v - \xi^n)^+ \geq \epsilon \right\}. \quad (5)$$

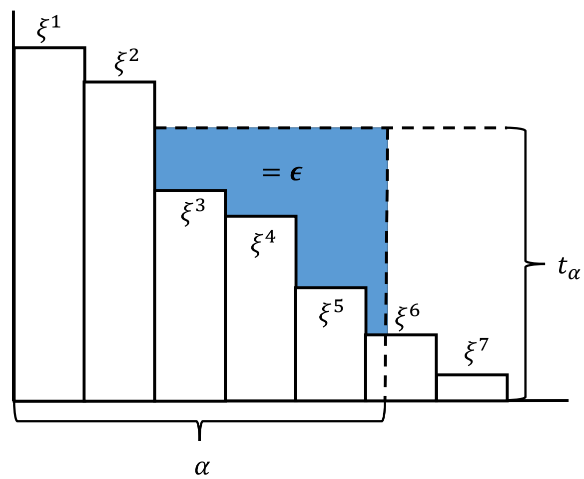

Figure 1 Illustration of the water-filling interpretation for the partial-sum inequality (4).

Let j^* be the largest index such that when the amount of water that fills the region of width α to the level ξ^{j^*} is no less than ϵ. That is,

$$j^* := \max \left\{ j \in \{1, \ldots, N\} : \ \xi^j \geq \frac{N\epsilon + \sum_{n=j+1}^{\alpha N} \xi^n}{N\alpha - j} \right\}. \quad (6)$$

For example, in Figure 1, if the water is filled up to the level as ξ^1 or ξ^2, the amount of water exceeds ϵ. In this example, $j^* = 2$. We note that such index j^* may not always exist, i.e., the problem (6) can be infeasible. This happens if the amount of water is strictly less than ϵ even when the water level reaches ξ^1. In this case, one can keep increasing the water level until the amount of water equals ϵ and let the worst-case VaR t_α equal the water level. Otherwise, the worst-case VaR can be obtained using the propositions below.

Proposition 2 **(Adapted from Theorem 2 in Ji and Lejeune (2021))**. *When Assumption A1 holds where the random vector ξ has a finite support with unknown mass probability, the worst-case VaR $t^d_\alpha = \xi^{j^*}$.*

Proposition 3 **(Adapted from Theorem 8 in Ji and Lejeune (2021)).** *When Assumption A2 holds where the random vector ξ has a continuum of realizations, the worst-case VaR*

$$t^c_\alpha = \frac{N\epsilon + \sum_{n=j^*+1}^{\alpha N} \xi^n}{N\alpha - j^*}.$$

It is easy to verify that $t^d_\alpha = \xi^{j^*} \geq t^c_\alpha$ given the definition of the critical index j^* in (6).

3. Risk-Adjustable DRCC

In this section, we develop tractable mixed-integer reformulations for the risk-adjustable DRCC (1) under the two assumptions of finite distribution A1 and continuous distribution A2. First, in Section 3.1, we provide the relation of the optimal values under the two distribution assumptions. In Section 3.2, we derive solution dominance under different allowed risk tolerance which later assists to develop tractable reformulations. Then, we develop tractable mixed-integer reformulations and valid inequalities under the finite and continuous distribution assumptions in Sections 3.3 and 3.4, respectively.

3.1. Relation of the Optimal Values under the Two Distribution Assumptions

Consider a function $t(\alpha)$ which maps the risk tolerance α to its corresponding worst-case VaR. If the function is known, the DRCC (1b) is equivalent to a linear constraint of x. Thus, the risk-adjustable DRCC problem (1) is rewritten as follows.

$$z(t(\alpha)) := \min_{x\in\mathcal{X}, \alpha\in[0,\bar{\alpha}]} \left\{ c^\top x + g(\alpha) : \ Tx \geq t(\alpha) \right\}, \tag{7}$$

where the optimal value depends on the choice of function $t(\alpha)$. Under Assumption A1 of the finite distribution, let $t^d(\alpha)$ be the worst-case VaR function and the optimal value of (7) be $z^d := z(t^d(\alpha))$. Similarly, under Assumption A2 of the continuum realizations, let $t^c(\alpha)$ be the worst-case VaR function and the optimal value of (7) be $z^c := z(t^c(\alpha))$. The next proposition presents the relation between the optimal values with finite and continuous distributions.

PROPOSITION 4. *The risk-adjustable DRCC problem* (1) *under the continuous distribution assumption A2 yields an optimal value no more than that under the finite distribution assumption A1, i.e.,* $z^d \geq z^c$.

The proposition is an immediate result from the fact that, for a given α, $t_\alpha^d \geq t_\alpha^c$.

3.2. Dominance of Risk Tolerance

The chance constrained programming literature (see, e.g., Prékopa 1990, Dentcheva et al. 2000, Ruszczyński 2002, Prékopa 2003) defines the concept of *non-dominated points*, or the so-called p-efficient points, where p refers to $1-\alpha$ in this paper.

DEFINITION 1 (p-EFFICIENT POINT). (Prékopa 2003, Dentcheva et al. 2000) Let $p \in (0,1)$. A point $v \in \mathbb{R}^m$ is a p-efficient point of the probability distribution f, $\mathbb{P}_f(v) \geq p$ and there is no $w \leq v,\ w \neq v$ such that $\mathbb{P}_f(w) \geq p$.

The concept can be extended to the DRO variant in the following definition.

DEFINITION 2 (DISTRIBUTIONALLY ROBUST p-EFFICIENT POINT). Let $p \in (0,1)$. A point $v \in \mathbb{R}^m$ is a distributionally robust p-efficient point of the ambiguity set $\mathcal{D}$, $\inf_{f\in\mathcal{D}} \mathbb{P}_f(v) \geq p$ and there is no $w \leq v,\ w \neq v$ such that $\inf_{f\in\mathcal{D}} \mathbb{P}_f(w) \geq p$.

In the case of individual chance constraint with an uncertain RHS, under the empirical distribution of $\{\xi^n\}_{n=1}^N$, the $(1-\alpha)$-efficient point is the $(1-\alpha)$-quantile (or $(1-\alpha)$-VaR) of the empirical distribution. The distributionally robust $(1-\alpha)$-efficient point coincides with the worst-case VaR t_α (obtained by assuming either the finite distribution or the continuous distribution), which is greater than the $(1-\alpha)$-quantile of the reference distribution in the Wasserstein ball $\mathcal{D}$. Similar to the $(1-\alpha)$-quantile, the worst-case VaR is nonincreasing in the risk tolerance, which is formally stated in the following proposition.

PROPOSITION 5. *Given $0 \le \alpha_1 < \alpha_2 \le \bar{\alpha}$, let t_{α_1} and t_{α_2} be the worst-case VaRs associated with α_1 and α_2, respectively. Then, $t_{\alpha_1} \ge t_{\alpha_2}$.*

The proof can be easily derived based on the water-filling interpretation in Section 2 and is omitted for brevity.

3.3. Finite Distribution

According to Propositions 2 and 3, the worst-case VaR t^d_α (if exists) under the finite distribution assumption is the smallest ξ^j which is no less than the worst-case VaR t^c_α under the continuous distribution assumption. That is, $t^d_\alpha = \min_{j\in\{1,\dots,N\}} \{\xi^j : \ \xi^j \ge t^c_\alpha\}$. We thus have the following result, which has already been anticipated in Proposition 2.

COROLLARY 1. *Under the finite distribution assumption A1 , for any risk tolerance $\alpha \in (0,1)$ such that $\xi^j \ge t^c_\alpha > \xi^{j+1}$ for some $j \in \{1,\dots,N\}$, the DRCC $\inf_{f\in\mathcal{D}} \mathbb{P}_f(Tx \ge \xi) \ge 1-\alpha$ is equivalent to a linear constraint:*

$$Tx \ge \xi^j.$$

Given any fixed $\alpha_1, \alpha_2 \in (0,1)$ such that $\xi^j \ge t^c_{\alpha_1} \ge t^c_{\alpha_2} > \xi^{j+1}$, the DRCCs of the two risk tolerances yield the same linear reformulation $Tx \ge \xi^j$ under the finite distribution assumption. Thus, in the risk-adjustable DRCC problem (1), it suffices to strengthen $\alpha \in (0, \bar{\alpha}]$ by restricting it to the risk tolerances α such that the corresponding worst-case VaR $t^d_\alpha \in (0,1)$ belongs to a discrete set:

$$t^d_\alpha \in \{\xi^1, \dots, \xi^N\}.$$

For each sample ξ^n, $n = 1,\dots,N$ in the discrete set, a risk tolerance α_n, which achieves the worst-case VaR at $t^d_{\alpha_n} = \xi^n$, can be obtained using a bisection search method as the flooded area is non-decreasing in the risk tolerance (see the water-filling interpretation in Section 1). We note that when ξ^n is too small, the corresponding risk tolerance may not exist. In this section, we assume that the corresponding risk tolerances exist for the first $N' \leq N$ largest samples $\{\xi^n\}_{n=1}^{N'}$ and denote their corresponding risk tolerances by $\alpha_n,\ n=1,\dots,N'$.

THEOREM 1. *Under the finite distribution assumption, the risk-adjustable DRCC problem* (1) *is equivalent to the following MILP formulation.*

$$z^d = \min_{x\in\mathcal{X},y} \quad c^\top x + \sum_{n=1}^{N'-1} \Delta_n y_n + \Delta_{N'} \tag{8a}$$

$$s.t. \quad Tx \geq \xi^n - M_n(1-y_n),\ n=1,\dots,N'-1 \tag{8b}$$

$$\sum_{n=1}^{N'-1} (\alpha_n - \alpha_{n+1}) y_n + \alpha_{N'} \in (0,\bar{\alpha}] \tag{8c}$$

$$y_n \in \{0,1\},\ n=1,\dots,N'-1, \tag{8d}$$

where $\Delta_n := g(\alpha_n) - g(\alpha_{n+1}) \leq 0,\ n=1,\dots,N'-1,\ \Delta_{N'} := g(\alpha_{N'})$ *and* M_n *is a big-M constant.*

Proof of Theorem 1: To see the equivalence, we need to show (1) $z^d \leq z_0$ and (2) $z^d \geq z_0$. Recall that z_0 is the optimal value of the risk-adjustable DRCC problem (1).

(1) $z^d \leq z_0$: Given an optimal solution (x_0,α_0) to the risk-adjustable DRCC problem (1), we will construct a feasible solution to MILP (8). Let $\bar{y}_n = 1$ if $Tx_0 \geq \xi^n$ and $\bar{y}_n = 0$ otherwise, for $n=1,\dots,N'$. Then the solution $(x_0, \bar{y}_n, n=1,\dots,N')$ satisfy constraints (8b) and (8d).

Let j^* be the smallest index such that $Tx_0 \geq \xi^{j^*}$. We will show that $\alpha_0 = \alpha_{j^*}$. When $j^*=1$, $t^d_{\alpha_0} = \xi^1$ and $\alpha_0 = \alpha_1$. When $j^* \geq 2$, we prove by contradiction by assuming two cases (i) $t^d_{\alpha_0} > \xi^{j^*}$ and (ii) $t^d_{\alpha_0} < \xi^{j^*}$. In the first case, $t^d_{\alpha_0} \leq \xi^{j^*-1}$. According to Proposition 5, $\alpha_0 > \alpha_{j^*-1}$. Then, (x_0, α_{j^*-1}) is feasible to the risk-adjustable DRCC (1) with a smaller objective value than z_0 as the function $g(\alpha)$ is increasing in α. In the second case, $\alpha_0 \geq \alpha_{j^*}$ due to Proposition 5 and (x_0, α_{j^*}) is a feasible solution with a smaller objective than z_0 in the risk-adjustable DRCC (1). Both cases result in a contradiction to the fact that z_0 is the optimal value of the risk-adjustable DRCC (1).

Since $\alpha_0 = \alpha_{j^*}$ and $\alpha_0 \in (0, \bar{\alpha}]$, $\alpha_{j^*} = \sum_{n=2}^{N'}(\alpha_{n-1} - \alpha_n)\bar{y}_{n-1} + \bar{y}_{N'}\alpha_{N'} \in (0, \bar{\alpha}]$ satisfies constraint (8c). Solution $(x_0, \bar{y})$ is feasible to (8) with $c^\top x_0 + g(\alpha_{j^*}) = z_0$. Thus, $z^d \le z_0$.

(2) $z^d \ge z_0$: Given an optimal solution $(\hat{x}, \hat{y})$ to problem (8), we construct a feasible solution to the risk-adjustable DRCC problem (1). Denote j the smallest index such that $T\hat{x} \ge \xi^j$, or, equivalently, the smallest index such that $\hat{y}_j = 1$. Let $\hat{\alpha} = \alpha_j$. It is easy to see that $(\hat{x}, \hat{\alpha})$ is feasible to the risk-adjustable DRCC problem (1) and its objective value is $c^\top \hat{x} + g(\alpha_j) = z^d$. So $z^d \ge z_0$.

Combining the two statements above completes the proof.

Remark 1. The big-M constant M_n is no less than $\xi^n - \xi^{N'}$.

Recall that the non-increasing order of samples: $\xi^1 \ge \xi^2 \ge \cdots \ge \xi^N$. Thus, given an optimal solution $\bar{x}$ to MILP (8), there exists a threshold index j^* such that $T\bar{x} \ge \xi^i$, for any $i \ge j^*$, and $T\bar{x} < \xi^i$, for any $i < j^*$. As the objective coefficient $\Delta_n,\ n = 1, \ldots, N'-1$ in (8a) are non-positive, in the optimal solution $(\bar{x}, \bar{y})$, we have $\bar{y}_n = 1$, for $i \ge j^*$, and $\bar{y}_n = 0$ for $i > j^*$. By exploiting this solution structure, the next proposition presents how the MILP formulation (8) can be strengthened.

Proposition 6.

i. The following inequalities are valid for the MILP (8)*:*

$$y_{n+1} \ge y_n,\ n = 1, \ldots, N'-1. \tag{9}$$

ii. The strengthened star inequality *(Luedtke et al. 2010) is valid for the MILP* (8)*:*

$$Tx \ge \xi^{N'} + \sum_{n=1}^{N'-1}(\xi^n - \xi^{n+1})y_n. \tag{10}$$

Proof of Proposition 6: The valid inequalities (9) follow from the discussion above. To see the second statement of the extended star inequalities, we introduce binary variable $z_n = 1 - y_n,\ n = 1, \ldots, N' - 1$. Without loss of generality, we assume that $\xi^n \geq 0$. Constraints (8b) and (8c) lead us to consider a *mixing set* (Atamtürk et al. 2000, Günlük and Pochet 2001, Luedtke et al. 2010):

$$P = \left\{ (t,z) \in \mathbb{R}_+ \times \{0,1\}^{N'-1} : \sum_{n=1}^{N'-1} (\alpha_n - \alpha_{n+1}) y_n + \alpha_{N'} \leq \bar{\alpha},\ t + z_n \xi^n \geq \xi^n,\ n = 1, \ldots, N' \right\} \tag{11}$$

where $t = Tx$. According to Theorem 2 in Luedtke et al. (2010), constraint (10) is facet-defining for conv(P). The proof is complete.

REMARK 2. When the distribution is known, a similar formulation for the stochastic chance-constrained problem can also be derived based on the Sample Average Approximation (Luedtke and Ahmed 2008). In this case, let α_n be the allowed risk tolerance when the VaR equals ξ^n and $\alpha_n = n/N$. The detailed MILP formulation for the stochastic chance-constrained formulation can be found in Appendix A. We note that the MILP formulation in Appendix A can be viewed as a hybrid of those in Shen (2014), Elçi et al. (2018).

3.4. Continuous Distribution

Unlike the case with finite distributions, under the continuous distribution assumption A2, the worst-case VaR cannot be restricted to a discrete set.

For a given risk tolerance α, constraint (4) is equivalent to

$$(\alpha N - j)(Tx - \xi^{k+1}) - \sum_{i=j+1}^{k} (\xi^i - \xi^{k+1}) \geq N\epsilon \tag{12}$$

where $k = \lfloor \alpha N \rfloor$ and j is the smallest index such that $Tx - \xi^{j+1} \geq 0$. For instance, in Figure 1, $j = 2$ and $k = 5$. When the risk tolerance α is not known, we introduce a binary variable $o_{jk} \in \{0,1\}$ to indicate if j and k are the two critical indices. Denote ξ^0 be an upper bound of ξ. We consider a mild assumption:

A3 For an optimal solution $\hat{x}$, $T\hat{x} \geq \xi^N$. That is, the optimal solution is restricted by the smallest realization of ξ.

THEOREM 2. *Under the continuous distribution assumption A2 and Assumption A3, the risk-adjustable DRCC problem is equivalent to the following mixed 0-1 conic formulation.*

$$
\begin{aligned}
z^c = \min_{x\in\mathcal{X},o,\alpha,u,w} \quad & c^\top x + g(\alpha) && \text{(13a)}\\
s.t. \quad & uw \geq \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} o_{jk} \sum_{i=j+1}^{k} (\xi^i - \xi^{k+1}) + N\epsilon && \text{(13b)}\\
& u \leq \alpha N - \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} j o_{jk} && \text{(13c)}\\
& w \leq Tx - \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} \xi^{k+1} o_{jk} && \text{(13d)}\\
& \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} \xi^j o_{jk} \geq Tx \geq \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} \xi^{j+1} o_{jk} && \text{(13e)}\\
& \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} (k+1) o_{jk} \geq \alpha N \geq \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} k o_{jk} && \text{(13f)}\\
& \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} o_{jk} = 1 && \text{(13g)}\\
& \alpha \in (0, \bar{\alpha}] && \text{(13h)}\\
& w \geq 0, \;\; u \geq 0 && \text{(13i)}\\
& o_{jk} \in \{0,1\}, \;\; 0 \leq j \leq k \leq N-1. && \text{(13j)}
\end{aligned}
$$

Proof of Theorem 2: To establish the equivalence, we first show that $z^c \leq z_0$ by constructing a feasible solution to problem (13) given an optimal solution to the risk-adjustable DRCC problem (1). Let (x_0, α_0) be an optimal solution to (1). Denote $k^* = \lfloor \alpha_0 N \rfloor$ and j^* as the smallest index such that $Tx_0 - \xi^{j^*+1} \geq 0$. Let $\bar{o}_{j^*k^*} = 1$, $\bar{o}_{jk} = 0$, $j \neq j^*$, $k \neq k^*$, $0 \leq j \leq k \leq N-1$, $\bar{u} = \alpha_0 N + 1 - j^*$, and $\bar{w} = Tx_0 - \xi^{k^*+1}$. It is easy to verify that $(x_0, \bar{o}, \alpha_0, \bar{u}, \bar{w})$ is a feasible solution and its objective value equals z_0. Thus, $z^c \leq z_0$.

To see the opposite direction $z^c \geq z_0$, consider an optimal solution $(\hat{x}, \hat{o}, \hat{\alpha}, \hat{u}, \hat{w})$ to problem (13). Since $\hat{o}$ is feasible, there exists $\hat{o}_{\hat{j}\hat{k}}$ such that $\hat{o}_{\hat{j}\hat{k}} = 1$ and $\hat{o}_{jk} = 0,\ j \neq \hat{j},\ k \neq \hat{k}$. Combining constraints (13b)–(13d), we obtain

$$(\hat{\alpha}N + 1 - \hat{j})(T\hat{x} - \xi^{\hat{k}+1}) - \sum_{i=\hat{j}+1}^{\hat{k}} (\xi^i - \xi^{\hat{k}+1}) \geq N\epsilon. \tag{14}$$

Constraints (13e) and (13f) are equivalent to

$$\xi^{\hat{j}} \geq T\hat{x} \geq \xi^{\hat{j}+1} \text{ and } \hat{k} + 1 \geq \alpha N \geq \hat{k}, \tag{15}$$

respectively. Constraints (14) and (15) imply that $(\hat{x}, \hat{\alpha})$ satisfies the DR chance constraint (1b). Thus, $(\hat{x}, \hat{\alpha})$ is feasible to the risk-adjustable DRCC problem (1) and $z^c \geq z_0$ as expected.

Remark 3. The mixed 0-1 conic reformulation (13) consists of $(N^2 - N)/2$ (additional) binary variables and two continuous variables. When the decision $x \in \mathcal{X} \subset \{0,1\}^d$ is restricted to binary variables, under the continuous distribution assumption, Zhang and Dong (2022) propose a MILP formulation (details are in Appendix B) by linearizing bilinear terms in the quadratic constraint (12) using McCormick inequalities (see, e.g., McCormick 1976). In addition to $(N^2 - N)/2$ binary variables as those in the conic reformulatio (13), the linearization introduces $(N^2 - N)(2d + 1)$ continuous variables, where d is the dimension of x. The MILP reformulation usually does not scale well when the problem size grows, partly due to the weaker relaxations caused by the big-M type constraints, and also due to a larger number of added variables and constraints. We will later show the computational comparison in Section 4.2.2.

In the mixed 0-1 conic reformulation (13), there is a rotated conic quadratic mixed 0-1 constraint (13b). Although the resulting mixed-integer conic reformulation can be directly solved by optimization solvers, mixed 0-1 conic programs are often time-consuming to solve, mainly due to the binary restrictions. In the following, we will develop valid inequalities for the mixed 0-1 conic reformulation (13) to help accelerate the branch-and-cut algorithm for solving (13). Specifically, we explore the *submodularity* structure of constraint (13b) as follows.

We first note that constraint (13b) can be rewritten in the following form

$$\sigma + \sum_{j=0}^{N-1} \sum_{k=j}^{N-1} d_{jk} o_{jk} \le uw, \tag{16}$$

where $\sigma = N\epsilon > 0$ and $d_{jk} = \sum_{i=j}^{k} (\xi^i - \xi^{k+1}) \ge 0,\ j = 0, \ldots, N-1,\ k = j, \ldots, N-1$. By introducing auxiliary variable $\tau \ge 0$, constraint (16) is equivalent to

$$\sqrt{\sigma + \sum_{j=0}^{N-1} \sum_{k=j}^{N-1} d_{jk} o_{jk}} \le \tau \tag{17a}$$

$$\sqrt{\tau^2 + (w-u)^2} \le w + u. \tag{17b}$$

The two inequalities (17a)–(17b) above are two second-order conic (SOC) constraints. In particular, the convex hull of the first constraint (17a) can be fully described utilizing extended polymatroid inequalities as the left-hand side of constraint (17a) is a *submodular* function (see, e.g., Atamtürk and Narayanan 2008, Atamtürk and Gómez 2020).

Definition 3 (Submodular Function). Define the collection of set $[(N^2-N)/2]$'s subsets $\mathcal{C} := \{S:\ \forall S \subset [(N^2-N)/2]\}$. Given a set function g: $\mathcal{C} \to \mathbb{R}$, g is submodular if and only if

$$g(S \cup \{j\}) - g(S) \ge g(R \cup \{j\}) - g(R),$$

for all subsets $S \subset R \subset \mathcal{C}$ and all elements $j \in \mathcal{C} \backslash R$.

We use $g(S)$ and $g(o)$ interchangeably, where $o \in \{0,1\}^{(N^2-N)/2}$ denotes the indicating vector of $S \subset \mathcal{C}$, i.e., $o_s = 1$ if $s \in S$ and $o_s = 0$ otherwise. The left-hand side of constraint (17a), $h(o) := \sqrt{\sigma + \sum_{j=0}^{N-1} \sum_{k=j}^{N-1} d_{jk} o_{jk}}$ is a submodular function, where o is a one dimensional vector consisting of $o_{jk},\ 0 \le j \le k \le N-1$.

Definition 4 (Extended Polymatroid). For a submodular function $g(S)$, the polyhedron

$$EP_g = \left\{\pi \in \mathbb{R}^{(N^2-N)/2}:\ \pi(S) \le g(S),\ \forall S \subset \mathcal{C}\right\}$$

is called an extended polymatroid associated with g, where $\pi(S) = \sum_{i \in S} \pi_i$.

For submodular function h, linear inequality

$$\pi^\top o \le z \tag{18}$$

is valid for the convex hull of the epigraph of h, i.e., $\text{conv}\{(o,z) \in \{0,1\}^{(N^2-N)/2} \times \mathbb{R}:\ z \ge h(o)\}$, if and only if π is in the extended polymatroid, i.e., $\pi \in EP_h$ (see Atamtürk and Narayanan 2008). The inequality (18) is called *extended polymatroid inequality.*

Although it suffices to only impose the extended polymatroid inequality at the extreme points of the extended polymatroid EP_h, there are an exponential number of them. Instead of adding all of them to the formulation (13), one can add them as needed in a branch-and-cut algorithm. Moreover, the separation of the valid inequality (18) can be done efficiently using a $O(n \log n)$ time greedy algorithm as follows. Given a solution $(\hat{o}, \hat{z}) \in [0,1]^{(N^2-N)/2} \times \mathbb{R}_+$, one can obtain a permutation $\{(1), \ldots, (N^2)\}$ such that the elements of o are sorted in a non-increasing order, $o_{(1)} \ge \ldots \ge o_{(N^2-N)/2}$. Let $S_{(i)} := \{(1), \ldots, (i)\},\ i = 1, \ldots, (N^2-N)/2$. Calculate $\hat{\pi}_{(1)} = h(S_{(1)})$ and $\hat{\pi}_{(i)} = h(S_{(i)}) - h(S_{(i-1)}),\ i = 2, \ldots, (N^2-N)/2$. If $\hat{\pi}^\top o \le \hat{z}$, the current solution $(\hat{o}, \hat{z})$ is optimal; otherwise, generate a valid inequality $\hat{\pi}^\top o \le z$.

4. Computational Study

In the computational study, we demonstrate the computational effectiveness of the proposed mixed integer programming formulations (with both discrete and continuous distributions) on instances of a DRCC counterpart of the transportation problem with random demand (Luedtke et al. 2010, Elçi et al. 2018). For continuous distributions, we also compute instances of demand response management using building load where the decisions are pure binary to compare the alternative MILP (which can be found in Appendix B) proposed in Zhang and Dong (2022) and our proposed mixed 0-1 conic reformulation. In Section 4.1, we describe the instance setup of the transportation problem and the demand response management problem. There are mainly two parts of results: (1) the computational performance (with CPU time, optimality gap, etc) of the different risk-adjustable DRCC models in Section 4.2, and (2) the solution details given by the models in Section 4.3. In particular, Section 4.2 demonstrates the computational efficacy of the proposed mixed integer formulations and valid inequalities. Section 4.3 shows that the risk-adjustable DRCC following the finite distribution assumption A1 provides the highest objective values compared to the risk-adjustable DRCC under the continuous distribution assumption A2 and the stochastic chance-constrained counterpart (which is presented in Appendix A).

4.1. Computational Setup

Transportation problem: There are I suppliers and D customers. The suppliers have limited capacity $M_i,\ i=1,\dots,I$. There occurs a transportation cost c_{ij} for shipping one unit from supplier i to customer j. The customer demands $\tilde{\xi}_j,\ j=1,\dots,D$ are random. Let f_j denote the distribution of $\tilde{\xi}_j$ and $\mathcal{D}_j$ be the Wasserstein ambiguity set regarding the distribution f_j. With a penalty cost p of risk tolerance α_j for every customer j, the risk-adjustable DRCC transportation problem is formulated as follows.

$$\min_{x\in\mathcal{X},\alpha}\left\{\sum_{i=1}^{I}\sum_{j=1}^{D}c_{ij}x_{ij}+p\sum_{j=1}^{D}\alpha_j:\inf_{f_j\in\mathcal{D}_j}\mathbb{P}_{f_j}(\sum_{i=1}^{I}x_{ij}\geq\tilde{\xi}_j)\geq 1-\alpha_j,\ 0\leq\alpha_j\leq\bar{\alpha},\ j=1,\dots,D\right\},\tag{19}$$

where $\mathcal{X}:=\{x\in\mathbb{R}_+^{I\times D}:\ \sum_{j=1}^{D}x_{ij}\leq M_i,\ i=1,\dots,I\}$. Following Elçi et al. (2018), the risk threshold's upper bound $\bar{\alpha}$ is set to 0.3 and p is set to 10^6. To break symmetry, a random perturbation is added to the penalty cost p following a uniform distribution on the interval $[0,100]$ for every $\alpha_j,\ j=1,\dots,D$. The radius of the Wasserstein ball $\mathcal{D}_j$ is 0.05. For other parameters (i.e., c_{ij}, M_i and samples of $\tilde{\xi}_j$), we use the data sets with $I = 40$ suppliers in Luedtke et al. (2010) with equal probabilities for all samples.

Building load control problem: There is an aggregate HVAC (i.e., heating, ventilation, and air conditioning) load of n buildings to absorb random local solar photovoltaic (PV) generation $\tilde{P}_t^{\mathrm{PV}}$ over T time periods throughout the day. Let $\mathcal{D}_t$ be the Wasserstein ambiguity regarding the distribution f of PV generation $\tilde{P}_t^{\mathrm{PV}}$ during period t. For each time period t, we solve the following risk-adjustable DRCC formulation for deciding the room temperature $x_{t,\ell}$ and HVAC ON/OFF decision $u_{t,\ell}$ of building ℓ.

$$\min_{(x,u)\in\mathcal{X}_t} \left\{ c_{\mathrm{sys}} \sum_{\ell=1}^{n} |x_{t,\ell} - x_{\mathrm{ref}}| + c_{\mathrm{switch}} \sum_{\ell=1}^{n} u_{t,\ell} + p\alpha_t : \inf_{f\in\mathcal{D}_t} \mathbb{P}_f \left(\sum_{\ell=1}^{n} P_\ell u_{t,\ell} \geq \tilde{P}_t^{\mathrm{PV}} \right) \geq 1 - \alpha_t,\ 0 \leq \alpha_t \leq \bar{\alpha} \right\}, \tag{20}$$

where $\mathcal{X}_t = \{(x_t, u_t) \in \mathbb{R}^n \times \{0,1\}^n :\ x_{t,\ell} = A_\ell x_{t-1,\ell} + B_\ell u_{t,\ell} + G_\ell v_\ell,\ x_{\min} \leq x_{t,\ell} \leq x_{\max},\ \ell = 1, \ldots, n\}$. The objective minimizes the cost of (1) the user's discomfort (indicated by the room temperature deviation from the set-point x_{ref}), (2) switching cycles, and (3) risk violation of PV tracking. The DRCC ensures that PV generation is absorbed by the HCAC fleet with probability $1 - \alpha_t$. The indoor temperature x_t and binary ON/OFF decision u_t need to satisfy the constraints of temperature comfort band and thermal dynamics in the feasible set $\mathcal{X}_t$. The parameters A_ℓ, B_ℓ, G_ℓ are obtained from the building's thermal resistances, thermal capacity, and cooling capacity. Parameter v_ℓ is a given system disturbance. We use all the parameters and data following Zhang and Dong (2022). In particular, the radius of the Wasserstein ball $\mathcal{D}_t$ is 0.02. To solve the ON/OFF decisions for a planning horizon of $T = 53$ periods throughout the day, one needs to sequentially solve 53 problems in the form of (20), one for each period.

The computations are conducted on a Windows 10 Pro machine with Intel(R) Core(TM) i7-8700 CPU3.20 GHz and 16 GB memory. All models and algorithms are implemented in Python 3.7.6 using Gurobi 10.0.1. The Gurobi default settings are used for optimizing all integer formulations except for the mixed integer conic formulation (13), for which the Gurobi parameter `MIPFocus` is set to 3. When implementing the branch-and-cut algorithm, we add the violated extended polymatroid inequalities using Gurobi `callback` class by `Model.cbLazy()` for integer solutions. For all the nodes in the branch-and-bound tree, we generate violated cuts at each node as long as any exists. The optimality gap tolerance is default as 10^{-4}. The time limit is set to 1800 seconds for computing the transportation problem instances and 100 seconds for solving the building load control problem in one period.

4.2. CPU and Optimality Gaps

Under the finite distribution assumption A1, we solve the MILP (8) with and without valid inequalities in Proposition 6. Under the continuous distribution assumption A2, the mixed 0-1 conic formulation (13) can be rewritten as a mixed 0-1 second-order cone programming (SOCP) formulation if constraint (16) is replaced by constraints (17). We solve the mixed 0-1 SOCP reformulation with and without valid the extended polymatroid inequalities. With only binary decisions, we also compare the mixed 0-1 SOCP reformulation with the alternative MILP reformulation in Appendix B. Our valid inequalities significantly reduce the solution time of directly solving the mixed integer models in Gurobi. The details are presented as follows.

4.2.1. Finite distributions We first optimize transportation problem instances with the finite distribution model using the MILP reformulation with and without strengthening techniques proposed in Proposition 6. Table 1 presents the CPU time (in seconds), "**Opt. Gap**" as the optimality gap, and "**Node**" as the total number of branching nodes. The CPU time includes the preprocessing time t_{BS} for calculating the violation risk α_n corresponding to every sample $\xi^n,\ n=1,\dots,N$ using the bisection search method, and the time t_{MILP} for solving the MILP reformulation (8) using Gurobi. In Table 1, we solve the transportation problem with $J \in \{100, 200\}$ customer demands with $N = \{50, 100, 200, 1000, 2000, 3000\}$ samples. For each (J, N) setting, five instances are solved. Table 1 presents the average CPU times, the average optimality gaps, and the average number of branching nodes. Details of each instance can be found in Appendix C.

In Table 1, with valid inequalities proposed in Proposition 6, all the instances are solved optimally at the root node within the time limit (thus optimality gap is zero and omitted). Whereas, if being solved without the valid inequalities, the instances of more samples ($N \geq 1000$) cannot be solved within the 18000-second time limit and ends with an optimality gap up to 5.47%. For larger-sized problems, solving the MILP (8) with valid inequalities is much faster than solving the MILP (8) directly due to the strength of the strengthened star inequality (10). With the valid inequalities, most of the CPU time spends on preprocessing (t_{BS}).

Table 1 **Comparison of CPU time (in seconds) and optimality gaps of finite distributions**

Demand	N	MILP + Valid Ineq.				MILP				
		t_{BS}	t_{MILP}	Time	Node	t_{BS}	t_{MILP}	Time	Opt. Gap	Node
100	50	0.00	0.04	0.04	1	0.01	0.31	0.31	N/A	1
100	100	0.02	0.07	0.09	1	0.02	1.02	1.04	N/A	158
100	200	0.07	0.09	0.15	1	0.06	6.45	6.52	N/A	2018
100	1000	1.26	0.33	1.60	1	1.21	LIMIT	LIMIT	0.09%	33454
100	2000	4.67	0.72	5.39	1	4.64	LIMIT	LIMIT	0.65%	9022
200	2000	9.18	1.60	10.79	1	9.39	LIMIT	LIMIT	2.58%	6678
200	3000	20.37	2.38	22.75	1	20.59	LIMIT	LIMIT	5.47%	6083

4.2.2. Continuous distributions We first focus on the computational performance of solving the building load control problem with the binary decision. We use the proposed mixed 0-1 SOCP formulation ("**MISOCP**") and the alternative MILP formulation

("**MILP-Binary**") from Zhang and Dong (2022). In particular, the MISOCP is obtained by replacing the rotated conic constraint (13b) with the SOC constraints (17a)-(17b). The left-hand side function $h(o)$ of (17a) is submodular and thus the extended polymatroid inequalities (18) is added in a branch-and-cut ("**B&C**") algorithm when being violated.

Table 2 reports, for each instance, the total CPU time of solving the building load control problem (20) for all 53 periods. If for any period, the problem cannot be solved within the time limit, we report "**#LIMIT**" as the number of periods which cannot be solved, and "**Avg. Gap**" as their average optimality gap. Owing to its stronger relaxations and fewer variables, the MISOCP (B&C) quickly solves all the instances, with an average of only 1.2 seconds per instance. The optimality gaps are all zeros and thus not reported in the table. In contrast, MILP-Binary fails to be solved within the 100-second time limit for each period, with an average of 17 periods not solved to optimal.

Table 2 Comparison of CPU time (in seconds) and optimality gaps of continuous distributions with binary variables

	MILP-Binary				MISOCP (B&C)		
Instance	Time	#LIMIT	Avg. Node	Avg. Gap	Time	#LIMIT	Avg. Node
1	2359.28	21	314880	0.34%	1.26	0	1
2	2124.01	18	282644	0.40%	1.32	0	11
3	2165.93	19	205147	0.59%	1.30	0	1
4	2293.98	19	286217	0.62%	1.50	0	14
5	1827.85	14	202231	0.55%	1.22	0	1
6	2206.03	18	246280	0.63%	1.20	0	1
7	2190.73	20	262080	0.62%	1.23	0	18
8	1893.50	15	223771	0.33%	1.00	0	1
9	1720.97	13	211767	0.63%	1.04	0	1
10	1899.43	16	220385	0.47%	1.41	0	60
avg.	2068.17	17	245540	0.52%	1.25	0	11

Next, we compare the branch-and-cut algorithm using the extended polymatroid inequalities (in column "**B&C**") with directly solving the MISOCP (13) (in column "**No Cuts**") on the transportation problem instances. If any instance cannot be solved within the 1800-second time limit, we report the average optimality gap and the number of unsolved instances in parentheses. In Table 3, the branch-and-cut algorithm solves the MISOCP faster than directly solving it in Gurobi.

Table 3 **Comparison of CPU time (in seconds) and optimality gaps of continuous distributions using MISOCP**

Demand	N	Instance	No Cuts			B&C		
			Time	Opt. Gap	Node	Time	Opt. Gap	Node
100	50	a	92.33	N/A	9186	7.75	N/A	1
100	50	b	66.09	N/A	16896	10.17	N/A	6
100	50	c	80.81	N/A	11245	10.49	N/A	335
100	50	d	38.65	N/A	6526	15.13	N/A	995
100	50	e	67.36	N/A	6565	8.31	N/A	1
		avg.	69.05	N/A	10084	10.37	N/A	268
100	100	a	64.15	N/A	1	40.33	N/A	1
100	100	b	86.54	N/A	46	39.79	N/A	1
100	100	c	367.28	N/A	8467	29.57	N/A	1
100	100	d	46.37	N/A	1	17.59	N/A	1
100	100	e	47.24	N/A	1	29.49	N/A	1
		avg.	122.32	N/A	1703	31.36	N/A	1
100	200	a	1455.53	N/A	1743	442.66	N/A	1246
100	200	b	LIMIT	0.12%	7211	434.15	N/A	3
100	200	c	LIMIT	0.32%	10108	LIMIT	0.18%	116520
100	200	d	LIMIT	0.08%	664	366.55	N/A	662
100	200	e	LIMIT	0.34%	28139	619.82	N/A	878
		avg.	1731.98	0.21% (4)	9573	732.73	0.18% (1)	23862

4.3. Solution Details of Models with Finite and Continuous Distributions

In this section, we focus on the solution details of the transportation problem, which are obtained by solving the risk-adjustable DRCC models (assuming finite distributions ("**Finite**") and continuous distributions ("**Continuous**")), as well as the risk-adjustable stochastic chance-constrained model ("**Stochastic**"). The detailed formulation of the stochastic chance-constrained model is available in Appendix A. In Section 4.2, we observe that the branch-and-cut algorithm does not scale as efficiently as the MILP (8) assuming finite distributions, particularly when the sample size increases. In this section, the solution details suggest that the MISOCP model assuming continuous distributions can be effectively approximated by the MILP model (8) for larger sample sizes. The details are presented below.

4.3.1. Optimal objective Values We compare the optimal objective values obtained from solving the three models: Finite, Continuous, and Stochastic. In Table 4, the relative difference (in columns "**Diff.**") is calculated as the relative gap with the Finite model. The positive relative differences of the Continuous models are as indicated by Proposition 4. All the relative differences for both Continuous and Stochastic models are positive, which indicates the conservatism of the Finite model compared to the other two models. Furthermore, the differences between the Continuous and the Finite models decrease as the sample size grows larger. For instance, with a sample size $N = 50$, the average difference between the Finite and Continuous models is 7.5%, which reduces to 1.7% when $N = 200$. This observation implies that solving the Finite model as a conservative approximation of the Continuous model becomes more suitable when the sample size is large and the MISOCP for the Continuous model is time-consuming to solve.

Table 4 Comparison of objective costs

Demand	N	Instance	Finite	Continuous		Stochastic	
			Obj.	Obj.	Diff.	Obj.	Diff.
100	50	a	37891812	34882415	7.9%	35877509	5.3%
100	50	b	39600119	36583112	7.6%	37580027	5.1%
100	50	c	40591537	37591717	7.4%	38591438	4.9%
100	50	d	39992224	36977377	7.5%	37972322	5.1%
100	50	e	41872481	38851630	7.2%	39851708	4.8%
		avg.	39989635	36977250	7.5%	37974601	5.0%
100	100	a	36300456	34797106	4.1%	35236038	2.9%
100	100	b	38370855	36931801	3.8%	37356067	2.6%
100	100	c	39353292	37849666	3.8%	38297139	2.7%
100	100	d	38786542	37271588	3.9%	37715295	2.8%
100	100	e	40741978	39244431	3.7%	39710167	2.5%
		avg.	38710625	37218919	3.9%	37662941	2.7%
100	200	a	35767904	35150000	1.7%	35202312	1.6%
100	200	b	37895369	37270088	1.7%	37318954	1.5%
100	200	c	38919823	38266132	1.7%	38322395	1.5%
100	200	d	38302041	37643158	1.7%	37668983	1.7%
100	200	e	40113429	39463734	1.6%	39516533	1.5%
		avg.	38199713	37558622	1.7%	37605836	1.6%

4.3.2. Allowed risk tolerance In this section, we look into the risk tolerance allowed by solving the three models. Recall that the transportation problem imposes a chance constraint for each demand location and with $D = 100$ customers, there are 100 allowed risk tolerances $\alpha_j,\ j = 1, \ldots, 100$. Figures 2-4 show the distributions of the risk tolerances obtained by solving the Finite, Continuous, and Stochastic models with sample size $N = \{50, 100, 200\}$. The Stochastic model assigns α's to smaller values than the two DRCC models. Additionally, as the sample size increases, there is more overlap between the distributions obtained from solving the Finite and Continuous models, which supports approximating the Continuous model with the Finite model when the sample size is large.

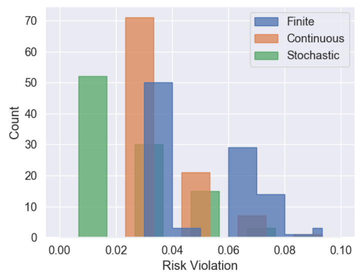

Figure 2 $N = 50$

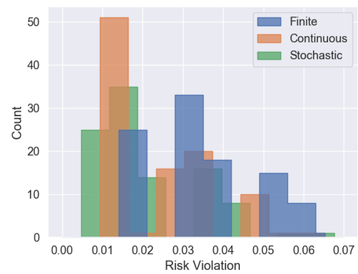

Figure 3 $N = 100$

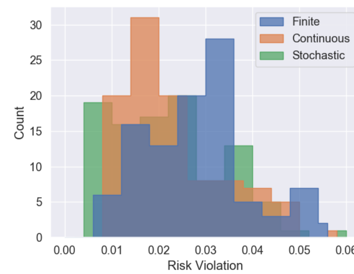

Figure 4 $N = 200$

5. Conclusions

In this paper, we investigated distributionally robust individual chance-constrained problems with a data-driven Wasserstein ambiguity set, where the uncertainty only affects the right-hand side and the risk tolerance is considered as a decision variable. The goal of the risk-adjustable DRCC is to trade-off between system costs and risk violation costs via penalizing the risk tolerance in the objective function. We considered two types of Wasserstein ambiguity sets with finite and continuous distributions. We provided a MILP reformulation of the risk-adjustable DRCC problem with finite distributions and a MISOCP reformulation for the continuous distribution case. Moreover, we derived valid inequalities for both reformulations. Via extensive numerical studies, we demonstrated that our valid inequalities accelerate solving the risk-adjustable DRCC models when compared to optimization solvers. Although the MISOCP reformulation does not scale well with larger sample size, the MILP reformulation can be used as an approximation of the MISOCP reformulation.

References

Atamtürk A, Gómez A (2020) Submodularity in conic quadratic mixed 0–1 optimization. *Operations Research* 68(2):609–630.

Atamtürk A, Narayanan V (2008) Polymatroids and mean-risk minimization in discrete optimization. *Operations Research Letters* 36(5):618–622.

Atamtürk A, Nemhauser GL, Savelsbergh MW (2000) The mixed vertex packing problem. *Mathematical Programming* 89(1):35–53.

Chen Z, Kuhn D, Wiesemann W (2022) Data-driven chance constrained programs over Wasserstein balls. *Operations Research* 0(0), URL `https://doi.org/10.1287/opre.2022.2330`.

Dentcheva D, Prékopa A, Ruszczynski A (2000) Concavity and efficient points of discrete distributions in probabilistic programming. *Mathematical Programming* 89(1):55–77.

Elçi Ö, Noyan N, Bülbül K (2018) Chance-constrained stochastic programming under variable reliability levels with an application to humanitarian relief network design. *Computers & Operations Research* 96:91–107.

Evers WH (1967) A new model for stochastic linear programming. *Management Science* 13(9):680–693.

Günlük O, Pochet Y (2001) Mixing mixed-integer inequalities. *Mathematical Programming* 90(3):429–457.

Ji R, Lejeune MA (2021) Data-driven distributionally robust chance-constrained optimization with wasserstein metric. *Journal of Global Optimization* 79(4):779–811.

Lejeune MA, Shen S (2016) Multi-objective probabilistically constrained programs with variable risk: Models for multi-portfolio financial optimization. *European Journal of Operational Research* 252(2):522–539.

Luedtke J, Ahmed S (2008) A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization* 19(2):674–699.

Luedtke J, Ahmed S, Nemhauser GL (2010) An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming* 122(2):247–272.

Ma H, Jiang R, Yan Z (2019) Distributionally robust co-optimization of power dispatch and do-not-exceed limits. *IEEE Transactions on Power Systems* 35(2):887–897.

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

Prékopa A (1990) Dual method for the solution of a one-stage stochastic programming problem with random rhs obeying a discrete probability distribution. *Zeitschrift für Operations Research* 34(6):441–461.

Prékopa A (2003) Probabilistic programming. *Handbooks in operations research and management science* 10:267–351.

Qiu F, Li Z, Wang J (2016) A data-driven approach to improve wind dispatchability. *IEEE Transactions on Power Systems* 32(1):421–429.

Rengarajan T, Dimitrov N, Morton DP (2013) Convex approximations of a probabilistic bicriteria model with disruptions. *INFORMS Journal on Computing* 25(1):147–160.

Rengarajan T, Morton DP (2009) Estimating the efficient frontier of a probabilistic bicriteria model. *Proceedings of the 2009 winter simulation conference (WSC)*, 494–504 (IEEE).

Ruszczyński A (2002) Probabilistic programming with discrete distributions and precedence constrained knapsack polyhedra. *Mathematical Programming* 93(2):195–215.

Shen S (2014) Using integer programming for balancing return and risk in problems with individual chance constraints. *Computers & operations research* 49:59–70.

Wang Z, Shen C, Liu F, Wang J, Wu X (2018) An adjustable chance-constrained approach for flexible ramping capacity allocation. *IEEE Transactions on Sustainable Energy* 9(4):1798–1811.

Zhang Y, Dong J (2022) Building load control using distributionally robust chance-constrained programs with right-hand side uncertainty and the risk-adjustable variants. *INFORMS Journal on Computing* 34(3):1531–1547, URL `https://doi.org/10.1287/ijoc.2021.1152`.

Appendix A: Stochastic Chance Constrained Problem: MILP Reformulation

Let $N' = \lceil \bar{\alpha} N \rceil$.

$$\min_{x\in\mathcal{X},y} \quad c^\top x + \sum_{n=1}^{N'-1} \Delta_n y_n + \Delta_{N'} \tag{21a}$$

$$\text{s.t.} \quad Tx \geq \xi^{N'} + \sum_{n=1}^{N'-1} (\xi^n - \xi^{n+1}) y_n \tag{21b}$$

$$y_{n+1} \geq y_n,\ n = 1,\dots,N'-2 \tag{21c}$$

$$\frac{1}{N}\left(N' - 1 - \sum_{n=1}^{N'-1} y_n\right) \in (0, \bar{\alpha}] \tag{21d}$$

$$y_n \in \{0,1\},\ n = 1,\dots,N'-1, \tag{21e}$$

where $\Delta_n := g(n/N) - g((n+1)/N),\ n = 1,\dots,N'-1$, and $\Delta_{N'} := g(N'/N)$.

Appendix B: Alternative MILP Reformulation for Risk-adjustable DRCC with Binary Variables

When all the decision variables are pure binary, i.e., $\mathcal{X} \subset \{0,1\}^d$, Zhang and Dong (2022) developed a MILP reformulation. The reformulation uses a binary variable o_{jk} to identify the critical indices j and k following the similar idea as in Section 3.4.

$$\sum_{j=0}^{N-1}\sum_{k=j}^{N-1} o_{jk}\left[(\alpha N - j)(Tx - \xi^{k+1}) - \sum_{i=j+1}^{k} (\xi^i - \xi^{k+1})\right] \geq N\epsilon \tag{22}$$

There are bilinear terms $o_{jk}x_\ell, \alpha o_{jk}$ and trilinear term $o_{jk}\alpha x_\ell$ in constraint (22). When the decisions $x_\ell,\ \ell = 1,\dots,d$ are pure binary, they can all be linearized using McCormick inequalities (McCormick 1976). The alternative MILP reformulation is as follows.

$$\min_{x\in\mathcal{X},o,\alpha,u,w} \quad c^\top x + g(\alpha) \tag{23a}$$

$$\text{s.t.} \quad \sum_{j=0}^{N-1}\sum_{k=j}^{N-1}\left[o_{jk}\sum_{i=j+1}^{k}\left(\xi^{k+1} - \xi^i\right) + NT(\delta_{jk} - j\tau_{jk}) - \xi^{k+1}(\varepsilon_{jk} - j o_{jk})\right] \geq N\epsilon \tag{23b}$$

$$\sum_{j=0}^{N-1}\sum_{k=j}^{N-1} \xi^j o_{jk} \geq Tx \geq \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} \xi^{j+1} o_{jk} \tag{23c}$$

$$\sum_{j=0}^{N-1}\sum_{k=j}^{N-1} (k+1) o_{jk} \geq \alpha N \geq \sum_{j=0}^{N-1}\sum_{k=j}^{N-1} k o_{jk} \tag{23d}$$

$$\sum_{j=0}^{N-1}\sum_{k=j}^{N-1} o_{jk} = 1 \tag{23e}$$

$$\alpha \in (0, \bar{\alpha}] \tag{23f}$$

$$\varepsilon_{jk} \leq o_{jk},\ \varepsilon_{jk} \leq \alpha,\ \varepsilon_{jk} \geq \alpha + o_{jk} - 1,\ \varepsilon_{jk} \geq 0,\ 0 \leq j \leq k \leq N-1 \tag{23g}$$

$$\delta_{\ell jk} \le \varepsilon_{jk},\ \delta_{\ell jk} \le x_\ell,\ \delta_{\ell jk} \ge \varepsilon_{jk} + x_\ell - 1,\ \delta_{\ell jk} \ge 0,\ 0 \le j \le k \le N-1,\ 1 \le \ell \le d \tag{23h}$$

$$\tau_{\ell jk} \le o_{jk},\ \tau_{\ell jk} \le x_\ell,\ \tau_{\ell jk} \ge o_{jk} + x_\ell - 1,\ \tau_{\ell jk} \ge 0,\ 0 \le j \le k \le N-1,\ 1 \le \ell \le d \tag{23i}$$

$$o_{jk} \in \{0,1\},\ 0 \le j \le k \le N-1. \tag{23j}$$

Appendix C: More results for CPU time and Optimality Gaps with Finite Distributions

Table 5 Comparison of CPU time (in seconds) and optimality gaps of finite distributions

			MILP + Valid Ineq.				MILP				
Demand	N	Instance	t_{BS}	t_{MILP}	Time	Node	t_{BS}	t_{MILP}	Time	Opt. Gap	Node
100	50	a	0.02	0.05	0.06	1	0.01	0.33	0.34	N/A	1
100	50	b	0.00	0.05	0.05	1	0.01	0.25	0.26	N/A	1
100	50	c	0.00	0.03	0.03	1	0.01	0.31	0.32	N/A	1
100	50	d	0.00	0.03	0.03	1	0.00	0.38	0.38	N/A	1
100	50	e	0.00	0.05	0.05	1	0.00	0.27	0.27	N/A	1
		avg.	0.00	0.04	0.04	1	0.01	0.31	0.31	N/A	1
100	100	a	0.02	0.06	0.08	1	0.02	0.91	0.93	N/A	85
100	100	b	0.02	0.06	0.08	1	0.02	0.84	0.87	N/A	1
100	100	c	0.02	0.08	0.09	1	0.02	1.06	1.08	N/A	223
100	100	d	0.02	0.07	0.09	1	0.03	1.37	1.40	N/A	480
100	100	e	0.02	0.08	0.10	1	0.02	0.91	0.93	N/A	1
		avg.	0.02	0.07	0.09	1	0.02	1.02	1.04	N/A	158
100	200	a	0.06	0.09	0.15	1	0.07	4.73	4.81	N/A	1420
100	200	b	0.06	0.09	0.14	1	0.06	5.41	5.47	N/A	1705
100	200	c	0.06	0.08	0.14	1	0.06	6.87	6.94	N/A	2605
100	200	d	0.08	0.08	0.16	1	0.06	8.21	8.28	N/A	2559
100	200	e	0.06	0.09	0.16	1	0.06	7.04	7.11	N/A	1800
		avg.	0.07	0.09	0.15	1	0.06	6.45	6.52	N/A	2018
100	1000	a	1.20	0.33	1.54	1	1.20	LIMIT	LIMIT	0.06%	55681
100	1000	b	1.23	0.35	1.58	1	1.19	LIMIT	LIMIT	0.05%	32960
100	1000	c	1.23	0.32	1.55	1	1.20	LIMIT	LIMIT	0.06%	25663
100	1000	d	1.47	0.31	1.78	1	1.23	LIMIT	LIMIT	0.23%	26593
100	1000	e	1.19	0.35	1.54	1	1.21	LIMIT	LIMIT	0.05%	26371
		avg.	1.26	0.33	1.60	1	1.21	LIMIT	LIMIT	0.09%	33454
100	2000	a	4.62	0.74	5.36	1	4.69	LIMIT	LIMIT	0.70%	7529
100	2000	b	4.70	0.75	5.45	1	4.70	LIMIT	LIMIT	0.59%	9567
100	2000	c	4.64	0.69	5.33	1	4.58	LIMIT	LIMIT	0.67%	8824
100	2000	d	4.72	0.70	5.42	1	4.61	LIMIT	LIMIT	0.64%	11390
100	2000	e	4.68	0.70	5.38	1	4.62	LIMIT	LIMIT	0.67%	7799
		avg.	4.67	0.72	5.39	1	4.64	LIMIT	LIMIT	0.65%	9022
200	2000	a	9.12	1.59	10.71	1	9.27	LIMIT	LIMIT	1.80%	6840
200	2000	b	9.25	1.61	10.86	1	9.34	LIMIT	LIMIT	1.86%	6835
200	2000	c	9.19	1.58	10.77	1	9.43	LIMIT	LIMIT	3.98%	6615
200	2000	d	9.26	1.67	10.93	1	9.48	LIMIT	LIMIT	3.24%	6479
200	2000	e	9.08	1.58	10.66	1	9.45	LIMIT	LIMIT	2.00%	6619
		avg.	9.18	1.60	10.79	1	9.39	LIMIT	LIMIT	2.58%	6678
200	3000	a	20.20	2.16	22.36	1	20.88	LIMIT	LIMIT	5.09%	6549
200	3000	b	20.08	2.58	22.66	1	20.22	LIMIT	LIMIT	5.50%	6510
200	3000	c	20.45	2.39	22.84	1	20.54	LIMIT	LIMIT	4.81%	6574
200	3000	d	20.73	2.36	23.09	1	20.81	LIMIT	LIMIT	6.00%	4232
200	3000	e	20.39	2.39	22.78	1	20.51	LIMIT	LIMIT	5.96%	6552
		avg.	20.37	2.38	22.75	1	20.59	LIMIT	LIMIT	5.47%	6083