

# Data-driven Chance Constrained Stochastic Program\*

Ruiwei Jiang and Yongpei Guan

Department of Industrial and Systems Engineering

University of Florida, Gainesville, FL 32611, USA

Email: rwjiang@ufl.edu; guan@ise.ufl.edu

Submitted on April 11, 2013

## Abstract

Chance constrained programming is an effective and convenient approach to control risk in decision making under uncertainty. However, due to unknown probability distributions of random parameters, the solution obtained from a chance constrained optimization problem can be biased. In addition, instead of knowing the true distributions of random parameters, in practice, only a series of historical data, which can be considered as samples taken from the true (while ambiguous) distribution, can be observed and stored. In this paper, we derive stochastic programs with data-driven chance constraints (DCCs) to tackle these problems and develop equivalent reformulations. For a given historical data set, we construct two types of confidence sets for the ambiguous distribution through nonparametric statistical estimation of its moments and density functions, depending on the amount of available data. We then formulate DCCs from the perspective of robust feasibility, by allowing the ambiguous distribution to run adversely within its confidence set. After deriving equivalent reformulations, we provide exact and approximate solution approaches for stochastic programs with DCCs under both moment-based and density-based confidence sets. In addition, we derive the relationship between the conservatism of DCCs and the sample size of historical data, which shows quantitatively what we call the value of data.

*Key words:* stochastic optimization; chance constraints; semi-infinite programming; S-Lemma

---

\*This paper is a generalized version of a paper under the same title (available online at [www.optimization-online.org/DB\\_HTML/2012/07/3525.html](http://www.optimization-online.org/DB_HTML/2012/07/3525.html)). This paper has been presented in the Simulation Optimization Workshop, Viña del Mar, Chile, March 21-23, 2013 and the 13<sup>th</sup> International Conference on Stochastic Programming, Bergamo, Italy, July 8-12, 2013. This paper has also been awarded the 2013 Stochastic Programming Society best student paper prize. This paper has been submitted for publication on April 11, 2013.

# 1 Introduction

## 1.1 Motivation and Literature Review

To assist decision making in uncertain environments, significant research progress has been made in stochastic optimization formulations and their solution approaches. One effective and convenient way of handling uncertainty arising in constraint parameters employs chance constraints. In a chance constrained optimization problem, decision makers are interested in satisfying a constraint, which is subject to uncertainty, by at least a pre-specified probability at the smallest cost,

$$\min_x \psi(x) \tag{1a}$$

$$\text{s.t. } \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha, \tag{1b}$$

$$x \in X, \tag{1c}$$

where  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  often represents a convex cost function,  $X$  represents a computable bounded convex set (e.g., a polytope) in  $\mathbb{R}^n$ ,  $\xi \in \mathbb{R}^K$  represents a  $K$ -dimensional random vector defined on a probability space  $(\Omega, \mathcal{F}, \text{Pr})$ , and the set function  $\mathbb{P}\{\cdot\}$  represents the probability distribution on  $\mathbb{R}^K$  induced by  $\xi$ , i.e.,  $\mathbb{P}\{C\} = \text{Pr}\{\xi^{-1}(C)\}$ ,  $\forall C \in \mathcal{B}(\mathbb{R}^K)$ . In addition, the linear inequality system  $A(\xi)x \leq b(\xi)$  represents the constraints to be satisfied, where  $A(\xi) \in \mathbb{R}^{m \times n}$  and  $b(\xi) \in \mathbb{R}^m$  denote the technology matrix and right-hand side subject to uncertainty, and  $x \in \mathbb{R}^n$  denotes the decision variable. Constraint (1b) is called a *single* chance constraint when  $m = 1$  (i.e., the matrix  $A(\xi)$  reduces to a row vector), and otherwise it is called a *joint* chance constraint. In addition,  $\alpha$  value represents the risk level (or tolerance of constraint violation) allowed by the decision makers, and usually  $\alpha$  is chosen to be small, e.g., 0.10 or 0.05.

Chance constraints emerge naturally as a modeling tool in various decision making circumstances. For example, decision makers in the finance industry may attempt to ensure that the return of their portfolio meets a target value with high probability. The study of chance constrained optimization problems has a long history dating back to Charnes et al. [9], Miller and Wagner [22], and Prékopa [30]. Unfortunately, constraint (1b) remains challenging to handle because of two key difficulties: (i) chance constraints are non-convex in general, and (ii) the probability associated with the chance constraints can be hard to compute since it requires a multi-dimensional integral. To address the first difficulty and recapture convexity, previous research identifies im-

portant cases under which chance constraints are nonlinear but convex (see, e.g., Charnes and Cooper [8], Prékopa [31], and Calafiore and El Ghaoui [7]), and provides conservative convex approximations (see, e.g., Pintér [28], Nemirovski and Shapiro [24], Rockafellar and Uryasev [32], and Chen et al. [10]). To address the second difficulty, previous research proposes scenario approximation approaches (see, e.g., Calafiore and Campi [5, 6], Nemirovski and Shapiro [23], Luedtke and Ahmed [20], and Pagnoncelli et al. [25]), which are computationally tractable and can guarantee to obtain a solution satisfying a chance constraint with high probability. In addition, integer programming (IP) techniques are successfully applied in exactly solving chance-constrained problems (see, e.g., Luedtke et al. [21], Küçükyavuz [17], Luedtke [19], and Lejeune [18]). An alternative of the chance constraint approach is the robust optimization (RO) approach (see, e.g., Soyster [37], Ben-Tal and Nemirovski [2], Bertsimas and Sim [3], Calafiore [4] and Ben-Tal et al. [1]), which requires the constraint  $A(\xi)x \leq b(\xi)$  to be satisfied for each  $\xi$  in a pre-defined uncertainty set  $U \subseteq \mathbb{R}^K$ , i.e.,

$$\sup_{\xi \in U} \{A(\xi)x - b(\xi)\} \leq 0, \quad (2)$$

where the operator  $\sup_{\xi \in U} \{\cdot\}$  is considered constraint-wise without loss of generality. One important merit of the RO approach is that, by a priori adjusting the uncertainty set  $U$ , one can ensure that the constraint  $\mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha_0$  for a pre-specified risk level  $\alpha_0$  is satisfied under any probability distribution  $\mathbb{P}$ . Hence, the RO approach can be viewed as a conservative approximation of chance constraints.

A basic, and perhaps the most challenging, question on chance constraints is the accessibility of the probability distribution  $\mathbb{P}$ . Most literature on chance constraints listed above assumes the decision makers have perfect knowledge of  $\mathbb{P}$ . In practice, however, it might be unrealistic to make such an assumption. Normally, decision makers have only a series of historical data points, which can be considered as samples taken from the true (while ambiguous) distribution. Based on the given data set, there are two potential issues for the classical chance constrained model (1): (i) it might be challenging to assume a specific probability distribution and to generate a large number of scenarios accordingly in the scenario approximations, and (ii) the solution might be sensitive to the ambiguous probability distribution and thus questionable in practice. To address these drawbacks, distributionally robust (or ambiguous) chance constrained (DRCC) models are proposed (see, e.g.,

Erdoğan and Iyengar [15], Calafiore and El Ghaoui [7], Nemirovski and Shapiro [24], Vandenberghe et al. [38], Zymler et al. [41], Xu et al. [39], and El Ghaoui et al. [14]). In DRCC models,  $\mathbb{P}$  is assumed to belong to a pre-defined confidence set  $\mathcal{D}$  with given first and second moment values (rather than being known with certainty), and the chance constraints are required to be satisfied under each probability distribution in  $\mathcal{D}$ :

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{A(\xi)x \leq b(\xi)\} \geq 1 - \alpha. \quad (3)$$

Most current literature proposes approximation approaches to solve DRCC, and to the best of our knowledge, only Vandenberghe et al. [38] and Zymler et al. [41] target deriving exact equivalent reformulations for the problem with the former providing an equivalent semidefinite programming reformulation and the latter providing a conditional value-at-risk (CVaR) approximation which is tight for the single chance constraint case. In this paper, we focus on data-driven approaches to construct the confidence set  $\mathcal{D}$ . Accordingly, we refer the constraints in the form of (3) data-driven chance constraints (DCCs). In our approach, we attempt to construct the confidence set  $\mathcal{D}$  based *only* on the historical data sampled from the true probability distribution and the statistical inferences obtained from the data. Intuitively, since real data are involved in its estimation,  $\mathcal{D}$  can get tighter around the true probability distribution  $\mathbb{P}$  with more data samples, and accordingly the DCC can become less conservative. In this paper, we propose *exact* approaches to handle both single and joint DCCs under different types of confidence sets by deriving their *equivalent* reformulations. Furthermore, we show the relationship between the conservatism of DCCs and the sample size of historical data, which depicts quantitatively the value of data.

## 1.2 Model Settings and Confidence Sets

In uncertain environments, people utilize historical data in various ways to help describe random parameters through statistical inference. For example, decision makers in the finance industry commonly describe the uncertainty in rate of return (RoR) of the investments by their mean and covariance matrix statistically inferred by the historical data. Hence, a confidence set  $\mathcal{D}$  can be naturally defined as all probability distributions whose first and second moments agree with the inference. For this case, we can denote the confidence set  $\mathcal{D}$  as  $\mathcal{D}_1$  shown as follows (cf. Delage and

Ye [12]):

$$\mathcal{D}_1 = \{\mathbb{P} \in \mathcal{M}_+ : (\mathbb{E}[\xi] - \mu)^\top \Lambda^{-1} (\mathbb{E}[\xi] - \mu) \leq \gamma_1, \mathbb{E}[(\xi - \mu)(\xi - \mu)^\top] \preceq \gamma_2 \Lambda\}, \quad (4)$$

where  $\mathcal{M}_+$  represents the set of all probability distributions,  $\mu$  and  $\Lambda \succ 0$  represent the inferred mean and covariance matrix respectively, and  $\gamma_1 > 0$  and  $\gamma_2 > 1$  are two parameters obtained from the process of inference. One advantage of  $\mathcal{D}_1$  is that we can construct it with a relatively smaller amount of data, since usually the first and second moments of  $\xi$  can be effectively estimated by the sample mean and covariance matrix. In addition,  $\mathcal{D}_1$  considers moment ambiguity (i.e., moment estimation errors are allowed) and develops nonparametric bounds on the mean and covariance matrix. One special case of  $\mathcal{D}_1$  can be constructed without considering moment ambiguity, e.g.,  $\mathcal{D}_1^0 = \{\mathbb{P} \in \mathcal{M}_+ : \mathbb{E}[\xi] = \mu, \mathbb{E}[\xi\xi^\top] = \mu\mu^\top + \Lambda\}$ . For this special confidence set setting, readers are referred to nice works accomplished recently by Vandenberghe et al. [38] and Zymler et al. [41].

Besides the moments, decision makers can also resort to the density function of the random vector  $\xi$ . For example, power system operators often describe random wind power available in a time unit at a wind farm by estimating its density function. A natural extension of such a “point” estimation is a “confidence set” estimation built around the point estimate, i.e., the decision makers might believe that although their estimate could suffer from some errors, the true density function is not too “far away” from it. One convenient and commonly used way of modeling the distance between density functions is by  $\phi$ -divergence, which is defined as

$$D_\phi(f||f_0) = \int_{\mathbb{R}^K} \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) f_0(\xi) d\xi,$$

where  $f$  and  $f_0$  denote the true density function and its estimate respectively, and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  is a convex function on  $\mathbb{R}^+$  such that

$$(C1) \quad \phi(1) = 0,$$

$$(C2) \quad 0\phi(x/0) := \begin{cases} x \lim_{p \rightarrow +\infty} \phi(p)/p & \text{if } x > 0, \\ 0 & \text{if } x = 0, \end{cases}$$

$$(C3) \quad \phi(x) = +\infty \text{ for } x < 0.$$

Three examples of  $\phi$ -divergence are as follows:

- Kullback-Leibler (KL) divergence:  $\phi(x) = x \log x - x + 1$  for  $x \geq 0$ ,

- $\chi$  divergence of order 2:  $\phi(x) = (x - 1)^2$  for  $x \geq 0$ ,
- Variation distance:  $\phi(x) = |x - 1|$  for  $x \geq 0$ .

For general  $\phi$ -divergence and other examples, interested readers are referred to Pardo [26] and Ben-Tal et al. [1]. Based on  $\phi$ -divergence, the decision makers can build a confidence set as follows:

$$\mathcal{D}_2 = \{\mathbb{P} \in \mathcal{M}_+ : D_\phi(f||f_0) \leq d, f = d\mathbb{P}/d\xi\}, \quad (5)$$

where the divergence tolerance  $d$  can be chosen by the decision makers to represent their risk-aversion level, or can be obtained from statistical inference. It can be observed that using  $\mathcal{D}_2$  as a confidence set for DCCs can perform better than using  $\mathcal{D}_1$ , because we can depict the profile of the ambiguous distribution  $\mathbb{P}$  more accurately by its density function than first two moments alone. Therefore, DCCs based on  $\mathcal{D}_2$  can be less conservative than those based on  $\mathcal{D}_1$ .

In this paper, we develop modeling and solution approaches for DCCs under both moment-based (e.g.,  $\mathcal{D}_1$ ) and density-based confidence sets (e.g.,  $\mathcal{D}_2$ ). We describe the construction of confidence sets, show how to equivalently reformulate DCCs, and discuss exact and approximate solution approaches for data-driven chance constrained programs (DCCPs) based on their equivalent reformulations. To the best of our knowledge, this paper provides the first study of DCCs under  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . The remainder of this paper is organized as follows. At the end of this section, we introduce notation and uncertainty settings to be used throughout this paper. We discuss the moment-based confidence sets ( $\mathcal{D}_1$ ) in Section 2. In Section 3, we discuss the density-based confidence set ( $\mathcal{D}_2$ ). In addition, in this section, we discover the relationship between the conservatism of DCCs and the sample size of historical data, which shows quantitatively the value of data. In Section 4, we execute a numerical study to compare the performances of  $\mathcal{D}_1$ ,  $\mathcal{D}_2$  and myopic approaches and verify the effectiveness of our proposed approaches. Finally, we summarize this paper in Section 5.

**Notation and Uncertainty Settings.** For a given vector  $x \in \mathbb{R}^K$ , we let  $\|x\|$  represent the  $L^2$ -norm of  $x$ , i.e.,  $\|x\| = \sqrt{x^\top x}$ , and  $\|x\|_\Lambda$  represent the norm of  $x$  induced by a symmetric and positive semidefinite matrix  $\Lambda \in \mathbb{S}_+^{K \times K}$ , i.e.,  $\|x\|_\Lambda = \sqrt{x^\top \Lambda x}$ . We specify the technology matrix  $A(\xi)$  and right-hand side  $b(\xi)$  by assuming that  $A(\xi)$  and  $b(\xi)$  are affinely dependent on  $\xi$  (a  $K$ -dimensional vector in the form  $\xi = (\xi_1, \dots, \xi_K)^\top$ ), i.e.,

$$A(\xi) = A_0 + \sum_{k=1}^K A_k \xi_k, \quad b(\xi) = b_0 + \sum_{k=1}^K b_k \xi_k, \quad (6)$$

where  $A_0$  and  $b_0$  represent the deterministic part of  $A(\xi)$  and  $b(\xi)$ , and each element in  $A(\xi)$  and  $b(\xi)$  is an affine function of  $\xi$ . This uncertainty setting has been adopted in the literature of stochastic programming and robust optimization (cf. Chen et al. [10] and Chen and Zhang [11]). Under this uncertainty setting, we can reformulate constraint  $A(\xi)x \leq b(\xi)$  as follows:

$$A(\xi)x \leq b(\xi) \Leftrightarrow A_0x + \sum_{k=1}^K (A_kx)\xi_k \leq b_0 + \sum_{k=1}^K b_k\xi_k \Leftrightarrow \bar{A}(x)\xi \leq \bar{b}(x),$$

where vector  $\bar{b}(x) = b_0 - A_0x$ , and  $\bar{A}(x)$  is an  $m \times K$  matrix defined as

$$\bar{A}(x) = [A_1x - b_1, \quad A_2x - b_2, \quad \dots, \quad A_Kx - b_K].$$

Accordingly, DCC (3) can be generally reformulated as

$$\inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P}\{\xi \in C\} \geq 1 - \alpha, \tag{7}$$

where  $C = \{\xi \in \mathbb{R}^K : \bar{A}\xi \leq \bar{b}\}$  is a polyhedron whose parameters  $\bar{A}$  and  $\bar{b}$  depend upon  $x$ . In the remainder of this paper, we use DCC (3) and its general reformulation (7) interchangeably for notation brevity.

## 2 DCC with Moment-based Confidence Set

In this section, we consider constraint (7) with  $\mathcal{D} = \mathcal{D}_1$ . In Section 2.1, we discuss the construction of  $\mathcal{D}_1$  by using a series of independent historical data  $\{\xi^i\}_{i=1}^N$  obeying the true distribution  $\mathbb{P}$ . In Section 2.2, we show a general equivalence relationship between a group of infinite nonconvex constraints and a group of linear matrix inequalities (LMIs). We apply this equivalence relationship to reformulate the single DCC under  $\mathcal{D}_1$ , i.e.,  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in C\} \geq 1 - \alpha$ , as LMIs with polyhedron  $C$  replaced by its interior,  $\text{int}(C) = \{\xi \in \mathbb{R}^K : \bar{A}\xi < \bar{b}\}$ . Furthermore, we extend the study to other single DCCs, e.g., DCCs with polytopic and marginal moment information, and show that they can also be reformulated as LMIs or even nicer second-order cone (SOC) constraints. In addition, we show the continuity of  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in C\} \geq 1 - \alpha$  over the polyhedron  $C$ , i.e.,

$$\begin{aligned} \min_{x \in X} \psi(x) &= \min_{x \in X} \psi(x) \\ \text{s.t. } \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in C\} &\geq 1 - \alpha & \text{s.t. } \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in \text{int}(C)\} &\geq 1 - \alpha, \end{aligned}$$

to accomplish the reformulation of (7). In Section 2.3, we further our study for joint DCCs. We observe that the joint DCCs under  $\mathcal{D}_1$  cannot be reformulated as LMIs although its left-hand side (i.e.,  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in \text{int}(C)\}$ ) can be evaluated in polynomial time, and we develop tractable approximation for the joint DCCs. Finally, in Section 2.4, we extend our study for the worst-case value-at-risk constraints.

## 2.1 Construction of Confidence Set

Given a series of independent historical data samples  $\{\xi^i\}_{i=1}^N$  drawn from the true distribution  $\mathbb{P}$  of the random vector  $\xi$ , we want to estimate the first and second moments of  $\xi$ . First, the point estimates of the mean and covariance matrix can be obtained by the sample moments

$$\mu = \frac{1}{N} \sum_{i=1}^N \xi^i, \quad \Lambda = \frac{1}{N} \sum_{i=1}^N (\xi^i - \mu)(\xi^i - \mu)^\top,$$

where  $\mu$  and  $\Lambda$  are maximum likelihood estimators of  $\mathbb{E}[\xi]$  and  $\text{cov}(\xi)$ , respectively. Second, we follow the procedure described in Delage and Ye [12] to construct a nonparametric confidence set for the mean and covariance matrix of  $\xi$  as follows:

$$\mathcal{D}_1 = \left\{ \mathbb{P} \in \mathcal{M}_+ : (\mathbb{E}[\xi] - \mu)^\top \Lambda^{-1} (\mathbb{E}[\xi] - \mu) \leq \gamma_1, \quad \mathbb{E}[(\xi - \mu)(\xi - \mu)^\top] \preceq \gamma_2 \Lambda \right\},$$

where the parameters  $\gamma_1 \geq 0$  and  $\gamma_2 > 1$  can be obtained from the process of inference (see Delage and Ye [12] for details).

## 2.2 Reformulation of the Single DCCs

In this subsection, we investigate the reformulation of the single DCC under  $\mathcal{D}_1$ ,

$$\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\} \geq 1 - \alpha, \tag{8}$$

where  $\bar{a} \in \mathbb{R}^K$  is a  $K$ -dimensional vector and  $\bar{b} \in \mathbb{R}$  is a scalar, both of which are affine functions of decision variable  $x$  as discussed in Section 1.2. Note that we reformulate the strict version of the DCCs (i.e.,  $\mathbb{P}\{\bar{a}^\top \xi < \bar{b}\}$ ) first, and then recover the soft version (i.e.,  $\mathbb{P}\{\bar{a}^\top \xi \leq \bar{b}\}$ ) in later Proposition 1. The reformulation relies on a general equivalence relationship between a group of infinite nonconvex constraints and a group of LMIs. Moreover, we find that this equivalence relationship can be extended to other single and even joint DCCs. We first show the equivalence relationship in Lemma 1.



**Lemma 1** For a given group of symmetric and positive semidefinite matrices  $M_i \in \mathbb{R}^{K \times K}$  (i.e.,  $M_i \in \mathbb{S}_+^{K \times K}$ ), vectors  $\bar{a}_i \in \mathbb{R}^K$ , and scalars  $\bar{b}_i \in \mathbb{R}$ ,  $i = 1, \dots, m$ , the following two claims are equivalent corresponding to a symmetric matrix  $H \in \mathbb{S}^{K \times K}$ , a vector  $p \in \mathbb{R}^K$ , and a scalar  $q$ :

(i)  $\xi^\top H \xi + p^\top \xi + q \leq I_C(\xi)$ ,  $\forall \xi \in \mathbb{R}^K$ , where  $I_C(\xi) = 1$  if  $\xi \in C$  and  $I_C(\xi) = 0$  otherwise,

(ii) There exist  $\{y_i\}_{i=1}^m \geq 0$ , such that

$$\begin{cases} \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq 0, \\ \begin{bmatrix} -H - y_i M_i & -\frac{1}{2}(p + y_i \bar{a}_i) \\ -\frac{1}{2}(p + y_i \bar{a}_i)^\top & y_i \bar{b}_i - q \end{bmatrix} \succeq 0, \quad \forall i = 1, \dots, m, \end{cases}$$

where  $C := \bigcap_{i=1}^m C_i$  with  $C_i := \{\xi \in \mathbb{R}^K : \xi^\top M_i \xi + \bar{a}_i^\top \xi < \bar{b}_i\}$ .

*Proof:* Following the definitions of  $C_i$  and  $C$ , we first observe that

$$I_C(\xi) = \prod_{i=1}^m I_{C_i}(\xi) = \prod_{i=1}^m I_{[\xi^\top M_i \xi + \bar{a}_i^\top \xi < \bar{b}_i]}(\xi), \quad \forall \xi \in \mathbb{R}^K,$$

and hence, statement (i) is equivalent to

$$\xi^\top H \xi + p^\top \xi + q \leq I_{[\xi^\top M_i \xi + \bar{a}_i^\top \xi < \bar{b}_i]}(\xi), \quad \forall \xi \in \mathbb{R}^K, \forall i = 1, \dots, m,$$

which is further equivalent to

$$\begin{cases} \xi^\top H \xi + p^\top \xi + q \leq 1, \quad \forall \xi \in \mathbb{R}^K, & (9a) \\ \xi^\top M_i \xi + \bar{a}_i^\top \xi \geq \bar{b}_i \Rightarrow \xi^\top H \xi + p^\top \xi + q \leq 0, \quad \forall \xi \in \mathbb{R}^K, \forall i = 1, \dots, m. & (9b) \end{cases}$$

Next we discuss the reformulation of constraint (9a) and implication (9b). First, we observe that constraint (9a) equivalently requires that the quadratic function  $\xi^\top H \xi + p^\top \xi + q - 1$  is nonpositive everywhere in  $\mathbb{R}^K$ , and hence, is equivalent to

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq 0. \quad (10)$$

Second, the implication (9b) is equivalent to the statement that, for each  $i = 1, \dots, m$ , the following quadratic system

$$\begin{cases} \xi^\top H \xi + p^\top \xi + q > 0 \\ \xi^\top M_i \xi + \bar{a}_i^\top \xi - \bar{b}_i \geq 0 \end{cases} \quad (11)$$

has no solution in  $\mathbb{R}^K$ . We claim that this is equivalent to the following (infinite) quadratic system having a nonnegative solution  $y_i$ :

$$-(\xi^\top H \xi + p^\top \xi + q) - y_i(\xi^\top M_i \xi + \bar{a}_i^\top \xi - \bar{b}_i) \geq 0, \quad \forall \xi \in \mathbb{R}^K. \quad (12)$$

To see this, we discuss the following cases:

**Case 1** If matrix  $M_i$  has a strictly positive eigenvalue, then there exists a  $\bar{\xi} \in \mathbb{R}^K$  such that  $\bar{\xi}^\top M_i \bar{\xi} + \bar{a}_i^\top \bar{\xi} - \bar{b}_i > 0$  since we can choose  $\bar{\xi}$  to make  $\bar{\xi}^\top M_i \bar{\xi} + \bar{a}_i^\top \bar{\xi} - \bar{b}_i$  arbitrarily large. Thus, in this case, the condition of the S-Lemma (cf. Yakubovich [40], Pólik and Terlaky [29], and Appendix A) is satisfied and accordingly the equivalence is guaranteed following the S-Lemma.

**Case 2** If all the eigenvalues of matrix  $M_i$  are nonpositive, then  $M_i = 0$  because  $M_i \succeq 0$  by assumption. We further discuss the following cases:

**Case 2.1** If  $\bar{a}_i \neq 0$ , we observe that there exists a  $\bar{\xi} \in \mathbb{R}^K$  such that  $\bar{a}_i^\top \bar{\xi} - \bar{b}_i > 0$  since we can choose  $\bar{\xi}$  to make  $\bar{a}_i^\top \bar{\xi}$  arbitrarily large. In this case, the equivalence is again guaranteed by the S-Lemma.

**Case 2.2** If  $\bar{a}_i = 0$  and  $\bar{b}_i \leq 0$ : First, suppose that (11) has no solution. Since  $\bar{a}_i^\top \bar{\xi} - \bar{b}_i = -\bar{b}_i \geq 0$  is satisfied, it follows that  $\bar{\xi}^\top H \bar{\xi} + p^\top \bar{\xi} + q > 0$  has no solution, i.e.,  $\bar{\xi}^\top H \bar{\xi} + p^\top \bar{\xi} + q \leq 0$  for all  $\bar{\xi} \in \mathbb{R}^K$ , which implies that (12) has a solution  $y_i = 0$ . Second, suppose that (12) has a solution  $y_i \geq 0$ . Since  $M_i = 0$ ,  $\bar{a}_i = 0$ , and  $\bar{b}_i \leq 0$ , it follows that  $-(\bar{\xi}^\top H \bar{\xi} + p^\top \bar{\xi} + q) \geq -y_i \bar{b}_i \geq 0$ , which implies that (11) has no solution. Hence, the equivalence is guaranteed in this case.

**Case 2.3** If  $\bar{a}_i = 0$  and  $\bar{b}_i > 0$ : Since  $\bar{a}_i^\top \bar{\xi} - \bar{b}_i = -\bar{b}_i < 0$ , (11) has no solution, and we only need to show that (12) has a nonnegative solution. But in view of (9a), we know that  $y_i = 1/\bar{b}_i > 0$  is a solution for (12).

We have proved the equivalence between (11) and (12). Another way of stating (12) is that there exists some  $y_i \geq 0$ , such that the quadratic function  $-\bar{\xi}^\top (H + y_i M_i) \bar{\xi} - (p + y_i \bar{a}_i)^\top \bar{\xi} + y_i \bar{b}_i - q$  is nonnegative everywhere in  $\mathbb{R}^K$ , which is equivalent to

$$\begin{bmatrix} -H - y_i M_i & -\frac{1}{2}(p + y_i \bar{a}_i) \\ -\frac{1}{2}(p + y_i \bar{a}_i)^\top & y_i \bar{b}_i - q \end{bmatrix} \succeq 0. \quad (13)$$

Therefore, we have equivalently reformulated statement (i) as LMIs (10) and (13) for  $i = 1, \dots, m$ , which completes the proof. ■

Second, we apply Lemma 1 to reformulate the single DCC (8) as LMIs.

**Theorem 1** Given a vector  $\bar{a}$  and a scalar  $\bar{b}$ , the DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\} \geq 1 - \alpha$  is equivalent to the following LMIs:

$$\begin{bmatrix} \gamma_2 \Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \leq \alpha \bar{y}, \quad (14a)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2} \bar{a} \\ \frac{1}{2} \bar{a}^\top & \bar{y} + \bar{a}^\top \mu - \bar{b} \end{bmatrix}, \quad (14b)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \bar{y} \geq 0, \quad (14c)$$

where operator “ $\cdot$ ” in constraint (14a) represents the Frobenius inner product.

*Proof:* First, we let  $z_{\mathcal{D}_1}$  represent the optimal objective value of the optimization problem on the left-hand side of the DCC, i.e.,  $z_{\mathcal{D}_1} = \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\}$ . We specify  $z_{\mathcal{D}_1}$  by the following optimization problem:

$$[\text{Primal}] \quad z_{\mathcal{D}_1} = \min_{\mathbb{P}} \int_{\mathbb{R}^K} I_{[\bar{a}^\top \xi < \bar{b}]}(\xi) d\mathbb{P} \quad (15a)$$

$$\text{s.t.} \quad \int_{\mathbb{R}^K} \begin{bmatrix} \Lambda & \xi - \mu \\ (\xi - \mu)^\top & \gamma_1 \end{bmatrix} d\mathbb{P} \succeq 0, \quad (15b)$$

$$\int_{\mathbb{R}^K} (\xi - \mu)(\xi - \mu)^\top d\mathbb{P} \preceq \gamma_2 \Lambda, \quad (15c)$$

$$\int_{\mathbb{R}^K} d\mathbb{P} = 1, \quad (15d)$$

where constraint (15b) describes the confidence set of  $\mathbb{E}[\xi]$  based on Schur complement, constraint (15c) describes the confidence set of  $\mathbb{E}[(\xi - \mu)(\xi - \mu)^\top]$ , and constraint (15d) guarantees that we are considering probability distributions on  $\mathbb{R}^K$ . We apply the duality theory for conic linear programming problems and dualize problem (15) as

$$[\text{Dual}] \quad z_{\mathcal{D}_1} = \max_{G, H, p, q, r} -\gamma_2 \Lambda \cdot G + r - \Lambda \cdot H - \gamma_1 q \quad (16a)$$

$$\text{s.t.} \quad (\xi - \mu)^\top (-G)(\xi - \mu) + 2p^\top (\xi - \mu) + r \leq I_{[\bar{a}^\top \xi < \bar{b}]}(\xi), \quad \forall \xi \in \mathbb{R}^K \quad (16b)$$

$$G \in \mathbb{S}_+^{K \times K}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad (16c)$$

where matrix  $\begin{bmatrix} H & p \\ p^\top & q \end{bmatrix}$ , matrix  $G$ , and scalar  $r$  represent the dual variables for constraints (15b), (15c), and (15d), respectively. Note here that strong duality holds for problems [Primal] and [Dual] based on established conic linear programming theory (cf. Isii [16], Smith [36], and Shapiro [34]). Second, we reformulate constraint (16b). By replacing  $\xi$  with  $\xi + \mu$ , we have constraint (16b)

equivalent to

$$\xi^\top(-G)\xi + 2p^\top\xi + r \leq I_{[\bar{a}^\top\xi < \bar{b} - \bar{a}^\top\mu]}(\xi), \quad \forall \xi \in \mathbb{R}^K,$$

which, in view of Lemma 1, is further equivalent to

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2}y\bar{a} \\ \frac{1}{2}y\bar{a}^\top & 1 + y(\bar{a}^\top\mu - \bar{b}) \end{bmatrix}, \quad \begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \succeq 0,$$

for some  $y \geq 0$ . Hence,  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top\xi < \bar{b}\} \geq 1 - \alpha$  is equivalent to

$$\begin{bmatrix} \gamma_2\Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \leq \alpha, \quad (17a)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2}y\bar{a} \\ \frac{1}{2}y\bar{a}^\top & 1 + y(\bar{a}^\top\mu - \bar{b}) \end{bmatrix}, \quad (17b)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad y \geq 0. \quad (17c)$$

Now we observe that  $y > 0$ , because otherwise (i.e.,  $y = 0$ ) by constraint (17b) we have

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \succeq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

and hence

$$\begin{bmatrix} \gamma_2\Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \geq \begin{bmatrix} \gamma_2\Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 1,$$

which violates constraint (17a) because  $\alpha \in (0, 1)$ . Hence, we let  $\bar{y} = 1/y \geq 0$ , replace matrices  $\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix}$  and  $\begin{bmatrix} H & p \\ p^\top & q \end{bmatrix}$  by  $y \begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix}$  and  $y \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix}$ , respectively in LMIs (17a)-(17c), and obtain the reformulation (14a)-(14c).  $\blacksquare$

Third, we extend the application of Lemma 1 to other single DCCs under various confidence sets. We present the extensions in the following corollaries, whose detailed proofs are provided in Appendices B-C for brevity.

**(Nonlinear inequality)** The above linear inequality  $\bar{a}^\top\xi < \bar{b}$  can be extended to a general quadratic one  $\xi^\top M\xi + \bar{a}^\top\xi < \bar{b}$  for any given  $M \in \mathbb{S}_+^{K \times K}$ . We omit the detailed proof for the following corollary because of its similarity to that for Theorem 1.

**Corollary 1** *The DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi^\top M \xi + \bar{a}^\top \xi < \bar{b}\} \geq 1 - \alpha$  is equivalent to the following LMIs:*

$$\begin{bmatrix} \gamma_2 \Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \leq \alpha \bar{y}, \quad (18a)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} \succeq \begin{bmatrix} M & \frac{1}{2}(\bar{a} + 2M\mu) \\ \frac{1}{2}(\bar{a} + 2M\mu)^\top & \bar{y} + \mu^\top M \mu + \bar{a}^\top \mu - \bar{b} \end{bmatrix}, \quad (18b)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \bar{y} \geq 0. \quad (18c)$$

**(Polytopic ambiguity)**  $\mathcal{D}_1$  incorporates the moment ambiguity by considering nonlinear convex sets (e.g., considering the mean within an ellipsoid). An alternative is to consider polytopic ambiguity. That is, decision makers can name a set of possible means and second moments  $\mathcal{P} := \{(\mu_i, \Sigma_i) \in \mathbb{R}^K \times \mathbb{S}_+^{K \times K} : \Sigma_i \succeq \mu_i \mu_i^\top, i = 1, \dots, I\}$ , where the condition  $\Sigma_i \succeq \mu_i \mu_i^\top$  is used to make  $\mathcal{P}$  well defined, and construct the probability distribution confidence set as follows:

$$\mathcal{D}_1^{\text{pol}} = \left\{ \mathbb{P} \in \mathcal{M}_+ : (\mathbb{E}[\xi], \mathbb{E}[\xi \xi^\top]) \in \text{conv}(\mathcal{P}) \right\},$$

where  $\text{conv}$  represents the convex hull, i.e., there exists  $\lambda \in \mathbb{R}_+^I$ , such that  $\mathbb{E}[\xi] = \sum_{i=1}^I \lambda_i \mu_i$ ,  $\mathbb{E}[\xi \xi^\top] = \sum_{i=1}^I \lambda_i \Sigma_i$ , and  $\sum_{i=1}^I \lambda_i = 1$ . The DCC under  $\mathcal{D}_1^{\text{pol}}$  can be reformulated as LMIs shown in the following corollary, whose detailed proof is provided in Appendix B.

**Corollary 2** *The DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1^{\text{pol}}} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\}$  is equivalent to the following LMIs:*

$$-r + (1 - q) \leq \alpha \bar{y}, \quad (19a)$$

$$\Sigma_i \cdot H + \mu_i^\top p - r \geq 0, \quad \forall i = 1, \dots, I, \quad (19b)$$

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2}\bar{a} \\ \frac{1}{2}\bar{a}^\top & \bar{y} - \bar{b} \end{bmatrix}, \quad (19c)$$

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \bar{y} \geq 0. \quad (19d)$$

**Remark 1** *One special case of corollary 2 is when  $I = 1$ , in which  $\mathcal{D}_1^{\text{pol}}$  reduces to be  $\{\mathbb{P} \in \mathcal{M}_+ : \mathbb{E}[\xi] = \mu_1, \mathbb{E}[\xi \xi^\top] = \Sigma_1\}$ , which is simpler than  $\mathcal{D}_1$  and does not allow moment ambiguity. The reformulation of the DCC under this special case is nicely described in Vandenberghe et al. [38] and Zymler et al. [41].*

**(Marginal moments)** Another variant of  $\mathcal{D}_1$  is to estimate the first two moments of each component  $\xi_k$  separately. That is, if we omit the correlation of each pair  $\xi_k$  and  $\xi_\ell$  and build confidence

intervals for  $\mathbb{E}[\xi_k]$  and  $\mathbb{E}[\xi_k^2]$  for  $k = 1, \dots, K$ , it results in

$$\mathcal{D}_1^{\text{mgn}} = \left\{ \mathbb{P} \in \mathcal{M}_+ : \mu_k^L \leq \mathbb{E}[\xi_k] \leq \mu_k^U, \mathbb{E}[\xi_k^2] \leq \sigma_k, \forall k = 1, \dots, K \right\},$$

where  $[\mu_k^L, \mu_k^U]$  is a confidence interval of  $\mathbb{E}[\xi_k]$  and  $\sigma_k$  is an upper bound of  $\mathbb{E}[\xi_k^2]$ , respectively. Also, we assume that  $\text{diag}(\sigma_1, \dots, \sigma_K) \succeq \mu\mu^\top$  for each  $\mu^L \leq \mu \leq \mu^U$  to make  $\mathcal{D}_1^{\text{mgn}}$  well defined. Note here that we can always increase  $\sigma_k$  values to make it happen. It is easier to construct  $\mathcal{D}_1^{\text{mgn}}$  than  $\mathcal{D}_1$  by using a smaller size data set, since we only consider the marginal distribution of the random vector  $\xi$ . We state the corollary as follows and the detailed proof is provided in Appendix C.

**Corollary 3** *The DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1^{\text{mgn}}} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\}$  is equivalent to the following SOC constraint:*

$$\sum_{k=1}^K (\mu_k^U p_k^U - \mu_k^L p_k^L + \sigma_k h_k) - q \leq \alpha \bar{y} - 1, \quad (20a)$$

$$q + \sum_{k=1}^K t_k \leq 1, \quad q + \bar{y} + \sum_{k=1}^K s_k \leq \bar{b} + 1, \quad (20b)$$

$$\left\| \begin{bmatrix} p_k^L - p_k^U \\ t_k - h_k \end{bmatrix} \right\| \leq t_k + h_k, \quad \forall k = 1, \dots, K, \quad (20c)$$

$$\left\| \begin{bmatrix} p_k^L - p_k^U + \bar{a}_k \\ s_k - h_k \end{bmatrix} \right\| \leq s_k + h_k, \quad \forall k = 1, \dots, K, \quad (20d)$$

$$p^L, p^U, h \geq 0, \quad \bar{y} \geq 0, t_k \geq 0, s_k \geq 0, \quad \forall k = 1, \dots, K. \quad (20e)$$

From the above corollary, we can observe that the resulting reformulation for the DCC under  $\mathcal{D}_1^{\text{mgn}}$  is a SOC constraint, which is more computationally tractable than LMIs.

Before closing this subsection, we show the continuity of the DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in C\} \geq 1 - \alpha$  over the polyhedron  $C$ . We state the following proposition and provide the detailed proof in Appendix D.

**Proposition 1** *Denote  $z(C)$  as the optimal objective value of an optimization problem with DCC, i.e.,*

$$\begin{aligned} z(C) &= \min_{x \in X} \psi(x) \\ &\text{s.t.} \quad \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi \in C\} \geq 1 - \alpha, \end{aligned}$$

where  $X$  is a convex set,  $\psi(\cdot)$  is a convex function,  $C = \{\xi \in \mathbb{R}^K : \bar{a}^\top \xi \leq \bar{b}\}$  with  $\bar{a}$  and  $\bar{b}$  being affine functions of  $x$ , and the Slater condition holds for the optimization problem. Then  $z(C) = z(\text{int}(C))$ .

**Remark 2** Proposition 1 is trivial if the probability distribution  $\mathbb{P}$  is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^K$  because in this case  $\mathbb{P}\{\bar{a}^\top \xi = \bar{b}\} = 0$ . However, this proposition holds for general distributions.

**Remark 3** Similar continuity results as Proposition 1 hold for other single DCCs discussed in corollaries 1-3. That is, Proposition 1 holds with polyhedron  $C$  replaced by  $\{\xi \in \mathbb{R}^K : \xi^\top M \xi + \bar{a}^\top \xi \leq \bar{b}\}$ , with  $\mathcal{D}_1$  replaced by  $\mathcal{D}_1^{pol}$ , and with  $\mathcal{D}_1$  replaced by  $\mathcal{D}_1^{mgn}$ , respectively.

### 2.3 Reformulation of the Joint DCCs

In this subsection, we study the joint DCCs. We observe that the joint DCCs under  $\mathcal{D}_1$  can no longer be reformulated as LMIs although its left-hand side (i.e.,  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi < \bar{b}\}$ ) can be evaluated in polynomial time, and we develop tractable conservative approximation and relaxation for the joint DCCs. First, we show the reformulation of the joint DCCs by the following corollary and omit the proof because of its similarity to that described in Lemma 1 and Theorem 1.

**Corollary 4** Given a matrix  $\bar{A}$  and a vector  $\bar{b}$ , the joint DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi < \bar{b}\} \geq 1 - \alpha$  is equivalent to the following constraints:

$$1 - \begin{bmatrix} \gamma_2 \Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} - \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \geq 1 - \alpha, \quad (21a)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2} y_i \bar{a}_i \\ \frac{1}{2} y_i \bar{a}_i^\top & 1 + y_i (\bar{a}_i^\top \mu - \bar{b}_i) \end{bmatrix}, \quad \forall i = 1, \dots, m, \quad (21b)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad y \geq 0. \quad (21c)$$

where  $\bar{a}_1^\top, \dots, \bar{a}_m^\top$  represent the  $m$  row vectors consisting of matrix  $\bar{A}$ .

Note that the reformulating constraints (21a)-(21c) are no longer LMIs because of the bilinear terms  $y_i \bar{a}_i$  in constraints (21b), where  $y_i$  represents a nonnegative multiplier and  $a_i$  is an affine function of decision variable  $x$ . However, with the multipliers  $y$  fixed, constraints (21a)-(21c) recover as LMIs. In this case, the worst-case probability bound  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi < \bar{b}\}$  equals the left-hand side of constraint (21a) and can be computed in polynomial time.

Second, we develop a tractable conservative approximation for the joint DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha$ . We observe that the difficulty of the joint DCC, as compared to the tractable single

DCCs, is that we have to satisfy multiple inequalities at the same time. To utilize the tractability of the single DCCs, one option is to reformulate the joint DCC as

$$\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} = \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\left\{\max_{i=1,\dots,m} \{\bar{a}_i^\top \xi - \bar{b}_i\} \leq 0\right\} \geq 1 - \alpha,$$

where the function  $\max_{i=1,\dots,m} \{\bar{a}_i^\top \xi - \bar{b}_i\}$  is convex and piece-wise linear over  $\xi$ . This observation motivates us to approximate this function by using a quadratic function in the form  $\xi^\top M\xi + c^\top \xi + d$ , in view that we can reformulate the resulting approximation as LMIs by using corollary 1. To make this approximation conservative, we ensure that  $\xi^\top M\xi + c^\top \xi + d \geq \bar{a}_i^\top \xi - \bar{b}_i$  for all  $\xi \in \mathbb{R}^K$  and all  $i = 1, \dots, m$ , which implies that

$$\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\left\{\max_{i=1,\dots,m} \{\bar{a}_i^\top \xi - \bar{b}_i\} \leq 0\right\} \geq \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\left\{\xi^\top M\xi + c^\top \xi + d \leq 0\right\} \geq 1 - \alpha.$$

In addition, to make this approximation as tight as possible, we can treat the approximation coefficients  $M$ ,  $c$ , and  $d$  as decision variables, and so we can find an ‘‘optimal’’ approximation on the way of optimizing over the joint DCC. We summarize the approximation result by the following proposition, whose detailed proof is provided in Appendix E.

**Proposition 2** *The joint DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha$  can be conservatively approximated by the following LMIs:*

$$\begin{bmatrix} \gamma_2 \Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \leq \alpha \bar{y}, \quad (22a)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} \succeq \begin{bmatrix} M & \frac{1}{2}(c + 2M\mu) \\ \frac{1}{2}(c + 2M\mu)^\top & \bar{y} + \mu^\top M\mu + c^\top \mu + d \end{bmatrix}, \quad (22b)$$

$$\begin{bmatrix} M & \frac{1}{2}(c - \bar{a}_i) \\ \frac{1}{2}(c - \bar{a}_i)^\top & d + \bar{b}_i \end{bmatrix} \succeq 0, \quad \forall i = 1, \dots, m, \quad (22c)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1 - r \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \bar{y} \geq 0. \quad (22d)$$

A relaxation of the joint DCC can be obtained by noting that

$$\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\left\{\bar{a}_i^\top \xi \leq \bar{b}_i\right\} \geq \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\left\{\max_{i=1,\dots,m} \{\bar{a}_i^\top \xi - \bar{b}_i\} \leq 0\right\} \geq 1 - \alpha, \quad \forall i = 1, \dots, m,$$

where the joint DCC is relaxed to be a single (and tractable) DCC. We can also tighten this relaxation by adding in single DCC relaxation for each  $i = 1, \dots, m$ , which is presented by the following proposition.



**Proposition 3** *The joint DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha$  can be relaxed to be the following LMIs:*

$$\begin{aligned} & \begin{bmatrix} \gamma_2 \Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix} \leq \alpha \bar{y}_i, \quad \forall i = 1, \dots, m, \\ & \begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2} \bar{a}_i \\ \frac{1}{2} \bar{a}_i^\top & \bar{y}_i + \bar{a}_i^\top \mu - \bar{b}_i \end{bmatrix}, \quad \forall i = 1, \dots, m, \\ & \begin{bmatrix} G_i & -p_i \\ -p_i^\top & 1 - r_i \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H_i & p_i \\ p_i^\top & q_i \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad \bar{y}_i \geq 0, \quad \forall i = 1, \dots, m. \end{aligned}$$

## 2.4 Reformulation of the Worst-Case Value-at-Risk Constraint

In this subsection, we extend our study on the DCCs to the worst-case value-at-risk constraint. First, value-at-risk (VaR) is a risk measure frequently used in portfolio optimization problems (see, e.g., Rockafellar and Uryasev [32], and El Ghaoui et al. [14]). Given a random vector  $\xi$  with probability distribution  $\mathbb{P}$ , a constraint on the VaR of random variable  $\bar{a}^\top \xi$  with risk level  $\alpha$  (denoted as  $\text{VaR}_\alpha(\bar{a}^\top \xi)$ ) can be defined as

$$\text{VaR}_\alpha(\bar{a}^\top \xi) := \inf \left\{ \ell \in \mathbb{R} : \mathbb{P}\{\bar{a}^\top \xi \leq \ell\} \geq 1 - \alpha \right\} \leq \bar{b},$$

where  $\bar{b}$  is a given upper bound for VaR. In practice,  $\mathbb{P}$  is not always perfectly known or accurately estimated, and a decision maker can thus consider an alternative worst-case value-at-risk (WVaR) constraint defined as

$$\text{VaR}_\alpha(\bar{a}^\top \xi) \leq \bar{b}, \quad \forall \mathbb{P} \in \mathcal{D}_1, \quad (24)$$

where  $\mathcal{D}_1$  can be estimated based on historical data samples of  $\xi$ , and constraint (24) is robust with regard to the underlying probability distribution of  $\xi$ . In the following proposition, we show the equivalence between constraint (24) and the single DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top \xi \leq \bar{b}\} \geq 1 - \alpha$ , and we provide the detailed proof in Appendix F.

**Proposition 4** *The WVaR constraint (24) is equivalent to the LMIs (14a)-(14c) defined in Theorem 1.*

**Remark 4** *From Proposition 4, it is very interesting to observe that the data-driven chance constraint and the worst-case value-at-risk constraint lead to the same reformulation. In other words, with the consideration of the data-driven robustness in the chance constraint, there is no need to consider the worst-case value-at-risk measure to improve the robustness.*

### 3 DCC with Density-based Confidence Set

In this section, we consider DCC (7) when density information is taken into account, i.e., with  $\mathcal{D} = \mathcal{D}_2$ . We first discuss the construction of  $\mathcal{D}_2$  in Section 3.1. Then, in Section 3.2, we show that the joint DCCs under  $\mathcal{D}_2$  is equivalent to a classical chance constraint (1b) with a deterministic probability distribution, which can be further solved by scenario approximation approaches. In addition, by deriving how the sample size of the data reflects the level of conservatism of DCCs, we depict quantitatively the value of data in Section 3.3.

#### 3.1 Construction of Confidence Set

In this subsection, we discuss the construction of confidence set  $\mathcal{D}_2$ , which is based on a general  $\phi$ -divergence defined as

$$D_\phi(f||f_0) = \int_{\mathbb{R}^K} \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) f_0(\xi) d\xi,$$

where  $f$  and  $f_0$  denote the density function (resp. probability mass function for discrete distribution) of  $\mathbb{P}$  and its estimate respectively, and the integral is with respect to Lebesgue measure on  $\mathbb{R}^K$  (resp. with respect to the counting measure for discrete distribution). For discrete distributions, histograms are useful in practice for depicting density function profiles. Suppose that we have a data set  $\{\xi^i\}_{i=1}^N$ . To draw a histogram, we first construct a nonempty partition  $\{B_j : j = 1, \dots, \bar{B}\}$  of the sample space, where  $\Omega = \bigcup_{j=1}^{\bar{B}} B_j$  and each  $B_j$  is called a “bin.” Second, we count the frequency  $N_j = \sum_{i=1}^N I_{[B_j]}(\xi^i)$  for each bin  $B_j$ , where  $I_{[B_j]}(\xi^i)$  equals one if  $\xi^i \in B_j$ , and zero otherwise. Finally, we can estimate the probability of landing in each bin, i.e.,  $\mathbb{P}\{B_j\}$ , by its empirical relative frequency  $N_j/N$ . For continuous distributions, we cannot directly use the histogram estimate for  $f$  since it is not absolutely continuous with regard to the Lebesgue measure. Instead, we replace the histogram estimate by its counterpart in estimating continuous density functions, called the kernel density estimator (KDE), which is defined as follows (see Rosenblatt [33] and Parzen [27]):

$$f_N(\xi) = \frac{1}{Nh_N^K} \sum_{i=1}^N H\left(\frac{\xi - \xi^i}{h_N}\right),$$

where  $h_N$  is a positive constant,  $K$  is the dimension of  $\xi$ , and  $H(\cdot)$  is a smooth function satisfying  $H(\cdot) \geq 0$ ,  $\int H(\xi) d\xi = 1$ ,  $\int \xi H(\xi) d\xi = 0$ , and  $\int \xi^2 H(\xi) d\xi > 0$ . One example for  $H(\cdot)$  is the standard normal density function. It is shown in Devroye and Györfi [13] that KDE converges to the true

density in  $L^1$ -norm, i.e., with probability one,

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

There are various ways to decide the divergence tolerance  $d$  in  $\mathcal{D}_2$ . First, decision makers can decide the value of  $d$  to reflect their risk-aversion preference, and adjust  $d$  to perform post-optimization sensitivity analysis. For example, decision makers may decide  $d = J/\ln(N)$ , where  $J$  is a constant and  $N$  is the total number of data samples, to reflect an intuition that  $\mathcal{D}_2$  becomes tighter as  $N$  increases. Second, for discrete distributions, we can estimate the value of  $d$  by using the histogram estimate. Pardo [26] (see, e.g., Theorem 3.1 in [26]) shows that  $\frac{2N}{\phi''(1)} D_\phi(f||f_0)$  converges in distribution to a Chi-square distributed random variable with  $\bar{B} - 1$  degrees of freedom as  $N$  goes to infinity, provided that  $\phi''(1)$  exists and is nonzero. This observation motivates us to approximate the divergence tolerance  $d$  by setting  $d = \phi''(1)\chi_{\bar{B}-1,1-\beta}^2/(2N)$  with large  $N$ , where  $\chi_{\bar{B}-1,1-\beta}^2$  represents the 100(1 -  $\beta$ )% (e.g.,  $\beta = 0.05$ ) percentile of the  $\chi_{\bar{B}-1}^2$  distribution. For continuous distributions, however, there are few literature discussing the estimation of  $d$  for a general  $\phi$ -divergence. One possible approach is to estimate  $d$  by  $\phi''(1)\chi_{\bar{B}-1,1-\beta}^2/(2N)$  for large  $N$ , as for the discrete case, motivated by the observation that both the histogram and the KDE converge to the true distribution as  $N$  goes to infinity. In general, in observation of multiple possible ways of estimating the  $\phi$ -divergence tolerance  $d$ , in this paper, we assume that  $d$  is in general a function of data sample size  $N$ , i.e.,  $d := d(N)$ ,  $d$  is nonincreasing as  $N$  increases, and  $d$  tends to zero as  $N$  tends to infinity, in our theoretical analysis.

### 3.2 Reformulation of DCC

In this subsection, we address the reformulation of the DCCs under  $\mathcal{D}_2$  with a general  $\phi$ -divergence measure considered. Before giving the main result, we review the definition of conjugate duality. Given a function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , the conjugate  $g^* : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  is defined as

$$g^*(t) = \sup_{x \in \mathbb{R}} \{tx - g(x)\}.$$

Also, we present some properties of the conjugate of function  $\phi$  which is used to define  $\phi$ -divergence measures. We provide the detailed proof of the following Lemma in Appendix G for completeness.

**Lemma 2** Let  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  be a convex function such that  $\phi(1) = 0$  and  $\phi(x) = +\infty$  for  $x < 0$ .

Then  $\phi^*$  satisfies the following properties:

(i)  $\phi^*$  is convex;

(ii)  $\phi^*$  is nondecreasing;

(iii)  $\phi^*(x) \geq x$  for all  $x \in \mathbb{R}$ ;

(iv) If  $\phi^*$  is a finite constant on an interval  $[a, b]$  for  $a, b \in \mathbb{R}$  and  $a < b$ , then  $\phi^*$  is a finite constant on the interval  $(-\infty, b]$ .

Based on properties (ii) and (iv), we have the following definition.

**Definition 1** Let  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  be a convex function such that  $\phi(1) = 0$  and  $\phi(x) = +\infty$  for  $x < 0$ .

Define  $\underline{m}(\phi^*) := \sup\{m \in \mathbb{R} : \phi^* \text{ is a finite constant on } (-\infty, m]\}$  and  $\bar{m}(\phi^*) := \inf\{m \in \mathbb{R} : \phi^*(m) = +\infty\}$ .

**Theorem 2** Given density estimate  $f_0$ , let  $\mathbb{P}_0$  represent the probability distribution defined by  $f_0$ .

Then the DCC  $\inf_{D_\phi(f||f_0) \leq d} \mathbb{P}\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha$  can be reformulated as

$$\mathbb{P}_0\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha'_+, \quad (25)$$

$$\text{where } \alpha' = 1 - \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} \left\{ \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)} \right\},$$

and  $x_+ = \max\{x, 0\}$  for  $x \in \mathbb{R}$ , if one of the following two conditions is satisfied:

(i)  $\lim_{x \rightarrow +\infty} \phi(x)/x = +\infty$ ,

(ii)  $f_0(\xi) > 0$  almost surely on  $\mathbb{R}^K$ .

*Proof:* Denoting set  $C = \{\xi \in \mathbb{R}^K : \bar{A}\xi \leq \bar{b}\}$ , we rewrite the left-hand side of the chance constraint as

$$z_{\mathcal{D}_2} = \min_f \int_{\mathbb{R}^K} I_C(\xi) f(\xi) d\xi \quad (26a)$$

$$\text{s.t. } \int_{\mathbb{R}^K} f_0(\xi) \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) d\xi \leq d, \quad (26b)$$

$$\int_{\mathbb{R}^K} f(\xi) d\xi = 1, \quad (26c)$$

$$f(\xi) \geq 0, \quad \forall \xi \in \mathbb{R}^K, \quad (26d)$$

where constraint (26b) bounds the  $\phi$ -divergence  $D_\phi(f||f_0)$  from above by  $d$ , and constraints (26c) and (26d) guarantee that  $f$  is a density function. Since problem (26) is, once again, a semi-infinite problem, we resort to duality. The Lagrangian dual of problem (26) can be written as

$$\begin{aligned} L &= \sup_{z \geq 0, z_0 \in \mathbb{R}} \inf_{f(\xi) \geq 0} \left\{ \int_{\mathbb{R}^K} \left( I_C(\xi) f(\xi) - z_0 f(\xi) + z f_0(\xi) \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) \right) d\xi + z_0 - zd \right\} \\ &= \sup_{z \geq 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd + \inf_{f(\xi) \geq 0} \left\{ \int_{\mathbb{R}^K} \left[ (I_C(\xi) - z_0) f(\xi) + z f_0(\xi) \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) \right] d\xi \right\} \right\}, \end{aligned} \quad (27)$$

where  $z$  and  $z_0$  represent the dual variables of constraints (26b) and (26c), respectively. Strong duality between problems (26) and (27) yields that  $z_{\mathcal{D}_2} = L$  and there exist  $z^* \geq 0$  and  $z_0^*$  such that the supremum in problem (27) is attained. We discuss the following cases for the values of  $z^*$ .

First, suppose that  $z^* > 0$ . Then we have

$$\begin{aligned} L &= \sup_{z > 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \sup_{f(\xi) \geq 0} \left\{ \int_{\mathbb{R}^K} \left[ \left( \frac{z_0 - I_C(\xi)}{z} \right) f(\xi) - f_0(\xi) \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) \right] d\xi \right\} \right\} \\ &= \sup_{z > 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \sup_{f(\xi) \geq 0} \left\{ \int_{[f_0(\xi) > 0] \cup [f_0(\xi) = 0]} \left[ \left( \frac{z_0 - I_C(\xi)}{z} \right) f(\xi) - f_0(\xi) \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) \right] d\xi \right\} \right\} \\ &= \sup_{z > 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \sup_{f(\xi) \geq 0} \left\{ \int_{[f_0(\xi) > 0]} \left[ \left( \frac{z_0 - I_C(\xi)}{z} \right) \left( \frac{f(\xi)}{f_0(\xi)} \right) - \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) \right] f_0(\xi) d\xi \right\} \right\} \end{aligned} \quad (28)$$

$$\begin{aligned} &= \sup_{z > 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \int_{[f_0(\xi) > 0]} \sup_{f(\xi)/f_0(\xi) \geq 0} \left\{ \left( \frac{z_0 - I_C(\xi)}{z} \right) \left( \frac{f(\xi)}{f_0(\xi)} \right) - \phi\left(\frac{f(\xi)}{f_0(\xi)}\right) \right\} f_0(\xi) d\xi \right\} \\ &= \sup_{z > 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \int_{[f_0(\xi) > 0]} \phi^* \left( \frac{z_0 - I_C(\xi)}{z} \right) f_0(\xi) d\xi \right\} \end{aligned} \quad (29)$$

$$= \sup_{z > 0, z_0 \in \mathbb{R}} \left\{ z_0 - zd - z \mathbb{P}_0\{C\} \phi^* \left( \frac{z_0 - 1}{z} \right) - z(1 - \mathbb{P}_0\{C\}) \phi^* \left( \frac{z_0}{z} \right) \right\}, \quad (30)$$

where equality (28) follows from the two assumptions because (i) if  $\lim_{x \rightarrow +\infty} \phi(x)/x = +\infty$ , then whenever  $f_0(\xi) = 0$ ,  $f(\xi)$  has to be zero to achieve optimality (see property (C2) of  $\phi$  function in Section 1.2), and (ii) if  $f_0(\xi) > 0$  almost surely on  $\mathbb{R}^K$ , then the equality is clear. Equality (29) follows from the definition of conjugate  $\phi^*$ , and equality (30) follows from conditional probability, conditioning on the value of  $I_C(\xi)$ .

Furthermore, to make the DCC satisfied, we observe that we should let the decision variables  $z_0$  and  $z$  be such that  $z_0/z \in [\underline{m}(\phi^*), \overline{m}(\phi^*)]$ . To see that, we discuss the following cases:

**Case 1.** Suppose that  $z_0/z > \overline{m}(\phi^*)$ . Then we have  $\phi^*(z_0/z) = +\infty$  and so

$$z_0 - zd - z \mathbb{P}_0\{C\} \phi^* \left( \frac{z_0 - 1}{z} \right) - z(1 - \mathbb{P}_0\{C\}) \phi^* \left( \frac{z_0}{z} \right) = -\infty,$$

in which case the DCC is violated.

**Case 2.** Suppose that  $z_0/z < \underline{m}(\phi^*)$ . Then  $(z_0 - 1)/z < \underline{m}(\phi^*)$  and hence there exists a finite constant  $m$  such that

$$\phi^* \left( \frac{z_0 - 1}{z} \right) = \phi^* \left( \frac{z_0}{z} \right) = m.$$

Also, since  $\phi^*(x) = m$  for each  $x < \underline{m}(\phi^*)$  by definition of  $\underline{m}(\phi^*)$ , and  $x \leq \phi^*(x) = m$  by property (iii) in Lemma 2, we have  $\underline{m}(\phi^*) \leq m$ . It follows that

$$\begin{aligned} & z_0 - zd - z\mathbb{P}_0\{C\}\phi^* \left( \frac{z_0 - 1}{z} \right) - z(1 - \mathbb{P}_0\{C\})\phi^* \left( \frac{z_0}{z} \right) \\ &= z_0 - zd - zm \\ &< \underline{m}(\phi^*)z - zd - zm \leq 0, \end{aligned}$$

and hence the DCC is violated.

Hence, we can add constraint  $z_0/z \in [\underline{m}(\phi^*), \bar{m}(\phi^*)]$  into problem (30), and the DCC  $z_{\mathcal{D}_2} \geq 1 - \alpha$  is equivalent to that there exist  $z > 0$  and  $z_0$  such that  $z_0/z \in [\underline{m}(\phi^*), \bar{m}(\phi^*)]$  and

$$\mathbb{P}_0\{C\} \geq \frac{1 - z_0 - \alpha + z\phi^* \left( \frac{z_0}{z} \right) + zd}{z\phi^* \left( \frac{z_0}{z} \right) - z\phi^* \left( \frac{z_0 - 1}{z} \right)}. \quad (31)$$

Therefore, the reformulation (25) is obtained by simultaneously replacing  $z$  and  $z_0$  by  $1/z$  and  $(z_0 + z)/z$ , respectively, in inequality (31).

Second, suppose that  $z^* = 0$ . Then we have  $L = \sup_{z_0 \in \mathbb{R}} \left\{ z_0 + \inf_{f(\xi) \geq 0} \int_{\mathbb{R}^K} (I_C(\xi) - z_0) f(\xi) d\xi \right\}$ . It follows that

**Case 1.** If  $C \neq \mathbb{R}^K$ , then at optimality  $z_0^* = 0$  with  $L = 0$  because we require  $I_C(\xi) - z_0^* \geq 0$  for each  $\xi \in \mathbb{R}^K$  to achieve optimality. It follows that the DCC  $z_{\mathcal{D}_2} \geq 1 - \alpha$  is violated because  $\alpha \in (0, 1)$ . On the other side, we notice that the optimal value of (30) is no larger than that of (27), which indicates that the reformulation (25) is violated as well.

**Case 2.** If  $C = \mathbb{R}^K$ , then at optimality  $z_0^* = 1$  with  $L = 1$  because we require  $I_C(\xi) - z_0^* = 1 - z_0^* \geq 0$  for each  $\xi \in \mathbb{R}^K$  to achieve optimality. It follows that the DCC  $z_{\mathcal{D}_2} = 1 \geq 1 - \alpha$  is satisfied because  $\alpha > 0$ . For this case, the reformulation (25) is also satisfied because  $\mathbb{P}_0\{\bar{A}\xi \leq \bar{b}\} = \mathbb{P}_0\{C\} = 1$  when  $C = \mathbb{R}^K$  and  $1 - \alpha'_+ \leq 1$ .  $\blacksquare$

**Remark 5** *The two conditions in Theorem 2 are mild because (i) many types of  $\phi$ -divergence measure satisfy  $\lim_{x \rightarrow +\infty} \phi(x)/x = +\infty$ , e.g., KL divergence (with  $\phi(x) := x \log x - x + 1$ ),  $\chi$  divergence of order  $\theta$  (with  $\phi(x) := |x - 1|^\theta$  and  $\theta > 1$ ), and Cressie and Read divergence (with  $\phi(x) := \frac{1-\theta+\theta x-x^\theta}{\theta(1-\theta)}$  and  $\theta > 1$ ), and (ii) any density estimate  $f_0$  can be slightly modified to satisfy  $f_0 > 0$ , and an important example is a KDE  $f_0$  by choosing  $H(\cdot)$  as the standard normal density.*

**Corollary 5** *The DCC reformulation  $\mathbb{P}_0\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha'_+$  is a conservative approximation of the nominal chance constraint  $\mathbb{P}_0\{\bar{A}\xi \leq \bar{b}\} \geq 1 - \alpha$ , i.e.,*

$$1 - \alpha' = \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} \left\{ \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)} \right\} \geq 1 - \alpha.$$

*Proof:* See Appendix H. ■

Henceforth, we call the chance constraint (25) a “reformulated DCC.” Theorem 2 and corollary 5 show that a reformulated DCC is a classical chance constraint with the ambiguous probability distribution  $\mathbb{P}$  replaced by its estimate  $\mathbb{P}_0$ , and the risk level  $\alpha$  decreased to  $\alpha'$ . As compared to the classical chance constraints, reformulated DCCs provide us the following theoretical merits:

1. In terms of modeling, unlike relying on an ambiguous probability distribution  $\mathbb{P}$ , we waive the “perfect information” assumption and resort to  $\mathbb{P}_0$  which can be estimated from the historical data. Meanwhile, we can make a more accurate  $\mathbb{P}_0$  estimate with more data on hand.
2. In terms of algorithm development, the estimate  $\mathbb{P}_0$  is more accessible than the ambiguous  $\mathbb{P}$ . First, the samples taken from  $\mathbb{P}_0$ , which is deterministic, is more trustable than from a guess of the ambiguous  $\mathbb{P}$ . Second, we can facilitate the sampling procedure by choosing density functions that are easier to sample from. For example, we can develop  $\mathbb{P}_0$  as a KDE and choose function  $H(\cdot)$  as a standard normal distribution. This observation motivates us to solve optimization problems with reformulated DCCs by using the scenario approximation approach.
3. In terms of conservatism, it is intuitive that when faced with probability distribution ambiguity, one can reduce  $\alpha$  to make a classical chance constraint more conservative. Theorem 2 can help to quantify how much  $\alpha$  needs to be reduced, and hence accurately depicts the relationship between the risk level  $\alpha$  and the conservatism.

For  $\mathcal{D}_2$  under general  $\phi$ -divergence measures, the perturbed risk level  $\alpha'$  in the reformulated DCC can be obtained by solving a two-dimensional nonlinear optimization problem. Many off-the-shelf optimization software (e.g., GAMS and AIMMS) and general-purpose optimization packages (e.g., MINOS and SNOPT) can be used. In this paper, we take three types of  $\phi$ -divergence measure, i.e.,  $\chi$  divergence of order 2, KL divergence, and variation distance, as examples to show how the perturbed risk level can be obtained. We find that the perturbed risk level for both  $\chi$  divergence of order 2 and variation distance can be obtained in a closed form. For the KL divergence, it seems hard to obtain a closed-form perturbed risk level. However, we can efficiently compute it by using bisection line search, because the computation effort can be shown to be equivalent to minimizing a univariate convex function. We summarize our results in the following three propositions, whose proofs are provided in Appendices I-K for brevity.

**Proposition 5** *Suppose that we develop  $\mathcal{D}_2$  by using the  $\chi$  divergence of order 2 with  $\phi(x) := (x - 1)^2$  and  $\alpha < 1/2$ . Then the perturbed risk level is*

$$\alpha' = \alpha - \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1 - 2\alpha)d}{2d + 2}.$$

**Proposition 6** *Suppose that we develop  $\mathcal{D}_2$  by using the variation distance with  $\phi(x) := |x - 1|$ . Then the perturbed risk level is*

$$\alpha' = \alpha - \frac{d}{2}.$$

**Proposition 7** *Suppose that we develop  $\mathcal{D}_2$  by using the KL divergence with  $\phi(x) := x \log x - x + 1$ . Then the perturbed risk level is*

$$\alpha' = 1 - \inf_{x \in (0,1)} \left\{ \frac{e^{-d} x^{1-\alpha} - 1}{x - 1} \right\},$$

*and can be computed by using bisection line search after  $\lceil \log_2(\frac{1}{\epsilon}) \rceil$  steps to achieve  $\epsilon$  accuracy.*

### 3.3 The Value of Data

Intuitively, as the sample size  $N$  increases we can depict the profile of the ambiguous probability distribution  $\mathbb{P}$  more accurately with a smaller  $\phi$ -divergence tolerance  $d$ . Based on Theorem 2, one may be interested in two questions:

- Will the perturbed risk level  $\alpha'$  increase to  $\alpha$  as  $d$  tends to zero (resp. as  $N$  tends to infinity)?



- How fast will  $\alpha'$  converge as  $d$  decreases (resp. as  $N$  increases)?

These two questions are important in practice. First, an affirmative answer to the first question indicates that the conservatism of the DCCs will vanish as  $N$  tends to infinity. Second, since collecting data frequently incurs cost, the convergence rate of  $\alpha'$  helps to estimate how a new set of data decreases the conservatism of DCC, i.e., the value of data. More specifically, for nominal risk level given  $\alpha$ , we can define the *value of data* as

$$\text{VoD}_\alpha = \frac{d\alpha'}{dN},$$

which represents the increase of  $\alpha'$  value if we marginally enlarge the data set. In this subsection, we answer the first question in the affirmative for  $\mathcal{D}_2$  under a general  $\phi$ -divergence. Moreover, we focus on two types of  $\phi$ -divergence, i.e., the  $\chi$  divergence of order 2, and the KL divergence, and discuss the convergence behavior of  $\alpha'$  and their corresponding  $\text{VoD}_\alpha$ . For discussion consistency, we assume that  $d$  is chosen to be a differential function of  $N$ , i.e.,  $d = d(N)$ ,  $d$  is nonincreasing as  $N$  increases, and  $d$  tends to zero as  $N$  tends to infinity. For example, we can follow Pardo [26] and set the divergence tolerance  $d = \phi''(1)\chi_{B-1,1-\beta}^2/(2N)$ . We first show the following proposition for the general  $\phi$ -divergence, whose detailed proof is provided in Appendix L.

**Proposition 8** *For a general  $\phi$ -divergence with  $x = 1$  as its unique minimizer, the perturbed risk level  $\alpha'$  defined as*

$$1 - \inf_{\substack{z>0, \\ \underline{m}(\phi^*) \leq z_0+z \leq \bar{m}(\phi^*)}} \left\{ \frac{\phi^*(z_0+z) - z_0 - \alpha z + d}{\phi^*(z_0+z) - \phi^*(z_0)} \right\}$$

*increases to  $\alpha$  as  $d$  decreases to zero (resp. as  $N$  increases to infinity).*

Next we show the convergence result of  $\alpha'$  and the corresponding  $\text{VoD}_\alpha$  for the  $\chi$  divergence of order 2 and the KL divergence, respectively. The detailed proofs are provided in Appendices M-N.

**Proposition 9** *Suppose that we develop  $\mathcal{D}_2$  by using the  $\chi$  divergence of order 2. Then  $\alpha'$  increases to  $\alpha$  as  $d$  decreases to zero. Furthermore, the value of data satisfies*

$$\text{VoD}_\alpha = \frac{1}{2(d(N)+1)^2} \left[ \frac{(2\alpha^2 - 2\alpha + 1)d(N) + 2(\alpha - \alpha^2)}{\sqrt{d(N)^2 + 4d(N)(\alpha - \alpha^2)}} - (1 - 2\alpha) \right] d'(N),$$

*where  $d'(N)$  represents the derivative of  $d(N)$  over  $N$ .*

**Proposition 10** *Suppose that we develop  $\mathcal{D}_2$  by using the KL divergence. Then  $\alpha'$  increases to  $\alpha$  as  $d$  decreases to zero. Furthermore, we can obtain*

$$d = \alpha \log\left(\frac{\alpha}{\alpha'}\right) + (1 - \alpha) \log\left(\frac{1 - \alpha}{1 - \alpha'}\right),$$

*and the value of data satisfies*

$$\text{VoD}_\alpha = \left[ \frac{\alpha'(1 - \alpha')}{\alpha' - \alpha} \right] d'(N),$$

*where  $d'(N)$  represents the derivative of  $d(N)$  over  $N$ .*

We remark that the value of  $\text{VoD}_\alpha$  depends on the choice of  $d(N)$ , i.e., the relationship between the  $\phi$ -divergence tolerance  $d$  and the data sample size  $N$ . For discrete distributions, the choice  $d(N) := \phi''(1)\chi_{\bar{B}-1, 1-\beta}^2/(2N)$  and accordingly  $\text{VoD}_\alpha$  becomes accurate as  $N$  increases to infinity. For continuous distributions,  $d(N)$  is used to approximate  $\mathcal{D}_2$ , and accordingly  $\text{VoD}_\alpha$  cannot guarantee to be accurate in general. In practice, we can use  $\text{VoD}_\alpha$  to help decision makers approximately estimate the value of a new set of data, and accurate estimators worth future studies. For illustration, we depict the relationships between  $1 - \alpha'$  and  $N$  and between  $\text{VoD}_\alpha$  and  $N$  in Figure 1 by using KL divergence under confidence levels with  $\alpha = 0.90$  and bin size  $\bar{B} = 30$  (e.g., for discrete distributions). We observe that with sample size  $N$  increasing, both  $1 - \alpha'$  and  $\text{VoD}_\alpha$  decay quickly. However, we need a large sample size (e.g.,  $N > 2000$ ) to guarantee  $\alpha'$  converging to  $\alpha$  and  $\text{VoD}_\alpha$  converging to zero. That is, we have to draw a large set of historical data to guarantee an almost exact description of the unknown probability distribution  $\mathbb{P}$ , and accordingly the risk level of the reformulated DCC,  $\alpha'$ , can be chosen to be near its deterministic counterpart  $\alpha$ . This observation makes sense because we need a large sample size to construct the histogram in the first place, especially when the dimension of random vectors becomes very large. In practice, we suggest using density-based confidence sets in an industry that has rich access to data and relies heavily on data to make decisions. As compared to density-based confidence sets, moment-based confidence sets are more conservative and only need small data sets for construction, and hence, are suitable to be used in an industry that has limited access to data.

## 4 Numerical Experiments

In this section, we conduct a simple numerical experiment to illustrate the application of DCCPs. We model DCCs with both moment-based and density-based confidence sets in a portfolio opti-

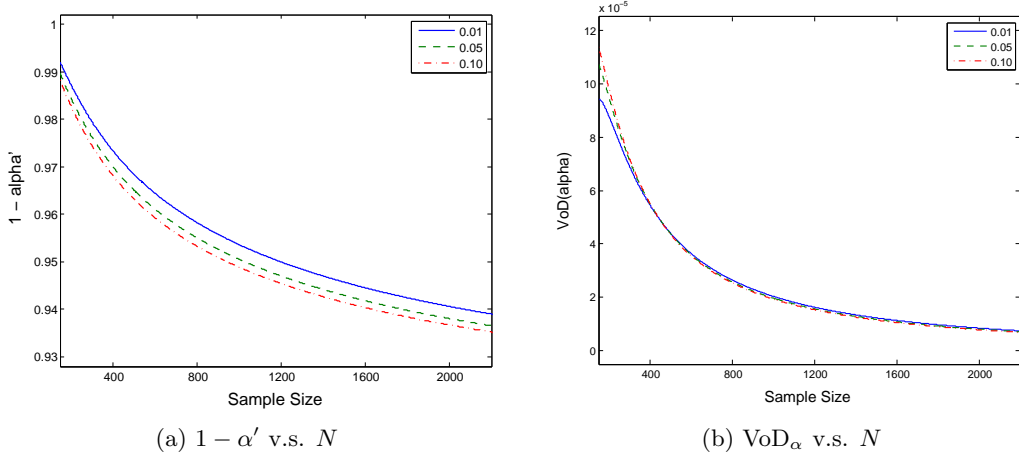


Figure 1: Evolution of values  $1 - \alpha'$  and  $\text{VoD}_\alpha$  against sample size under risk level  $\alpha = 0.90$  and confidence levels  $\beta = 0.01, 0.05, 0.10$

mization problem. In this experiment, a generic DCCP for the portfolio optimization problem can be formulated as

$$\begin{aligned}
 \text{[DCPO]} \quad & \max_{x \geq 0} \sum_{i=1}^n \mathbb{E}[\xi_i] x_i \\
 \text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{D}} \mathbb{P} \left\{ \begin{array}{l} \sum_{i=1}^n \xi_i x_i \geq T_0(\sum_{i=1}^n x_i), \\ \sum_{i \in N_j} \xi_i x_i \geq T_j(\sum_{i \in N_j} x_i), \forall j = 1, \dots, J \end{array} \right\} \geq 1 - \alpha, \quad \sum_{i=1}^n x_i = 1,
 \end{aligned}$$

where  $n$  represents the total number of investments,  $\xi_i$  represents the rate of return (RoR) of investment  $i$ ,  $x_i$  represents the share of investment  $i$ ,  $N_1, \dots, N_J$  represent different portfolio segments with  $\bigcup_{j=1}^J N_j = \{1, \dots, n\}$  (e.g.,  $N_1$  consists of stocks,  $N_2$  consists of bonds, and so on), and  $T_0, \dots, T_J$  represent the investment targets of different portfolio segments. To specify which confidence set we use, we denote [DCPO-M] when using a moment-based confidence set (e.g.,  $\mathcal{D} = \mathcal{D}_1$ ), and denote [DCPO-D] when using a density-based confidence set (e.g.,  $\mathcal{D} = \mathcal{D}_2$  under the KL divergence). Inspired by Delage and Ye [12], we evaluate [DCPO] in this experiment by using a historical data set of 30 assets from years 2008 to 2011, obtained from the Yahoo! Finance website<sup>1</sup>. In each experiment, we randomly choose 4 assets, randomly assign them into  $J = 2$  portfolio segments, and build a dynamic portfolio with these assets. The assets in the portfolio are updated every thirty transaction days, through years 2008 to 2011, by adopting optimal investment decisions obtained

<sup>1</sup>The 30 assets are AAR Corp., AT&T, Avery Denison Corp., Boeing Corp., Bristol-Myers-Squibb, Cisco Systems, Dell Computer Corp., Dow Chemical, Duke Energy Company, Du Pont, Eli Lilly and Co., Exelon Corp., FMC Corp., General Electric, Hewlett Packard, Hitachi, Honeywell, IBM Corp., Ingersoll Rand, Intel Corp., Lockheed Martin, Merck and Co., Microsoft, Motorola, Northern Telecom, Oracle, Pinnacle West, Texas Instruments, United Technologies and a 0%-interest-rate deposit.

from [DCPO]. During any day of the experiment, we collect the most recent 2000 days of RoR data to construct both confidence sets  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . In this experiment, we estimate each mean RoR, i.e.,  $\mathbb{E}[\xi_i]$  for each  $i = 1, \dots, n$ , by the sample mean RoR based on the most recent 2000 days data. We employ the conservative approximation discussed in Section 2.3 to solve [DCPO-M], and the scenario approximation approach to solve [DCPO-D].

In this experiment, we evaluate the performance of the investment decisions obtained from [DCPO] during each trading day against the real data in the following thirty days. That is, after making the investment decision during each trading day, we will hold the assets for thirty days and see how it performs in the real market. To set a benchmark, we compare [DCPO-M] and [DCPO-D] to a myopic model, which maximizes an average return over the last 2000 days. In this experiment, we run 100 replications in total by choosing four assets (among the 30 assets) in each replication, and summarize the results obtained from all replications in Table 1 and Figure 2.

	Avg.	St. dev	10th Perc.	90th Perc.
DCPO-D	1.118	0.194	0.969	1.415
DCPO-M	1.010	0.166	0.823	1.245
Myopic	0.991	0.507	0.499	1.394

Table 1: Comparison of average end wealth and risk in 100 replications through years 2008-2011

From Table 1 and Figure 2, we observe that both [DCPO-D] and [DCPO-M] outperform the Myopic approach in both end wealth and risk control for the years 2008-2011. In particular, [DCPO-D] largely outperforms the other two approaches with at least 9.6% more in end wealth and a much smaller standard deviation as compared to the Myopic approach. It indicates that the density-based DCCPs can make robust and profitable portfolio selection. [DCPO-M] slightly outperforms the Myopic approach in terms of end wealth, but has the smallest standard deviation. From Figure 2, we can also observe that [DCPO-M] clearly outperforms the other two approaches during years 2008-2009 when the market is depressed. This observation makes sense because [DCPO-M] considers moment ambiguity and we employ a conservative approximation to solve [DCPO-M].

## 5 Conclusion

In this paper, we developed exact and approximate approaches for DCCPs. Starting from the historical data, we described how to construct moment-based and density-based confidence sets

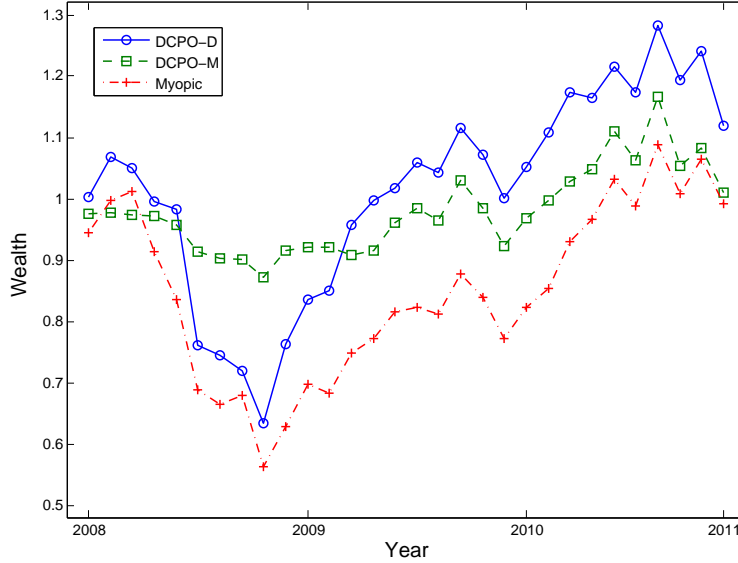


Figure 2: Comparison of wealth evolution in 100 replications through years 2008-2011

for the ambiguous probability distributions, how to equivalently reformulate DCCs, and how to effectively solve DCCPs. In general, in this study, we proposed a framework to provide robust decisions based on the available data set information. Besides guaranteeing the robustness, our framework ensures that the proposed approach is less conservative when more data information is on hand. Possible future research directions include the study of DCCs under different confidence sets and their solution approaches. It is also interesting to study the accurate estimators of the value of data for general  $\phi$ -divergence.

## Acknowledgments

This research was partly supported by the U.S. National Science Foundation under CAREER Award CMMI0942156 and the U.S. Office of Naval Research under Young Investigator Award N000141010749.

## References

- [1] A. Ben-Tal, D. den Hertog, A. De Waegenare, B. Melenberg, and G. Rennen. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357, 2013.

- [2] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.
- [3] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
- [4] G.C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18(3):853–877, 2007.
- [5] G.C. Calafiore and M.C. Campi. Uncertain convex programs: Randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.
- [6] G.C. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
- [7] G.C. Calafiore and L. El Ghaoui. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1):1–22, 2006.
- [8] A. Charnes and W.W. Cooper. Deterministic equivalents for optimizing and satisficing under chance constraints. *Operations Research*, 11(1):18–39, 1963.
- [9] A. Charnes, W.W. Cooper, and G.H. Symonds. Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil. *Management Science*, 4(3):235–263, 1958.
- [10] W. Chen, M. Sim, J. Sun, and C.P. Teo. From CVaR to uncertainty set: Implications in joint chance constrained optimization. *Operations Research*, 58(2):470–485, 2010.
- [11] X. Chen and Y. Zhang. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research*, 57(6):1469–1482, 2009.
- [12] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.
- [13] L. Devroye and L. Györfi. *Nonparametric Density Estimation: The  $\ell_1$  View*. John Wiley & Sons Inc, 1985.
- [14] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.
- [15] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1):37–61, 2006.
- [16] K. Isii. The extrema of probability determined by generalized moments (I) bounded random variables. *Annals of the Institute of Statistical Mathematics*, 12(2):119–134, 1960.
- [17] S. Küçükyavuz. On mixing sets arising in chance-constrained programming. *Mathematical Programming*, 132(1–2):31–56, 2010.
- [18] M. A. Lejeune. Pattern-based modeling and solution of probabilistically constrained optimiza-

- tion problems. *Operations Research*, 60(6):1356–1372, 2012.
- [19] J. Luedtke. A branch-and-cut decomposition algorithm for solving general chance-constrained mathematical programs. *Preprint available at www.optimization-online.org*, 2011.
- [20] J. Luedtke and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization*, 19(2):674–699, 2008.
- [21] J. Luedtke, S. Ahmed, and G.L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122(2):247–272, 2010.
- [22] B.L. Miller and H.M. Wagner. Chance constrained programming with joint constraints. *Operations Research*, 13(6):930–945, 1965.
- [23] A. Nemirovski and A. Shapiro. Scenario approximations of chance constraints. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*, pages 3–47. Springer, 2006.
- [24] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007.
- [25] 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.
- [26] L. Pardo. *Statistical Inference Based on Divergence Measures*, volume 185. CRC Press, 2006.
- [27] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- [28] J. Pintér. Deterministic approximations of probability inequalities. *Mathematical Methods of Operations Research*, 33(4):219–239, 1989.
- [29] I. Pólik and T. Terlaky. A survey of the S-Lemma. *SIAM Review*, 49(3):371–418, 2007.
- [30] A. Prékopa. On probabilistic constrained programming. In *Proceedings of the Princeton Symposium on Mathematical Programming*, pages 113–138. Citeseer, 1970.
- [31] A. Prékopa. *Stochastic Programming*. Springer, 1995.
- [32] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.
- [33] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.
- [34] A. Shapiro. On duality theory of conic linear problems. In *Semi-Infinite Programming*, pages 135–165. Kluwer Academic Publishers, 2000.

- [35] A. Shapiro and K. Scheinberg. Duality and optimality conditions. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, volume 27 of *International Series in Operations Research & Management Science*, pages 67–110. Springer, 2000.
- [36] J. Smith. Generalized Chebychev inequalities: Theory and applications in decision analysis. *Operations Research*, 43(5):807–825, 1995.
- [37] A.L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.
- [38] L. Vandenberghe, S. Boyd, and K. Comanor. Generalized Chebyshev bounds via semidefinite programming. *SIAM Review*, 49(1):52–64, 2007.
- [39] H. Xu, C. Caramanis, and S. Mannor. Optimization under probabilistic envelope constraints. *Operations Research*, 60(3):682–699, 2012.
- [40] V. Yakubovich. S-procedure in nonlinear control theory. *Vestnik Leningrad University*, 4:73–93, 1977.
- [41] S. Zymmler, D. Kuhn, and B. Rustem. Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, 137(1-2):167–198, 2013.



## Appendix A S-Lemma

We state the S-Lemma as follows:

**Lemma 3** (*S-Lemma, Yakubovich [40]*) *Let  $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$  be quadratic functions and suppose that there is an  $\bar{x} \in \mathbb{R}^n$  such that  $g(\bar{x}) < 0$ . Then the following two statements are equivalent.*

(i) *There is no  $x \in \mathbb{R}^n$  such that*

$$\begin{cases} f(x) < 0 \\ g(x) \leq 0. \end{cases}$$

(ii) *There is a nonnegative number  $y \geq 0$  such that*

$$f(x) + yg(x) \geq 0, \quad \forall x \in \mathbb{R}^n.$$

## Appendix B Proof of Corollary 2

*Proof:* First, we let  $z_{\mathcal{D}_1^{\text{pol}}}$  represent the optimal objective value of the optimization problem on the left-hand side of the DCC, i.e.,  $z_{\mathcal{D}_1^{\text{pol}}} = \inf_{\mathbb{P} \in \mathcal{D}_1^{\text{pol}}} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\}$ . We specify  $z_{\mathcal{D}_1^{\text{pol}}}$  by the following optimization problem:

$$[\text{Primal}_{\text{pol}}] \quad z_{\mathcal{D}_1^{\text{pol}}} = \min_{\mathbb{P}, \lambda \geq 0} \int_{\mathbb{R}^K} I_{[\bar{a}^\top \xi < \bar{b}]}(\xi) d\mathbb{P} \quad (32a)$$

$$\text{s.t.} \quad \int_{\mathbb{R}^K} \xi d\mathbb{P} = \sum_{i=1}^I \lambda_i \mu_i, \quad (32b)$$

$$\int_{\mathbb{R}^K} \xi \xi^\top d\mathbb{P} = \sum_{i=1}^I \lambda_i \Sigma_i, \quad (32c)$$

$$\int_{\mathbb{R}^K} d\mathbb{P} = 1, \quad (32d)$$

$$\sum_{i=1}^I \lambda_i = 1, \quad (32e)$$

where constraints (32b), (32c), and (32e) guarantee that  $(\mathbb{E}[\xi], \mathbb{E}[\xi \xi^\top]) \in \text{conv}(\mathcal{P})$ . We apply the duality theory for conic linear programming problems and dualize problem (32) as

$$[\text{Dual}_{\text{pol}}] \quad z_{\mathcal{D}_1^{\text{pol}}} = \max_{H, p, q, r} \quad r + q \quad (33a)$$

$$\text{s.t.} \quad \Sigma_i \cdot H + \mu_i^\top p - r \geq 0, \quad \forall i = 1, \dots, I, \quad (33b)$$

$$\xi^\top H \xi + p^\top \xi + q \leq I_{[\bar{a}^\top \xi < \bar{b}]}(\xi), \quad \forall \xi \in \mathbb{R}^K, \quad (33c)$$

where  $p, H \in \mathbb{S}^{K \times K}$ ,  $q$ , and  $r$  are dual variables for constraints (32b), (32c), (32d), and (32e) respectively. Constraint (33c) can be reformulated as the following LMIs by using Lemma 1:

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2}y\bar{a} \\ \frac{1}{2}y\bar{a}^\top & 1 - y\bar{b} \end{bmatrix}, \quad (34a)$$

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad (34b)$$

where  $y \geq 0$ . Again, note here that strong duality holds for problems [Primal<sub>pol</sub>] and [Dual<sub>pol</sub>] based on established conic linear programming theory. Hence, the DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1^{\text{pol}}} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\} \geq 1 - \alpha$  can be reformulated as

$$-r + (1 - q) \leq \alpha, \quad (35a)$$

$$\Sigma_i \cdot H + \mu_i^\top p - r \geq 0, \quad \forall i = 1, \dots, I, \quad (35b)$$

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2}y\bar{a} \\ \frac{1}{2}y\bar{a}^\top & 1 - y\bar{b} \end{bmatrix}, \quad (35c)$$

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \in \mathbb{S}_+^{(K+1) \times (K+1)}, \quad y \geq 0. \quad (35d)$$

Now we observe that  $y > 0$ , because otherwise (i.e.,  $y = 0$ ) we have

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

by constraint (35c), and so

$$\begin{aligned} -r + (1 - q) &\geq -\Sigma_i \cdot H - \mu_i^\top p + (1 - q) && \text{(for any } i = 1, \dots, I, \text{ by constraint (35b))} \\ &= \begin{bmatrix} \Sigma_i & \mu_i \\ \mu_i^\top & 1 \end{bmatrix} \cdot \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \\ &\geq \begin{bmatrix} \Sigma_i & \mu_i \\ \mu_i^\top & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 1, \end{aligned} \quad (36)$$

where the second inequality holds because

$$\begin{bmatrix} \Sigma_i & \mu_i \\ \mu_i^\top & 1 \end{bmatrix} \succeq 0 \quad \text{and} \quad \begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \succeq 0.$$

Observation (36) violates constraint (35a) because  $\alpha \in (0, 1)$ . Hence, we let  $\bar{y} = 1/y$  and replace  $H, p, q, r$ , and  $(1 - q)$  by their  $y$ -multipliers, i.e.,  $yH, yp, yq, yr$ , and  $y(1 - q)$  in LMIs (35), and obtain the reformulation as LMIs (19).  $\blacksquare$

## Appendix C Proof of Corollary 3

*Proof:* First, we let  $z_{\mathcal{D}_1^{\text{mgn}}}$  represent the optimal objective value of the optimization problem on the left-hand side of the DCC, i.e.,  $z_{\mathcal{D}_1^{\text{mgn}}} = \inf_{\mathbb{P} \in \mathcal{D}_1^{\text{mgn}}} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\}$ . We specify  $z_{\mathcal{D}_1^{\text{mgn}}}$  by the following optimization problem:

$$[\text{Primal}_{\text{mgn}}] \quad z_{\mathcal{D}_1^{\text{mgn}}} = \min_{\mathbb{P}} \int_{\mathbb{R}^K} I_{[\bar{a}^\top \xi < \bar{b}]}(\xi) d\mathbb{P} \quad (37a)$$

$$\text{s.t.} \quad \mu^L \leq \int_{\mathbb{R}^K} \xi d\mathbb{P} \leq \mu^U, \quad (37b)$$

$$\int_{\mathbb{R}^K} \xi_k^2 d\mathbb{P} \leq \sigma_k, \quad \forall k = 1, \dots, K, \quad (37c)$$

$$\int_{\mathbb{R}^K} d\mathbb{P} = 1, \quad (37d)$$

where constraints (37b) describe the confidence interval of  $\mathbb{E}[\xi]$  and constraints (37c) describe the upper bound of  $\mathbb{E}[\xi_k^2]$  for each  $k = 1, \dots, K$  respectively. We apply the duality theory for conic linear programming problems and dualize problem (37) as

$$[\text{Dual}_{\text{mgn}}] \quad z_{\mathcal{D}_1^{\text{mgn}}} = \max_{p, h, q} \sum_{k=1}^K (p_k^L \mu_k^L - p_k^U \mu_k^U - \sigma_k h_k) + q \quad (38a)$$

$$\text{s.t.} \quad -\sum_{k=1}^K h_k \xi_k^2 + (p^L - p^U)^\top \xi + q \leq I_{[\bar{a}^\top \xi < \bar{b}]}(\xi), \quad \forall \xi \in \mathbb{R}^K, \quad (38b)$$

$$p^L, p^U, h \geq 0, \quad (38c)$$

where vectors  $p^L$  and  $p^U$ , vector  $h$ , and scalar  $q$  represent the dual variables for constraints (37b), (37c), and (37d), respectively. Again, note here that strong duality holds for problems  $[\text{Primal}_{\text{mgn}}]$  and  $[\text{Dual}_{\text{mgn}}]$  based on established conic linear programming duality theory. To apply Lemma 1, we first simplify the notation by defining

$$p = p^L - p^U \quad \text{and} \quad H = \text{diag}(-h_1, \dots, -h_K). \quad (39)$$

Now by Lemma 1, constraint (38b) is equivalent to the following constraints

$$\begin{bmatrix} -H & -\frac{1}{2}p \\ -\frac{1}{2}p^\top & 1 - q \end{bmatrix} \succeq 0, \quad (40)$$

$$\begin{bmatrix} -H & -\frac{1}{2}(p + y\bar{a}) \\ -\frac{1}{2}(p + y\bar{a})^\top & y\bar{b} - q \end{bmatrix} \succeq 0, \quad (41)$$

$$y \geq 0.$$

We now show that (40) is second-order cone representable, and the reformulation of (41) follows similarly:

$$\begin{aligned}
& \begin{bmatrix} h_1 & & & \\ & \ddots & & -\frac{1}{2}p \\ & & h_K & \\ & -\frac{1}{2}p^\top & & 1-q \end{bmatrix} \succeq 0 \\
\Leftrightarrow & \sum_{k=1}^K h_k \xi_k^2 - \sum_{k=1}^K p_k \xi_k \xi_{K+1} + (1-q) \xi_{K+1}^2 \geq 0, \quad \forall \xi_1, \dots, \xi_K, \xi_{K+1} \in \mathbb{R} \\
\Leftrightarrow & \begin{cases} h_k = 0 \Rightarrow p_k = 0, \quad \forall k = 1, \dots, K, \\ \sum_{\substack{k=1 \\ h_k \neq 0}}^K h_k \left[ \xi_k - \left( \frac{p_k}{2h_k} \right) \xi_{K+1} \right]^2 + \left( 1 - q - \sum_{\substack{k=1 \\ h_k \neq 0}}^K \frac{p_k^2}{4h_k} \right) \xi_{K+1}^2 \geq 0, \quad \forall \xi_1, \dots, \xi_K, \xi_{K+1} \in \mathbb{R} \end{cases} \\
\Leftrightarrow & \begin{cases} h_k = 0 \Rightarrow p_k = 0, \quad \forall k = 1, \dots, K, \\ h_k \geq 0, \\ 1 - q - \sum_{\substack{k=1 \\ h_k \neq 0}}^K \frac{p_k^2}{4h_k} \geq 0 \end{cases} \\
\Leftrightarrow & \exists t_k \geq 0, \quad \forall k = 1, \dots, K, \text{ such that} \\
& \begin{cases} p_k^2 \leq 4h_k t_k, \quad \forall k = 1, \dots, K, \\ q + \sum_{k=1}^K t_k \leq 1 \end{cases} \\
\Leftrightarrow & \exists t_k \geq 0, \quad \forall k = 1, \dots, K, \text{ such that} \\
& \begin{cases} \left\| \begin{bmatrix} p_k \\ t_k - h_k \end{bmatrix} \right\| \leq t_k + h_k, \quad \forall k = 1, \dots, K, \\ q + \sum_{k=1}^K t_k \leq 1. \end{cases} \tag{42}
\end{aligned}$$

By substituting the simplification (39) back into equation (42), we obtain the reformulation of constraint (40) as follows

$$\begin{cases} \left\| \begin{bmatrix} p_k^L - p_k^U \\ t_k - h_k \end{bmatrix} \right\| \leq t_k + h_k, \quad \forall k = 1, \dots, K, \\ \sum_{k=1}^K t_k \leq 1 - q \\ t_k \geq 0, \quad \forall k = 1, \dots, K. \end{cases}$$

The formulation of constraint (41) can be similarly obtained as follows

$$\begin{cases} \left\| \begin{bmatrix} p_k^L - p_k^U + y\bar{a}_k \\ s_k - h_k \end{bmatrix} \right\| \leq s_k + h_k, \quad \forall k = 1, \dots, K, \\ q - y\bar{b} + \sum_{k=1}^K s_k \leq 0 \\ s_k \geq 0, \quad \forall k = 1, \dots, K. \end{cases}$$

Hence, the DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1^{\text{mgn}}} \mathbb{P}\{\bar{a}^\top \xi < \bar{b}\} \geq 1 - \alpha$  is equivalent to

$$\sum_{k=1}^K (p_k^U \mu_k^U - p_k^L \mu_k^L + \sigma_k h_k) + (1 - q) \leq \alpha, \quad (43a)$$

$$\left\| \begin{bmatrix} p_k^L - p_k^U \\ t_k - h_k \end{bmatrix} \right\| \leq t_k + h_k, \quad \forall k = 1, \dots, K, \quad (43b)$$

$$\sum_{k=1}^K t_k \leq 1 - q, \quad (43c)$$

$$\left\| \begin{bmatrix} p_k^L - p_k^U + y\bar{a}_k \\ s_k - h_k \end{bmatrix} \right\| \leq s_k + h_k, \quad \forall k = 1, \dots, K, \quad (43d)$$

$$1 + \sum_{k=1}^K s_k \leq y\bar{b} + (1 - q), \quad (43e)$$

$$y \geq 0, t_k \geq 0, s_k \geq 0, \quad \forall k = 1, \dots, K. \quad (43f)$$

Now we observe that  $y > 0$ , because otherwise (i.e.,  $y = 0$ ) we have

$$\begin{bmatrix} -H & -\frac{1}{2}(p^L - p^U) \\ -\frac{1}{2}(p^L - p^U)^\top & 1 - q \end{bmatrix} \succeq \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

by (41), and it follows that

$$\begin{aligned} \sum_{k=1}^K (p_k^U \mu_k^U - p_k^L \mu_k^L + \sigma_k h_k) + (1 - q) &= \begin{bmatrix} \Sigma_{\text{Diag}} & \mu^L \\ (\mu^L)^\top & 1 \end{bmatrix} \cdot \begin{bmatrix} -H & -\frac{1}{2}(p^L - p^U) \\ -\frac{1}{2}(p^L - p^U)^\top & 1 - q \end{bmatrix} + (\mu^U - \mu^L)^\top p^U \\ &\geq \begin{bmatrix} \Sigma & \mu^L \\ (\mu^L)^\top & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 1, \end{aligned}$$

where  $\Sigma_{\text{Diag}} := \text{diag}(\sigma_1, \dots, \sigma_K)$ , and so constraint (43a) is violated because  $\alpha \in (0, 1)$ . Hence, we let  $\bar{y} = 1/y \geq 0$  and replace  $p^U$ ,  $p^L$ ,  $h$ ,  $t$ ,  $s$ , and  $(1 - q)$  with their  $y$ -multipliers, i.e.,  $yp^U$ ,  $yp^L$ ,  $yh$ ,  $yt$ ,  $ys$ , and  $y(1 - q)$ , respectively. This replacement shows that constraints (43) are equivalent to constraints (20), and completes the proof.  $\blacksquare$

## Appendix D Proof of Proposition 1

*Proof:* Following Theorem 1, we observe that the optimization problem under DCC can be reformulated as a convex semidefinite program as follows:

$$z(\text{int}(C)) = \min_{x \in X} \psi(x) \quad (44a)$$

$$\text{s.t.} \quad \begin{bmatrix} \gamma_2 \Lambda & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} + \begin{bmatrix} \Lambda & 0 \\ 0 & \gamma_1 \end{bmatrix} \cdot \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \preceq \alpha \bar{y}, \quad (44b)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \succeq \begin{bmatrix} 0 & \frac{1}{2} \bar{a} \\ \frac{1}{2} \bar{a}^\top & \bar{y} + \bar{a}^\top \mu - \bar{b} \end{bmatrix}, \quad (44c)$$

$$\begin{bmatrix} G & -p \\ -p^\top & 1-r \end{bmatrix} \in \mathbb{S}^{(K+1) \times (K+1)}, \quad \begin{bmatrix} H & p \\ p^\top & q \end{bmatrix} \in \mathbb{S}^{(K+1) \times (K+1)}, \quad \bar{y} \geq 0, \quad (44d)$$

where  $\bar{a}$  and  $\bar{b}$  are affine functions of  $x$  as discussed in Section 1.2. For any given  $\delta > 0$ , let  $z(\text{int}(C + \delta))$  represent the optimal objective value of problem (44a)-(44d) with scalar  $\bar{b}$  replaced by  $\bar{b} + \delta$ . It follows that

$$z(\text{int}(C)) \leq z(C) \leq z(\text{int}(C + \delta)), \quad (45)$$

because

$$\mathbb{P}\{\xi \in \text{int}(C)\} \leq \mathbb{P}\{\xi \in C\} \leq \mathbb{P}\{\xi \in \text{int}(C + \delta)\},$$

where  $C + \delta := \{\xi \in \mathbb{R}^K : \bar{a}^\top \xi \leq \bar{b} + \delta\}$ . By an established continuity result for convex semidefinite program (see Theorem 4.1.9 in Shapiro and Scheinberg [35]),  $z(\text{int}(C + \delta))$  is continuous at  $\delta = 0$ , i.e.,  $\lim_{\delta \downarrow 0} z(\text{int}(C + \delta)) = z(\text{int}(C))$ , because  $X$  is a convex set,  $\psi(\cdot)$  is a convex function, and the Slater condition holds for problem (44a)-(44d) by assumption. Therefore, the proof is complete in view of (45).  $\blacksquare$

**Remark 6** *It can be shown that a sufficient condition of the Slater condition for problem (44a)-(44d) is as follows: there exists an  $\alpha' < \alpha$ , such that with  $\alpha$  replaced by  $\alpha'$  there exists a feasible solution  $(x', G', H', p', r', q', \bar{y}')$  to problem (44a)-(44d) with  $\bar{y}' > 0$ .*

## Appendix E Proof of Proposition 2

*Proof:* First, we reformulate the nonlinear DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\xi^\top M \xi + c^\top \xi + d \leq 0\} \geq 1 - \alpha$  by using Corollary 1 and Proposition 1, and obtain LMIs (22a), (22b), and (22d).

Second, to ensure a conservative approximation, we ensure that  $\xi^\top M \xi + c^\top \xi + d \geq \bar{a}_i^\top \xi - \bar{b}_i$  for all  $\xi \in \mathbb{R}^K$  and all  $i = 1, \dots, m$ , which implies that each quadratic  $\xi^\top M \xi + c^\top \xi + d - \bar{a}_i^\top \xi + \bar{b}_i$  is

nonnegative on  $\mathbb{R}^K$ . Hence, this statement can be reformulated as LMIs (22c).  $\blacksquare$

## Appendix F Proof of Proposition 4

*Proof:* We establish the equivalence between WVaR constraint (24) and the single DCC  $\inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top \xi \leq \bar{b}\} \geq 1 - \alpha$  as follows:

$$\inf \left\{ \ell \in \mathbb{R} : \mathbb{P}\{\bar{a}^\top \xi \leq \ell\} \geq 1 - \alpha \right\} \leq \bar{b}, \quad \forall \mathbb{P} \in \mathcal{D}_1 \quad (46)$$

$$\Leftrightarrow \mathbb{P}\{\bar{a}^\top \xi \leq \bar{b}\} \geq 1 - \alpha, \quad \forall \mathbb{P} \in \mathcal{D}_1 \quad (47)$$

$$\Leftrightarrow \inf_{\mathbb{P} \in \mathcal{D}_1} \mathbb{P}\{\bar{a}^\top \xi \leq \bar{b}\} \geq 1 - \alpha.$$

We remark the equivalence between (46) and (47) as follows. Denote set  $L := \{\ell \in \mathbb{R} : \mathbb{P}\{\bar{a}^\top \xi \leq \ell\} \geq 1 - \alpha\}$  in (46). First, if (46) is valid, then  $\bar{b}$  is no smaller than the infimum of set  $L$  and thus  $\bar{b} \in L$  because the function  $F(\ell) := \mathbb{P}\{\bar{a}^\top \xi \leq \ell\}$  is nondecreasing and right-continuous. It then follows that (47) holds. Second, if (47) is valid, then  $\bar{b} \in L$ , and hence (46) holds by the definition of infimum. The proof is complete by noticing Theorem 1 and Proposition 1.  $\blacksquare$

## Appendix G Proof of Lemma 2

*Proof:*

(i) By definition,  $\phi^*$  is a supremum of linear functions and hence convex.

(ii) For any  $x_1, x_2 \in \mathbb{R}$  such that  $x_1 < x_2$ , we have

$$x_1 t - g(t) \leq x_2 t - g(t), \quad \forall t \geq 0.$$

Also, since  $\phi(t) = +\infty$  for  $t < 0$ , we have

$$\phi^*(x) = \sup_{t \in \mathbb{R}} \{x t - \phi(t)\} = \sup_{t \geq 0} \{x t - \phi(t)\},$$

and so

$$\phi^*(x_1) = \sup_{t \geq 0} \{x_1 t - \phi(t)\} \leq \sup_{t \geq 0} \{x_2 t - \phi(t)\} = \phi^*(x_2).$$

(iii) Since  $\phi(1) = 0$ , we have

$$\phi^*(x) = \sup_{t \geq 0} \{x t - \phi(t)\} \geq x.$$

(iv) We prove by contradiction. Suppose that  $\phi^*(x) = m$  on the interval  $[a, b]$  and  $\phi^*(y) = m' \neq m$  for some  $y < a$ . First, we observe that  $m' < m$  because  $\phi^*$  is nondecreasing. Second, there exists some  $\lambda \in [0, 1]$  such that  $a = \lambda y + (1 - \lambda)b$ . It follows that

$$\phi^*(a) \leq \lambda\phi^*(y) + (1 - \lambda)\phi^*(b) = \lambda m' + (1 - \lambda)m < m,$$

which gives a desirable contradiction. ■

## Appendix H Proof of Corollary 5

*Proof:* For any given  $z_0$  and  $z > 0$ , we have  $\phi^*(z_0 + z) \geq z_0 + z$  and  $\phi^*(z_0) \geq z_0$  by property (iii) in Lemma 2. It follows that

$$\begin{aligned} & \alpha\phi^*(z_0 + z) + (1 - \alpha)\phi^*(z_0) \geq \alpha(z_0 + z) + (1 - \alpha)z_0 \\ \Rightarrow & \alpha\phi^*(z_0 + z) + (1 - \alpha)\phi^*(z_0) \geq z_0 + \alpha z - d \quad (\text{since } d > 0) \\ \Rightarrow & \phi^*(z_0 + z) - z_0 - \alpha z + d \geq (1 - \alpha)(\phi^*(z_0 + z) - \phi^*(z_0)) \\ \Rightarrow & \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)} \geq 1 - \alpha. \end{aligned}$$

■

## Appendix I Proof of Proposition 5

*Proof:* First, since  $\phi(x) = (x - 1)^2$ , we have

$$\phi^*(x) = \begin{cases} -1, & \text{if } x \leq -2, \\ \frac{1}{4}x^2 + x, & \text{if } x \geq -2. \end{cases}$$

Hence,  $\underline{m}(\phi^*) = -2$  and  $\overline{m}(\phi^*) = +\infty$ . Second, we solve the problem

$$\inf_{\substack{z > 0, \\ z_0 + z \geq -2}} \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)} = \inf_{\substack{z > 0, \\ z_0 \geq -2}} \frac{\phi^*(z_0) - z_0 + (1 - \alpha)z + d}{\phi^*(z_0) - \phi^*(z_0 - z)}$$

to optimality, where we make a transform by replacing  $z_0$  by  $z_0 - z$ . We let  $f(z_0, z)$  represent the objective function and discuss the following cases:

(1) If  $z_0 - z \leq -2$ , then  $\phi^*(z_0 - z) = -1$  and  $\phi^*(z_0) = \frac{1}{4}z_0^2 + z_0$ . It follows that

$$f(z_0, z) = \frac{(\frac{1}{4}z_0^2 + z_0) - z_0 + (1 - \alpha)z + d}{(\frac{1}{4}z_0^2 + z_0) - (-1)} = \frac{\frac{1}{4}z_0^2 + (1 - \alpha)z + d}{(\frac{1}{2}z_0 + 1)^2},$$



and so

$$\frac{\partial f(z_0, z)}{\partial z_0} = \frac{\frac{1}{2}z_0 - (1 - \alpha)z - d}{\left(\frac{1}{2}z_0 + 1\right)^3}.$$

Since  $z_0 \leq z - 2$ ,  $z_0 \geq -2$  and  $\alpha < 1/2$  by assumption, we have  $(1/2)z_0 - (1 - \alpha)z - d \leq (\alpha - 1/2)z - d - 1 < 0$  and  $\frac{1}{2}z_0 + 1 \geq 0$ . Hence,  $\partial f(z_0, z)/\partial z_0 < 0$  for any fixed  $z$  and it is optimal to choose  $z_0^* = z - 2$ . It follows that

$$\inf_{\substack{z > 0, \\ z_0 \geq -2}} f(z_0, z) = \inf_{z > 0} f(z - 2, z) = \inf_{z > 0} 4(d + 1) \left(\frac{1}{z}\right)^2 - 4\alpha \left(\frac{1}{z}\right) + 1.$$

Therefore, it is optimal to choose  $z^* = 2(d + 1)/\alpha$  and

$$\inf_{\substack{z > 0, \\ z_0 \geq -2}} f(z_0, z) = 1 - \frac{\alpha^2}{d + 1}.$$

(2) If  $z_0 - z \geq -2$ , then  $\phi^*(z_0 - z) = \frac{1}{4}(z_0 - z)^2 + (z_0 - z)$  and  $\phi^*(z_0) = \frac{1}{4}z_0^2 + z_0$ . It follows that

$$f(z_0, z) = \frac{\left(\frac{1}{4}z_0^2 + z_0\right) - z_0 + (1 - \alpha)z + d}{\left(\frac{1}{4}z_0^2 + z_0\right) - \left(\frac{1}{4}(z_0 - z)^2 + (z_0 - z)\right)} = \frac{\frac{1}{4}z_0^2 + (1 - \alpha)z + d}{\frac{1}{2}zz_0 + z - \frac{1}{4}z^2},$$

and so

$$\frac{\partial f(z_0, z)}{\partial z_0} = \frac{z(z_0^2 + (4 - z)z_0 - 4(1 - \alpha)z - 4d)}{8\left(\frac{1}{2}zz_0 + z - \frac{1}{4}z^2\right)^2}.$$

For fixed  $z$ , we set  $\partial f(z_0, z)/\partial z_0 = 0$  and obtain

$$z_0 = \frac{(z - 4) \pm \sqrt{z^2 + 8(1 - 2\alpha)z + 16(d + 1)}}{2}.$$

Since  $z_0 \geq z - 2$ , we rule out the negative root and so

$$z_0^* = \frac{(z - 4) + \sqrt{z^2 + 8(1 - 2\alpha)z + 16(d + 1)}}{2}$$

is a stationary point of  $f(z_0, z)$  with  $z$  fixed and the corresponding objective value

$$f(z_0^*, z) = \frac{1}{2} \sqrt{16(d + 1) \left(\frac{1}{z}\right)^2 + 8(1 - 2\alpha) \left(\frac{1}{z}\right) + 1} + \frac{1}{2} \left(1 - 4 \left(\frac{1}{z}\right)\right).$$

Now we show that  $z_0^*$  is an optimal solution for  $\inf_{z_0 \geq z - 2} f(z_0, z)$  with  $z$  fixed. We compare the value of  $f(z_0^*, z)$  with  $f(+\infty, z)$  and  $f(z - 2, z)$  because  $+\infty$  and  $z - 2$  are the end points of the feasible region of  $z_0$ . We observe that  $f(+\infty, z) = +\infty$ . Also, we have

$$f(z - 2, z) = \frac{\frac{1}{4}(z - 2)^2 + (1 - \alpha)z + d}{\frac{1}{2}z(z - 2) + z - \frac{1}{4}z^2} = 4(d + 1) \left(\frac{1}{z}\right)^2 - 4\alpha \left(\frac{1}{z}\right) + 1,$$

and  $f(z-2, z) \geq f(z_0^*, z)$ . To see that, we compare the values of  $f(z-2, z)$  and  $f(z_0^*, z)$  by the following inequalities, where the inequalities below imply those above.

$$\begin{aligned}
& f(z-2, z) \geq f(z_0^*, z) \\
\Leftrightarrow & 8(d+1) \left(\frac{1}{z}\right)^2 - 8\alpha \left(\frac{1}{z}\right) + 2 \geq \sqrt{16(d+1) \left(\frac{1}{z}\right)^2 + 8(1-2\alpha) \left(\frac{1}{z}\right) + 1} + \left(1 - 4 \left(\frac{1}{z}\right)\right) \\
\Leftrightarrow & \left[8(d+1) \left(\frac{1}{z}\right)^2 + 4(1-2\alpha) \left(\frac{1}{z}\right) + 1\right]^2 \geq 16(d+1) \left(\frac{1}{z}\right)^2 + 8(1-2\alpha) \left(\frac{1}{z}\right) + 1 \\
\Leftrightarrow & 16 \left(\frac{1}{z}\right)^2 \left[2(d+1) \left(\frac{1}{z}\right) + (1-2\alpha)\right]^2 \geq 0.
\end{aligned}$$

Hence,  $\inf_{z_0 \geq z-2} f(z_0, z) = f(z_0^*, z)$  with  $z$  fixed. Therefore, we have

$$\inf_{z>0, z_0 \geq z-2} f(z_0, z) = \inf_{z>0} \frac{1}{2} \sqrt{16(d+1)z^2 + 8(1-2\alpha)z + 1} + \frac{1}{2}(1-4z),$$

where we have  $1/z$  replaced by  $z$ . Similarly, we set

$$\frac{\partial f(z_0^*, z)}{\partial z} = \frac{8(d+1)z + 2(1-2\alpha)}{\sqrt{16(d+1)z^2 + 8(1-2\alpha)z + 1}} - 2 = 0,$$

and obtain

$$z^* = \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1-2\alpha)d}{4d(d+1)}.$$

Therefore, we have

$$f(z_0^*, z^*) = 1 - \alpha + \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1-2\alpha)d}{2(d+1)}.$$

Again, we shall compare the value of  $f(z_0^*, z^*)$  with  $f(z_0^*, +\infty)$  and  $f(z_0^*, 0)$  since  $+\infty$  and  $0$  are the end points of the feasible region of  $z$ . We observe that  $f(z_0^*, +\infty) = +\infty$  and  $f(z_0^*, 0) = 1 \geq f(z_0^*, z^*)$ , and hence

$$\inf_{z>0, z_0 \geq z-2} f(z_0, z) = 1 - \alpha + \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1-2\alpha)d}{2(d+1)}.$$

Finally, we compare the optimal value of  $f(z_0, z)$  in the two cases. We claim that the optimal value obtained in the latter case is smaller (and hence globally optimal). To see that, we compare the two values by the following inequalities, where the inequalities below imply those above.

$$\begin{aligned}
& 1 - \frac{\alpha^2}{d+1} \geq 1 - \alpha + \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1-2\alpha)d}{2(d+1)} \\
\Leftrightarrow & d + 2\alpha - 2\alpha^2 \geq \sqrt{d^2 + 4d(\alpha - \alpha^2)} \\
\Leftrightarrow & (d + 2\alpha - 2\alpha^2)^2 \geq d^2 + 4d(\alpha - \alpha^2) \\
\Leftrightarrow & 4\alpha^2(\alpha - 1)^2 \geq 0.
\end{aligned}$$

Therefore, the perturbed risk level is

$$\alpha' = \alpha - \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1 - 2\alpha)d}{2d + 2}.$$

■

## Appendix J Proof of Proposition 6

*Proof:* First, Since  $\phi(x) = |x - 1|$ , we have

$$\phi^*(x) = \begin{cases} -1, & \text{if } x < -1, \\ x, & \text{if } -1 \leq x \leq 1, \\ +\infty, & \text{if } x > 1. \end{cases}$$

Hence,  $\underline{m}(\phi^*) = -1$  and  $\overline{m}(\phi^*) = 1$ . Second, we solve the problem

$$\inf_{z > 0, -1 \leq z_0 + z \leq 1} f(z_0, z) := \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)}$$

to optimality. We discuss the following cases:

(1) If  $z_0 \leq -1$ , then  $\phi^*(z_0) = -1$  and  $\phi^*(z_0 + z) = z_0 + z$ . It follows that

$$f(z_0, z) = \frac{(1 - \alpha)z + d}{z_0 + z + 1}.$$

Note here that for any given  $z$ ,  $f(z_0, z)$  is a nonincreasing function of  $z_0$ , due to the fact that  $z_0 + z + 1 \geq 0$ . Meanwhile,  $z_0 + z \leq 1$ . Hence, it is optimal to choose  $z_0^* = \min\{1 - z, -1\}$  and so

$$f(z_0^*, z) = \begin{cases} \frac{(1 - \alpha)z + d}{2}, & \text{if } z \geq 2, \\ \frac{(1 - \alpha)z + d}{z}, & \text{if } z \leq 2. \end{cases}$$

Therefore,  $f(z_0^*, z)$  is nonincreasing on  $z$  on the interval  $(0, 2]$  and nondecreasing on  $z$  on the interval  $[2, +\infty)$ , and so  $f(z_0^*, z^*) = 1 - \alpha + \frac{d}{2}$ .

(2) If  $-1 \leq z_0 \leq 1$ , then  $\phi^*(z_0) = z_0$ . Also, we have  $z \leq 2$  and  $\phi^*(z_0 + z) = z_0 + z$  because  $-1 \leq z_0 + z \leq 1$ . Hence,

$$f(z_0, z) = \frac{(1 - \alpha)z + d}{z} = 1 - \alpha + \frac{d}{z} \geq 1 - \alpha + \frac{d}{2},$$

and the lower bound is attained at  $z^* = 2$ . Therefore,  $f(z_0^*, z^*) = 1 - \alpha + \frac{d}{2}$ .

To sum up, we have  $1 - \alpha' = f(z_0^*, z^*) = 1 - \alpha + \frac{d}{2}$ , or equivalently  $\alpha' = \alpha - \frac{d}{2}$ .

■

## Appendix K Proof of Proposition 7

*Proof:* We divide the proof into two parts. In the first part, we show that the perturbed risk level

$$\alpha' = 1 - \inf_{x \in (0,1)} \left\{ \frac{e^{-d}x^{1-\alpha} - 1}{x - 1} \right\}. \quad (48)$$

In the second part, we show how to compute  $\alpha'$  by using bisection line search.

**(Risk level)** First, since  $\phi(x) = x \log x - x + 1$ , we have  $\phi^*(x) = e^x - 1$ . Hence,  $\underline{m}(\phi^*) = -\infty$  and  $\overline{m}(\phi^*) = +\infty$ . Second, we solve the problem

$$\inf_{z > 0} f(z_0, z) := \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)} = \inf_{z > 0} \frac{e^z + (d - \alpha z - z_0 - 1)e^{-z_0}}{e^z - 1}$$

to optimality. Since

$$\frac{\partial f}{\partial z_0} = -\frac{d - \alpha z - z_0}{e^{z_0}(e^z - 1)},$$

we have  $z_0^* = d - \alpha z$  by setting  $\partial f / \partial z_0 = 0$ , and so

$$\begin{aligned} \inf_{z > 0} f(z_0, z) &= \inf_{z > 0} f(z_0^*, z) \\ &= \inf_{z > 0} \frac{e^z - e^{\alpha z - d}}{e^z - 1} \\ &= \inf_{z > 0} \frac{1 - e^{-d}(1/e^z)^{1-\alpha}}{1 - (1/e^z)} \\ &= \inf_{x \in (0,1)} \frac{e^{-d}x^{1-\alpha} - 1}{x - 1}, \end{aligned} \quad (49)$$

where equation (49) follows by replacing  $(1/e^z)$  with  $x$ . Therefore, we have proved equation (48).

**(Computation)** We compute  $\alpha'$  by searching the optimal solution to the minimization problem

$$\inf_{x \in (0,1)} \left\{ \frac{e^{-d}x^{1-\alpha} - 1}{x - 1} \right\}.$$

First, by denoting  $1 - \alpha' = \inf_{x \in (0,1)} h(x)$ , we have

$$h'(x) = \frac{1 - e^{-d}\alpha x^{1-\alpha} - e^{-d}(1-\alpha)x^{-\alpha}}{(x-1)^2}, \quad \forall x \in (0,1).$$

It is clear that  $(x-1)^2$  decreases as  $x$  increases. Meanwhile, since  $x < 1$  and  $x^{-\alpha-1} > x^{-\alpha}$ , we have

$$(1 - e^{-d}\alpha x^{1-\alpha} - e^{-d}(1-\alpha)x^{-\alpha})'_x = e^{-d}\alpha(1-\alpha)(x^{-\alpha-1} - x^{-\alpha}) > 0.$$

Therefore,  $h'(x)$  increase as  $x$  increases in  $(0,1)$ , and hence the function  $h(x)$  is convex over  $x$  in  $(0,1)$ . Because  $\lim_{x \rightarrow 0^+} h'(x) = -\infty$  and  $\lim_{x \rightarrow 1^-} h'(x) = +\infty$ , we have:

$$\text{the infimum of } h(x) \text{ is attained in the interval } (0,1). \quad (50)$$

We can compute the optimal  $x^*$  by forcing

$$\frac{1 - e^{-d}\alpha(x^*)^{1-\alpha} - e^{-d}(1-\alpha)(x^*)^{-\alpha}}{(x^* - 1)^2} = 0,$$

i.e.,  $(x^*)^\alpha = e^{-d}\alpha x^* + e^{-d}(1-\alpha)$ . The intersection of functions  $x^\alpha$  and  $e^{-d}\alpha x + e^{-d}(1-\alpha)$  can be easily computed by a bisection line search. Finally, to achieve  $\epsilon$  accuracy, i.e.,  $|\hat{x} - x^*| \leq \epsilon$ , of the incumbent probing value  $\hat{x}$ , we only have to conduct  $S$  steps of bisection, such that  $2^{-S} \leq \epsilon$ . It follows that  $S \geq \lceil \log_2(\frac{1}{\epsilon}) \rceil$ .  $\blacksquare$

## Appendix L Proof of Proposition 8

*Proof:* We divide the proof into two parts. We show the monotonicity in the first part, and the convergence in the second part.

**(Monotonicity)** Suppose that we are given  $d_1 > d_2 > 0$ . First, since  $\phi^*$  is nondecreasing by property (ii) in Lemma 2, we have  $\phi^*(z_0 + z) \geq \phi^*(z_0)$  for any given  $z_0$  and  $z > 0$ . Second, it can be shown that  $\phi^*(z_0 + z) > \phi^*(z_0)$ , because otherwise the DCC will be violated (the proof is similar to the one for Theorem 2 on ruling out the case  $z_0/z < \underline{m}(\phi^*)$ , and note that  $z_0/z$  in the proof for Theorem 2 is equivalent to  $z_0 + z$  here because of variable replacement). Hence, we have

$$\frac{\phi^*(z_0 + z) - z_0 - \alpha z + d_1}{\phi^*(z_0 + z) - \phi^*(z_0)} \geq \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d_2}{\phi^*(z_0 + z) - \phi^*(z_0)},$$

and so

$$1 - \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} \left\{ \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d_1}{\phi^*(z_0 + z) - \phi^*(z_0)} \right\} \leq 1 - \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} \left\{ \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d_2}{\phi^*(z_0 + z) - \phi^*(z_0)} \right\}.$$

Therefore, the value of  $\alpha'$  increases as  $d$  decreases.

**(Convergence)** Now suppose that  $d \downarrow 0$  and we want to prove that  $\alpha' \uparrow \alpha$ . Since  $1 - \alpha'$  can be defined as

$$\inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} f(z_0, z) := \left\{ \frac{\phi^*(z_0 + z) - z_0 - \alpha z + d}{\phi^*(z_0 + z) - \phi^*(z_0)} \right\},$$

and  $1 - \alpha'$  decreases as  $d$  decreases, we have

$$\lim_{d \downarrow 0} \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} f(z_0, z) = \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} \inf_{d > 0} f(z_0, z) = \inf_{\substack{z > 0, \\ \underline{m}(\phi^*) \leq z_0 + z \leq \bar{m}(\phi^*)}} \left\{ \frac{\phi^*(z_0 + z) - z_0 - \alpha z}{\phi^*(z_0 + z) - \phi^*(z_0)} \right\}.$$

We show that for  $z_0 = 0$ ,  $\inf_{z > 0} f(0, z) \leq 1 - \alpha$  and hence the conclusion follows. First, we observe that  $\phi^*(0) = 0$ , because  $\phi^*(0) = \sup_x \{-\phi(x)\} = -\inf_x \phi(x)$  and  $x = 1$  is a global minimizer for  $\phi$

by assumption. It follows that

$$f(0, z) = \frac{\phi^*(z) - \alpha z}{\phi^*(z)} = 1 - \frac{\alpha z}{\phi^*(z)}.$$

Second, since  $x = 1$  is the unique minimizer for  $\phi$ , we have  $(\phi^*)'(0) = \operatorname{argmax}_x \{-\phi(x)\} = 1$  by a property of convex conjugates. Hence, we have  $\phi^*(z) = z + o(z)$  based on Taylor series, where  $\lim_{z \downarrow 0} o(z)/z = 0$ . It follows that

$$\frac{\phi^*(z)}{z} \rightarrow 1 \text{ as } z \downarrow 0,$$

and so

$$f(0, z) = 1 - \frac{\alpha z}{\phi^*(z)} \rightarrow 1 - \alpha \text{ as } z \downarrow 0,$$

noting that  $\underline{m}(\phi^*) \leq 0$  based on the fact that  $\phi^*(x) \leq x$  for  $x > 0$  (see property (iii) in Lemma 2). Therefore,  $\inf_{z>0} f(0, z) \leq 1 - \alpha$  is proved, and the proof is complete by noting that  $d$  tends to zero as  $N$  tends to infinity by assumption.  $\blacksquare$

## Appendix M Proof of Proposition 9

*Proof:* First, we define

$$f(d) := \frac{\sqrt{d^2 + 4d(\alpha - \alpha^2)} - (1 - 2\alpha)d}{2d + 2}.$$

From Proposition 5, we have

$$\alpha' = \alpha - f(d).$$

We show that  $f(d)$  decreases as  $d$  decreases. To this end, we have

$$f'(d) = \frac{1}{2(d+1)^2} \left[ \frac{(2\alpha^2 - 2\alpha + 1)d + 2(\alpha - \alpha^2)}{\sqrt{d^2 + 4d(\alpha - \alpha^2)}} - (1 - 2\alpha) \right]. \quad (51)$$

To show that  $f'(d) > 0$ , we compare  $f'(d)$  and zero by the following inequalities, where the inequalities below imply those above.

$$\begin{aligned} & f'(d) > 0 \\ \Leftrightarrow & \frac{(2\alpha^2 - 2\alpha + 1)d + 2(\alpha - \alpha^2)}{\sqrt{d^2 + 4d(\alpha - \alpha^2)}} > (1 - 2\alpha) \\ \Leftrightarrow & (2\alpha^2 - 2\alpha + 1)d + 2(\alpha - \alpha^2) > (1 - 2\alpha)\sqrt{d^2 + 4d(\alpha - \alpha^2)} \\ \Leftrightarrow & [(2\alpha^2 - 2\alpha + 1)d + 2(\alpha - \alpha^2)]^2 > (1 - 2\alpha)^2 [d^2 + 4d(\alpha - \alpha^2)] \\ \Leftrightarrow & 4\alpha^2(1 - \alpha)^2(d + 1)^2 > 0. \end{aligned}$$

Hence,  $f'(d) > 0$  and so  $\alpha'$  increases as  $d$  decreases. Furthermore, since  $\lim_{d \downarrow 0} f(d) = 0$ , we have

$$\lim_{d \downarrow 0} \alpha' = \alpha - \lim_{d \downarrow 0} f(d) = \alpha,$$

and so  $\alpha'$  increases to  $\alpha$  as  $d$  decreases to zero.

Second, since  $d = d(N)$  by assumption, we have

$$\begin{aligned} \text{VoD}_\alpha &= \frac{d\alpha'}{dN} = \left( \frac{d\alpha'}{dd} \right) \left( \frac{dd(N)}{dN} \right) \\ &= f'(d) \left( \frac{dd(N)}{dN} \right). \end{aligned} \quad (52)$$

The proof is complete by substituting the definition of  $f'(d)$  in (51) into equation (52). ■

## Appendix N Proof of Proposition 10

*Proof:* We divide the proof into three parts. We show the convergence in the first part, develop the relationship between  $\alpha'$  and  $d$  in the second part, and compute the value of data in the last part.

**(Convergence)** From Proposition 7, we know that

$$1 - \alpha' = \inf_{x \in (0,1)} \left\{ \frac{e^{-d}x^{1-\alpha} - 1}{x - 1} \right\}. \quad (53)$$

It is clear that  $e^{-d} \uparrow 1$  as  $d \downarrow 0$ , and so we have

$$\alpha' = 1 - \inf_{x \in (0,1)} \left\{ \frac{e^{-d}x^{1-\alpha} - 1}{x - 1} \right\} \uparrow 1 - \inf_{x \in (0,1)} \left\{ \frac{x^{1-\alpha} - 1}{x - 1} \right\} = 1 - (1 - \alpha) = \alpha,$$

which proves that  $\alpha'$  increases to  $\alpha$  as  $d$  decreases to zero.

**(Relationship between  $\alpha'$  and  $d$ )** From Proposition 7, we know that the optimal value of the embedded optimization problem in equality (53) can be attained by some  $\bar{x} \in (0, 1)$  (based on claim (50) in the proof of Proposition 7), which is the stationary point of the objective function.

It follows that

$$\begin{cases} \frac{e^{-d}\bar{x}^{1-\alpha} - 1}{\bar{x} - 1} = 1 - \alpha' \\ \bar{x}^\alpha = e^{-d}\alpha\bar{x} + e^{-d}(1 - \alpha). \end{cases} \quad (54)$$

Solving this nonlinear equation system, we reformulate the first equation and then substitute the second equation into the first as follows:

$$\begin{aligned} e^{-d}\bar{x} - (1 - \alpha')\bar{x}\bar{x}^\alpha &= \alpha'\bar{x}^\alpha \\ \Rightarrow e^{-d}\bar{x} - (1 - \alpha')\bar{x}\left(e^{-d}\alpha\bar{x} + e^{-d}(1 - \alpha)\right) &= \alpha'\left(e^{-d}\alpha\bar{x} + e^{-d}(1 - \alpha)\right) \\ \Rightarrow (\bar{x} - 1)\left(\alpha(1 - \alpha')\bar{x} - \alpha'(1 - \alpha)\right) &= 0. \end{aligned}$$

Ruling out the solution  $\bar{x} = 1$ , we have  $\bar{x} = \frac{\alpha'(1-\alpha)}{\alpha(1-\alpha')} \in (0, 1)$ . Finally, we substitute the solution of  $\bar{x}$  back into the second equation in (54) and obtain

$$e^{-d} = \bar{x}^\alpha / (\alpha\bar{x} + 1 - \alpha) = (\alpha'/\alpha)^\alpha ((1-\alpha')/(1-\alpha))^{1-\alpha}. \quad (55)$$

Finally, by taking the natural logarithm on both sides of equation (55), we obtain that

$$d = \alpha \log\left(\frac{\alpha}{\alpha'}\right) + (1-\alpha) \log\left(\frac{1-\alpha}{1-\alpha'}\right). \quad (56)$$

**(Value of data)** From equation (56) we have

$$\frac{dd}{d\alpha'} = -\frac{\alpha}{\alpha'} + \frac{1-\alpha}{1-\alpha'} = \frac{\alpha' - \alpha}{\alpha'(1-\alpha')}.$$

It is easy to observe that  $dd/d\alpha'$  is a monotone function of  $\alpha'$  and  $dd/d\alpha' \neq 0$ . Hence, we have

$$\frac{d\alpha'}{dd} = 1 / \left( \frac{dd}{d\alpha'} \right) = \frac{\alpha'(1-\alpha')}{\alpha' - \alpha}.$$

Therefore,

$$\text{VoD}_\alpha = \left( \frac{d\alpha'}{dd} \right) \left( \frac{dd(N)}{dN} \right) = \left[ \frac{\alpha'(1-\alpha')}{\alpha' - \alpha} \right] d'(N).$$

■