

# Risk-Averse Two-Stage Stochastic Program with Distributional Ambiguity\*

Ruiwei Jiang<sup>†</sup> and Yongpei Guan<sup>‡</sup>

<sup>†</sup>Department of Systems and Industrial Engineering  
University of Arizona, Tucson, AZ 85721

<sup>‡</sup>Department of Industrial and Systems Engineering  
University of Florida, Gainesville, FL 32611

Email: ruiweijiang@email.arizona.edu; guan@ise.ufl.edu

## Abstract

In this paper, we develop a risk-averse two-stage stochastic program (RTSP) which explicitly incorporates the distributional ambiguity covering both discrete and continuous distributions. Starting from a set of historical data samples, we construct a confidence set for the ambiguous probability distribution through nonparametric statistical estimation of its density function. We then formulate RTSP from the perspective of distributional robustness by hedging against the worst-case distribution within the confidence set and considering the corresponding expected total cost. In particular, we derive an equivalent reformulation for RTSP which explicitly reflects its linkage with a full spectrum of coherent risk measures under various risk-averseness levels. Our reformulation result can also be extended to other interesting stochastic programming models with expectation, conditional value-at-risk, chance, and stochastic dominance constraints. Furthermore, we perform convergence analysis to show that the risk-averseness of RTSP vanishes as the data sample size grows to infinity, in the sense that the optimal objective value and the set of optimal solutions of RTSP converge to those of TSP. Finally, we develop a solution algorithm for the reformulation of RTSP based on the sample average approximation method, and numerical experiments on newsvendor and lot-sizing problems explain and demonstrate the effectiveness of our proposed framework.

*Key words:* stochastic optimization; data-driven decision making; distributional ambiguity

---

\*This paper was first submitted for publication on June 17th, 2014.

# 1 Introduction

Following the seminal works (see, e.g., [5] and [12]) dating back to the 1950s, significant research progress has been made on a specific class of stochastic programming models called two-stage stochastic programming (TSP) in the form of

$$(TSP) \quad \min_{x \in X} c^\top x + \mathbb{E}_{\mathbb{P}} [Q(x, \xi)], \quad (1)$$

where  $Q(x, \xi)$  represents the value function of an optimization problem for given  $x$  and realized random vector  $\xi$ , i.e.,

$$Q(x, \xi) = \min_y q(\xi)^\top y \quad (2a)$$

$$\text{s.t.} \quad W(\xi)y + T(\xi)x \leq s(\xi). \quad (2b)$$

In the above formulation,  $\xi$  is assumed a  $K$ -dimensional random vector defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\Omega \subseteq \mathbb{R}^K$  represents the sample space for  $\xi$ ,  $\mathcal{F}$  represents a  $\sigma$ -algebra on  $\Omega$ , and  $\mathbb{P}$  represents a known probability distribution. TSP divides the decision making process into two stages, where  $x$  represents a set of “here-and-now” decisions which have to be made before the random vector  $\xi$  is realized and  $y$  represents a set of “wait-and-see” decisions which can be made as a recourse after the realization of  $\xi$ . To formally describe the costs incurred and constraints to respect, we let  $c \in \mathbb{R}^{n_1}$  represent the cost vector for the first-stage problem,  $X := \{x \in \mathbb{R}^{(n_1-p)} \times \mathbb{Z}^p : Ax \leq b\}$  represent the feasible region for the first-stage decision  $x$  with  $A \in \mathbb{R}^{m_1 \times n_1}$ ,  $b \in \mathbb{R}^{m_1}$ , and  $p \in [0, n_1] \cap \mathbb{Z}$  (e.g.,  $p > 0$  implies there are integer variables in  $x$ ),  $q(\xi) \in \mathbb{R}^{n_2}$  represent the cost vector for the second-stage problem,  $W(\xi) \in \mathbb{R}^{m_2 \times n_2}$  represent the recourse matrix,  $T(\xi) \in \mathbb{R}^{m_2 \times n_1}$  represent the technology matrix, and  $s(\xi) \in \mathbb{R}^{m_2}$  represent the right-hand side vector. From the standpoint of a decision maker, TSP provides a flexible approach to assist decision making under an uncertain environment, where we make crucial decisions while at least part of the problem parameters are not known or cannot be accurately estimated. An advantage of TSP also stems from recourse decisions  $y$  which can help achieve feasibility and cost effectiveness of the stochastic programming model. Since proposed, TSP has received wide attention in terms of exploring model structures and developing solution algorithms (see, e.g., [39], [18], [22], [26], and [4]), and has been applied in a wide range of practical problems including power system operations (see, e.g. [38] and [34]), telecommunication network design (see, e.g. [33]), and supply chain planning (see, e.g., [31]), just to name a few. For a complete survey of theory and applications of TSP, readers are referred to [20], [10], [36] and references therein.

Although TSP emerges naturally as a modeling tool and is widely applied in practice, there are still challenges. For example, the traditional TSP assumes that the decision maker is risk-neutral and evaluates the uncertain cost function by its expectation. This assumption does not apply to the risk-averse decision makers that pay additional attention to those “risky” scenarios where higher-than-expectation costs are incurred, and so the solution to TSP can become suboptimal in a risk-averse circumstance. More importantly, TSP assumes that we can accurately estimate the underlying probability distribution  $\mathbb{P}$  when  $x$  is to be decided. This assumption is not necessarily practical because our knowledge of  $\mathbb{P}$  is often ambiguous, when we only have a series of data samples  $\{\xi^n\}_{n=1}^N$  which can be collected from  $\mathbb{P}$ . Based on these samples, a point estimator  $\mathbb{P}_0$  of  $\mathbb{P}$  can be biased and accordingly the solution to TSP can become suboptimal without considering the distributional ambiguity. In this paper, we propose a risk-averse TSP (RTSP) model to address these challenges by explicitly incorporating the distributional ambiguity into a classical TSP model. Starting from a set of historical data samples, we construct a confidence set  $\mathcal{D} \subseteq \mathcal{M}_+$  for the ambiguous probability distribution  $\mathbb{P}$  (both discrete and continuous) through nonparametric statistical estimation of its density function, where  $\mathcal{M}_+$  represents the space of all probability distributions on  $(\Omega, \mathcal{F})$ . We then formulate RTSP from the perspective of distributional robustness by hedging against the worst-case distribution within the confidence set and considering the corresponding expected total cost, i.e.,

$$\text{(RTSP)} \quad \min_{x \in X} c^\top x + \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [Q(x, \xi)]. \quad (3)$$

RTSP belongs to the general minimax stochastic program domain where various classes of confidence sets have been proposed and studied. One stream of research proposed moment-based confidence sets. For example, in [35], the confidence set  $\mathcal{D}$  is defined by a set of generalized moment constraints, e.g.,  $\mathcal{D} := \{\mathbb{P} \in \mathcal{M}_+ : \mathbb{E}_{\mathbb{P}}[\varphi_i(\xi)] \leq \bar{\varphi}_i, \forall i\}$  where each  $\varphi_i(\cdot)$  represents a measurable functional on  $(\Omega, \mathcal{F})$ , and each  $\bar{\varphi}_i \in \mathbb{R}$ . In [13], the confidence set  $\mathcal{D}$  is constructed based on the first two moments of  $\xi$ , e.g.,  $\mathcal{D} := \{\mathbb{P} \in \mathcal{M}_+ : \mathbb{E}_{\mathbb{P}}[\xi] \in \text{ES}(\mu_0), \mathbb{E}_{\mathbb{P}}[(\xi - \mu_0)(\xi - \mu_0)^\top] \preceq \gamma_2 \Sigma_0\}$ , where  $\mu_0 \in \mathbb{R}^K$  and  $\Sigma_0 \in \mathbb{R}^{K \times K}$  are given, and  $\text{ES}(\mu_0)$  represents an ellipsoid centered at  $\mu_0$ . Readers are referred to other moment-based confidence sets in similar forms described in [32], [41], [42], [43], [27], and [2], among others. An alternative approach to construct the confidence set  $\mathcal{D}$  is based on divergence measures or metrics on  $\mathcal{M}_+$ . For example, in [11], the confidence set  $\mathcal{D}$  is defined based on the Kullback-Leibler divergence. In [21], the confidence set  $\mathcal{D}$  is defined based on the  $\chi^2$  goodness-of-fit test. In [6] and [24],  $\mathcal{D}$  is defined based on the more general  $\phi$ -divergence. For metrics on  $\mathcal{M}_+$ , in [16],  $\mathcal{D}$  is defined based on the Prohorov metric. In [29] and [27], Kantorovich

distance is utilized to define  $\mathcal{D}$ .

In this paper, we propose to model the confidence set  $\mathcal{D}$  based on the  $L^1$ -norm on  $\mathcal{M}_+$ . This modeling choice is motivated by the following two reasons: (i) this confidence set can be conveniently developed by using historical data, and is hence data-driven (see Section 2.1), and (ii) the objective function of RTSP based on this confidence set can be reformulated as a coherent risk measure (see Section 2.2), and accordingly RTSP can be solved by the sample average approximation (SAA) method (see Section 3). From a modeling perspective, the proposed RTSP model in this paper is akin to those proposed in [6] and [24]. However, [6] and [24] focused on discrete probability distributions while the results in this paper apply to general distributions (including both discrete and continuous distributions). Also, our reformulation results discover a direct and intuitive linkage between RTSP and coherent risk measures. We summarize the main contributions of this paper as follows:

- To the best of our knowledge, we develop an initial study on reformulation and algorithm developments for two-stage stochastic programs with distributional ambiguity for general distributions including both discrete and continuous cases.
- We derive an equivalent reformulation for the RTSP model which shows an explicit linkage between the RTSP model and a full spectrum of coherent risk optimization models under various risk-averseness levels. Also, we extend the reformulation result to other interesting stochastic programming models with expectation, conditional value-at-risk, chance, and stochastic dominance constraints.
- Our convergence analysis shows that the optimal objective value and optimal solution set of the RTSP model converge to their counterparts of TSP as the historical data size grows to infinity, which indicates that the RTSP model can be applied in a data-driven decision making scheme where a decision maker can adjust her risk-averseness according to the data samples available on hand. We also develop an SAA method to solve the RTSP model, corresponding to a given historical data set. We show that the optimal objective value and solution set of the SAA problem converge to those of the RTSP model as the SAA sample size grows to infinity.
- We conduct numerical experiments to apply RTSP for the newsvendor problem and the capacitated lot-sizing problem, respectively. By comparing the risk-averse optimal solutions to the perfect-information solutions under various sample sizes and distributional settings,

we illustrate the convergence of the risk-averse optimal solutions and demonstrate that the risk-averse solutions perform reasonably well based on a moderate amount of data.

The remainder of this paper is organized as follows. We describe our main results on the equivalent reformulation of the RTSP model and the convergence analysis in Section 2. In Section 3, we analyze the SAA-based solution approach to solve the RTSP model. In Section 4, we perform computational studies for the risk-averse newsvendor and capacitated lot-sizing problems. Finally, we summarize the paper in Section 5.

## 2 Risk-Averse Two-Stage Stochastic Program

In this section, we investigate the construction of confidence set  $\mathcal{D}$ , and the reformulation and convergence properties of RTSP. In Section 2.1, we discuss how  $\mathcal{D}$  can be constructed in a data-driven manner. For the reformulation of RTSP, as pointed out in several pioneering works (see, e.g., [3], [14], and [17]), it can be shown that the objective function (3) is equivalent to

$$\min_{x \in X} c^\top x + \rho [Q(x, \xi)]$$

for a coherent risk measure  $\rho(\cdot)$  under some regularity conditions<sup>1</sup>. In Section 2.2, we specify  $\rho$  as a coherent risk measure by deriving an equivalent reformulation of (3). In Section 2.3, we establish convergence analysis for the risk-averseness of RTSP as the sample size grows to infinity. In this paper, we assume that RTSP satisfies the following assumption without loss of generality:

$$(A1) \quad \sup_{x \in X, \xi \in \Omega} |Q(x, \xi)| < \infty.$$

In practice, we can add slack variables for constraints (2b) to ensure  $\sup_{x \in X, \xi \in \Omega} Q(x, \xi) < \infty$ . Meanwhile, if  $Q(x_0, \xi_0) = -\infty$  for some first-stage decision  $x_0 \in X$  and a scenario  $\xi_0 \in \Omega$ , then we have a  $-\infty$  cost (or equivalently  $+\infty$  revenue) in the corresponding scenario, which rarely happens.

### 2.1 Confidence Set Construction

In this subsection, we discuss the construction of sample space  $\Omega$  and confidence set  $\mathcal{D}$ . Throughout the discussion, we assume a set of independent data samples  $\{\xi^n\}_{n=1}^N$  from the ambiguous distribution  $\mathbb{P}$  are available.

There are various ways to estimate sample space  $\Omega$  of the random vector  $\xi$  if it follows a continuous distribution. For example, decision makers in practice often use the three-sigma rule

---

<sup>1</sup>We refer the interested readers to the aforementioned papers and the references therein for the definition of coherent risk measures and the regularity conditions.

to estimate the range of each component of  $\xi$ , i.e.,  $\xi_k$  for  $k = 1, \dots, K$ . Let  $\bar{\mu}_k$  and  $s_k$  represent the sample mean and sample standard deviation of  $\xi_k$  respectively, i.e.,  $\bar{\mu}_k = \frac{1}{N} \sum_{n=1}^N \xi_k^n$  and  $s_k^2 = \frac{1}{N-1} \sum_{n=1}^N (\xi_k^n - \bar{\mu}_k)^2$ . Then the rule designates  $\Omega_k := [\bar{\mu}_k - 3s_k, \bar{\mu}_k + 3s_k]$  as a sample space estimate of  $\xi_k$ . It follows that  $\Omega = \prod_{k=1}^K \Omega_k$  serves as a sample space estimate of  $\xi$  where operator  $\Pi$  represents the Cartesian product. An alternative way of constructing  $\Omega$  is through the confidence region of  $\xi$ . More specifically, we let  $\bar{\mu}$  and  $\bar{\Sigma}$  represent the sample mean and covariance matrix of  $\xi$  respectively, i.e.,  $\bar{\mu} = \frac{1}{N} \sum_{n=1}^N \xi^n$  and  $\bar{\Sigma} = \frac{1}{N} \sum_{n=1}^N (\xi^n - \bar{\mu})(\xi^n - \bar{\mu})^\top$ , and assume that  $\bar{\Sigma}$  is positive definite. Then by the Markov's inequality we have

$$\mathbb{P} \left\{ (\xi - \bar{\mu})^\top \bar{\Sigma}^{-1} (\xi - \bar{\mu}) \leq \varphi \right\} \geq 1 - \frac{\mathbb{E}_{\mathbb{P}} [(\xi - \bar{\mu})^\top \bar{\Sigma}^{-1} (\xi - \bar{\mu})]}{\varphi} = 1 - \frac{\bar{\Sigma}^{-1} : \mathbb{E}_{\mathbb{P}} [(\xi - \bar{\mu})(\xi - \bar{\mu})^\top]}{\varphi},$$

where  $\varphi$  is a positive real number and operator  $:$  represents the Frobenius product. A statistical upper bound  $\gamma_2 \bar{\Sigma}$  for  $\mathbb{E}_{\mathbb{P}} [(\xi - \bar{\mu})(\xi - \bar{\mu})^\top]$  is provided in [13] (see, e.g., Theorem 2), where  $\gamma_2$  is a positive real number that can further be estimated from the data set. By using this statistical upper bound, we have

$$\mathbb{P} \left\{ (\xi - \bar{\mu})^\top \bar{\Sigma}^{-1} (\xi - \bar{\mu}) \leq \varphi \right\} \geq 1 - \frac{\gamma_2 \bar{\Sigma}^{-1} : \bar{\Sigma}}{\varphi} = 1 - \frac{\gamma_2 K}{\varphi}.$$

Therefore, if we choose  $\varphi \geq (\gamma_2 K)/\alpha$  for  $\alpha$  being, e.g., 0.05 or 0.10, then  $\Omega := \{\xi : (\xi - \bar{\mu})^\top \bar{\Sigma}^{-1} (\xi - \bar{\mu}) \leq \varphi\}$  forms an approximate  $100(1 - \alpha)\%$  confidence region, and can serve as a sample space estimate of  $\xi$ . If  $\xi$  follows a discrete distribution, the three-sigma rule and the confidence region described above can both be used to estimate the sample space, and we also need to take the discrete structure of  $\xi$  distribution into account. For example, if  $\xi$  is known to take integral values, then the intersection  $\Omega \cap \mathbb{Z}^K$  is a sample space estimate of  $\xi$  in this case.

In practice, histograms are often used to estimate or to identify the profile of  $\mathbb{P}$  if it is discrete. Suppose that  $\Omega = \{\tilde{\xi}^1, \dots, \tilde{\xi}^R\}$  represents a sample space estimate of  $\xi$  containing  $R$  masses in total. Then, we can use a histogram with frequency  $f_0^r = \sum_{n=1}^N I_n^r / N$  to estimate the probability mass function at  $\tilde{\xi}^r$  (denoted as  $f^r$ ), where  $I_n^r = 1$  if  $\xi^n = \tilde{\xi}^r$  and  $I_n^r = 0$  otherwise for each  $r = 1, \dots, R$  and  $n = 1, \dots, N$ . By the Lindeberg-Lévy central limit theorem we have  $\sqrt{N}(f^r - f_0^r)$  converges to  $\mathcal{N}(0, \sigma_r^2)$  in distribution as  $N$  goes to infinity, where  $\mathcal{N}(0, \sigma_r^2)$  represents a normal random variable with mean 0 and variance  $\sigma_r^2 = f^r(1 - f^r)$ . Hence, for large  $N$  and by using  $f_0^r(1 - f_0^r)$  to estimate  $\sigma_r^2$ , interval  $[0, (z_{\alpha/2}/\sqrt{N})\sqrt{f_0^r(1 - f_0^r)}]$  is an approximate  $1 - \alpha$  confidence interval for  $|f^r - f_0^r|$ , where  $z_{\alpha/2}$  represents the  $100(1 - \alpha/2)$ th percentile of the  $\mathcal{N}(0, 1)$  random variable. This observation motivates us to construct a confidence set  $\mathcal{D}$  around  $f_0$  based on the

$L^1$ -norm as follows:

$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{M}_+ : \sum_{r=1}^R |f^r - f_0^r| \leq d, f = d\mathbb{P}/d\xi \right\},$$

where  $d := (z_{\alpha/2}/\sqrt{N}) \sum_{r=1}^R \sqrt{f_0^r(1-f_0^r)}$ .

If  $\mathbb{P}$  is continuous, however, histograms are not usually accurate because they are not absolutely continuous with regard to the Lebesgue measure. For the case when  $\mathbb{P}$  is continuous, i.e., there exists a density function  $f : \mathbb{R}^K \rightarrow \mathbb{R}$  such that  $f = d\mathbb{P}/d\xi$ , we propose to estimate  $\mathbb{P}$  by using a counterpart of the histogram, called the kernel density estimator (KDE) defined as

$$f_0(\xi) = \frac{1}{N|H|^{1/2}} \sum_{n=1}^N P\left(H^{-1/2}(\xi - \xi^n)\right),$$

where  $f_0 : \mathbb{R}^K \rightarrow \mathbb{R}^+$  is a point estimator of  $f$  and represents the density function of  $\mathbb{P}_0$  on  $\mathbb{R}^K$  (i.e.,  $f_0 = d\mathbb{P}_0/d\xi$  with  $\mathbb{P}_0$  corresponding to a point estimator of  $\mathbb{P}$ ),  $H \in \mathbb{R}^{K \times K}$  represents a symmetric and positive definite bandwidth matrix, and  $P : \mathbb{R}^k \rightarrow \mathbb{R}^+$  represents a symmetric kernel function satisfying  $P(\cdot) \geq 0$ ,  $\int P(\xi)d\xi = 1$ , and  $\int \xi^2 P(\xi)d\xi > 0$ . One example for  $P$  is the standard multivariate normal (SMN) density, i.e.,  $H$  represents an identity matrix and  $P(\xi) = (2\pi)^{-K/2} \exp\{-(1/2)\xi^\top \xi\}$ . It is shown in [15] that  $f_0$  converges to  $f$  in  $L^1$ -norm with probability one, i.e.,

$$\int_{\mathbb{R}^K} |f(\xi) - f_0(\xi)|d\xi \rightarrow 0 \text{ as } N \rightarrow \infty.$$

This feature is desirable in a data-driven decision making environment because it indicates that the proposed point estimator  $f_0$  based on KDE eventually converges to the true density. Another advantage of using  $f_0$  is that it is easy to resample from, and so sampling algorithms based on  $f_0$  can be easily implemented. Note that the  $L^1$ -norm convergence conclusion is independent from the kernel function  $P(\xi)$ , and hence we can select  $P(\xi)$  that are more convenient to sample from than others. For example, if  $P(\xi)$  is selected to be the SMN, in order to resample from  $f_0$ , we first choose a data sample  $\bar{\xi}$  uniformly from set  $\{\xi^n\}_{n=1}^N$ , and then generate a sample from the SMN distribution whose mean value is shifted to  $\bar{\xi}$ . These observations motivate us to construct a confidence set  $\mathcal{D}$  around  $\mathbb{P}_0$  based on the  $L^1$ -norm as follows:

$$\mathcal{D} = \left\{ \mathbb{P} \in \mathcal{M}_+ : \int_{\mathbb{R}^K} |f(\xi) - f_0(\xi)|d\xi \leq d, f = d\mathbb{P}/d\xi \right\}, \quad (4)$$

where  $\mathcal{M}_+$  represents the set of all probability distributions, and  $\mathcal{D}$  incorporates all the probability distributions whose  $L^1$ -metric distance away from the point estimator  $\mathbb{P}_0$  is bounded by a distance tolerance  $d$ . In practice, the value of  $d$  can be decided in various ways. First, a decision maker can

decide the value of  $d$  to reflect her risk-averseness level, and adjust  $d$  to perform post-optimization sensitivity analysis. For example, she can decide  $d = C_1 N^{-1}$  to indicate the  $L^1$ -norm convergence, where  $C_1$  represents a positive constant and  $N$  represents the size of available data samples. Second, literature in nonparametric statistics usually measure the risk of KDE by the mean squared error  $\text{MSE} := \mathbb{E} [\int_{\mathbb{R}^K} (f(\xi) - f_0(\xi))^2 d\xi]$  where expectation  $\mathbb{E}$  is with regard to the random sample  $\{\xi^n\}_{n=1}^N$ . For example, if bandwidth matrix  $H$  is diagonal and optimally chosen then it can be shown that  $\text{MSE} = \mathcal{O}(N^{-4/(4+K)})$  (see, e.g., [40]). That is, there exists a positive constant  $C_2$  such that  $\text{MSE} \leq C_2 N^{-4/(4+K)}$  for large  $N$ . Furthermore, the  $L^1$ -norm in definition of  $\mathcal{D}$  can be bounded by the  $L^2$ -norm in definition of MSE by using the Hölder's inequality, i.e.,

$$\int_{\mathbb{R}^K} |f(\xi) - f_0(\xi)| d\xi \leq \sqrt{V(\Omega) \int_{\mathbb{R}^K} (f(\xi) - f_0(\xi))^2 d\xi}, \quad (5)$$

where  $\Omega$  represents the sample space of  $\xi$  and  $V(\Omega)$  represents the volume of  $\Omega$  in  $\mathbb{R}^K$ . Note that the bound in (5) is only meaningful when  $\Omega$  is bounded, and this condition is often satisfied in practice. For example, the construction based on the three-sigma rule and confidence regions discussed above result in bounded sample space  $\Omega$ . In this case, in view of MSE, we are motivated to choose  $d = \sqrt{C_2 V(\Omega)} N^{-2/(4+K)}$ .

In the remainder of this paper, we assume that the state space  $\Omega$  of  $\xi$  is bounded and  $f_0(\cdot)$  is defined on  $\Omega$ , i.e.,  $f_0(\xi) = 0$  for  $\xi \notin \Omega$ . Note here that the second assumption can be satisfied by considering the truncated version of  $f_0(\cdot)$  on  $\Omega$ , i.e.,

$$f_0^{\text{trc}}(\xi) := \begin{cases} f_0(\xi) / \int_{\Omega} f_0(\xi) d\xi, & \xi \in \Omega, \\ 0, & \xi \notin \Omega, \end{cases} \quad \forall \xi \in \mathbb{R}^K.$$

## 2.2 Equivalent Reformulation

In this subsection, we derive an equivalent reformulation for RTSP. We derive a reformulation for the worst-case expectation

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [g(x, \xi)]$$

for a general cost function  $g(x, \xi)$ , and then apply this general reformulation result to RTSP. For the convenience of analysis, we define the essential supremum of a general cost function

$$\text{ess sup}_{\xi \in \Omega} g(x, \xi) := \inf \{a \in \mathbb{R} : \mu(\{\xi \in \Omega : g(x, \xi) > a\}) = 0\}, \quad (6)$$



where  $\mu(\cdot)$  represents the Lebesgue measure. In other words,  $\text{ess sup}_{\xi \in \Omega} g(x, \xi)$  represents the almost surely worst-case cost  $g(x, \xi)$  over set  $\Omega$  with regard to the Lebesgue measure. We first describe the general reformulation result as follows.

**Theorem 1** *For any fixed  $x \in \mathbb{R}^{n_1}$  and a general function  $g(x, \xi)$ , we have*

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [g(x, \xi)] = (1 - d/2) \text{CVaR}_{d/2}^{\mathbb{P}_0} [g(x, \xi)] + (d/2) \text{ess sup}_{\xi \in \Omega} g(x, \xi)$$

if  $d \in [0, 2)$ , where  $\text{CVaR}_{d/2}^{\mathbb{P}_0} [g(x, \xi)]$  represents the conditional value-at-risk of  $g(x, \xi)$  with regard to  $\mathbb{P}_0$  at confidence level  $d/2$ , and

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [g(x, \xi)] = \text{ess sup}_{\xi \in \Omega} g(x, \xi)$$

if  $d \geq 2$ .

*Proof:* We begin by stating  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [g(x, \xi)]$  as the following optimization problem:

$$\sup_{f(\xi) \geq 0} \int_{\Omega} g(x, \xi) f(\xi) d\xi \tag{7a}$$

$$\text{s.t.} \int_{\Omega} |f(\xi) - f_0(\xi)| d\xi \leq d \tag{7b}$$

$$\int_{\Omega} f(\xi) d\xi = 1, \tag{7c}$$

where constraints (7b)-(7c) describe  $\mathbb{P} \in \mathcal{D}$ ,  $f$  represents the probability density function of  $\mathbb{P}$  if  $\xi$  follows a continuous distribution (i.e.,  $f = d\mathbb{P}/d\xi$ ), or the probability mass function if  $\xi$  follows a discrete distribution. The Lagrangian dual of problem (7) can be written as

$$\inf_{z \geq 0, z_0} \sup_{f(\xi) \geq 0} \int_{\Omega} g(x, \xi) f(\xi) d\xi + z_0 \left( 1 - \int_{\Omega} f(\xi) d\xi \right) - z \left( \int_{\Omega} |f(\xi) - f_0(\xi)| d\xi - d \right)$$

where  $z$  and  $z_0$  represent the dual variables for constraints (7b) and (7c) respectively. We then proceed to the reformulation as follows:

$$\begin{aligned} & \inf_{z \geq 0, z_0} \sup_{f(\xi) \geq 0} \int_{\Omega} g(x, \xi) f(\xi) d\xi + z_0 \left( 1 - \int_{\Omega} f(\xi) d\xi \right) - z \left( \int_{\Omega} |f(\xi) - f_0(\xi)| d\xi - d \right) \\ &= \inf_{z \geq 0, z_0} z_0 + zd + \sup_{f(\xi) \geq 0} \int_{\Omega} \left[ (g(x, \xi) - z_0) f(\xi) - z |f(\xi) - f_0(\xi)| \right] d\xi \\ &= \inf_{z \geq 0, z_0} z_0 + zd + \int_{\Omega} \sup_{f(\xi) \geq 0} \left[ (g(x, \xi) - z_0) f(\xi) - z |f(\xi) - f_0(\xi)| \right] d\xi \end{aligned} \quad (8a)$$

$$= \inf_{z \geq 0, z_0} z_0 + zd + \int_{\Omega \cap [g(x, \xi) - z_0 < -z]} [-z f_0(\xi)] d\xi + \int_{\Omega \cap [-z \leq g(x, \xi) - z_0 \leq z]} (g(x, \xi) - z_0) f_0(\xi) d\xi \quad (8b)$$

$$\text{s.t. } \text{ess sup}_{\xi \in \Omega} g(x, \xi) \leq z_0 + z \quad (8c)$$

$$= \inf_{z \geq 0, z_0} z_0 + zd + \int_{[g(x, \xi) - z_0 < -z]} [-z f_0(\xi)] d\xi + \int_{[-z \leq g(x, \xi) - z_0]} (g(x, \xi) - z_0) f_0(\xi) d\xi \quad (8d)$$

$$\text{s.t. } \text{ess sup}_{\xi \in \Omega} g(x, \xi) \leq z_0 + z$$

$$\begin{aligned} &= \inf_{z \geq 0, z_0} z_0 + zd + \int_{[g(x, \xi) - z_0 < -z]} [-z f_0(\xi)] d\xi + \int_{\mathbb{R}^K} (g(x, \xi) - z_0) f_0(\xi) d\xi \\ &\quad - \int_{[g(x, \xi) - z_0 < -z]} (g(x, \xi) - z_0) f_0(\xi) d\xi \end{aligned} \quad (8e)$$

$$\text{s.t. } \text{ess sup}_{\xi \in \Omega} g(x, \xi) \leq z_0 + z$$

$$= \inf_{z \geq 0, z_0} zd + \mathbb{E}_{\mathbb{P}_0} [g(x, \xi)] + \int_{[z_0 - z - g(x, \xi) > 0]} (z_0 - z - g(x, \xi)) f_0(\xi) d\xi \quad (8f)$$

$$\text{s.t. } \text{ess sup}_{\xi \in \Omega} g(x, \xi) \leq z_0 + z$$

$$= \inf_{z \geq 0} zd + \mathbb{E}_{\mathbb{P}_0} [g(x, \xi)] + \int_{[m^* - 2z - g(x, \xi) \geq 0]} (m^* - 2z - g(x, \xi)) f_0(\xi) d\xi, \quad (8g)$$

where  $m^* := \text{ess sup}_{\xi \in \Omega} g(x, \xi)$ . Equality (8b) follows by solving the one-dimensional problem

$$V^*(z_0, z) := \sup_{f(\xi) \geq 0} \left[ (g(x, \xi) - z_0) f(\xi) - z |f(\xi) - f_0(\xi)| \right]$$

for each  $\xi \in \Omega$ , which has the optimal objective value

$$V^*(z_0, z) = \begin{cases} -z f_0(\xi), & \text{if } g(x, \xi) - z_0 < -z, \\ (g(x, \xi) - z_0) f_0(\xi), & \text{if } -z \leq g(x, \xi) - z_0 \leq z, \\ +\infty, & \text{if } g(x, \xi) - z_0 > z, \end{cases}$$

and hence, without loss of optimality, we can add constraints (8c) into the dual problem based on the definition of essential supremum (6). In fact, if  $\text{ess sup}_{\xi \in \Omega} g(x, \xi) > z_0 + z$ , then  $\mu(\{\xi \in \Omega : g(x, \xi) > z_0 + z\}) > 0$  by definition (6). It follows that, by setting  $f(\xi) = +\infty$  for each

$\xi \in \{\xi \in \Omega : g(x, \xi) > z_0 + z\}$ , we have  $V^*(z_0, z) = +\infty$  for each  $\xi \in \{\xi \in \Omega : g(x, \xi) > z_0 + z\}$  and so the objective value of the optimization problem (8a) is  $+\infty$ . Thus, we can add constraints (8c) without loss of optimality. Equality (8d) is due to constraints (8c) and the fact that  $f_0(\xi) = 0$  for each  $\xi \notin \Omega$ . Equality (8e) follows from the partition  $\mathbb{R}^K = [-z \leq g(x, \xi) - z_0] \cup [g(x, \xi) - z_0 < -z]$ . Equality (8f) is due to

$$\int_{\mathbb{R}^K} (g(x, \xi) - z_0) f_0(\xi) d\xi = \int_{\mathbb{R}^K} g(x, \xi) f_0(\xi) d\xi - \int_{\mathbb{R}^K} z_0 f_0(\xi) d\xi = \mathbb{E}_{\mathbb{P}_0}[g(x, \xi)] - z_0,$$

where  $\mathbb{P}_0$  represents the probability measure induced by  $f_0(\xi)$  (i.e.,  $\mathbb{P}_0(A) = \int_A f_0(\xi) d\xi$  for each  $A \subseteq \mathbb{R}^K$ ). Equality (8g) holds because  $z_0^* = m^* - z$  is an optimal solution to the optimization problem (8f) over variable  $z_0$  for a fixed  $z$ . To see this, we consider two feasible solutions  $z_0^2 > z_0^1 \geq m^* - z$ . Then we have

$$\begin{aligned} & \int_{[z_0^2 - z - g(x, \xi) > 0]} (z_0^2 - z - g(x, \xi)) f_0(\xi) d\xi - \int_{[z_0^1 - z - g(x, \xi) > 0]} (z_0^1 - z - g(x, \xi)) f_0(\xi) d\xi \\ &= \int_{[z_0^1 - z - g(x, \xi) > 0]} (z_0^2 - z - g(x, \xi)) f_0(\xi) d\xi + \int_{[z_0^1 - z \leq g(x, \xi) < z_0^2 - z]} (z_0^2 - z - g(x, \xi)) f_0(\xi) d\xi \\ & \quad - \int_{[z_0^1 - z - g(x, \xi) > 0]} (z_0^1 - z - g(x, \xi)) f_0(\xi) d\xi \end{aligned} \quad (9)$$

$$= \int_{[z_0^1 - z - g(x, \xi) > 0]} (z_0^2 - z_0^1) f_0(\xi) d\xi + \int_{[z_0^1 - z \leq g(x, \xi) < z_0^2 - z]} (z_0^2 - z - g(x, \xi)) f_0(\xi) d\xi \geq 0, \quad (10)$$

where equality (9) is due to the partition  $[z_0^2 - z - g(x, \xi) > 0] = [z_0^1 - z - g(x, \xi) > 0] \cup [z_0^1 - z \leq g(x, \xi) < z_0^2 - z]$ , and inequality (10) is due to  $z_0^2 - z_0^1 > 0$ . Hence,  $z_0 = m^* - z$  is optimal to problem (8f), noting that  $z_0 \geq m^* - z$  based on constraint (8c).

It remains to solve the one-dimensional optimization problem (8g). Let  $h(z)$  represent its objective function and we have

$$h'(z) = d - 2 \int_{[m^* - 2z - g(x, \xi) \geq 0]} f_0(\xi) d\xi = d - 2\mathbb{P}_0\{g(x, \xi) \leq m^* - 2z\},$$

which is nondecreasing in  $z$  and hence  $h(z)$  is convex in  $z$ . We discuss the following two cases:

- (i) If  $d \geq 2$ , we have  $d/2 \geq 1$  and hence  $h'(z) = 2[d/2 - \mathbb{P}_0\{g(x, \xi) \leq m^* - 2z\}] \geq 0$ . Therefore,  $z^* = 0$  is optimal to problem (8g) whose optimal objective value is

$$\begin{aligned} h(0) &= \mathbb{E}_{\mathbb{P}_0}[g(x, \xi)] + \int_{[m^* - g(x, \xi) \geq 0]} (m^* - g(x, \xi)) f_0(\xi) d\xi \\ &= \mathbb{E}_{\mathbb{P}_0}[g(x, \xi)] + \int_{\mathbb{R}^K} (m^* - g(x, \xi)) f_0(\xi) d\xi \\ &= \mathbb{E}_{\mathbb{P}_0}[g(x, \xi)] + m^* - \mathbb{E}_{\mathbb{P}_0}[g(x, \xi)] = m^*, \end{aligned} \quad (11)$$

where equality (11) is due to  $g(x, \xi) \leq m^*$  almost everywhere in  $\Omega$  and  $f_0(\xi) = 0$  for each  $\xi \notin \Omega$ . This proves that  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [g(x, \xi)] = \text{ess sup}_{\xi \in \Omega} g(x, \xi)$  when  $d \geq 2$ .

(ii) If  $d \in [0, 2)$ , fixing  $h'(z^*) = 0$  yields

$$\mathbb{P}_0\{g(x, \xi) \leq m^* - 2z^*\} = d/2,$$

which is equivalent to

$$m^* - 2z^* = \inf \{\gamma : \mathbb{P}_0\{g(x, \xi) \leq \gamma\} \geq d/2\} = \text{VaR}, \quad (12)$$

where  $\text{VaR} := \text{VaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)]$  represents the value-at-risk of  $g(x, \xi)$  with regard to  $\mathbb{P}_0$  with confidence level  $d/2$ , for notation brevity.

It follows that

$$\begin{aligned} h(z^*) &= z^*d + \mathbb{E}_{\mathbb{P}_0}[g(x, \xi)] + \int_{[g(x, \xi) \leq \text{VaR}]} (\text{VaR} - g(x, \xi)) f_0(\xi) d\xi \\ &= (d/2) m^* - (d/2) \text{VaR} + \int_{\mathbb{R}^K} g(x, \xi) f_0(\xi) d\xi \\ &\quad + \int_{[g(x, \xi) \leq \text{VaR}]} \text{VaR} f_0(\xi) d\xi - \int_{[g(x, \xi) \leq \text{VaR}]} g(x, \xi) f_0(\xi) d\xi \end{aligned} \quad (13a)$$

$$\begin{aligned} &= (d/2) m^* + \text{VaR} (\mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\} - d/2) \\ &\quad + \int_{[g(x, \xi) > \text{VaR}]} g(x, \xi) f_0(\xi) d\xi \end{aligned} \quad (13b)$$

$$\begin{aligned} &= (d/2) m^* + (1 - d/2) \left[ \frac{\mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\} - d/2}{1 - d/2} \text{VaR} \right. \\ &\quad \left. + \frac{\mathbb{P}_0\{g(x, \xi) > \text{VaR}\}}{1 - d/2} \text{CVaR}_{d/2}^{\mathbb{P}_0+}[g(x, \xi)] \right] \end{aligned} \quad (13c)$$

$$= (d/2) m^* + (1 - d/2) \text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)], \quad (13d)$$

where equality (13a) follows from equation (12) and the definition of  $\mathbb{E}_{\mathbb{P}_0}[g(x, \xi)]$ , equality (13b) combines terms and uses the fact that  $\int_{[g(x, \xi) \leq \text{VaR}]} f_0(\xi) d\xi = \mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\}$ , equality (13c) is due to the definition of

$$\text{CVaR}_{d/2}^{\mathbb{P}_0+}[g(x, \xi)] = \int_{[g(x, \xi) > \text{VaR}]} g(x, \xi) f_0(\xi) d\xi / \mathbb{P}_0\{g(x, \xi) > \text{VaR}\}, \quad (14)$$

following equation (10) in [30], and equality (13d) is due to the definition of

$$\text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)] = \alpha_{d/2} \text{VaR} + (1 - \alpha_{d/2}) \text{CVaR}_{d/2}^{\mathbb{P}_0+}[g(x, \xi)] \quad (15)$$

with  $\alpha_{d/2} = (\mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\} - d/2)/(1 - d/2)$ , following equation (21) in [30]. This completes the proof for the case  $d \in [0, 2)$ .  $\blacksquare$

**Remark 1** *Theorem 1 specifies the coherent risk measure  $\rho(g(x, \xi))$  as a convex combination of the conditional value-at-risk and the worst-case cost. The weight of CVaR is  $1 - \min\{d/2, 1\}$  and that of the worst-case cost is  $\min\{d/2, 1\}$ .*

**Remark 2** *Theorem 1 develops a full spectrum of risk measures from the worst-case cost at the most conservative extreme to expectation at the least conservative extreme. Specifically, it indicates that (i) when the sample size is limited (and accordingly  $d$  is sufficiently large), we have to protect the worst-case scenarios of  $\xi$ , (ii) when the sample size increases and  $d$  becomes smaller, the weight of CVaR grows and that of the worst-case cost reduces, and (iii) as the sample size goes to infinity and  $d$  goes to zero, the risk measure converges to the expectation (note here that CVaR converges to expectation as  $d$  goes to zero, based on the definitions of VaR in (12), CVaR<sup>+</sup> in (14), and CVaR in (15)).*

**Remark 3** *It can be shown, by a similar proof, that Theorem 1 holds if probability distributions  $\mathbb{P}$ ,  $\mathbb{P}_0$ , and confidence set  $\mathcal{D}$  are discrete. In that case,  $\Omega$  is a set of points in  $\mathbb{R}^K$ , and  $\mathcal{D}$  and density functions  $f$ ,  $f_0$  are with respect to the counting measure on  $\mathbb{R}^K$ .*

To obtain more insights of Theorem 1, we derive the structure of the worst-case distribution of  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)]$ . In the following proposition, we develop a series of probability distributions with regard to which the value of  $\mathbb{E}[g(x, \xi)]$  converges to  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)]$ .

**Proposition 1** *For any fixed  $x \in \mathbb{R}^{n_1}$  and a general function  $g(x, \xi)$ , letting parameters  $\beta_1 = \mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\} - d/2$  and  $\beta_2 = \mathbb{P}_0\{g(x, \xi) = \text{VaR}\}$ , define*

$$(i) \quad f_1(\xi) = \begin{cases} f_0(\xi), & \text{if } g(x, \xi) > \text{VaR}, \\ 0, & \text{o.w.}, \end{cases}$$

$$(ii) \quad f_2(\xi) = \begin{cases} \beta_1 f_0(\xi) / \beta_2, & \text{if } g(x, \xi) = \text{VaR}, \\ 0, & \text{o.w.}, \end{cases} \quad \text{if } \beta_2 > 0,$$

$$(iii) \quad f_3^{(n)}(\xi) = \begin{cases} d/(2\mu(\Gamma_n)), & \text{if } \xi \in \Gamma_n, \\ 0, & \text{o.w.}, \end{cases} \quad \text{for each } n \in \mathbb{N}, \text{ and}$$

$$(iv) \quad f_4^{(n)}(\xi) = \begin{cases} 1/\mu(\Gamma_n), & \text{if } \xi \in \Gamma_n, \\ 0, & \text{o.w.}, \end{cases}$$

where  $\Gamma_n := \{\xi \in \Omega : g(x, \xi) > \text{ess sup}_{\xi \in \Omega} g(x, \xi) - 1/n\}$  for each  $n \in \mathbb{N}$ .

Then

$$\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)] = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_*^{(n)}(\xi) d\xi, \quad (16)$$

where for each  $n \in \mathbb{N}$ ,

1.  $f_*^{(n)}(\xi) = f_1(\xi) + f_3^{(n)}(\xi)$  when  $d \in [0, 2)$  and  $\beta_2 = 0$ ;
2.  $f_*^{(n)}(\xi) = f_1(\xi) + f_2(\xi) + f_3^{(n)}(\xi)$  when  $d \in [0, 2)$  and  $\beta_2 > 0$ ;
3.  $f_*^{(n)}(\xi) = f_4^{(n)}(\xi)$  when  $d \geq 2$ .

*Proof:* See Appendix A for the detailed proof. ■

**Remark 4** *Intuitively, the worst-case distribution that attains  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)]$  is a mix between a continuous distribution (e.g.,  $f_1(\xi)$ ) whose density agrees with  $f_0(\xi)$  on the event  $[g(x, \xi) \geq \text{VaR}]$  (with an exception when  $\mathbb{P}_0\{g(x, \xi) = \text{VaR}\} > 0$  which corresponds to  $f_1(\xi) + f_2(\xi)$ ), and a discrete mass (e.g., the limits of  $f_3^{(n)}(\xi)$  or  $f_4^{(n)}(\xi)$ ) where  $\text{ess sup}_{\xi \in \Omega} g(x, \xi)$  is attained.*

In many practical cases, the value of  $\text{ess sup}_{\xi \in \Omega} g(x, \xi)$  agrees with the value of  $\sup_{\xi \in \Omega} g(x, \xi)$  which can further be obtained by solving a robust optimization problem. We introduce two important sufficient conditions for the equivalence in the following lemma.

**Lemma 1** *For each  $x \in X$ ,  $\text{ess sup}_{\xi \in \Omega} g(x, \xi) = \sup_{\xi \in \Omega} g(x, \xi)$  when  $\mu(\Omega) > 0$  if one of the following two conditions holds:*

1.  $g(x, \xi)$  is continuous in variable  $\xi$  on  $\Omega$ ;
2.  $g(x, \xi)$  is piecewise continuous in variable  $\xi$  on  $\Omega$ , i.e., there exist disjoint subsets  $\{\Omega_i\}_{i \in \Psi}$  with index set  $\Psi$  finite or countable such that  $\Omega = \bigcup_{i \in \Psi} \Omega_i$  with  $\mu(\Omega_i) > 0$  and  $g(x, \xi)$  continuous in variable  $\xi$  on  $\Omega_i$  for each  $i \in \Psi$ .

*Proof:* See Appendix B for the detailed proof. ■

Now we apply Theorem 1 and Lemma 1 to RTSP which leads to the following proposition.

**Proposition 2** *If  $d \in [0, 2)$ , then RTSP is equivalent to the following optimization problem*

$$\min_{x \in X, \zeta \in \mathbb{R}} c^\top x + (1 - d/2)\zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi + (d/2) \sup_{\xi \in \Omega} Q(x, \xi). \quad (17a)$$

*Otherwise, if  $d \geq 2$ , then RTSP is equivalent to*

$$\min_{x \in X} c^\top x + \sup_{\xi \in \Omega} Q(x, \xi). \quad (17b)$$

*Proof:* See Appendix C for the detailed proof. ■

In addition, we can extend Theorem 1 to other interesting stochastic programming models with expectation, CVaR, chance, and stochastic dominance constraints, which can be summarized as follows:

**Proposition 3 (Expectation constraint)** *For a general function  $F(x, \xi)$  that is piecewise continuous in variable  $\xi$  on  $\Omega$ , the ambiguous expectation constraint  $\sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[F(x, \xi)] \leq 0$  is equivalent to*

$$(1 - d/2)\zeta + \int_{\mathbb{R}^K} (F(x, \xi) - \zeta)^+ f_0(\xi) d\xi + (d/2) \sup_{\xi \in \Omega} F(x, \xi) \leq 0$$

if  $d \in [0, 2)$ , and

$$\sup_{\xi \in \Omega} F(x, \xi) \leq 0$$

if  $d \geq 2$ .

*Proof:* The proof is similar to that for Proposition 2 and thus omitted. ■

**Proposition 4 (CVaR constraint)** *For a general function  $F(x, \xi)$  that satisfies Assumption (A1) and is piecewise continuous in variable  $\xi$  on  $\Omega$ , the ambiguous conditional value-at-risk constraint  $\sup_{\mathbb{P} \in \mathcal{D}} \text{CVaR}_{\alpha}^{\mathbb{P}}[F(x, \xi)] \leq r_0$  with  $\alpha \in (0, 1)$  and  $r_0 \in \mathbb{R}$  is equivalent to*

$$(1 - \alpha)\zeta + (1 - d/2)\zeta_0 + \int_{\mathbb{R}^K} [(F(x, \xi) - \zeta)^+ - \zeta_0]^+ f_0(\xi) d\xi + (d/2) \left( \sup_{\xi \in \Omega} F(x, \xi) - \zeta \right)^+ \leq (1 - \alpha)r_0$$

if  $d \in [0, 2)$ , and

$$\sup_{\xi \in \Omega} F(x, \xi) \leq r_0$$

if  $d \geq 2$ .

*Proof:* See Appendix D for the detailed proof. ■

**Proposition 5 (Chance constraint)** *For a given  $\alpha \in (0, 1)$ , the ambiguous chance constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{F(x, \xi) \leq 0\} \geq 1 - \alpha$  for a general function  $F(x, \xi)$ , is equivalent to*

$$\mathbb{P}_0 \{F(x, \xi) \leq 0\} \geq 1 - \alpha + d/2$$

if  $d \in [0, 2\alpha]$ , and

$$\text{ess sup}_{\xi \in \Omega} F(x, \xi) \leq 0$$

if  $d > 2\alpha$ . Furthermore, if  $F(x, \xi)$  is piecewise continuous in variable  $\xi$  on  $\Omega$ , then the ambiguous chance constraint is equivalent to  $\sup_{\xi \in \Omega} F(x, \xi) \leq 0$  if  $d > 2\alpha$ .

*Proof:* See Appendix E for the detailed proof. ■

**Proposition 6 (Stochastic dominance constraint)** For a given  $X_F(\xi) := \{x : F(x, \xi) \leq 0\}$  for a general function  $F(x, \xi)$ , the ambiguous stochastic dominance constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{F(x, \xi) \leq \eta\} \geq F_Y(\eta)$  for each  $\eta \in [c_1, c_2]$ , where  $F_Y(\cdot)$  represents the distribution function of random variable  $Y$  and  $c_1, c_2 \in \mathbb{R}$  such that  $0 < F_Y(c_1) \leq F_Y(c_2) < 1$ , is equivalent to

$$\mathbb{P}_0 \{F(x, \xi) \leq \eta\} \geq F_Y(\eta) + d/2, \quad \forall \eta \in [c_1, c_2]$$

if  $d \in [0, 2(1 - F_Y(c_2))]$ , and

$$\begin{cases} \mathbb{P}_0 \{F(x, \xi) \leq \eta\} \geq F_Y(\eta) + d/2, & \forall \eta \in [c_1, F_Y^{-1}(1 - d/2)], \\ \text{ess sup}_{\xi \in \Omega} F(x, \xi) \leq F_Y^{-1}(1 - d/2), \end{cases}$$

if  $d > 2(1 - F_Y(c_2))$ , where  $F_Y^{-1}(x) := \sup_{\eta \in \mathbb{R}} \{F_Y(\eta) \leq x\}$  for  $x \in [0, 1]$ .

*Proof:* See Appendix F for the detailed proof. ■

Finally, we explore structural results for a risk-averse newsvendor problem with distributional ambiguity, i.e., a variant of the classical newsvendor problem in a distributionally ambiguous environment, to illustrate the usage of Theorem 1. As in the classical setting, a newsvendor orders  $x$  newspapers to satisfy an uncertain future demand  $\xi$  that is not realized at the moment of making the order. A newspaper can be ordered at a cost of  $c$  dollars, sold at a price of  $p$  dollars, and salvaged at a price of  $s$  dollars if it is not sold. To avoid trivial cases, we assume that  $0 \leq s \leq c \leq p$ . Supposing that the uncertain demand  $\xi$  follows a probability distribution  $\mathbb{P}$ , a risk-neutral newsvendor aims to minimize the expected total cost, i.e.,

$$\min_{x \in \mathbb{R}_+} cx + \mathbb{E}_{\mathbb{P}} [Q(x, \xi)],$$

where  $Q(x, \xi) := -p \min\{x, \xi\} - s \max\{x - \xi, 0\}$ . We consider a distributionally ambiguous variant, where the newsvendor's knowledge on  $\mathbb{P}$  is imperfect and only includes the range of  $\xi$ , e.g.,  $\Omega :=$



$[\xi^L, \xi^U]$  for  $\xi^L, \xi^U \in \mathbb{R}_+$ , a point estimator  $\mathbb{P}_0$  with density function  $f_0(\cdot)$ , and a confidence set  $\mathcal{D}$  as defined in (4). Accordingly, the risk-averse newsvendor can hedge against the distributional ambiguity by considering

$$\min_{x \in \mathbb{R}_+} F^{nv}(x) := cx + \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[Q(x, \xi)].$$

Based on Theorem 1, we can solve this optimization problem and obtain the following results.

**Proposition 7 (Newsvendor problem with distributional ambiguity)** *The objective of the risk-averse newsvendor problem*

$$F^{nv}(x) = cx + (1 - d/2)^+ CVaR_{d/2}^{\mathbb{P}_0}[Q(x, \xi)] + \min\{d/2, 1\} \sup_{\xi \in \Omega} Q(x, \xi).$$

Moreover, if  $\mathbb{P}_0$  is absolutely continuous, then an optimal solution  $x^* \in \operatorname{argmin}_{x \in \mathbb{R}_+} F^{nv}(x)$  is such that

$$x^* = F_0^{-1} \left( \frac{p-c}{p-s} - \frac{d}{2} \right)$$

if  $(p-c)/(p-s) \geq d/2$ , and  $x^* = \xi^L$  if  $(p-c)/(p-s) < d/2$ , where  $F_0^{-1}(x) = \sup_{\eta \in \mathbb{R}} \{\mathbb{P}_0\{\xi \leq \eta\} \leq x\}$ .

*Proof:* See Appendix G for the detailed proof. ■

The risk-averse optimal solution presented in Proposition 7 illustrates an interesting linkage between the distance tolerance  $d$  and the conservatism of decision making. We note that a risk-neutral optimal solution is  $F_0^{-1}((p-c)/(p-s))$ . As the value of  $d$  decreases, the risk-averse optimal solution increases from the most conservative extreme (i.e.,  $\xi^L$ ) and becomes less and less conservative. Eventually, the risk-averse optimal solution converges to the risk-neutral one (i.e.,  $F_0^{-1}((p-c)/(p-s))$ ) as  $d$  goes to zero. In the following Section 2.3, we establish a convergence analysis to show this linkage for the general RTSP in a rigorous manner. This linkage will also be demonstrated by numerical experiments in Section 4.1.

### 2.3 Convergence Analysis

In this subsection, we analyze the convergence property of RTSP as the sample size  $N$  grows to infinity (and accordingly  $d$  decreases to zero). We find that as  $N$  goes to infinity, the optimal objective value and the set of optimal solutions of RTSP converge to those of TSP with regard to the ambiguous distribution  $\mathbb{P}$ , respectively, under mild conditions. This result confirms an intuition that the risk-averseness of RTSP vanishes as the sample size grows to infinity. Before stating the main result, we describe the following lemmas to be used later to prove the convergence property.

**Lemma 2** *Expected value functions  $\mathbb{E}[Q(x, \xi) - \zeta]^+$  with regard to  $\mathbb{P}$  and  $\mathbb{P}_0$  are continuous in variables  $(x, \zeta) \in \mathbb{R}^{n_1+1}$ .*

*Proof:* See Appendix H for the detailed proof. ■

**Lemma 3** *Worst-case value function  $\sup_{\xi \in \Omega} Q(x, \xi)$  is convex and continuous in variable  $x \in \mathbb{R}^{n_1}$ .*

*Proof:* See Appendix I for the detailed proof. ■

**Lemma 4** *TSP with regard to  $\mathbb{P}_0$  is equivalent to the following optimization problem:*

$$\min_{x \in X, \zeta \in \mathbb{R}} c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi.$$

*Meanwhile, TSP with regard to  $\mathbb{P}$  is equivalent to the following optimization problem:*

$$\min_{x \in X, \zeta \in \mathbb{R}} c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f(\xi) d\xi.$$

*Proof:* See Appendix J for the detailed proof. ■

Now we are ready to state our main convergence result. We let  $z(d)$  and  $U(d)$  represent the optimal objective value and the set of optimal solutions of RTSP with regard to a distance tolerance  $d \geq 0$ , respectively, where  $d := d(N)$  is a function of sample size  $N$ . Accordingly, we let  $z(0)$  and  $U(0)$  represent the optimal objective value and the set of optimal solutions of TSP with regard to  $\mathbb{P}$ , respectively. The main result can be stated in the following theorem.

**Theorem 2** *Suppose that (i)  $X$  is compact, and (ii)  $f_0$  converges to  $f$  in  $L^1$ -norm with probability one, i.e.,*

$$\int_{\mathbb{R}^K} |f(\xi) - f_0(\xi)| d\xi \rightarrow 0 \text{ as } N \rightarrow \infty.$$

*Then  $z(d)$  converges to  $z(0)$  as  $N$  increases to infinity with probability one, i.e.,*

$$\lim_{N \rightarrow \infty} z(d(N)) = z(0).$$

*Furthermore,  $U(d)$  converges to  $U(0)$  as  $N$  increases to infinity with probability one in the sense that  $\lim_{N \rightarrow \infty} \sup_{(x, \zeta) \in U(d(N))} \text{dist}((x, \zeta), U(0)) = 0$ .*

*Proof:* See Appendix K for the detailed proof. ■

**Remark 5** Assumption (i) for  $X$  being compact is satisfied when (a) first-stage decision variables  $x$  are finite and integral, and (b) first-stage decision variables  $x$  are pure-continuous or mixed-integer, and are bounded (e.g., in a polytope). In general, this condition is mild because it is mostly satisfied in practice. Meanwhile, Assumption (ii) for  $L^1$  convergence is satisfied if  $f_0$  is obtained from a KDE as discussed in Section 2.1.

### 3 Solution Approaches

In this section, we investigate a sample average approximation (SAA) approach to solve RTSP (17) for a given sample size  $N$  (or equivalently for a given distance  $d$ ). When  $d \geq 2$ , RTSP (17b) reduces to a two-stage (adjustable) robust optimization problem, for which various solution approaches have been proposed in literature (see, e.g., [7], [23], and [8]). In this section, we focus on the case in which  $d \in [0, 2)$  and RTSP (17a) is considered. Throughout this section, we make the following assumption on the sample space  $\Omega$ :

$$(A2) \quad \Omega \text{ is a knapsack polytope, i.e., there exist real numbers } \{\xi_k^L\}_{k=0}^K \text{ and } \{\xi_k^U\}_{k=0}^K \text{ such that } \xi_k^L < \xi_k^U \text{ for } k = 0, \dots, K \text{ and } \Omega = \left\{ \xi \in \mathbb{R}^K : \xi_0^L \leq \sum_{k=1}^K \xi_k \leq \xi_0^U, \xi_k^L \leq \xi_k \leq \xi_k^U, \forall k = 1, \dots, K \right\}.$$

Note here that the sample space  $\Omega$  in the form stated in Assumption (A2) has its advantages. First, it can be conveniently constructed based on historical data (see Section 2.1). It can also conservatively approximate other sample spaces. For instance, sample spaces in the form  $\Omega' := \{\xi \in \mathbb{R}^K : (\xi - \bar{\mu})^\top \bar{\Sigma}^{-1} (\xi - \bar{\mu}) \leq \varphi\}$  can be conservatively approximated by  $\Omega$  if

$$\xi_k^L \leq \inf_{\xi \in \Omega'} \xi_k, \quad \xi_k^U \geq \sup_{\xi \in \Omega'} \xi_k, \quad \xi_0^L \leq \inf_{\xi \in \Omega'} \sum_{k=1}^K \xi_k, \quad \text{and} \quad \xi_0^U \geq \sup_{\xi \in \Omega'} \sum_{k=1}^K \xi_k.$$

Furthermore, Assumption (A2) allows us to reformulate  $\sup_{\xi \in \Omega} Q(x, \xi)$  in RTSP (17a) as a mixed-integer linear program (see below Proposition 8). Finally, we note that the sample space plays a less important role as the data sample size  $N$  increases, corresponding to  $d \rightarrow 0$ , in view of Theorem 2. In Section 3.3, we will extend the discussion to the case when Assumption (A2) is dropped.

To solve RTSP (17a), we first summarize the reformulation of  $\sup_{\xi \in \Omega} Q(x, \xi)$  in the following proposition. The proof is similar to that for Proposition 5 in [19] and thus omitted here for brevity.

**Proposition 8** *The value of  $\sup_{\xi \in \Omega} Q(x, \xi)$  can be obtained by solving a mixed-integer program.*

Second, we resample from  $\mathbb{P}_0$  to obtain a series of samples  $\{\xi^s\}_{s=1}^S$ , and then formulate an SAA

problem for RTSP (17a)

$$z_S(d) = \min_{x \in X, \zeta \in \mathbb{R}} c^\top x + (1 - d/2)\zeta + \frac{1}{S} \sum_{s=1}^S (Q(x, \xi^s) - \zeta)^+ + (d/2) \sup_{\xi \in \Omega} Q(x, \xi), \quad (18)$$

where we approximate  $\mathbb{E}_{\mathbb{P}_0}[Q(x, \xi) - \zeta]^+$  by the sample average  $(1/S) \sum_{s=1}^S (Q(x, \xi^s) - \zeta)^+$ . A number of theoretical and computational studies have been developed for the (risk-neutral) TSP (see, e.g., [26], [1], and [22]) and the chance-constrained stochastic programming models (see, e.g., [25] and [28]). In this section, we develop and analyze an SAA approach as a counterpart for solving RTSP. We analyze the convergence properties of the SAA problem in Section 3.1 and describe the solution validation procedure and summarize the implementation of the SAA approach in Section 3.2. In Section 3.3, we discuss an alternative SAA approach for solving RTSP (17a) with a general sample space, i.e., when Assumption (A2) is dropped.

### 3.1 Convergence Properties of the SAA Problem

We analyze the limiting behavior of the SAA problem (18) in approximating RTSP (corresponding to a given distance  $d$ ), as the sample size  $S$  increases. We show that both the optimal objective value (e.g.,  $z_S(d)$ ) and the set of optimal solutions of the SAA problem (18) (defined as  $U_S(d)$ ) converge to those of RTSP (e.g.,  $z(d)$  and  $U(d)$ ). We summarize the main result in the following theorem.

**Theorem 3** *Suppose that  $X$  is compact. Then the optimal objective value of the SAA problem (18) converges to that of RTSP (17a) as  $S$  increases to infinity with probability 1, i.e.,  $\lim_{S \rightarrow \infty} z_S(d) = z(d)$ . Furthermore,  $U_S(d)$  converges to  $U(d)$  as  $S$  increases to infinity with probability 1 in the sense that  $\lim_{S \rightarrow \infty} \sup_{x \in U_S(d)} \text{dist}(x, U(d)) = 0$ .*

*Proof:* For each  $(x, \zeta) \in X \times \mathbb{R}$ , it is clear that

$$\lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S (Q(x, \xi^s) - \zeta)^+ = \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi$$

by the strong law of large numbers. Now letting  $\hat{h}(x, \zeta, d)$  and  $\hat{h}_S(x, \zeta, d)$  represent the objective values of RTSP (17a) and the SAA problem (18) corresponding to each  $(x, \zeta) \in X \times \mathbb{R}$ , respectively, we have  $\hat{h}_S(x, \zeta, d)$  converges to  $\hat{h}(x, \zeta, d)$  pointwise, i.e.,

$$\lim_{S \rightarrow \infty} \hat{h}_S(x, \zeta, d) = \hat{h}(x, \zeta, d)$$

with probability 1. The conclusion then follows from Theorems 5.3 and 7.48 in [36].  $\blacksquare$

### 3.2 Solution Validation and the SAA Algorithm

In implementation of the SAA approach, we construct the following lower and upper bounds for  $z(d)$  to validate the solution quality.

**Lower Bound.** To obtain lower bounds for  $z(d)$  corresponding to a given sample size  $S$ , we first establish the following proposition as a simple variant of Theorem 1 in [26].

**Proposition 9** *Let  $\mathbb{P}_0^S$  represent the product probability distribution of samples  $\{\xi^s\}_{s=1}^S$  from  $\mathbb{P}_0$ . Then  $z(d) \geq \mathbb{E}_{\mathbb{P}_0^S}[z_S(d)]$ .*

*Proof:* Define

$$Q'(x, \zeta, \xi) := c^\top x + (1 - d/2)\zeta + (Q(x, \xi) - \zeta)^+.$$

Then RTSP can be restated as  $\min_{x \in X, \zeta \in \mathbb{R}} \{\mathbb{E}_{\mathbb{P}_0}[Q'(x, \zeta, \xi)] + \sup_{\xi \in \Omega} Q(x, \xi)\}$ . It follows that

$$\begin{aligned} z(d) &= \min_{x \in X, \zeta \in \mathbb{R}} \left\{ \mathbb{E}_{\mathbb{P}_0}[Q'(x, \zeta, \xi)] + \sup_{\xi \in \Omega} Q(x, \xi) \right\} \\ &= \min_{x \in X, \zeta \in \mathbb{R}} \left\{ \mathbb{E}_{\mathbb{P}_0^S} \left[ \frac{1}{S} \sum_{s=1}^S Q'(x, \zeta, \xi^s) + \sup_{\xi \in \Omega} Q(x, \xi) \right] \right\} \\ &\geq \mathbb{E}_{\mathbb{P}_0^S} \left[ \min_{x \in X, \zeta \in \mathbb{R}} \left\{ \frac{1}{S} \sum_{s=1}^S Q'(x, \zeta, \xi^s) + \sup_{\xi \in \Omega} Q(x, \xi) \right\} \right] = \mathbb{E}_{\mathbb{P}_0^S}[z_S(d)]. \end{aligned}$$

$\blacksquare$

Hence, a statistical lower bound of  $z(d)$  corresponding to a given sample size  $S$ , can be obtained by independently constructing and solving SAA problem (18) for  $M$  replications, each of whose optimal objective values are denoted by  $z_S^m(d)$  for  $m = 1, \dots, M$ . By computing the sample mean of these optimal objective values, i.e.,  $\bar{z} := (1/M) \sum_{m=1}^M z_S^m(d)$  and their sample variance, i.e.,  $\sigma_z^2 := (1/(M-1)) \sum_{m=1}^M (z_S^m(d) - \bar{z})^2$ , we obtain that  $\bar{z} - t_{M-1, \beta} \sigma_z / \sqrt{M}$  is a true lower bound of  $z(d)$  with at least probability  $(1 - \beta)$ , where  $\beta \in (0, 1)$  and  $t_{M-1, \beta}$  represents the  $100(1 - \beta)$  percentile of the  $t$ -distribution with  $M - 1$  degrees of freedom.

**Upper Bound.** Suppose that a solution  $(\bar{x}, \bar{\zeta}) \in X \times \mathbb{R}$  is obtained from solving SAA problem (18). It is clear that the corresponding objective value

$$c^\top \bar{x} + (1 - d/2)\bar{\zeta} + \mathbb{E}_{\mathbb{P}_0} [Q(\bar{x}, \xi) - \bar{\zeta}]^+ + (d/2) \sup_{\xi \in \Omega} Q(\bar{x}, \xi)$$

is an upper bound of  $z(d)$ . Accordingly, by independently sampling  $\{\xi^s\}_{s=1}^{S'}$  from  $\mathbb{P}_0$ , the objective value of  $(\bar{x}, \bar{\zeta})$  can be estimated based on each sample:

$$\hat{U}(\bar{x}, \bar{\zeta}, \xi^s) := c^\top \bar{x} + (1 - d/2) \bar{\zeta} + (Q(\bar{x}, \xi^s) - \bar{\zeta})^+ + (d/2) \sup_{\xi \in \Omega} Q(\bar{x}, \xi).$$

By computing the sample mean of these objective values, i.e.,  $\bar{U} := (1/S') \sum_{s=1}^{S'} \hat{U}(\bar{x}, \bar{\zeta}, \xi^s)$  and their sample variance, i.e.,  $\sigma_U^2 := (1/(S'-1)) \sum_{s=1}^{S'} \left( \hat{U}(\bar{x}, \bar{\zeta}, \xi^s) - \bar{U} \right)^2$ , we obtain that  $\bar{U} + t_{S'-1, \beta} \sigma_U / \sqrt{S'}$  is a true upper bound of  $z(d)$  with at least probability  $(1 - \beta)$ .

**SAA Algorithm.** We summarize the SAA algorithm as follows.

1. For  $m = 1, \dots, M$ , repeat the following steps:
  - (1) Generate independent samples  $\{\xi^s\}_{s=1}^{S'}$  from  $\mathbb{P}_0$ .
  - (2) Solve SAA problem (18), and record its optimal objective value  $z_{S'}^m(d)$  and an optimal solution  $(\bar{x}, \bar{\zeta})$ .
  - (3) Estimate the corresponding upper bound  $U(\bar{x}, \bar{\zeta}, d) := \bar{U} + t_{S'-1, \beta} \sigma_U / \sqrt{S'}$ .
2. Estimate a lower bound  $L(d) := \bar{z} - t_{M-1, \beta} \sigma_z / \sqrt{M}$ , with  $\bar{z} = (1/M) \sum_{m=1}^M z_{S'}^m(d)$ .
3. Pick an optimal solution  $(\bar{x}^*, \bar{\zeta}^*)$  corresponding to the smallest upper bound  $U(\bar{x}^*, \bar{\zeta}^*, d)$ .
4. Estimate the optimality gap by  $\text{GAP} = (U(\bar{x}^*, \bar{\zeta}^*, d) - L(d)) / L(d) \times 100\%$ .
5. If the optimality gap satisfies predetermined termination tolerances (e.g.,  $|\text{GAP}| < \epsilon_1$  and  $t_{M-1, \beta}, t_{S'-1, \beta} < \epsilon_2$ ), terminate with the current solution. Otherwise, update  $S \leftarrow 2S$ ,  $M \leftarrow 2M$ , and  $S' \leftarrow 2S'$ , and go to Step 1.

### 3.3 Discussion on an Alternative SAA Approach When $\Omega$ is a General Sample Space

The SAA approach can be extended to solve RTSP (17a) with a general sample space  $\Omega$ , i.e., when Assumption (A2) is dropped. For this case, apart from the samples  $\{\xi^s\}_{s=1}^{S'}$  drawn from  $\mathbb{P}_0$  for the expected value part in the objective function, we also consider an alternative probability distribution  $\mathbb{P}_\Omega$  on  $\Omega$  and sample from  $\mathbb{P}_\Omega$  to obtain a series of independent samples  $\{\hat{\xi}^s\}_{s=1}^{S_\Omega}$  for the worst-case cost part in the objective function, which are also independent from the samples  $\{\xi^s\}_{s=1}^{S'}$  from  $\mathbb{P}_0$ . An alternative SAA problem is formulated as

$$z_{S, S_\Omega}(d) = \min_{x \in X, \zeta \in \mathbb{R}} c^\top x + (1 - d/2) \zeta + \frac{1}{S} \sum_{s=1}^S (Q(x, \xi^s) - \zeta)^+ + (d/2) \max_{s=1, \dots, S_\Omega} Q(x, \hat{\xi}^s), \quad (20)$$

where the worst-case cost  $\sup_{\xi \in \Omega} Q(x, \xi)$  is approximated by  $\max_{s=1, \dots, S_\Omega} Q(x, \hat{\xi}^s)$ , the sample maximum. It can be shown that the optimal objective value and the set of optimal solutions to the alternative SAA problem (20) converge to those of RTSP (17a) under mild assumptions.

**Theorem 4** *Suppose that (i)  $X$  is compact and (ii) the alternative probability distribution satisfies*

$$\mathbb{P}_\Omega \left\{ Q(x, \xi) > \sup_{\xi \in \Omega} Q(x, \xi) - \epsilon \right\} > 0$$

*for any  $\epsilon > 0$ . Then the optimal objective value of the SAA problem (20) converges to that of RTSP (17a) as  $S, S_\Omega$  increase to infinity with probability 1, i.e.,  $\lim_{S, S_\Omega \rightarrow \infty} z_{S, S_\Omega}(d) = z(d)$ . Furthermore,  $U_{S, S_\Omega}(d)$  converges to  $U(d)$  as  $S, S_\Omega$  increase to infinity with probability 1 in the sense that  $\lim_{S, S_\Omega \rightarrow \infty} \sup_{x \in U_{S, S_\Omega}(d)} \text{dist}(x, U(d)) = 0$ .*

*Proof:* See Appendix L for the detailed proof. ■

Note here that this condition can be satisfied when (i)  $\mathbb{P}_\Omega$  has a positive density everywhere in  $\Omega$  (e.g., when  $\mathbb{P}_\Omega$  is a kernel density estimator on  $\Omega$ ), or (ii)  $\mathbb{P}_\Omega$  is discrete and has a positive mass on each extreme point of  $\Omega$ . Similar to the convergence properties, the solution validation and solution algorithm for the alternative SAA approach can be adapted from Section 3.2.

## 4 Computational Experiments

In this section, we provide two numerical experiments to illustrate the applications of RTSP. The first example illustrates the risk-averse newsvendor problem with distributional ambiguity as described in Section 2.2. We apply Proposition 7 to obtain the risk-averse optimal solution and compare it with the perfect-information solution in terms of order quantity and average total cost. The second example illustrates a two-stage risk-averse capacitated lot-sizing problem with distributional ambiguity. We solve this problem based on the SAA algorithm described in Section 3.2, and compare the obtained risk-averse optimal solution with the perfect-information solution in terms of setup decisions and average total cost.

### 4.1 Risk-Averse Newsvendor Problem

In this experiment, we illustrate the risk-averse newsvendor problem with distributional ambiguity as described in Section 2.2. We compute the risk-averse optimal solutions under various distributional settings, by assuming that the demand follows four classes of (true) probability distributions namely truncated normal, log-normal, Weibull, and exponential distributions, respectively. Under each distributional setting, we first collect a total amount of  $N$  independent data samples,  $\{\xi^1, \dots, \xi^N\}$ , from the true distribution to estimate the sample space  $\Omega := [\xi^L, \xi^U]$  and confidence set  $\mathcal{D}$ . Then we compute the risk-averse solution by Proposition 7 based on the estimated  $\Omega$  and  $\mathcal{D}$ .

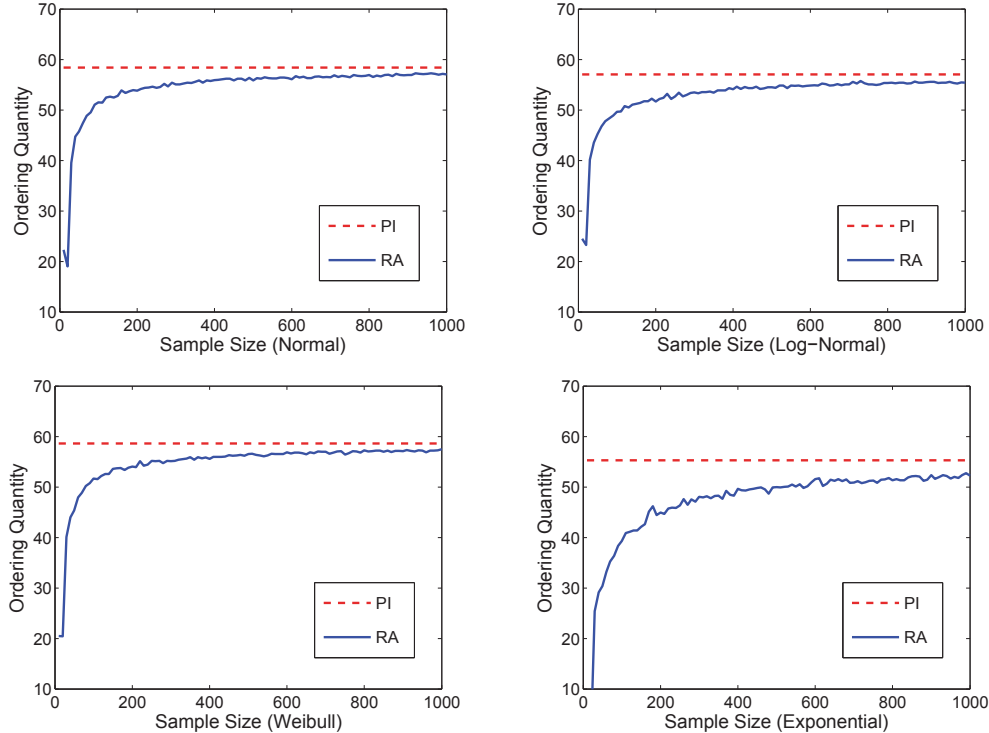


Figure 1: Comparison of the risk-averse optimal solution (RA) and the perfect-information solution (PI) for the newsvendor problem

Finally, we compare the risk-averse solution with the perfect-information solution, i.e., the optimal solution a risk-neutral newsvendor would make if she knows the true probability distribution. For parameter settings, we set ordering cost  $c = \$1.0$ , selling price  $p = \$3.0$ , and salvage value  $s = \$0.5$ . For distribution estimation, we apply the three-sigma rule to estimate  $\Omega$ , and develop the point estimate  $\mathbb{P}_0$  as a KDE based on the set of data samples as described in Section 2.1. Also, we set  $d := 20N^{-1}$  to represent the distance tolerance  $d$  as a function of sample size  $N$ . To evaluate the performance of the risk-averse solution, we compare it with the perfect-information solution under various sample sizes evolving from 10 to 1000. We also compare their out-of-sample average total costs based on 10,000 random demand samples drawn from the true probability distribution under each distributional setting.

The experiment results are summarized in Figures 1 and 2. Under various sample sizes and distributional settings, we compare the risk-averse solution with the perfect-information solution in Figure 1, and the average total costs of these solutions in Figure 2. We first observe that under all distributional settings and sample sizes, the risk-averse solution always orders less than the perfect-information solution, and accordingly has a higher average total cost. Second, we observe that the



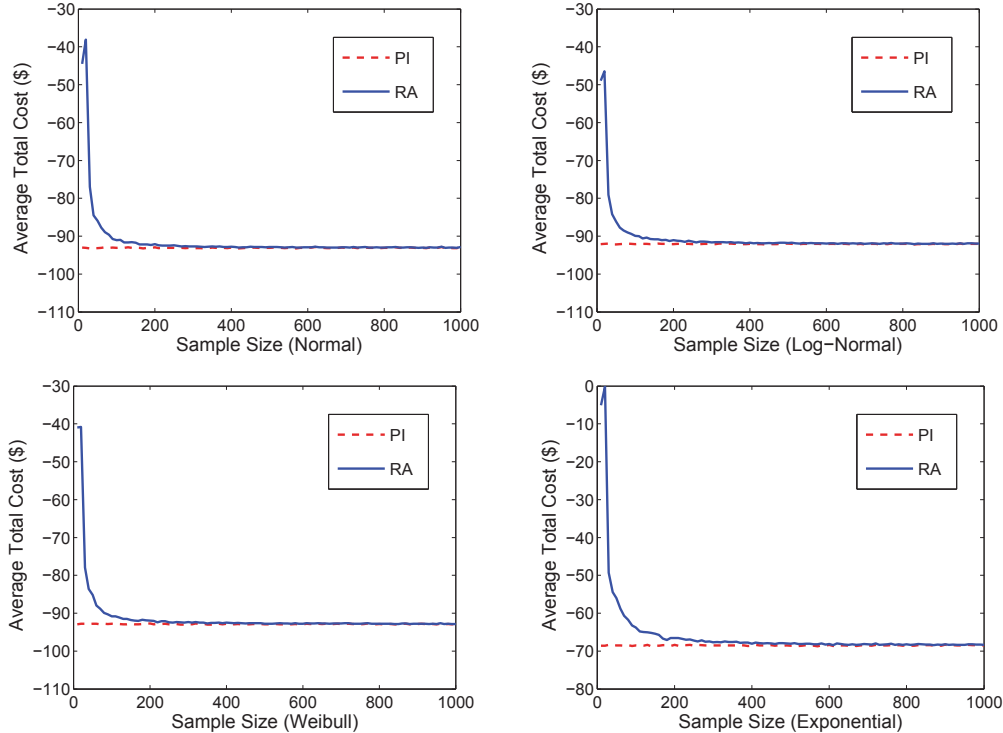


Figure 2: Comparison of the average total cost of the risk-averse optimal solution (RA) and that of the perfect-information solution (PI) for the newsvendor problem

risk-averse solution quickly gets close to the perfect-information solution (e.g., within 5-6% for 350 data samples) as the sample size grows under all distributional settings. Accordingly, the average total cost of the risk-averse solution quickly converges to that of the perfect-information solution (e.g., within 0.5% for 300 data samples). It implies that a risk-averse newsvendor can quickly decrease her decision conservatism by learning from the historical data. Third, by comparing the convergence rates in Figures 1 and 2, we observe that the average cost performance converges faster than the solution does. As compared to taking more than 1000 data samples for the optimal solution to converge, it takes around 600 data samples for the average cost to converge in all distributional settings. It indicates that a risk-averse newsvendor with a moderate amount of data samples can already work well and the solution is close to the perfect-information case. The value of collecting additional data samples (beyond moderate) is not significant.

## 4.2 Risk-Averse Lot-Sizing Problem

In this experiment, we apply the RTSP model to the capacitated lot-sizing problem, in which a decision maker schedules production to satisfy uncertain demands in a finite operational interval  $\{1, \dots, T\}$ . In each time period  $t \in \{1, \dots, T\}$ , production amount can be made up to a capacity

$C_t$  at a setup cost (or fixed cost)  $c_t$  and a unit production cost  $q_t$ . A random demand  $\xi_t$  arrives at the beginning of each time period  $t$ , and can be satisfied by either immediate production or inventory left from previous periods, or outsourcing (demand is not satisfied by production and/or inventory). Positive inventory at the end of period  $t$  is charged with a unit holding cost  $h_t$ . Each outsourcing incurs a unit cost  $M_t$ . In this experiment, we consider a two-stage lot-sizing problem, where the binary setup decisions (i.e., whether or not to produce) are determined in the first stage, while the production amounts (i.e., how much to produce) are determined in the second stage. Supposing that the demand uncertainty follows a probability distribution  $\mathbb{P}$ , a risk-neutral decision maker minimizes the expected total cost

$$\min_{x \in \{0,1\}^T} \sum_{t=1}^T c_t x_t + \mathbb{E}_{\mathbb{P}} [Q(x, \xi)],$$

where

$$\begin{aligned} Q(x, \xi) = & \min_{y(\xi), v(\xi), p(\xi) \geq 0} \sum_{t=1}^T (q_t y_t(\xi) + h_t v_t(\xi) + M_t p_t(\xi)) \\ \text{s.t. } & v_{t-1}(\xi) + y_t(\xi) + p_t(\xi) = \xi_t(\xi) + v_t(\xi), \quad \forall t = 1, \dots, T, \\ & y_t(\xi) \leq C_t x_t(\xi), \quad \forall t = 1, \dots, T, \\ & v_0(\xi) = v_T(\xi) = 0, \end{aligned}$$

where variables  $y_t(\xi)$ ,  $v_t(\xi)$ , and  $p_t(\xi)$  represent the production, inventory, and outsourcing amounts in each time period  $t$  corresponding to demand outcome scenario  $\xi$ , respectively. We study a risk-averse variant by considering the distributional ambiguity, where the decision maker's knowledge about  $\mathbb{P}$  is imperfect and only a series of historical data samples  $\{\xi^1, \dots, \xi^N\}$  are available. Based on these data samples, a decision maker can obtain a sample space estimation of  $\xi$  (e.g.,  $\Omega$ ), a point estimate  $\mathbb{P}_0$ , and a confidence set  $\mathcal{D}$  as defined in (4). Accordingly, the risk-averse lot-sizing problem can be modeled as

$$\min_{x \in \{0,1\}^T} \sum_{t=1}^T c_t x_t + \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [Q(x, \xi)].$$

In this experiment, we set  $T = 20$ , unit holding cost  $h_t = \$1$ , setup cost  $c_t = \$300$ , and unit production cost  $q_t = \$3$  for each  $t \in \{1, \dots, T\}$ . For the random demand  $\xi = (\xi_1, \dots, \xi_T)$ , we set  $\mathbb{E}_{\mathbb{P}}[\xi_t] = 40(1 + 0.2 \sin(6t/T))$  and  $\sigma(\xi_t) = (1/4)\mathbb{E}_{\mathbb{P}}[\xi_t]$  for each  $t$ . Also, we consider four distributional settings by assuming that  $\xi$  follows truncated normal, log-normal, Weibull, and exponential distributions under each setting, respectively. For distribution estimation, we apply

the three-sigma rule to obtain  $\Omega$ , and develop  $\mathbb{P}_0$  as a KDE based on the set of historical data as described in Section 2.1. We set  $d := 40N^{-1}$  to represent the distance tolerance  $d$  as a function of sample size  $N$ . We solve this problem by using the SAA algorithm summarized in Section 3.2 and obtain the risk-averse optimal solution. To evaluate the performance of the risk-averse setup (i.e., the first-stage solution), we compare it with the perfect-information setup under data sample size evolving from 20 to 4000, and the four distributional settings. The experiment results are

Sample Size	Time Periods																				Diff
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
20	1	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	11
40	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	10
100	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	10
500	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	0	8
1000	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	0	8
1500	1	0	1	0	1	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	3
2000	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
2500	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
3000	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
4000	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
PI	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0

Table 1: Comparison of the risk-averse setup and the perfect-information setup (PI) in the lot-sizing problem under various sample sizes and Weibull distribution

summarized in Table 1 and Figures 3 and 4. In Table 1, we report the Weibull distribution case to compare the risk-averse setup with the perfect-information setup under various sample sizes. We display the perfect-information setup in the last row as a benchmark, and we measure the difference between these setups by using the taxicab geometry, i.e.,

$$\text{Diff} = \sum_{t=1}^T |x_t^* - x_t^{\text{PI}}|,$$

where  $x^*$  represents the risk-averse setup, and  $x^{\text{PI}}$  represents the perfect-information setup. In Table 1, we observe that the Diff value (shown in the last column) decreases as the sample size increases and converges to zero once sample size exceeds 2000. In Figure 3, we depict Diff values under various distributional settings, and notice that similar convergence behaviors appear under all distributional settings. In Figure 4, we compare the risk-averse setup and perfect-information setup in terms of out-of-sample average total cost, based on 10,000 random samples drawn from the true probability distribution under each distributional setting. For example, to evaluate the risk-averse setup, we fix the setup decision as  $x^*$  and obtain  $Q(x^*, \xi^m)$  by solving the second-stage problem for each  $m = 1, \dots, 10000$ . The average total cost is computed as  $c^\top x^* + (1/10000) \sum_{m=1}^{10000} Q(x^*, \xi^m)$ . In Figure 4, we observe that the average total cost of the risk-averse setup is always higher than that

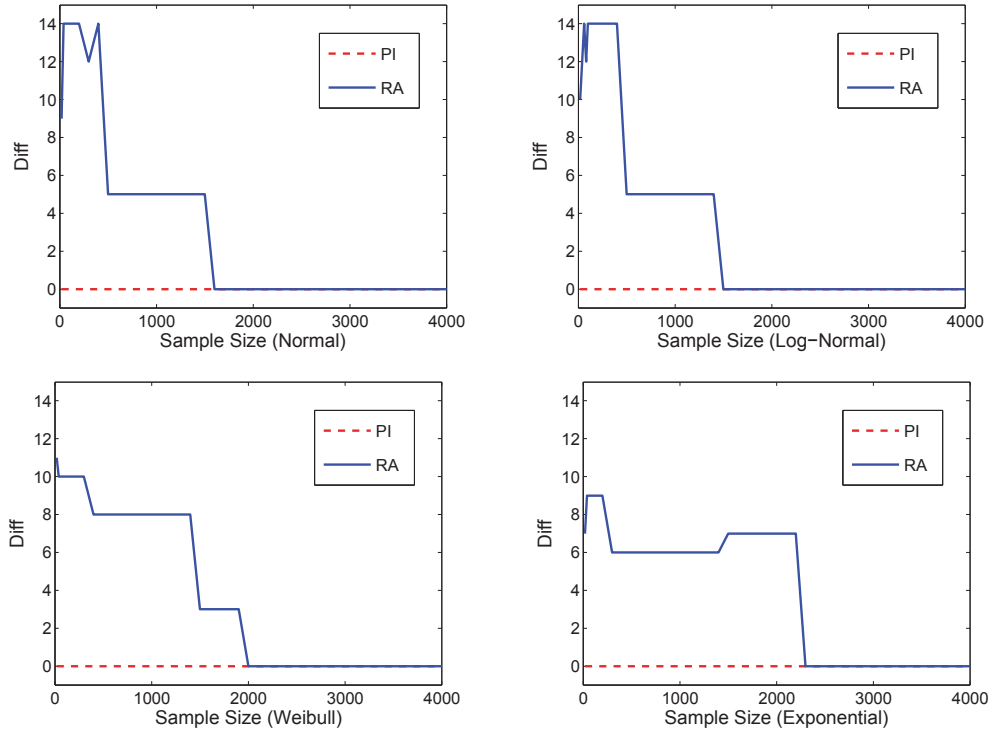


Figure 3: Comparison of the risk-averse optimal setup (RA) and the perfect-information setup (PI) for the lot-sizing problem

of the perfect-information setup under all distributional settings. As sample size grows, the average total cost of the risk-averse setup quickly gets close to that of the perfect-information setup (e.g., within 5-6%) once the sample size exceeds 500. Also, the two average total cost values converge as the sample size exceeds 2500. These observations indicate that a risk-averse setup performs reasonably well with a moderate amount of data, and it improves to the perfect-information setup if the data amount becomes sufficient.

## 5 Conclusions

In this paper, we proposed a risk-averse counterpart for the classical two-stage stochastic program by explicitly incorporating the distributional ambiguity, considering both discrete and continuous distribution cases, within a data-driven decision making environment. Through deriving an equivalent reformulation of the proposed model, our result provides the insights on the relationship between the objective function of the proposed model and coherent risk measures. In addition, our convergence analysis results show that the conservatism of the proposed model vanishes as the size of historical data increases to infinity. Furthermore, for each RTSP corresponding to a given

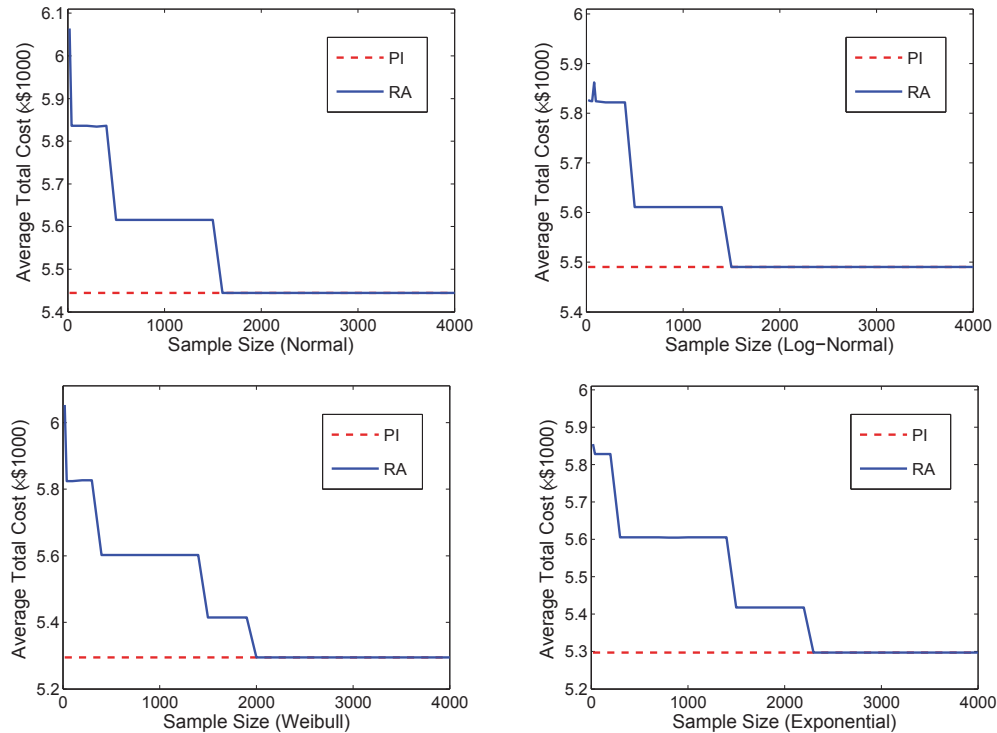


Figure 4: Comparison of the average total cost of the risk-averse optimal setup (RA) and that of the perfect-information setup (PI) for the lot-sizing problem

historical data set, an SAA-based solution algorithm was developed to solve the model efficiently. Considering the wide range of decision making under uncertainty problems, our proposed modeling and solution framework has great potential to be applied to solve many practical problems.

## Acknowledgement

The authors would like to thank Professor Alexander Shapiro for the comments and suggestions on an early version of this paper.

## References

- [1] S. Ahmed and A. Shapiro. The sample average approximation method for stochastic programs with integer recourse. Technical report, Department of Industrial and Systems Engineering, Georgia Institute of Technology, 2002.
- [2] J. Ang, F. Meng, and J. Sun. Two-stage stochastic linear programs with incomplete information on uncertainty. *European Journal of Operational Research*, 233(1):16–22, 2014.
- [3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

- [4] G. Bayraksan and D. Morton. A sequential sampling procedure for stochastic programming. *Operations Research*, 59(4):898–913, 2011.
- [5] E. Beale. On minimizing a convex function subject to linear inequalities. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):173–184, 1955.
- [6] A. Ben-Tal, D. den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.
- [7] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.
- [8] D. Bertsimas and V. Goyal. On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical programming*, 134(2):491–531, 2012.
- [9] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, New York, 1999.
- [10] J. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, 1997.
- [11] G.C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.
- [12] G. Dantzig. Linear programming under uncertainty. *Management science*, 1(3-4):197–206, 1955.
- [13] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
- [14] F. Delbaen. Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics*, pages 1–37. Springer, 2002.
- [15] L. Devroye and L. Györfi. *Nonparametric Density Estimation: The  $\ell_1$  View*. John Wiley & Sons Inc, 1985.
- [16] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61, 2006.
- [17] H. Föllmer and A. Schied. Stochastic finance: An introduction in discrete time. *The Mathematical Intelligencer*, 26(4):67–68, 2004.
- [18] J. Higle and S. Sen. Stochastic decomposition: An algorithm for two-stage linear programs with recourse. *Mathematics of Operations Research*, 16(3):650–669, 1991.
- [19] R. Jiang, M. Zhang, G. Li, and Y. Guan. Two-stage network constrained robust unit commitment problem. *European Journal of Operational Research*, 234(3):751–762, 2014.
- [20] P. Kall and S. Wallace. *Stochastic Programming*. John Wiley and Sons, 1994.
- [21] D. Klabjan, D. Simchi-Levi, and M. Song. Robust stochastic lot-sizing by means of histograms. *Production and Operations Management*, 22(3):691–710, 2013.
- [22] A. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

- [23] D. Kuhn, W. Wiesemann, and A. Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.
- [24] D. Love and G. Bayraksan. Phi-divergence constrained ambiguous stochastic programs. Technical report, University of Arizona, 2012.
- [25] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.
- [26] W.-K. Mak, D. Morton, and R.K. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56, 1999.
- [27] S. Mehrotra and H. Zhang. Models and algorithms for distributionally robust least squares problems. *Mathematical Programming*, in press, 2013.
- [28] B. Pagnoncelli, S. Ahmed, and A. Shapiro. Sample average approximation method for chance constrained programming: theory and applications. *Journal of Optimization Theory and Applications*, 142(2):399–416, 2009.
- [29] G. Pflug and D. Wozabal. Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442, 2007.
- [30] R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26(7):1443–1471, 2002.
- [31] T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167(1):96–115, 2005.
- [32] H. Scarf. A min-max solution of an inventory problem. In K. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, pages 201–209. Stanford University Press, 1958.
- [33] S. Sen, R. Doverspike, and S. Cosares. Network planning with random demand. *Telecommunication Systems*, 3(1):11–30, 1994.
- [34] S. Sen, L. Yu, and T. Genc. A stochastic programming approach to power portfolio optimization. *Operations Research*, 54(1):55–72, 2006.
- [35] A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14(4):1237–1249, 2004.
- [36] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9. SIAM, 2009.
- [37] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.
- [38] S. Takriti, J. Birge, and E. Long. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems*, 11(3):1497–1508, 1996.
- [39] R. Van Slyke and R. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663, 1969.

- [40] L. Wasserman. *All of Nonparametric Statistics*. Springer, 2006.
- [41] J. Žáčková. On minimax solutions of stochastic linear programming problems. *Časopis pro pěstování matematiky*, 91(4):423–430, 1966.
- [42] S. Zymler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.
- [43] S. Zymler, D. Kuhn, and B. Rustem. Worst-case value at risk of nonlinear portfolios. *Management Science*, 59(1):172–188, 2013.



## Appendix A Proof of Proposition 1

*Proof:* We discuss the following cases on the values of  $d$  and  $\beta_2$ . For each case, we show that (i)  $f_*^{(n)}(\xi)$  is a well-defined density function in  $\mathcal{D}$ , i.e.,  $\int_{\mathbb{R}^K} f_*^{(n)}(\xi)d\xi = 1$ , and  $\int_{\mathbb{R}^K} |f_*^{(n)}(\xi) - f_0(\xi)|d\xi \leq d$ , and (ii) equality (16) holds.

1. When  $d \in [0, 2)$  and  $\beta_2 = 0$ , we have  $\mathbb{P}_0\{g(x, \xi) \leq \gamma\}$  is a continuous function in  $\gamma$  at  $\gamma = \text{VaR}$ . It follows that  $\mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\} = d/2$  by the definition of VaR. Hence,

$$\begin{aligned} \int_{\mathbb{R}^K} f_1(\xi)d\xi &= \int_{[g(x, \xi) > \text{VaR}]} f_0(\xi)d\xi \\ &= \mathbb{P}_0\{g(x, \xi) > \text{VaR}\} = 1 - d/2. \end{aligned}$$

Also,

$$\int_{\mathbb{R}^K} f_3^{(n)}(\xi)d\xi = \int_{\Gamma_n} \frac{d}{2\mu(\Gamma_n)}d\xi = d/2. \quad (21)$$

Thus,  $\int_{\mathbb{R}^K} f_*^{(n)}(\xi)d\xi = \int_{\mathbb{R}^K} f_1(\xi)d\xi + \int_{\mathbb{R}^K} f_3^{(n)}(\xi)d\xi = 1$ . In addition, when  $n$  is sufficiently large, we have

$$\begin{aligned} \int_{\mathbb{R}^K} |f_*^{(n)}(\xi) - f_0(\xi)|d\xi &= \int_{\mathbb{R}^K} |f_1(\xi) + f_3^{(n)}(\xi) - f_0(\xi)|d\xi \\ &= \int_{[g(x, \xi) \leq \text{VaR}]} f_0(\xi)d\xi + \int_{\Gamma_n} f_3^{(n)}(\xi)d\xi \\ &= \mathbb{P}_0\{g(x, \xi) \leq \text{VaR}\} + d/2 \\ &= d/2 + d/2 = d, \end{aligned}$$

where the second equality is due to the definitions of  $f_1(\xi)$ ,  $f_3^{(n)}(\xi)$ , and  $\Gamma_n$ , and the third equality is due to equality (21). Thus,  $f_*^{(n)}(\xi)$  is a well-defined density function.

Now we show that equality (16) holds. We first have

$$\begin{aligned} \int_{\mathbb{R}^K} g(x, \xi)f_1(\xi)d\xi &= \int_{[g(x, \xi) > \text{VaR}]} g(x, \xi)f_0(\xi)d\xi \\ &= \mathbb{P}_0\{g(x, \xi) > \text{VaR}\} \text{CVaR}_{d/2}^{\mathbb{P}_0^+}[g(x, \xi)] \\ &= (1 - d/2)\text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)], \end{aligned} \quad (22)$$

where the second equality is due to the definition of  $\text{CVaR}_{d/2}^{\mathbb{P}_0^+}[g(x, \xi)]$  following Equation (10) in [30], and the last equality is due to  $\mathbb{P}_0\{g(x, \xi) > \text{VaR}\} = 1 - d/2$  and the property of  $\text{CVaR}_{d/2}^{\mathbb{P}_0^+}[g(x, \xi)]$  following Proposition 5 in [30].

Then, we show that

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi)f_3^{(n)}(\xi)d\xi = (d/2) \text{ess sup}_{\xi \in \Omega} g(x, \xi). \quad (23)$$

In fact, on one hand, we have, when  $n$  is sufficiently large,

$$\int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi \leq \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi) \int_{\mathbb{R}^K} f_3^{(n)}(\xi) d\xi = (d/2) \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi)$$

by the definition of  $\operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi)$ , and so  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi \leq (d/2) \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi)$ .

On the other hand, we have

$$\int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi \geq \left( \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi) - \frac{1}{n} \right) \int_{\mathbb{R}^K} f_3^{(n)}(\xi) d\xi = (d/2) \left( \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi) - \frac{1}{n} \right)$$

by the definition of  $f_3^{(n)}(\xi)$ , and so by letting  $n \rightarrow \infty$  we have  $\lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi \geq (d/2) \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi)$ . Therefore, we have proved equality (23) and so

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_*^{(n)}(\xi) d\xi &= \int_{\mathbb{R}^K} g(x, \xi) f_1(\xi) d\xi + \lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi \\ &= (1 - d/2) \operatorname{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)] + (d/2) \operatorname{ess\,sup}_{\xi \in \Omega} g(x, \xi) \\ &= \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)] \end{aligned}$$

based on equalities (22), (23) and Theorem 1.

2. When  $d \in [0, 2)$  and  $\beta_2 > 0$ , we have

$$\begin{aligned} \int_{\mathbb{R}^K} (f_1(\xi) + f_2(\xi)) d\xi &= \int_{[g(x, \xi) > \operatorname{VaR}]} f_0(\xi) d\xi + \int_{[g(x, \xi) = \operatorname{VaR}]} \frac{\beta_1}{\beta_2} f_0(\xi) d\xi \\ &= \mathbb{P}_0\{g(x, \xi) > \operatorname{VaR}\} + \beta_1 \\ &= 1 - d/2. \end{aligned}$$

Then,

$$\int_{\mathbb{R}^K} f_*^{(n)}(\xi) d\xi = \int_{\mathbb{R}^K} (f_1(\xi) + f_2(\xi)) d\xi + \int_{\mathbb{R}^K} f_3^{(n)}(\xi) d\xi = (1 - d/2) + d/2 = 1,$$

where the second equality is due to equality (21). In addition, we have

$$\begin{aligned} &\int_{\mathbb{R}^K} |f_*^{(n)}(\xi) - f_0(\xi)| d\xi \\ &= \int_{\mathbb{R}^K} |f_1(\xi) + f_2(\xi) + f_3^{(n)}(\xi) - f_0(\xi)| d\xi \\ &= \int_{[g(x, \xi) < \operatorname{VaR}]} f_0(\xi) d\xi + \int_{[g(x, \xi) = \operatorname{VaR}]} |f_2(\xi) - f_0(\xi)| d\xi + \int_{\Gamma_n} f_3^{(n)}(\xi) d\xi \\ &= \mathbb{P}_0\{g(x, \xi) < \operatorname{VaR}\} + \int_{[g(x, \xi) = \operatorname{VaR}]} \frac{d/2 - \mathbb{P}_0\{g(x, \xi) < \operatorname{VaR}\}}{\beta_2} f_0(\xi) d\xi + d/2 \\ &= \mathbb{P}_0\{g(x, \xi) < \operatorname{VaR}\} + d/2 - \mathbb{P}_0\{g(x, \xi) < \operatorname{VaR}\} + d/2 \\ &= d, \end{aligned}$$

where the first and second equalities are due to the definition of  $f_*^{(n)}(\xi)$ , and the third equality is due to the definition of  $f_2(\xi)$  and equality (21). Hence,  $f_*^{(n)}(\xi)$  is a well-defined density function in  $\mathcal{D}$ . Now we show that equality (16) holds. We first have

$$\begin{aligned} & \int_{\mathbb{R}^K} g(x, \xi) (f_1(\xi) + f_2(\xi)) d\xi \\ &= \int_{[g(x, \xi) > \text{VaR}]} g(x, \xi) f_0(\xi) d\xi + \int_{[g(x, \xi) = \text{VaR}]} \frac{\beta_1}{\beta_2} g(x, \xi) f_0(\xi) d\xi \\ &= \mathbb{P}_0\{g(x, \xi) > \text{VaR}\} \text{CVaR}_{d/2}^{\mathbb{P}_0^+}[g(x, \xi)] + \beta_1 \text{VaR} \end{aligned} \quad (24a)$$

$$= (1 - d/2) \text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)], \quad (24b)$$

where equality (24a) is due to the definition of  $\text{CVaR}_{d/2}^{\mathbb{P}_0^+}[g(x, \xi)]$  following Equation (10) in [30], and equality (24b) is due to the definition of  $\text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)]$  following Equation (21) in [30]. Then,

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_*^{(n)}(\xi) d\xi &= \int_{\mathbb{R}^K} g(x, \xi) (f_1(\xi) + f_2(\xi)) d\xi + \lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi \\ &= (1 - d/2) \text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)] + (d/2) \text{ess sup}_{\xi \in \Omega} g(x, \xi) \\ &= \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)], \end{aligned}$$

where the second equality is due to equalities (24b) and (23), and the last equality is due to Theorem 1.

3. When  $d \geq 2$ , we notice that  $f_*^{(n)}(\xi) = f_4^{(n)}(\xi) = (2/d)f_3^{(n)}(\xi)$  and so

$$\int_{\mathbb{R}^K} f_*^{(n)}(\xi) d\xi = (2/d) \int_{\mathbb{R}^K} f_3^{(n)}(\xi) d\xi = (2/d)(d/2) = 1, \quad (25)$$

where the second equality is due to equality (21). In addition, we have

$$\int_{\mathbb{R}^K} |f_*^{(n)}(\xi) - f_0(\xi)| d\xi \leq \int_{\mathbb{R}^K} (f_*^{(n)}(\xi) + f_0(\xi)) d\xi = 2 \leq d.$$

Hence,  $f_*^{(n)}(\xi)$  is a well-defined density function in  $\mathcal{D}$ . Furthermore, we have

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_*^{(n)}(\xi) d\xi = (2/d) \lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_3^{(n)}(\xi) d\xi = \text{ess sup}_{\xi \in \Omega} g(x, \xi), \quad (26)$$

where the last equality is due to equality (23). Therefore, we have

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^K} g(x, \xi) f_*^{(n)}(\xi) d\xi = \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)]$$

based on Theorem 1. ■

## Appendix B Proof of Lemma 1

*Proof:* To prove the first condition, we define  $\theta_1 := \text{ess sup}_{\xi \in \Omega} g(x, \xi)$  and  $\theta_2 := \sup_{\xi \in \Omega} g(x, \xi)$ , and assume for contradiction that  $\theta_1 < \theta_2$ . By definition (6), we have  $\mu(\{\xi \in \Omega : g(x, \xi) > \theta_1\}) = 0$  and so  $\mu(g^{-1}(\theta_1, \theta_2)) = 0$  where  $g^{-1}(\theta_1, \theta_2) := \{\xi \in \Omega : \theta_1 < g(x, \xi) < \theta_2\}$ . But since  $(\theta_1, \theta_2)$  is nonempty and open and  $g(x, \xi)$  is continuous in variable  $\xi$ , we have  $g^{-1}(\theta_1, \theta_2)$  is nonempty and open. It follows that  $\mu(g^{-1}(\theta_1, \theta_2)) > 0$  which is a contradiction with  $\mu(g^{-1}(\theta_1, \theta_2)) = 0$ .

To prove the second condition, we define  $\theta_1^i := \text{ess sup}_{\xi \in \Omega_i} g(x, \xi)$  and  $\theta_2^i := \sup_{\xi \in \Omega_i} g(x, \xi)$  for each  $i \in \Psi$ . Since  $g(x, \xi)$  is continuous in variable  $\xi$  on  $\Omega_i$  and  $\mu(\Omega_i) > 0$  for each  $i \in \Psi$ , we have  $\theta_1^i = \theta_2^i$  by the first condition and so  $\sup_{i \in \Psi} \theta_1^i = \sup_{i \in \Psi} \theta_2^i$ . Since  $\theta_2 = \sup_{i \in \Psi} \theta_2^i$ , it remains to show  $\theta_1 = \sup_{i \in \Psi} \theta_1^i$ . First, we have  $\sup_{i \in \Psi} \theta_1^i \leq \theta_1$  by the definitions of  $\theta_1$  and  $\theta_1^i$ . Second, we assume  $\sup_{i \in \Psi} \theta_1^i < \theta_1$  for contradiction. Let  $\hat{\theta} = (\sup_{i \in \Psi} \theta_1^i + \theta_1)/2$ , then  $\sup_{i \in \Psi} \theta_1^i < \hat{\theta} < \theta_1$ . It follows that  $\mu(\{\xi \in \Omega : g(x, \xi) > \hat{\theta}\}) > 0$  and  $\mu(\{\xi \in \Omega_i : g(x, \xi) > \hat{\theta}\}) = 0$  for each  $i \in \Psi$  by definition (6). But

$$\{\xi \in \Omega : g(x, \xi) > \hat{\theta}\} = \bigcup_{i \in \Psi} \{\xi \in \Omega_i : g(x, \xi) > \hat{\theta}\}$$

because  $\Omega = \bigcup_{i \in \Psi} \Omega_i$ , and so

$$0 < \mu(\{\xi \in \Omega : g(x, \xi) > \hat{\theta}\}) = \sum_{i \in \Psi} \mu(\{\xi \in \Omega_i : g(x, \xi) > \hat{\theta}\}) = 0$$

because  $\Psi$  is finite or countable. This is a contradiction as desired.  $\blacksquare$

## Appendix C Proof of Proposition 2

*Proof:* Since  $Q(x, \xi)$  is continuous in variable  $\xi$  on  $\Omega$  based on Assumption (A1), it follows from Lemma 1 that  $\text{ess sup}_{\xi \in \Omega} Q(x, \xi) = \sup_{\xi \in \Omega} Q(x, \xi)$ .

If  $d \geq 2$ , the conclusion follows immediately from Theorem 1.

If  $d \in [0, 2)$ , by Theorem 1, RTSP is equivalent to the following optimization problem

$$\min_{x \in X} c^\top x + (1 - d/2) \text{CVaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)] + (d/2) \sup_{\xi \in \Omega} Q(x, \xi).$$

The conclusion follows from the fact that

$$\text{CVaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)] = \min_{\zeta \in \mathbb{R}} \zeta + \frac{1}{1 - d/2} \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi,$$

which is due to Theorem 10 in [30]. This completes the proof.  $\blacksquare$

## Appendix D Proof of Proposition 4

*Proof:* First, we restate the ambiguous CVaR constraint as

$$\sup_{\mathbb{P} \in \mathcal{D}} \min_{\zeta \in \mathbb{R}} \left\{ \zeta + \frac{1}{1-\alpha} \int_{\mathbb{R}^K} (F(x, \xi) - \zeta)^+ f(\xi) d\xi \right\} \leq r_0 \quad (27)$$

based on Theorem 10 in [30]. For each  $\mathbb{P} \in \mathcal{D}$  and  $\zeta \in \mathbb{R}$ , we let  $J_{\mathbb{P}}(\zeta)$  represent the objective function of the max-min optimization problem embedded in constraint (27). Since  $F(x, \xi)$  satisfies Assumption (A1), there exist  $\underline{M}$  and  $\overline{M}$  such that  $\underline{M} \leq F(x, \xi) \leq \overline{M}$  for each  $x \in \mathbb{R}^{n_1}$  and  $\xi \in \Omega$ . It follows that, for each probability distribution  $\mathbb{P} \in \mathcal{D}$  and  $\zeta \leq \underline{M}$ ,

$$J_{\mathbb{P}}(\zeta) = - \left( \frac{\alpha}{1-\alpha} \right) \zeta + \frac{1}{1-\alpha} \mathbb{E}_{\mathbb{P}}[F(x, \xi)] \geq - \left( \frac{\alpha}{1-\alpha} \right) \underline{M} + \frac{1}{1-\alpha} \mathbb{E}_{\mathbb{P}}[F(x, \xi)] = J_{\mathbb{P}}(\underline{M});$$

and for each  $\zeta \geq \overline{M}$ ,

$$J_{\mathbb{P}}(\zeta) = \zeta \geq \overline{M} = J_{\mathbb{P}}(\overline{M}).$$

That is,  $J_{\mathbb{P}}(\zeta)$  is nonincreasing in  $\zeta$  in interval  $(-\infty, \underline{M}]$  and nondecreasing in  $\zeta$  in interval  $[\overline{M}, +\infty)$ . Hence, we can restrict the feasible region of  $\zeta$  in constraint (27) to be within interval  $[\underline{M}, \overline{M}]$ , and accordingly the ambiguous CVaR constraint is equivalent to

$$\sup_{\mathbb{P} \in \mathcal{D}} \min_{\zeta \in [\underline{M}, \overline{M}]} \left\{ \zeta + \frac{1}{1-\alpha} \int_{\mathbb{R}^K} (F(x, \xi) - \zeta)^+ f(\xi) d\xi \right\} \leq r_0. \quad (28)$$

By Sion's minimax theorem [37], we can exchange the order of operators sup and min in constraint (28) because  $J_{\mathbb{P}}(\zeta)$  is convex in  $\zeta$  and concave (actually linear) in  $\mathbb{P}$ , and  $\zeta$  has a compact feasible region. It follows that

$$\begin{aligned} & \sup_{\mathbb{P} \in \mathcal{D}} \min_{\zeta \in [\underline{M}, \overline{M}]} \zeta + \frac{1}{1-\alpha} \int_{\mathbb{R}^K} (F(x, \xi) - \zeta)^+ f(\xi) d\xi \\ &= \min_{\zeta \in [\underline{M}, \overline{M}]} \zeta + \frac{1}{1-\alpha} \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}} [F(x, \xi) - \zeta]^+ \end{aligned} \quad (29a)$$

$$= \min_{\zeta \in [\underline{M}, \overline{M}]} \zeta + \frac{1}{1-\alpha} \left\{ (1-d/2)^+ \text{CVaR}_{d/2}^{\mathbb{P}_0} [F(x, \xi) - \zeta]^+ + \min\{d/2, 1\} \sup_{\xi \in \Omega} (F(x, \xi) - \zeta)^+ \right\} \quad (29b)$$

$$= \min_{\zeta \in [\underline{M}, \overline{M}]} \zeta + \frac{(1-d/2)^+}{1-\alpha} \text{CVaR}_{d/2}^{\mathbb{P}_0} [F(x, \xi) - \zeta]^+ + \frac{\min\{d/2, 1\}}{1-\alpha} \left( \sup_{\xi \in \Omega} F(x, \xi) - \zeta \right)^+, \quad (29c)$$

where equality (29a) follows from the Sion's minimax theorem, equality (29b) follows Theorem 1, and equality (29c) follows the fact that

$$\sup_{\xi \in \Omega} (F(x, \xi) - \zeta)^+ = \left( \sup_{\xi \in \Omega} F(x, \xi) - \zeta \right)^+.$$

We discuss the following two cases:

**Case 1.** If  $d \in [0, 2)$ , then we can reformulate the ambiguous CVaR constraint based on (29c) as

$$\min_{\zeta \in [\underline{M}, \overline{M}]} \zeta + \frac{1-d/2}{1-\alpha} \left\{ \zeta_0 + \frac{1}{1-d/2} \mathbb{E}_{\mathbb{P}_0} [(F(x, \xi) - \zeta)^+ - \zeta_0]^+ \right\} + \frac{d/2}{1-\alpha} \left( \sup_{\xi \in \Omega} F(x, \xi) - \zeta \right)^+ \leq r_0.$$

Without loss of generality, we can replace the feasible region of  $\zeta$  in the embedded optimization problem by  $\mathbb{R}$  following a similar argument, as we did for function  $J_{\mathbb{P}}(\zeta)$ , to show that the objective function is nonincreasing in  $\zeta$  in the interval  $(-\infty, \underline{M}]$  and nondecreasing in  $\zeta$  in the interval  $[\overline{M}, +\infty)$ . The conclusion for this case follows by removing the constraint  $\zeta \in [\underline{M}, \overline{M}]$ .

**Case 2.** If  $d \geq 2$ , then we can reformulate the ambiguous CVaR constraint based on (29c) as

$$\min_{\zeta \in [\underline{M}, \overline{M}]} \zeta + \frac{1}{1-\alpha} \left( \sup_{\xi \in \Omega} F(x, \xi) - \zeta \right)^+ \leq r_0.$$

It is clear that an optimal solution to the embedded optimization problem is  $\zeta^* = \sup_{\xi \in \Omega} F(x, \xi)$  with the optimal objective value  $\sup_{\xi \in \Omega} F(x, \xi)$ . Therefore, the conclusion is proved for this case.  $\blacksquare$

## Appendix E Proof of Proposition 5

*Proof:* First, we define  $X_F(\xi) := \{x : F(x, \xi) \leq 0\}$  for a general function  $F(x, \xi)$ . We also define set  $\Delta(x) := \{\xi \in \mathbb{R}^K : F(x, \xi) > 0\}$  for any given  $x \in \mathbb{R}^{n_1}$ , and accordingly function  $g(x, \xi) = I_{\Delta(x)}(\xi)$  where indicator function  $I_{\Delta(x)}(\cdot)$  is defined as

$$I_{\Delta(x)}(\xi) = \begin{cases} 1, & \text{if } F(x, \xi) > 0, \\ 0, & \text{o.w.} \end{cases}$$

It follows that

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} = 1 - \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{F(x, \xi) > 0\} = 1 - \sup_{\mathbb{P} \in \mathcal{D}} \mathbb{E}_{\mathbb{P}}[g(x, \xi)] \quad (30)$$

$$= \begin{cases} 1 - (1-d/2)\text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)] - (d/2)\text{ess sup}_{\xi \in \Omega} g(x, \xi), & \text{if } d \in [0, 2) \end{cases} \quad (31a)$$

$$= \begin{cases} 1 - \text{ess sup}_{\xi \in \Omega} g(x, \xi), & \text{if } d \geq 2, \end{cases} \quad (31b)$$

where the last equality in (30) follows from the property of the indicator function, and equality (31) is due to Theorem 1.

Second, for each  $x \in \mathbb{R}^{n_1}$ , we define indicator  $\bar{I}(x) = \text{ess sup}_{\xi \in \Omega} g(x, \xi)$ . It follows that  $\bar{I}(x) = 1$  if  $\mu(\{\xi \in \Omega : F(x, \xi) > 0\}) > 0$  and  $\bar{I}(x) = 0$  otherwise.

Third, since  $g(x, \xi)$  can only be zero or one, by the definition of value-at-risk we have

$$\text{VaR} = \begin{cases} 1, & \text{if } \mathbb{P}_0\{F(x, \xi) \leq 0\} < d/2, \\ 0, & \text{o.w.} \end{cases}$$

It follows from the definition of conditional value-at-risk that

$$\text{CVaR}_{d/2}^{\mathbb{P}_0}[g(x, \xi)] = \text{VaR} + \frac{1}{1 - d/2} \int_{\mathbb{R}^K} (g(x, \xi) - \text{VaR})^+ f_0(\xi) d\xi$$

$$= \begin{cases} 1, & \text{if } \mathbb{P}_0\{F(x, \xi) \leq 0\} < d/2, \\ \frac{\mathbb{P}_0\{F(x, \xi) > 0\}}{1 - d/2}, & \text{o.w.} \end{cases} \quad (32a)$$

$$(32b)$$

Hence, in view of (31) and the definition of  $\bar{I}(x)$  we have

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} = \begin{cases} 0, & \text{if } d \in [0, 2) \text{ and } \mathbb{P}_0\{F(x, \xi) \leq 0\} < d/2, \\ \mathbb{P}_0\{F(x, \xi) \leq 0\} - (d/2)\bar{I}(x), & \text{if } d \in [0, 2) \text{ and } \mathbb{P}_0\{F(x, \xi) \leq 0\} \geq d/2, \\ 1 - \bar{I}(x), & \text{if } d \geq 2. \end{cases} \quad (33a)$$

$$(33b)$$

$$(33c)$$

Note here (33a) holds because  $\mathbb{P}_0\{F(x, \xi) \leq 0\} < d/2 < 1$  indicates that  $\mathbb{P}_0\{F(x, \xi) > 0\} > 0$  and so  $\mu(\{\xi \in \Omega : F(x, \xi) > 0\}) > 0$ , which implies  $\bar{I}(x) = 1$ . Thus, the conclusion holds based on (31a) and (32a). Finally, we discuss the following three cases:

**Case 1.** If  $d \geq 2$ , then we need to prove that constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha$  is equivalent to  $\text{ess sup}_{\xi \in \Omega} F(x, \xi) \leq 0$ . The necessity is clear. We prove the sufficiency by contradiction. Suppose that  $\text{ess sup}_{\xi \in \Omega} F(x, \xi) > 0$ , i.e.,  $\mu(\{\xi \in \Omega : F(x, \xi) > 0\}) > 0$ , then  $\bar{I}(x) = 1$  and so  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} = 0$  due to (33), which contradicts constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha$  because  $\alpha < 1$ .

**Case 2.** If  $d \in (2\alpha, 2)$ , then we need to prove that constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha$  is equivalent to  $\text{ess sup}_{\xi \in \Omega} F(x, \xi) \leq 0$ . The necessity is clear. We prove the sufficiency by contradiction. Suppose that  $\text{ess sup}_{\xi \in \Omega} F(x, \xi) > 0$ , i.e.,  $\mu(\{\xi \in \Omega : F(x, \xi) > 0\}) > 0$ , then  $\bar{I}(x) = 1$ . Since  $d < 2$ , only cases (33a) and (33b) can be taken. If (33a) is taken, then  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} = 0$  which contradicts constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha > 0$  because  $\alpha < 1$ . If (33b) is taken, then  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} = \mathbb{P}_0\{F(x, \xi) \leq 0\} - d/2 < 1 - \alpha$  due to the fact that  $d/2 > \alpha$ , which contradicts constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha$ .

**Case 3.** If  $d \in [0, 2\alpha]$ , then we need to prove that constraint  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha$  is equivalent to  $\mathbb{P}_0\{F(x, \xi) \leq 0\} \geq 1 - \alpha + d/2$ . To prove the necessity, suppose that  $\mathbb{P}_0\{F(x, \xi) \leq 0\} \geq 1 - \alpha + d/2$ . It follows that  $\mathbb{P}_0\{F(x, \xi) \leq 0\} \geq d/2$  because  $\alpha < 1$ . Based on (33b), we have

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} &= \mathbb{P}_0\{F(x, \xi) \leq 0\} - (d/2)\bar{I}(x) \\ &\geq 1 - \alpha + d/2 - (d/2)\bar{I}(x) \geq 1 - \alpha, \end{aligned} \quad (34)$$

where inequalities in (34) are due to (33b) and the definition of  $\bar{I}(x)$ . To prove the sufficiency, suppose that  $\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{x \in X_F(\xi)\} \geq 1 - \alpha$ . It follows from (33b) that  $\mathbb{P}_0\{F(x, \xi) \leq 0\} - (d/2)\bar{I}(x) \geq 1 - \alpha$ , which implies

$$\mathbb{P}_0\{F(x, \xi) \leq 0\} \geq 1 - \alpha + (d/2)\bar{I}(x).$$

If  $\bar{I}(x) = 1$ , then we have  $\mathbb{P}_0\{F(x, \xi) \leq 0\} \geq 1 - \alpha + d/2$  and the sufficiency is proved. Otherwise, if  $\bar{I}(x) = 0$ , then we have  $\mathbb{P}_0\{F(x, \xi) \leq 0\} = 1 \geq 1 - \alpha + d/2$  because  $d \leq 2\alpha$ , and so the sufficiency holds. ■

## Appendix F Proof of Proposition 6

*Proof:* The proofs for cases when  $d \in [0, 2(1 - F_Y(c_2))]$  and when  $d > 2(1 - F_Y(c_2))$  and  $\eta \in [c_1, F_Y^{-1}(1 - d/2)]$  immediately follow from Proposition 5 and are thus omitted. We focus on the case when  $d > 2(1 - F_Y(c_2))$  and  $\eta \in (F_Y^{-1}(1 - d/2), c_2]$ .

For each  $\eta \in (F_Y^{-1}(1 - d/2), c_2]$ , the ambiguous stochastic dominance constraint is equivalent to

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{F(x, \xi) - \eta \leq 0\} \geq 1 - (1 - F_Y(\eta)). \quad (35)$$

Since  $\eta > F_Y^{-1}(1 - d/2)$ , we have  $F_Y(\eta) > 1 - d/2$  and so  $d > 2[1 - F_Y(\eta)]$ . It follows that constraint (35) is equivalent to  $\text{ess sup}_{\xi \in \Omega} F(x, \xi) \leq \eta$  based on Proposition 5. Since this constraint holds for each  $\eta > F_Y^{-1}(1 - d/2)$ , we have  $\text{ess sup}_{\xi \in \Omega} F(x, \xi) \leq F_Y^{-1}(1 - d/2)$  and the proof is complete. ■

## Appendix G Proof of Proposition 7

*Proof:* The first part of the conclusion immediately follows from Theorem 1 and Lemma 1 because  $Q(x, \xi)$  is continuous in variable  $\xi$  on  $\Omega$ . We focus on the second part, i.e., obtaining  $x^*$ .

First, if  $d \geq 2$ , then  $F^{\text{nv}}(x) = cx + \sup_{\xi \in \Omega} Q(x, \xi)$ . Because

$$Q(x, \xi) = \begin{cases} -(p - s)\xi - sx & \text{if } \xi < x, \\ -px & \text{o.w.}, \end{cases} \quad (36)$$

$Q(x, \xi)$  is nonincreasing in  $\xi$  for any given  $x$ . It follows that  $\sup_{\xi \in \Omega} Q(x, \xi) = Q(x, \xi^L)$  and so

$$F^{\text{nv}}(x) = cx + Q(x, \xi^L) = (c - s)x - (p - s) \min\{x, \xi^L\}.$$

Hence,  $x^* = \xi^L$ .

Second, if  $d \in (0, 2)$ , then  $F^{\text{nv}}(x) = cx + (1 - d/2)\text{CVaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)] + (d/2)\sup_{\xi \in \Omega} Q(x, \xi)$ . Similarly, we have  $\sup_{\xi \in \Omega} Q(x, \xi) = Q(x, \xi^L) = -sx - (p - s) \min\{x, \xi^L\}$ . To derive  $\text{CVaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)]$ , we let  $\hat{\zeta}$  represent  $\text{VaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)]$ , i.e.,  $\hat{\zeta} = \inf_{\eta \in \mathbb{R}} \{\mathbb{P}_0\{Q(x, \xi) \leq \eta\} \geq d/2\}$ . We discuss the following two cases for the first derivative of  $F^{\text{nv}}(x)$  in  $x$ :



**Case 1.** If  $\mathbb{P}_0\{\xi \geq x\} \geq d/2$ , then  $\hat{\zeta} = -px$  by the definition of  $Q(x, \xi)$  in (36) and the fact that  $Q(x, \xi)$  is nonincreasing in  $\xi$ . It follows that

$$\text{CVaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)] = -px + \frac{1}{1-d/2} \int_{-\infty}^x (p-s)(x-\xi) d\mathbb{P}_0(\xi).$$

Hence, we have

$$F^{\text{nv}}(x) = \begin{cases} [c - (1 - \frac{d}{2})p - (\frac{d}{2})s]x + \int_{-\infty}^x (p-s)(x-\xi) d\mathbb{P}_0(\xi) - (\frac{d}{2})(p-s)\xi^L & \text{if } x \geq \xi^L, \\ (c-p)x + \int_{-\infty}^x (p-s)(x-\xi) d\mathbb{P}_0(\xi) & \text{if } x < \xi^L, \end{cases}$$

and so

$$\frac{dF^{\text{nv}}(x)}{dx} = \begin{cases} -(p-c) + (p-s)[\mathbb{P}_0\{\xi \leq x\} + d/2] & \text{if } x \geq \xi^L \\ -(p-c) + (p-s)\mathbb{P}_0\{\xi \leq x\} = -(p-c) & \text{if } x < \xi^L, \end{cases}$$

because  $\mathbb{P}_0\{\xi \leq x\} = 0$  for  $x < \xi^L$ .

**Case 2.** If  $\mathbb{P}_0\{\xi \geq x\} < d/2$ , then  $x > \xi^L$  because  $\mathbb{P}_0\{\xi \geq \xi^L\} = 1 > d/2$ , and so  $\sup_{\xi \in \Omega} Q(x, \xi) = -sx - (p-s)\xi^L$ . Also,  $\hat{\zeta} > -px$  by the definition of  $\hat{\zeta}$ . It follows that

$$\begin{aligned} \hat{\zeta} &= \inf_{\eta \in \mathbb{R}} \{ \mathbb{P}_0\{-(p-s)\xi - sx \leq \eta\} \geq d/2 \} \\ &= \inf_{\eta \in \mathbb{R}} \left\{ \mathbb{P}_0 \left\{ \xi \geq -\frac{sx + \eta}{p-s} \right\} \geq d/2 \right\} \\ &= \inf_{\eta \in \mathbb{R}} \left\{ \mathbb{P}_0 \left\{ \xi \leq -\frac{sx + \eta}{p-s} \right\} \leq 1 - d/2 \right\} \end{aligned} \quad (37a)$$

$$= -(p-s)F_0^{-1}(1-d/2) - sx, \quad (37b)$$

where equality (37a) follows from the absolute continuity of  $\mathbb{P}_0$  and equality (37b) follows from the definition of  $F_0^{-1}(\cdot)$ . Hence, we have

$$\text{CVaR}_{d/2}^{\mathbb{P}_0}[Q(x, \xi)] = -(p-s)F_0^{-1}(1-d/2) - sx + \frac{1}{1-d/2} \int_{\mathbb{R}} (p-s) [F_0^{-1}(1-d/2) - \xi]^+ d\mathbb{P}_0(\xi),$$

and so

$$\begin{aligned} F^{\text{nv}}(x) &= (c-s)x - (1-d/2)(p-s)F_0^{-1}(1-d/2) - (d/2)(p-s)\xi^L + \\ &\quad \int_{\mathbb{R}} (p-s) [F_0^{-1}(1-d/2) - \xi]^+ d\mathbb{P}_0(\xi). \end{aligned}$$

Hence,  $dF^{\text{nv}}(x)/dx = c-s$ .

Based on Cases 1 and 2, we have  $dF^{\text{nv}}(x)/dx \leq 0$  in the interval  $[0, \xi^L)$  and  $dF^{\text{nv}}(x)/dx \geq 0$  if  $\mathbb{P}_0\{\xi \geq x\} < d/2$ . It follows that we can only consider  $x^* \in [\xi^L, F_0^{-1}(1-d/2)]$  (i.e.,  $x^* \geq \xi^L$  and  $\mathbb{P}_0\{\xi \geq x^*\} \geq d/2$ ) because  $F^{\text{nv}}(x)$  is continuous on  $\mathbb{R}_+$ . Based on the conclusion of Case 1, we can discuss the following two cases:

**Case (a)** If  $(p-c)/(p-s) < d/2$ , then  $dF^{\text{nv}}(x)/dx = (p-s)(d/2 - (p-c)/(p-s) + \mathbb{P}_0\{\xi \leq x\}) \geq 0$  in the interval  $[\xi^L, F_0^{-1}(1-d/2)]$ . It follows that  $x^* = \xi^L$  is optimal.

**Case (b)** If  $(p - c)/(p - s) \geq d/2$ , then by forcing  $dF^{\text{nv}}(x)/dx \leq 0$  we have  $\mathbb{P}_0\{\xi \leq x\} \leq (p - c)/(p - s) - (d/2)$ . It follows that  $x^* = F_0^{-1}((p - c)/(p - s) - (d/2))$  is optimal. Note that  $x^* \geq \xi^{\text{L}}$  because  $F_0(\xi^{\text{L}}) = 0$  and  $x^* \leq F_0^{-1}(1 - d/2)$  because  $F_0^{-1}(\cdot)$  is nondecreasing. ■

## Appendix H Proof of Lemma 2

*Proof:* We prove the continuity result for the case  $\mathbb{E}_{\mathbb{P}_0}[Q(x, \xi) - \zeta]^+$ . The case for  $\mathbb{P}$  can be similarly proved and is thus omitted for brevity.

Consider any point  $(\bar{x}, \bar{\zeta}) \in \mathbb{R}^{n_1+1}$  and any sequence  $\{(x_i, \zeta_i)\}_{i \in \mathbb{N}}$  in  $\mathbb{R}^{n_1+1}$  converging to  $(\bar{x}, \bar{\zeta})$ . First, for each  $\xi \in \Omega$ , since (i)  $Q(x, \xi)$  is convex in  $x$  on  $\mathbb{R}^{n_1}$ , (ii)  $Q(x, \xi) < +\infty$  for each  $x \in \mathbb{R}^{n_1}$  by Assumption (A1), and (iii) function  $g(x) := (x)^+$  is convex and nondecreasing, we have  $(Q(x, \xi) - \zeta)^+$  is convex and continuous in  $(x, \zeta)$  on  $\mathbb{R}^{n_1+1}$ , i.e.,

$$\lim_{i \rightarrow \infty} (Q(x_i, \xi) - \zeta_i)^+ = (Q(\bar{x}, \xi) - \bar{\zeta})^+, \quad \forall \xi \in \Omega. \quad (38)$$

It follows that there exists an  $\bar{M} \in \mathbb{R}_+$  such that  $(Q(x_i, \xi) - \zeta_i)^+ \leq \bar{M}$  for each  $i \in \mathbb{N}$  and each  $\xi \in \Omega$  because  $\sup_{\xi \in \Omega} (Q(\bar{x}, \xi) - \bar{\zeta})^+ < +\infty$  by Assumption (A1). Then, we have

$$\lim_{i \rightarrow \infty} \mathbb{E}_{\mathbb{P}_0}[Q(x_i, \xi) - \zeta_i]^+ = \mathbb{E}_{\mathbb{P}_0} \left[ \lim_{i \rightarrow \infty} (Q(x_i, \xi) - \zeta_i)^+ \right] \quad (39a)$$

$$= \mathbb{E}_{\mathbb{P}_0}[Q(\bar{x}, \xi) - \bar{\zeta}]^+, \quad (39b)$$

where equality (39a) follows from the dominated convergence theorem, and equality (39b) follows from equation (38). Therefore, the proof is complete. ■

## Appendix I Proof of Lemma 3

*Proof:* First, since  $Q(x, \xi)$  is convex in  $x$  on  $\mathbb{R}^{n_1}$  for each  $\xi \in \Omega$ , worst-case value function  $\sup_{\xi \in \Omega} Q(x, \xi)$  is also convex in  $x$  on  $\mathbb{R}^{n_1}$  because the maximization operator preserves convexity.

Second, since  $\sup_{\xi \in \Omega} Q(x, \xi)$  is convex in  $x$  on  $\mathbb{R}^{n_1}$  and  $\sup_{\xi \in \Omega} Q(x, \xi) < +\infty$  for each  $x \in \mathbb{R}^{n_1}$  by Assumption (A1), continuity of  $\sup_{\xi \in \Omega} Q(x, \xi)$  follows from convexity. ■

## Appendix J Proof of Lemma 4

*Proof:* We prove the case for TSP with regard to  $\mathbb{P}_0$ . The case for TSP with regard to  $\mathbb{P}$  can be similarly proved and is thus omitted for brevity.

First, let  $\zeta^* = \min_{x \in X, \xi \in \Omega} Q(x, \xi)$ . Since  $Q(x, \xi) > -\infty$  for any  $x \in X$  and  $\xi \in \Omega$  by Assumption (A1),  $\zeta^* > -\infty$ . Then

$$\min_{x \in X, \zeta \in \mathbb{R}} c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi \quad (40a)$$

$$\leq \min_{x \in X} c^\top x + \zeta^* + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta^*)^+ f_0(\xi) d\xi \quad (40b)$$

$$= \min_{x \in X} c^\top x + \zeta^* + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta^*) f_0(\xi) d\xi \quad (40c)$$

$$= \min_{x \in X} c^\top x + \int_{\mathbb{R}^K} Q(x, \xi) f_0(\xi) d\xi,$$

where inequality (40b) holds because  $\zeta = \zeta^*$  is only a feasible solution for (40a), and equality (40c) holds because  $\zeta^* \leq Q(x, \xi)$  for each  $x \in X$  and  $\xi \in \Omega$  based on the definition of  $\zeta^*$ .

Second, in view that  $(Q(x, \xi) - \zeta)^+ \geq Q(x, \xi) - \zeta$ , we have

$$\min_{x \in X, \zeta \in \mathbb{R}} c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi$$

$$\geq \min_{x \in X, \zeta \in \mathbb{R}} c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta) f_0(\xi) d\xi$$

$$= \min_{x \in X} c^\top x + \int_{\mathbb{R}^K} Q(x, \xi) f_0(\xi) d\xi,$$

which completes the proof. ■

## Appendix K Proof of Theorem 2

*Proof:* We use  $N \rightarrow \infty$  and  $d \rightarrow 0$  interchangeably throughout the proof because  $d = d(N) \rightarrow 0$  as  $N \rightarrow \infty$ .

First, we can assume that  $0 \leq d \leq 2$  without loss of generality because we analyze the case when  $d$  goes to zero. We introduce additional notation for presentation brevity. We let  $\hat{h}(x, \zeta, d)$  represent the objective function of the RTSP reformulation with regard to distance tolerance  $d$ , i.e.,

$$\hat{h}(x, \zeta, d) := c^\top x + (1 - d/2)\zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi + (d/2) \sup_{\xi \in \Omega} Q(x, \xi),$$

$h_0(x, \zeta)$  represent the objective function of TSP with regard to  $\mathbb{P}_0$  by Lemma 4, i.e.,

$$h_0(x, \zeta) := c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi,$$

and  $h(x, \zeta)$  represent the objective function of TSP with regard to  $\mathbb{P}$ , i.e.,

$$h(x, \zeta) := c^\top x + \zeta + \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f(\xi) d\xi.$$

Second, since  $\mathbb{E}_{\mathbb{P}_0}[Q(x, \xi) - \zeta]^+$  is continuous in  $(x, \zeta)$  on  $\mathbb{R}^{n_1+1}$  by Lemma 2,  $h_0(x, \zeta)$  is continuous in  $(x, \zeta)$  on  $X \times \mathbb{R}$ . Then for any  $\epsilon > 0$ , there exists an open neighborhood  $U_1(\bar{x}, \bar{\zeta})$  for each  $(\bar{x}, \bar{\zeta}) \in X \times \mathbb{R}$  such that  $U_1(\bar{x}, \bar{\zeta}) \subseteq X \times \mathbb{R}$  and

$$\sup_{(x, \zeta) \in U_1(\bar{x}, \bar{\zeta})} |h_0(x, \zeta) - h_0(\bar{x}, \bar{\zeta})| \leq \epsilon/3. \quad (41)$$

Similarly, by the continuity of  $h(x, \zeta)$ , there exists an open neighborhood  $U_2(\bar{x}, \bar{\zeta})$  for each  $(\bar{x}, \bar{\zeta}) \in X \times \mathbb{R}$  such that  $U_2(\bar{x}, \bar{\zeta}) \subseteq X \times \mathbb{R}$  and

$$\sup_{(x, \zeta) \in U_2(\bar{x}, \bar{\zeta})} |h(x, \zeta) - h(\bar{x}, \bar{\zeta})| \leq \epsilon/3. \quad (42)$$

In addition, since  $\sup_{\xi \in \Omega} Q(x, \xi)$  is continuous in  $x$  on  $X$  by Lemma 3, plus  $\mathbb{E}_{\mathbb{P}_0}[Q(x, \xi) - \zeta]^+$  is continuous in  $(x, \zeta)$  on  $\mathbb{R}^{n_1+1}$ ,  $\hat{h}(x, \zeta, 2)$  is continuous in  $(x, \zeta)$  on  $X \times \mathbb{R}$  when  $d$  is fixed at 2. Then for any  $\epsilon > 0$ , there exists an open neighborhood  $U_3(\bar{x}, \bar{\zeta})$  for each  $(\bar{x}, \bar{\zeta}) \in X \times \mathbb{R}$ , such that  $U_3(\bar{x}, \bar{\zeta}) \subseteq X \times \mathbb{R}$  and

$$\sup_{(x, \zeta) \in U_3(\bar{x}, \bar{\zeta})} |\hat{h}(x, \zeta, 2) - \hat{h}(\bar{x}, \bar{\zeta}, 2)| \leq \epsilon/3. \quad (43)$$

Furthermore, we define

$$U(\bar{x}, \bar{\zeta}) = U_1(\bar{x}, \bar{\zeta}) \cap U_2(\bar{x}, \bar{\zeta}) \cap U_3(\bar{x}, \bar{\zeta}) \quad (44)$$

for each  $(\bar{x}, \bar{\zeta}) \in X \times \mathbb{R}$ . Since  $\hat{h}(x, \zeta, d) = (1 - d/2)h_0(x, \zeta) + (d/2)\hat{h}(x, \zeta, 2)$ , we have  $\hat{h}(x, \zeta, d) \leq \max\{h_0(x, \zeta), \hat{h}(x, \zeta, 2)\}$  for each  $d \in [0, 2]$  and so

$$\begin{aligned} & \sup_{(x, \zeta) \in U(\bar{x}, \bar{\zeta})} |\hat{h}(x, \zeta, d) - \hat{h}(\bar{x}, \bar{\zeta}, d)| \\ & \leq \max \left\{ \sup_{(x, \zeta) \in U(\bar{x}, \bar{\zeta})} |h_0(x, \zeta) - h_0(\bar{x}, \bar{\zeta})|, \sup_{(x, \zeta) \in U(\bar{x}, \bar{\zeta})} |\hat{h}(x, \zeta, 2) - \hat{h}(\bar{x}, \bar{\zeta}, 2)| \right\} \\ & \leq \max\{\epsilon/3, \epsilon/3\} = \epsilon/3, \end{aligned} \quad (45)$$

where inequality (45) follows from inequalities (41), (43), and the definition of  $U(\bar{x}, \bar{\zeta})$  in (44).

Third, since  $X$  is compact,  $\bar{M} := \sup_{x \in X, \xi \in \Omega} Q(x, \xi) < +\infty$  and  $\underline{M} := \min_{x \in X, \xi \in \Omega} Q(x, \xi) > -\infty$  by Assumption (A1). Hence, we can restrict that  $\zeta \in [\underline{M}, \bar{M}]$  for both TSP and RTSP without loss of generality. Then  $\sup_{x \in X, \xi \in \Omega} (Q(x, \xi) - \zeta)^+ < +\infty$ . Also, since  $f_0$  converges to  $f$  in  $L^1$ -norm by Assumption (ii), we have

$$\begin{aligned} \left| \int_{\xi \leq \bar{\xi}} (f(\xi) - f_0(\xi)) d\xi \right| & \leq \int_{\xi \leq \bar{\xi}} |f(\xi) - f_0(\xi)| d\xi \\ & \leq \int_{\mathbb{R}^K} |f(\xi) - f_0(\xi)| d\xi \end{aligned}$$

for each  $\bar{\xi} \in \mathbb{R}^K$ , and so  $F_0(\bar{\xi}) := \int_{\xi \leq \bar{\xi}} f_0(\xi) d\xi$  converges to  $F(\bar{\xi}) := \int_{\xi \leq \bar{\xi}} f(\xi) d\xi$  for each  $\bar{\xi} \in \mathbb{R}^K$ . Hence, we have

$$\lim_{N \rightarrow \infty} \int_{\mathbb{R}^K} (Q(\bar{x}, \xi) - \bar{\zeta})^+ f_0(\xi) d\xi = \int_{\mathbb{R}^K} (Q(\bar{x}, \xi) - \bar{\zeta})^+ f(\xi) d\xi$$

based on the Helly-Bray Theorem [9]. Then

$$\lim_{d \rightarrow 0} \hat{h}(\bar{x}, \bar{\zeta}, d) = h(\bar{x}, \bar{\zeta}) \quad (46)$$

for each  $(\bar{x}, \bar{\zeta}) \in X \times [\underline{M}, \overline{M}]$  by the definition of  $\hat{h}$  and  $h$ . Thus, there exists a  $\bar{d}(\bar{x}, \bar{\zeta}) > 0$  such that

$$|\hat{h}(\bar{x}, \bar{\zeta}, d) - h(\bar{x}, \bar{\zeta})| \leq \epsilon/3, \quad \forall d \in [0, \bar{d}(\bar{x}, \bar{\zeta})]. \quad (47)$$

Fourth, since  $X \times [\underline{M}, \overline{M}]$  is compact, there exists a finite set of points  $\{(\bar{x}_m, \bar{\zeta}_m)\}_{m=1}^M$  such that

$$X \times [\underline{M}, \overline{M}] \subseteq \bigcup_{m=1}^M U(\bar{x}_m, \bar{\zeta}_m), \quad (48)$$

i.e., for each  $(x, \zeta) \in X \times [\underline{M}, \overline{M}]$ , there exists an  $m' \in \{1, \dots, M\}$  such that  $(x, \zeta) \in U(\bar{x}_{m'}, \bar{\zeta}_{m'})$ . Note here that  $m'$  depends on  $(x, \zeta)$  and we mark the dependency by a prime for notation brevity.

Therefore, by choosing  $\bar{d} = \min_{m=1, \dots, M} \bar{d}(\bar{x}_m, \bar{\zeta}_m) > 0$  (noting here that selecting a positive  $\bar{d}$  is guaranteed by the finite coverage (48)), we have

$$\begin{aligned} & |\hat{h}(x, \zeta, d) - h(x, \zeta)| \\ & \leq |\hat{h}(x, \zeta, d) - \hat{h}(\bar{x}_{m'}, \bar{\zeta}_{m'}, d)| + |\hat{h}(\bar{x}_{m'}, \bar{\zeta}_{m'}, d) - h(\bar{x}_{m'}, \bar{\zeta}_{m'})| + |h(\bar{x}_{m'}, \bar{\zeta}_{m'}) - h(x, \zeta)| \end{aligned} \quad (49a)$$

$$\leq \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon \quad (49b)$$

for each  $(x, \zeta) \in X \times [\underline{M}, \overline{M}]$  and each  $d \in [0, \bar{d}]$ , where inequality (49a) follows from the triangle inequality (used twice) and coverage (48), and inequality (49b) follows from (45), (47), (44), (42), and the definition of  $\bar{d}$ . Thus  $\hat{h}(x, \zeta, d)$  uniformly converges to  $h(x, \zeta)$  on  $X \times [\underline{M}, \overline{M}]$  as  $d$  goes to zero.

Fifth, we let  $(x^*(d), \zeta^*(d))$  represent an optimal solution to RTSP, and  $(x^*(0), \zeta^*(0))$  represent an optimal solution to TSP. Then

$$z(d) - z(0) \leq \hat{h}(x^*(0), \zeta^*(0), d) - h(x^*(0), \zeta^*(0)) \leq \epsilon$$

and

$$z(0) - z(d) \leq h(x^*(d), \zeta^*(d)) - \hat{h}(x^*(d), \zeta^*(d), d) \leq \epsilon,$$

for each  $d \in [0, \bar{d}]$ , which follows from the uniform convergence property established above, and the sub-optimality of  $(x^*(0), \zeta^*(0))$  for RTSP and  $(x^*(d), \zeta^*(d))$  for TSP, respectively. Hence,

$$|z(d) - z(0)| \leq \max \left\{ \hat{h}(x^*(0), \zeta^*(0), d) - h(x^*(0), \zeta^*(0)), h(x^*(d), \zeta^*(d)) - \hat{h}(x^*(d), \zeta^*(d), d) \right\} \leq \epsilon$$

for each  $d \in [0, \bar{d}]$ . Therefore, the convergence property of  $z(d)$  to  $z(0)$  is proved.

Finally, we prove the convergence property of  $U(d)$  to  $U(0)$  by contradiction. Suppose that  $\sup_{(x,\zeta) \in U(d)} \text{dist}((x, \zeta), U(0))$  does not converge to zero as  $d \rightarrow 0$ , then there exists a positive number  $\epsilon_0$ , a sequence of positive numbers  $\{d_n\}_{n \in \mathbb{N}}$  converging to zero, and a sequence  $\{(x_n, \zeta_n)\}_{n \in \mathbb{N}}$  in  $X \times [\underline{M}, \overline{M}]$  with  $(x_n, \zeta_n) \in U(d_n)$ , such that  $\text{dist}((x_n, \zeta_n), U(0)) > \epsilon_0$  for each  $n \in \mathbb{N}$ . Since  $X \times [\underline{M}, \overline{M}]$  is compact, there exists a  $(\bar{x}_1, \bar{\zeta}_1) \in X \times [\underline{M}, \overline{M}]$  and an  $\mathbb{N}_1 \subseteq \mathbb{N}$  such that subsequence  $\{(x_{n_1}, \zeta_{n_1})\}_{n_1 \in \mathbb{N}_1}$  converges to  $(\bar{x}_1, \bar{\zeta}_1)$  by the Bolzano-Weierstrass theorem. Then

$$\lim_{n_1 \rightarrow \infty} \hat{h}(x_{n_1}, \zeta_{n_1}, d(n_1)) = \lim_{n_1 \rightarrow \infty, d \rightarrow 0} \hat{h}(x_{n_1}, \zeta_{n_1}, d) \quad (50a)$$

$$= \lim_{d \rightarrow 0} \hat{h}(\bar{x}_1, \bar{\zeta}_1, d) \quad (50b)$$

$$= h(\bar{x}_1, \bar{\zeta}_1), \quad (50c)$$

where equality (50a) holds because  $d \rightarrow 0$  as  $N \rightarrow \infty$ , equality (50b) follows the continuity of  $\hat{h}$  in  $(x, \zeta)$ , and equality (50c) follows the convergence property (46). By the fact that  $(x_n, \zeta_n) \in U(d_n)$  and the convergence property  $\lim_{d \rightarrow 0} z(d) = z(0)$ , we have  $\lim_{n_1 \rightarrow \infty} \hat{h}(x_{n_1}, \zeta_{n_1}, d(n_1)) = z(0)$  and so

$$h(\bar{x}_1, \bar{\zeta}_1) = z(0). \quad (51)$$

Thus,  $(\bar{x}_1, \bar{\zeta}_1) \in U(0)$  and so  $\text{dist}((\bar{x}_1, \bar{\zeta}_1), U(0)) = 0$ . However, since  $\{(x_{n_1}, \zeta_{n_1})\}_{n_1 \in \mathbb{N}_1}$  converges to  $(\bar{x}_1, \bar{\zeta}_1)$  and  $\text{dist}((x_{n_1}, \zeta_{n_1}), U(0)) > \epsilon_0$  for each  $n_1 \in \mathbb{N}_1$ , we have  $\text{dist}((\bar{x}_1, \bar{\zeta}_1), U(0)) \geq \epsilon_0 > 0$ , which is a contradiction. This completes the proof.  $\blacksquare$

## Appendix L Proof of Theorem 4

*Proof:* Without loss of generality, we assume  $S_\Omega = S$  in this proof. Similar to the proof for Theorem 3, we let  $\hat{h}(x, \zeta, d)$  represent the objective function of RTSP (17a) and  $\hat{h}_S(x, \zeta, d)$  represent the objective function of SAA problem (20). We prove that the pointwise convergence holds between  $\hat{h}_S(x, \zeta, d)$  and  $\hat{h}(x, \zeta, d)$ , i.e., for each  $(x, \zeta) \in X \times \mathbb{R}$ ,

$$\lim_{S \rightarrow \infty} \hat{h}_S(x, \zeta, d) = \hat{h}(x, \zeta, d)$$

with probability 1. The conclusion then follows from Theorems 5.3 and 7.48 in [36].

For each  $(x, \zeta) \in X \times \mathbb{R}$ , it is clear that

$$\lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S (Q(x, \xi^s) - \zeta)^+ = \int_{\mathbb{R}^K} (Q(x, \xi) - \zeta)^+ f_0(\xi) d\xi$$

by the strong law of large numbers. It remains to show that the sample maximum converges to the worst-case cost, i.e.,

$$\lim_{S \rightarrow \infty} \max_{s=1, \dots, S} Q(x, \hat{\xi}^s) = \sup_{\xi \in \Omega} Q(x, \xi). \quad (52)$$

Indeed, by condition (ii) we have that for any  $\epsilon > 0$ ,

$$\begin{aligned} & \mathbb{P}_\Omega \left\{ \max_{s=1, \dots, S} Q(x, \hat{\xi}^s) > \sup_{\xi \in \Omega} Q(x, \xi) - \epsilon \right\} \\ &= 1 - \mathbb{P}_\Omega \left\{ \max_{s=1, \dots, S} Q(x, \hat{\xi}^s) \leq \sup_{\xi \in \Omega} Q(x, \xi) - \epsilon \right\} \\ &= 1 - \prod_{s=1, \dots, S} \mathbb{P}_\Omega \left\{ Q(x, \hat{\xi}^s) \leq \sup_{\xi \in \Omega} Q(x, \xi) - \epsilon \right\} \\ &= 1 - (1 - \delta)^S, \end{aligned}$$

where  $\delta := \mathbb{P}_\Omega \{ Q(x, \hat{\xi}^s) > \sup_{\xi \in \Omega} Q(x, \xi) - \epsilon \} > 0$ . Since  $1 \geq \delta > 0$ , we have

$$\sum_{S=1}^{\infty} \mathbb{P}_\Omega \left\{ \left| \max_{s=1, \dots, S} Q(x, \hat{\xi}^s) - \sup_{\xi \in \Omega} Q(x, \xi) \right| \geq \epsilon \right\} = \sum_{S=1}^{\infty} (1 - \delta)^S < \infty.$$

Therefore, the convergence (52) follows from the Borel-Cantelli Lemma, which completes the proof.

■