

Distributionally Robust Optimization under Distorted Expectations

Jun Cai[†], Jonathan Yu-Meng Li^{††}, Tiantian Mao^{†††}

[†]Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

^{††} Telfer School of Management
University of Ottawa, Ottawa, ON, Canada

^{†††} Department of Statistics and Finance, School of Management
University of Science and Technology of China
Hefei, Anhui, China

March 30, 2020

Abstract

Distributionally robust optimization (DRO) has arose as an important paradigm to address the issue of distributional ambiguity in decision optimization. In its standard form, DRO seeks an optimal solution against the worst-possible expected value evaluated based on a set of candidate distributions. In the case where a decision maker is not risk neutral, the most common scheme applied in DRO to capture one’s risk attitude is employing an expected utility functional. In this paper, we propose to address a decision maker’s risk attitude in DRO by following an alternative scheme known as “dual expected utility”. In this scheme, a distortion function is applied to convert physical probabilities to subjective probabilities so that the resulting expectation, also known as distorted expectation, captures the decision maker’s risk attitude. Unlike an expected utility functional which is linear in probability, in the dual scheme a distorted expectation is generally non-linear in probability. We distinguish DRO based on distorted expectations by terming it “Distributionally Robust Risk Optimization” (DRRO), and show that DRRO can be equally, if not more, tractable to solve than DRO based on utility functionals. Our tractability results hold for *any* distortion function, and hence our scheme provides more flexibility to capture more realistic forms of risk attitudes. These include, as an important example, the inverse S-shaped distortion functionals that play a prominent role in Cumulative Prospect Theory (CPT), and several other non-convex risk measures developed more recently. Central

to our development is the characterization of worst-case distributions based on the notion of convex envelope, which enables us to discover “hidden convexity” in DRRO. We demonstrate through a numerical example that a production manager who overly weights “very good” and “very bad” outcomes may act as if (s)he is risk-averse when taking into account distributional ambiguity. Worst-case distributions are presented that can provide further explanation of such risk-averse behaviour.

Keywords: Distributionally robust optimization, distortion risk measure, convex risk measure, convex envelope.

1 Introduction

Distributional ambiguity refers to the situation where the probability distribution of uncertain outcomes is unknown or cannot be uniquely identified. This issue arises when one only has limited information to infer the “true” distribution, which is typically the case in most applications where only sample data is available. The question of how to account for distributional ambiguity in decision making has been of central interest in a number of fields including Economics, Finance, Control System, and Operations Research/Management Science. One modelling paradigm that has been successfully adopted in all these fields to address the issue is distributionally robust optimization (DRO). In its standard form, DRO takes the following formulation of a minimax optimization problem

$$\min_{\vec{w} \in \mathcal{W}} \sup_{F \in \mathcal{F}} \mathbb{E}[f(\vec{w}, \vec{X})], \quad (1.1)$$

where \vec{w} denotes a decision vector constrained by the feasible set \mathcal{W} , \vec{X} denotes a random vector characterized by the distribution F , and f is a cost function. The power of DRO lies in its flexibility to characterize one’s (partial) knowledge about the distribution F through specifying the set \mathcal{F} . The set is also known as ambiguity set, which consists of all possible distributions that are consistent with one’s knowledge. To ensure that the decision is robust against distributional ambiguity, DRO generates a solution that is optimal with respect to the worst-case distribution, i.e. a distribution from the set \mathcal{F} that gives the largest possible expected cost. One of the most attractive features of DRO is its computational tractability, namely that the optimization problem (1.1) can often be solved efficiently in large scale with various kinds of ambiguity set \mathcal{F} and the cost function f . In particular, one common way of defining the set \mathcal{F} is through specifying the moments of the distribution. The earlier works of Popescu (2007), Bertsimas et al. (2010), Delage and Ye (2010), and Natarajan et al. (2010) show that in the case where the ambiguity set is specified through the first two moments, the DRO (1.1) is tractable to solve for a large class of cost functions. More recently, Wiesemann et al. (2014) provides general tractability results for the case where the ambiguity set is described through supports and higher order moments. While the focus of this paper is on the moment-based ambiguity sets, we should point out here that the ambiguity set can also be defined

according to some distance functions over distributions. It remains an active stream of research to study the tractability of DRO in this case, but since this is not the focus of this paper we refer interested readers to the recent work of Ben-Tal et al. (2013), Jiang and Guan (2016), Esfahani and Kuhn (2018) and the references therein.

DRO provides a means to capture one’s aversion towards ambiguity, but care has to be taken to distinguish one’s attitude towards risk from ambiguity. This can possibly be best demonstrated if a linear cost function $f(\vec{w}, \vec{x}) = \vec{w}^\top \vec{x}$ is considered in (1.1), and the set \mathcal{F} denotes a set of distributions characterized by the first moment $\vec{\mu}$, i.e. the mean. In this case, DRO simply reduces to minimizing $\vec{w}^\top \vec{\mu}$, which essentially reflects no concern about the uncertainty. To properly capture one’s risk attitude, the literature of DRO often suggests the adoption of the expected utility framework, i.e. solving instead the following DRO problem

$$\min_{\vec{w} \in \mathcal{W}} \sup_{F \in \mathcal{F}} \mathbb{E}[u(f(\vec{w}, \vec{X}))], \quad (1.2)$$

where u denotes some disutility function. For a large class of disutility functions, Popescu (2007) shows that the above DRO problem (1.2) can be solved as a parametric quadratic program in the case where the cost function f is linear and the ambiguity set \mathcal{F} is characterized by the first two moments. Chen et al. (2011) considers a further special case of the above DRO problem in the setting of robust portfolio selection, and shows that the problem can either be solved analytically or by a particularly efficient procedure when the utility function takes the form of lower partial moment or S-shaped function. Many others have considered the case where the disutility function takes a general piecewise linear form and showed how the corresponding DRO problems can be reduced to finite-dimensional convex or conic programs that are polynomially solvable (see e.g. Delage and Ye (2010), Natarajan et al. (2010), Bertsimas et al. (2010), and Wiesemann et al. (2014)).

Adopting expected utility framework in DRO appears to be natural given the long history of expected utility theory and its rigorous axiomatic foundation. There are however noticeable shortcomings about its use. While there are a number of “textbook” utility functions, it has not been clear which utility function one should employ when it comes to actual practice. This is partly due to the difficulty of interpreting the physical meaning of utility values and is further complicated by the need to determine the domain of a utility function. Moreover, empirical finding has shown that one of the axioms, namely the independence axiom, is often violated in real-life decision-making.

In this paper, our goal is to pursue an alternative route to address risk in DRO. In particular, we invoke “dual utility theory” that was first developed by Yaari (1987) and has been found lately that it is closely connected to the modern theory of risk measure (Artzner et al. (1999), Föllmer and Schied (2002)). Similar to expected utility theory, Yaari (1987) established the dual theory by proposing an alternative set of axioms, and proved that there always exists a “distortion” function that converts a physical probability distribution to a subjective one so that the resulting expectation, also known as distorted expectation, will capture risk preference satisfying the axioms. Applying

the theory, we consider the following formulation of DRO, which is termed Distributionally Robust Risk Optimization (DRRO) throughout this paper

$$\min_{\vec{w} \in \mathcal{W}} \sup_{F \in \mathcal{F}} \rho_h(f(\vec{w}, \vec{X})), \quad (1.3)$$

where h is a distortion function, and ρ_h denotes the expectation based on the “distorted” probabilities (see Definition 2.2). The function ρ_h is also called distortion risk measure in the literature of risk measures. Aside from its theoretical justification, the use of distorted expectation can be particularly appealing from a practical point of view. Namely, a distortion function can be conveniently specified as a function that maps from unit interval to unit interval, i.e. with fixed domain and range, and its value has a clear probability interpretation. The advantage of distorted expectation in practical use can probably be best highlighted by the fact that it consists of several risk measures that are now commonly applied in practice, namely Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), Spectral Risk Measures, Wang’s transform, to name a few.

Yarri’s theory is considered “dual” to expected utility theory because in contrast to expected utility functional, which is nonlinear in random variables and linear in probability distributions, distorted expectation is linear in quantile functions but nonlinear in probability distributions. This delineates also the important differences between the DRRO problem (1.3) and the classical DRO problem (1.2). For example, it is well known that worst-case distributions obtained from the classical DRO problem (1.2) with a moment-based ambiguity set always take the form of a discrete distribution with the number of supports being less or equal to the number of moment constraints. This structure of worst-case distributions, which has often been criticized as unrealistic or degenerate, is a direct consequence of the linearity of expectation functional in probability distributions. The fact that distorted expectation is nonlinear in probability distributions implies that the worst-case distributions can take potentially a richer form, which may depend more heavily on the structure of the distortion function.

While there are special cases of distortion risk measures that have been considered in the DRO literature (e.g. Value-at-Risk in El Ghaoui et al. (2003) and Zymmler et al. (2013b), Conditional Value-at-Risk in Chen et al. (2011) and Natarajan et al. (2010)), to date there is no unifying approach to solving DRO with general distortion risk measures. One potential exception is the work of Li (2018), who studies the case of spectral risk measures (Acerbi (2002)) in full generality but the result is limited to the case of linear cost function and ambiguity sets characterized by the first two moments. Spectral risk measures represent an important class of distortion risk measures that are characterized by convex distortion functions and justified by the axioms of convex risk measures (Föllmer and Schied (2002)). The property of convexity, while useful from a “normative” perspective, i.e. specifying how risk should be perceived, can however be unrealistic from a “descriptive” perspective, i.e. describing how risk is actually perceived. In particular, as pointed out by the leading descriptive theory of risky choices, namely cumulative prospect theory (CPT), the distortion pattern observed in many empirical settings is neither convex nor concave but rather

takes an inverse S-shaped form. This is because individuals tend to be more sensitive to “very good” and “very bad” outcomes and overweight their likelihood. Distortion functions proposed to capture the pattern includes the one-parameter inverse S-shaped function applied in the seminal work of Tversky and Kahneman (1992) and several other popular variants with more parameters to fit the pattern. The detailed discussions and their applications in economics, finance, and other fields can be found in Tversky and Kahneman (1992), Tversky and Fox (1995), Prelec (1998), Jin and Zhou (2008), Xu and Zhou (2013), and the references therein. Besides the consideration of empirical distortion patterns, there are several other reasons why there is a need to go beyond the scope of convex distortion function. In particular, it is now well recognized that there is a conflict between the convexity, more precisely subadditivity, and robustness of risk measures (Cont et al. (2010)). One alternative that has received much attention is Range Value-at-Risk (RVaR) proposed by Cont et al. (2010), which is a more robust risk measure than VaR and CVaR. RVaR is a distortion risk measure characterized by a S-shaped distortion function. There is also a very practical need from the industry to strike a balance between VaR and CVaR, as the former tends to underestimate the risk exposure whereas the latter has often been found overly-conservative. Belles-Sampera et al. (2014) introduced the GlueVaR risk measure, which is a weighted sum of VaR and CVaR and can be interpreted in terms of one’s risk attitude. GlueVaR is also a distortion risk measure with non-convex distortion function, and its application in non-financial problems such as health, safety, environmental, or catastrophic risk management can be found in Belles-Sampera et al. (2014). Despite the need to adopt the aforementioned distortion risk measures, little progress has been made regarding their use in risk minimization. This can be well related to the fact that the problem of minimizing a special case of non-convex distortion risk measure, namely VaR, is already known as a NP-hard problem (Benati and Rizzi (2007)).

The main result of this paper is to show that distortion risk measures can be tractably optimized for a large class of DRRO problems. This includes the case where the ambiguity set is defined based on support and bounded moments (see e.g. Wiesemann et al. (2014)), and the case the set is defined based on fixed mean and covariance (see e.g. Bertsimas et al. (2010)). More specifically, we show that in these cases, as long as the cost function $f(\vec{w}, \vec{x})$ is concave in \vec{x} , one can always solve a DRRO problem by formulating an alternative DRRO problem characterized by some convex distortion function, and the latter can often be solved in good precision as a convex optimization problem. This applies for example to any two-stage stochastic programming problem with cost uncertainty (see e.g. Delage et al. (2014)). From the methodological point of view, our results might be particularly interesting to the community of robust optimization as we reveal “hidden convexity” (Ben-Tal et al. (2015), Mak et al. (2015)) exists for a large class of non-convex DRRO problems. Prior to our work, in the literature of DRO this hidden convexity has only been observed in the case of VaR. Namely, Zymler et al. (2013a) (see also Yang and Xu (2016)) proves that worst-case CVaR in fact provides a tight approximation to worst-case VaR for any convex or quadratic convex function in the case where the ambiguity set is characterized by fixed mean and covariance.

We prove, under very general moment conditions, that any worst-case distortion risk measure can be exactly approximated by an alternative worst-case distortion risk measure that replaces the original distortion function by its convex envelope. The result in Zymler et al. (2013a) indeed is a special case of ours since the distortion function of CVaR is the convex envelope of the distortion function of VaR. We should note that we are not the first to consider the use of convex envelope in analyzing worst-case risk measure. Wang et al. (2015) and Cai et al. (2018) consider the settings where only marginal distributions are known, and they show that worst-case distortion risk measures in this case can be exactly approximated by the use of convex envelope for asymptotically large random sums. In addition, Cornilly and Vanduffel (2019) and Zhu and Shao (2018) prove the exactness of the approximation for a finite combination of random variables when the first moment and a higher order moment are given. Unlike these works which focus on deriving analytical solutions and hence require fairly specific structure of cost function, i.e. linearity, and moment condition, i.e. maximally two moments are known, we establish our results in the general setting of DRO so that they are immediately applicable to existing DRO problems. Moreover, since our final solution is calculated based on convex optimization technique, we are also able to characterize the worst-case distributions in a far general fashion. One interesting class of DRO problems that we have paid a particular attention to in this paper is the two-stage production and transportation planning problem that has been studied for example in Natarajan et al. (2010) and Yang and Xu (2016). We conduct a numerical experiment in this paper to demonstrate how DRRO can be applied to answer the question of how a production manager may act if (s)he overly-weights "very good" and "very bad" outcomes, as depicted by CPT, when facing distributional ambiguity. The numerical results provide a precise description of how the optimal decisions can become more risk-averse, and how the worst-case distributions that the manager tries to hedge can vary, as more weights are put on the "extreme" outcomes.

The rest of the paper is organized as follows. In Section 2, we provide necessary background about distorted expectation (distortion risk measures) and clarify the relationship between several definitions commonly found in the literature. We discuss several non-convex distortion functions and their applications. We then investigate in Section 3 the tractability of solving DRRO problems based on two moment-based ambiguity sets. The notion of convex envelope is introduced, together with its application in our analysis. We discuss also the special case of linear cost function, which leads to closed-form results for worst-case distortion risk measures. Section 4 presents preliminary numerical results. The conclusions are given in Section 5. The proofs of the preliminary lemmas and most of the main results are provided in Section 6 as an appendix.

2 Distortion Risk Measures and Their Representations

In this section, we start off by defining distorted expectations, or equivalently distortion risk measures, in the most general fashion, and then provide a few alternative representations (or defini-

tions) that are more accessible. We consider all the random variables are defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and each random variable $X : \Omega \rightarrow \mathbb{R}$ is understood as some form of loss, i.e. the larger the worse. The expectation of a random variable X can be written as $\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}$. It was postulated in Yaari's dual theory of choice under risk (Yaari (1987)) that in the case where the decision maker is not risk neutral, the risk attitude can be captured by some form of expectation with the probability \mathbb{P} re-weighted, or distorted, according to some distortion function.

Definition 2.1. A function $h : [0, 1] \rightarrow [0, 1]$ is a distortion function if it is non-decreasing and satisfies $h(0) = h(0+) = 0$, and $h(1) = h(1-) = 1$ ¹.

Namely, as an analogy to expectation, a distorted expectation can be defined through the integral $\rho_h(X) = \int_{\Omega} X d(h \circ \mathbb{P})$. In general, the distorted measure $h \circ \mathbb{P}$ is non-additive, and therefore the integral needs to be defined in terms of Choquet integral (see e.g. Denneberg (1994a), and Denneberg (1994b)), which leads to the following standard definition of distortion risk measure.

Definition 2.2. Given a random variable X and a distortion function h , the functional

$$\rho_h(X) \triangleq \int_0^{\infty} h^*(\mathbb{P}(X > x)) dx - \int_{-\infty}^0 (1 - h^*(\mathbb{P}(X > x))) dx \quad (2.1)$$

$$= \int_0^{\infty} (1 - h(\mathbb{P}(X \leq x))) dx - \int_{-\infty}^0 h(\mathbb{P}(X \leq x)) dx, \quad (2.2)$$

where $h^*(p) = 1 - h(1 - p)$, $p \in [0, 1]$, is called a distortion risk measure.

While (2.1) follows the standard definition of Choquet integral, in the rest of this paper we will use (2.2) as our definition of distortion risk measure given its convenient interpretation of distorting the cumulative distribution function (CDF) of a random variable. In particular, we should point out that although the definition (2.2) is not an expectation with respect to the original probability space, it can be viewed, in the strict sense, as the expectation if one considers each random variable separately. To make our point precise, we provide the following alternative representation of (2.2), which is perhaps more accessible to readers who are less familiar with Choquet integral

$$\rho_h(X) = \int_{\mathbb{R}} x dh(F_X(x)), \quad (2.3)$$

where $F_X(x) = \mathbb{P}(X \leq x)$ is the CDF of X and the integral follows the standard definition of Lebesgue-Stieltjes integral². That is, distortion risk measures are expected values calculated

¹The conditions of $h(0) = h(0+) = 0$ and $h(1) = h(1-) = 1$ are necessary for the integrals in (2.1) to be finite when the random variable X is unbounded or has a support \mathbb{R} .

² Throughout the paper, for an increasing function $g : \mathbb{R} \rightarrow \mathbb{R}$ and a function $f : \mathbb{R} \rightarrow \mathbb{R}$, the Lebesgue-Stieltjes integral $\int_{\mathbb{R}} f(x) dg(x)$ is defined (see, for instance, Merkle et al. (2014)) as $\int_{\mathbb{R}} f(x) dg_+(x)$ or $\int_{\mathbb{R}} f(x) \mu_g(dx)$, where $g_+(x) = g(x+)$ and μ_g is a measure defined by $\mu_g([a, b]) = g(b+) - g(a-)$ for any $a \leq b$. That is to say that if an increasing function $g : \mathbb{R} \rightarrow \mathbb{R}$ is not right-continuous, $g(x)$ in the Lebesgue-Stieltjes integral $\int_{\mathbb{R}} f(x) dg(x)$ is treated as its right-continuous copy $g(x+)$. In this way, the integral is well defined.

based on “distorted” CDFs $F_X^h(x)$ with distortion functions of h , where $F_X^h(x) \triangleq h(F_X(x+)) = \lim_{y \downarrow x} h(F_X(y))$ is the right-continuous copy of $h(F_X(x))$ and $F_X^h(x)$ is a CDF for any distortion function h ³. Hence, the representation (2.3), although may not be considered as standard as (2.2), provides perhaps more clear justification of why distortion risk measures are considered as (distorted) expectations.

This unified view of distorted expectation (DE) is often considered “dual” to classical expected utility (EU). While EU essentially extends the expectation by “distorting” the values of random variables through a utility function, DE distorts the CDFs instead. As shown in Yaari (1987), if one replaces the controversial axiom of independence in EU theory by the axiom of dual independence⁴, then any preference relation \succeq ⁵ satisfying the axioms can be captured by some DE, i.e. $X \succeq Y$ if and only if there exists a distortion function h such that $\rho_h(X) \geq \rho_h(Y)$. For this reason, the functional ρ_h is also often referred to as dual utility functional. What makes DE as a potentially more attractive alternative to EU is that it further satisfies the axioms of risk measures. Recall that a functional ρ is called a monetary risk measure if it satisfies

1. Monotonicity: If $X \geq Y$, then $\rho(X) \geq \rho(Y)$.
2. Translation invariance: $\rho(X + c) = \rho(X) + c$ for any $c \in \mathbb{R}$.

Moreover, DE is the only candidate if one seeks a monetary risk measure satisfying the following axiom motivated by the common risk management practice.

3. Comonotonic additivity: $\rho(X + Y) = \rho(X) + \rho(Y)$ for any X and Y that are comonotonic.

Namely, the fact that two random variables are comonotonic, meaning that they move in the same direction for every possible outcome, implies that there should be no diversification benefit. It is well-known that any Choquet integral satisfies the three axioms, which explains why DE defined originally from Choquet integral also satisfies the three axioms.

As important examples of distortion risk measures, we review here two most well known risk measures, namely Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) (or Expected Shortfall (ES)). From here on, we write $F_X^{-1}(\alpha) \triangleq \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}$, $\alpha \in (0, 1)$ to denote the inverse CDF that is left-continuous and $F_X^{-1+}(\alpha) \triangleq \inf\{x \in \mathbb{R} : F_X(x) > \alpha\}$, $\alpha \in (0, 1)$ to denote the right-continuous one. Interestingly, there are two definitions of VaR appearing in the literature. The standard one of VaR at level α is simply $\text{VaR}_\alpha(X) = F_X^{-1}(\alpha)$ whereas the definition $\text{VaR}_\alpha^+(X) = F_X^{-1+}(\alpha)$ is applied in the literature of worst-case risk measure (El Ghaoui et al. (2003)). In the rest of this paper, we call $\text{VaR}_\alpha(X)$ the left-continuous Value-at-Risk (VaR) and

³Indeed, (2.3) and (2.2) are identical since $(2.3) = \int_{\mathbb{R}} x dF_X^h(x) = \int_0^\infty (1 - F_X^h(x)) dx - \int_{-\infty}^0 F_X^h(x) dx = \int_0^\infty (1 - h(F_X(x))) dx - \int_{-\infty}^0 h(F_X(x)) dx = (2.2)$.

⁴That is, if X, Y, Z are pairwise comonotonic and $X \succeq Y$, then we have $pX + (1-p)Z \succeq pY + (1-p)Z$, $\forall p \in [0, 1]$.

⁵defined over random variables whose values in the unit interval

$\text{VaR}_\alpha^+(X)$ the right-continuous Value-at-Risk (VaR). Recall that CVaR (or ES) at level α takes the form

$$\text{CVaR}_\alpha(X) = \text{ES}_\alpha(X) \triangleq \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(X) du. \quad (2.4)$$

It is straightforward to confirm that $\text{VaR}_\alpha(X)$, $\text{VaR}_\alpha^+(X)$, and $\text{CVaR}_\alpha(X)$ are distortion risk measures with distortion functions $h_{\text{VaR}}(x) = I_{\{\alpha \leq x \leq 1\}}$, $h_{\text{VaR}^+}(x) = I_{\{\alpha < x \leq 1\}}$, and $h_{\text{CVaR}}(x) = (x - \alpha)_+ / (1 - \alpha)$, respectively. Here $(x)_+ = \max\{x, 0\}$.

There are two other definitions of distortion risk measures commonly found in the literature. For the purpose of this paper, it is important to clarify the relationship between these definitions and the definition (2.2). Namely, given a distortion function h , they are written as the following Lebesgue-Stieltjes integrals of the $\text{VaR}_\alpha(X)$ and $\text{VaR}_\alpha^+(X)$ with respect to the distortion function $h(u)$ for $u \in (0, 1)$:

$$\int_{(0,1)} \text{VaR}_\alpha(X) dh(\alpha) \triangleq \int_0^1 \text{VaR}_\alpha(X) dh(\alpha), \quad (2.5)$$

and

$$\int_{(0,1)} \text{VaR}_\alpha^+(X) dh(\alpha) \triangleq \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha). \quad (2.6)$$

While these two definitions are often used interchangeably with the definition (2.2), they are not equivalent in general. More importantly, they can have different implications when it comes to studying their worst-case counterparts. Before discussing this point further, we first give in the following lemma the precise relationships between the distortion risk measure ρ_h defined by (2.2) and the functionals (2.5) and (2.6).

Lemma 2.1. *Let h be a distortion function and let $h_+(x) \triangleq h(x+)$ and $h_-(x) \triangleq h(x-)$ be the right-continuous and left-continuous copies of h , respectively. Then, the following assertions hold.*

(i) $h_- \leq h \leq h_+$, and

$$\rho_{h_+}(X) \leq \rho_h(X) \leq \rho_{h_-}(X),$$

where

$$\begin{aligned} \rho_{h_+}(X) &= \int_0^\infty (1 - h_+(\mathbb{P}(X \leq x))) dx - \int_{-\infty}^0 h_+(\mathbb{P}(X \leq x)) dx \\ \rho_{h_-}(X) &= \int_0^\infty (1 - h_-(\mathbb{P}(X \leq x))) dx - \int_{-\infty}^0 h_-(\mathbb{P}(X \leq x)) dx. \end{aligned}$$

(ii) $\rho_{h_+}(X)$ and $\rho_{h_-}(X)$ respectively have the following Lebesgue-Stieltjes integral representations:

$$\rho_{h_+}(X) = \int_0^1 \text{VaR}_\alpha(X) dh(\alpha) \quad (2.7)$$

and

$$\rho_{h_-}(X) = \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha). \quad (2.8)$$

(iii) If h is right continuous, then $\rho_h(X) = \int_0^1 \text{VaR}_\alpha(X) dh(\alpha)$; if h is left continuous, then $\rho_h(X) = \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha)$; and if h is continuous, then

$$\rho_h(X) = \int_0^1 \text{VaR}_\alpha(X) dh(\alpha) = \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha). \quad (2.9)$$

The main takeaway from the above lemma is that in the case where the distortion function h is discontinuous, the risk measure ρ_h might not have the representation (2.5) (or (2.6)). As a perhaps obvious example, while $\text{VaR}_\alpha^+(X)$ is a distortion risk measure with the distortion function $h_{\text{VaR}^+}(x) = I_{\{\alpha < x \leq 1\}}$, it does not have a representation (2.5). This stresses the importance of applying the definition (2.2) (or (2.3)) rather than (2.5) (or (2.6)) throughout this work. Moreover, another reason why emphasis needs to be put on this technical detail is that the discontinuity of h in fact has important implication on the existence of worst-case distributions. As an important example, the well known result of worst-case Value-at-Risk (El Ghaoui et al. (2003)) is derived based on the definition of $\text{VaR}_\alpha^+(X)$ rather than the standard definition of Value-at-Risk, i.e. $\text{VaR}_\alpha(X)$. What appears less noticed however is that if one replaces the definition of Value-at-Risk in El Ghaoui et al. (2003) by the standard definition $\text{VaR}_\alpha(X)$, there actually does not exist a distribution that attains the worst-case risk (more details can be found in Remark 3.11 of Section 3.3). This is important to point out, not only because the description about the worst-case distributions should be made precise but also because the analysis of worst-case risk measures often hinges heavily on the exact behaviour of the worst-case distributions.

In what follows, we discuss several important examples of distortion risk measures. Figures 1-3 demonstrate some of the distortion functions and the corresponding distorted CDFs where the original CDF takes the form of an empirical distribution. It is well-known (see, for example, Mao and Cai (2018)) that a distortion risk measure ρ_h is a coherent risk measure if and only if the distortion function h is convex.

Example 2.1. (Convex distortion functions) Given that h is a convex distortion function, it has derivative almost everywhere on $[0, 1]$. Denote by $D \subseteq [0, 1]$ the set of points where h is differentiable. Then for any function ϕ defined on $[0, 1]$ satisfying $\phi(\alpha) = h'(\alpha)$ for $\alpha \in D$, we have $\mu(\{\alpha \in [0, 1], \phi(\alpha) \neq h'(\alpha)\}) = 0$, where μ is the Lebesgue measure on $[0, 1]$. It is thus not hard to confirm that any distortion risk measure with a convex distortion function is equivalent to the following definition of spectral risk measure (Acerbi (2002))

$$\rho_\phi(X) = \int_0^1 \text{VaR}_\alpha(X) \phi(\alpha) d\alpha,$$

where $\phi : [0, 1] \rightarrow \mathbb{R}_+$ is right-continuous, monotonically nondecreasing and satisfies $\int_0^1 \phi(\alpha) d\alpha = 1$. Among several examples of convex distortion risk measures, the most well known ones (see e.g. Wang (2000) and Wozabal (2014)) are Wang transform where $h(\alpha) = 1 - \Phi(\Phi^{-1}(1 - \alpha) + \lambda)$ and the proportional hazard transform where $h(\alpha) = 1 - (1 - \alpha)^r$. \square

The most prominent example of distortion function used in behavioural economics and finance is the inverse S-shaped distortion function. A distortion function h is called inverse S-shaped (S-shaped) if there exists a point $t^* \in (0, 1)$ such that h is concave (convex) on $[0, t^*)$ and is convex (concave) on $(t^*, 1]$.

Example 2.2. (Inverse S-shaped distortion functions)

- (i) The following inverse S-shaped distortion is applied in the seminal work of Tversky and Kahneman (1992):

$$h(t) = \frac{t^\alpha}{(t^\alpha + (1-t)^\alpha)^{1/\alpha}}, \quad \alpha^* \leq \alpha < 1, \quad (2.10)$$

where $\alpha^* = 0.279$.

- (ii) A two-parameter inverse S-shaped distortion function is proposed in Guegan and Hassani (2015) as a polynomial function of degree 3 with the following form:

$$h(t) = a \left(\frac{t^3}{6} - \frac{\delta}{2} t^2 + \left(\frac{\delta^2}{2} + \beta \right) t \right), \quad (2.11)$$

where $0 < \delta < 1$, $\beta \in \mathbb{R}$, and

$$a = \left(\frac{1}{6} - \frac{\delta}{2} + \frac{\delta^2}{2} + \beta \right)^{-1}.$$

- (iii) The inverse S-shaped distortion function used in Xu and Zhou (2013) has the following form:

$$h(t) = \begin{cases} 2t - 2t^2, & 0 \leq t \leq \frac{1}{2}, \\ 2t^2 - 2t + 1, & \frac{1}{2} < t \leq 1. \end{cases} \quad (2.12)$$

For more examples of inverse S-shaped distortion functions proposed in behavioural economics and finance, we refer readers to Prelec (1998), Wu and Gonzalez (1996), Bleichrodt and Pinto (2000) and the references therein. \square

Since the work of Cont et al. (2010), increasing attention has been paid to the issue of robustness of a risk measure. Range Value-at-Risk (RVaR) proposed by Cont et al. (2010) has now been considered a useful risk measure that can resolve the intrinsic conflict between the sub-additivity and robustness of a risk measure. RVaR is closely related to, and in fact is a special case of, the risk measure GlueVaR proposed more recently by Belles-Sampera et al. (2014). GlueVaR is more attractive than VaR and CVaR in that it can strike a fine balance between the two and helps different parties reach a consensus.

Example 2.3. (GlueVaR and RVaR) A distortion risk measure ρ_g is called a GlueVaR risk measure, denoted as $\text{GlueVaR}_{\alpha,\beta}^{h_1,h_2}$, if its distortion function g has the following expression ⁶:

$$g(u) = g_{\alpha,\beta}^{h_1,h_2}(u) = \begin{cases} 0, & 0 \leq u < \alpha, \\ 1 - h_1 + \frac{h_2-h_1}{\beta-\alpha}(u - \beta), & \alpha \leq u < \beta, \\ 1 + \frac{h_1(u-1)}{1-\beta}, & \beta \leq u \leq 1, \end{cases} \quad (2.13)$$

where the constants α , β , h_1 and h_2 satisfy

$$0 < \alpha < \beta < 1 \quad \text{and} \quad 0 \leq h_1 \leq h_2 \leq 1. \quad (2.14)$$

Obviously, if $h_1 = h_2 = 0$, the distortion function $g_{\alpha,\beta}^{0,0}(u)$ is reduced to the distortion function $h_{\text{VaR}}(u) = I_{\{\alpha \leq u \leq 1\}}$ of the risk measure VaR_α and if $h_1 = h_2 = 1$, the distortion function $g_{\alpha,\beta}^{1,1}(u)$ is reduced to the distortion function $h_{\text{CVaR}}(u) = (u - \beta)_+ / (1 - \beta)$ of the risk measure CVaR_β . Furthermore, if $h_1 = 0$ and $h_2 = 1$, the distortion function $g_{\alpha,\beta}^{0,1}(u)$ is reduced to

$$h_{\alpha,\beta}(t) = \begin{cases} 0 & 0 \leq u \leq \alpha, \\ \frac{u-\alpha}{\beta-\alpha}, & \alpha < u \leq \beta, \\ 1, & \beta < u \leq 1. \end{cases} \quad (2.15)$$

and the corresponding distortion risk measure reduces to the Range Value-at-Risk introduced by Cont et al. (2010), denoted as $\text{RVaR}_{\alpha,\beta}$, which is defined as

$$\text{RVaR}_{\alpha,\beta}(X) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \text{VaR}_t(X) dt, \quad 0 < \alpha < \beta < 1. \quad (2.16)$$

Therefore, all the three commonly used distortion risk measures VaR , CVaR , and RVaR are the special cases of the GlueVaR. Moreover, $g_{\alpha,\beta}^{h_1,h_2}(u)$ is equal to zero on $(0, \alpha)$ and thus $g_{\alpha,\beta}^{h_1,h_2}(u)$ is both convex and concave on $(0, \alpha)$. In addition, $g_{\alpha,\beta}^{h_1,h_2}(u)$ is concave on $(\alpha, 1)$ if $\frac{h_2-h_1}{\beta-\alpha} \geq \frac{h_1}{\beta}$ and is convex on $(\alpha, 1)$ if $\frac{h_2-h_1}{\beta-\alpha} \leq \frac{h_1}{\beta}$. Therefore, $\text{GlueVaR}_{\alpha,\beta}^{h_1,h_2}$ is an S-shaped distortion risk measure if $\frac{h_2-h_1}{\beta-\alpha} \geq \frac{h_1}{\beta}$ and an inverse S-shaped distortion risk measure if $\frac{h_2-h_1}{\beta-\alpha} \leq \frac{h_1}{\beta}$. \square

The following Beta distortion function proposed by Wirth (1999) has wide applications in actuarial science, finance, and insurance management. It can be applied to easily fit various forms of risk attitudes.

⁶ We point out that the GlueVaR risk measure defined in Belles-Sampera et al. (2014) or Definition 3.1 of Cai et al. (2017) is based on (2.2) with the following distortion function:

$$g^*(u) = \begin{cases} \frac{h_1}{1-\beta}u, & 0 \leq u \leq 1 - \beta, \\ h_1 + \frac{h_2-h_1}{\beta-\alpha}(u - (1 - \beta)), & 1 - \beta < u \leq 1 - \alpha, \\ 1, & 1 - \alpha < u \leq 1, \end{cases}$$

where the constants α , β , h_1 and h_2 satisfy $0 < \alpha < \beta < 1$ and $0 \leq h_1 \leq h_2 \leq 1$. Thus, based on (2.1), the distortion function for the GlueVaR is $g(u) = 1 - g^*(1 - u)$, which is (2.13).

Example 2.4. (Beta distortion function) The Beta distortion function is defined by the incomplete beta function as follows

$$h(t) = \frac{1}{B(\alpha, \beta)} \int_0^t u^{\alpha-1} (1-u)^{\beta-1} du, \quad \alpha, \beta > 0, \quad (2.17)$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function (Wirth, 1999). The beta distortion model is quite flexible as with different choices of α and β the distortion function could be convex ($\alpha \geq 1, \beta \leq 1$), concave ($\alpha \leq 1, \beta \geq 1$), S-shaped ($\alpha, \beta \geq 1$) or inverse S-shaped ($\alpha, \beta \leq 1$) distortion functions. It reduces to the proportional hazard transform when $\alpha = 1$.

3 Tractability of Distributionally Robust Risk Optimization

DRO formulated based on distorted expectation is more involved than classical DRO in that distorted expectation might be non linear and non-convex in probability distribution. This makes it not possible to apply common techniques such as duality theory for moment problem to analyze the tractability of the problem. We begin our investigation of the tractability by making first the following assumption, which is fairly standard in the literature of DRO.

Assumption 3.1. The feasible set \mathcal{W} is convex and the cost function $f(\vec{w}, \vec{x}) : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ is convex in \vec{w} and concave in \vec{x} .

This assumption ensures that in the special case where the distortion risk measure reduces to an expectation, the problem of DRRO reduces to a classical DRO problem that admits a convex optimization formulation. As noted for example in Wiesemann et al. (2014) there is a wide range of applications that can be formulated as DRO satisfying the above assumption. One popular instance is the case of portfolio management where $f(\vec{w}, \vec{x})$ represents a long-only portfolio that consists of assets whose returns are convex in the underlying uncertain variables. In this case, $f(\vec{w}, \vec{x})$ takes the form of a weighted sum of concave functions. Another important class of applications is the two-stage decision making problems under cost uncertainty that have been studied extensively for example in Delage et al. (2014). In these problems, the cost function can be generally written as $f(\vec{w}, \vec{x}) = c^\top \vec{w} + h(\vec{w}, \vec{x})$ where $c^\top \vec{w}$ represents the first-stage (deterministic) cost and $h(\vec{w}, \vec{x})$ represents the second-stage (recourse) cost after the uncertain cost \vec{x} is realized and a recourse decision \vec{y} is made, e.g.

$$\begin{aligned} h(\vec{w}, \vec{x}) := & \min_{\vec{y} \in \mathcal{Y}} \quad \vec{x}^\top C \vec{y} \\ & \text{subject to} \quad A\vec{w} + B\vec{y} \leq b. \end{aligned}$$

In what follows, we first base our discussion of the tractability of DRRO on two ambiguity sets that are commonly applied in the literature of DRO (Sections 3.1 and 3.2) and then pay a particular attention to the special case of linear cost function (Section 3.3).

3.1 The Case of Support and Bounded Moments

One common scheme in DRO to specify an ambiguity set is to describe “the maximum level of dispersion” of distributions from their mean (see e.g. Wiesemann et al. (2014)). In particular, the dispersion can be characterized by a set of support and moments defined based on convex functions. The following ambiguity set, which was first proposed in Delage et al. (2014), provides a general description of the case

$$\mathcal{F}_{\vec{m}, \mathcal{G}} = \left\{ F \in \mathcal{F}_0(\mathbb{R}^d) \left| \begin{array}{l} \int \vec{X} dF = \vec{m} = (m_1, \dots, m_d)^\top, \\ \int g(\vec{X}) dF \leq 0, \quad g \in \mathcal{G}, \\ \text{where } \mathcal{G} \text{ is a set of convex functions on } \mathbb{R}^d, \\ \int_{\{\vec{X} \in G\}} dF = 1, \text{ where } G \in \mathbb{R}^d \text{ is a convex set,} \end{array} \right. \right\}, \quad (3.1)$$

where $\mathcal{F}_0(\mathbb{R}^d)$ denotes the probability distributions on \mathbb{R}^d . The above set encompasses many examples of ambiguity sets from the literature and is highly expressive in that the set \mathcal{G} could include infinitely many convex functions. For instance, one could consider the following set

$$\mathcal{G} = \left\{ g : \mathbb{R}^d \rightarrow \mathbb{R} \mid \exists \vec{z} \in \mathbb{R}^d, g(\vec{x}) = h(\vec{z}^\top(\vec{x} - \vec{m})) - b(\vec{z}) \right\},$$

where $h : \mathbb{R} \rightarrow \mathbb{R}$ denotes some convex function. Let Σ denote the covariance of \vec{X} . One can then recover the mean-absolute-deviation ambiguity set (Postek et al. (2018)) by setting $h(z) = |z|$ and $b(\vec{z}) = \sqrt{\vec{z}^\top \Sigma \vec{z}}$, and the mean-covariance ambiguity set proposed in Delage and Ye (2010), i.e.

$$\left\{ F \in \mathcal{F}_0(\mathbb{R}^d) \mid \mathbb{P}(\vec{X} \in G) = 1, \mathbb{E}_F[\vec{X}] = \vec{m}, \mathbb{E}_F[(\vec{X} - \vec{m})(\vec{X} - \vec{m})^\top] \preceq \Sigma \right\} \quad (3.2)$$

by setting $h(z) = z^2/2$ and $b(\vec{z}) = \vec{z}^\top \Sigma \vec{z}$. One can also choose any $b(\vec{z}) : \mathbb{R}^d \rightarrow \mathbb{R}$ to represent the upper bound of dispersion along the direction \vec{z} (Hanasusanto et al. (2017)) and recover the mean semi-variance ambiguity set (Delage et al. (2014)) by setting $h(z) = (\max\{0, z\})^2$ and the mean Huber ambiguity set (Wiesemann et al. (2014)) by setting $h(z) = z^2/2$ if $|z| \leq \delta$ and $h(z) = \delta(|z| - \delta/2)$ for some $\delta > 0$.

In the remainder of this section, we investigate the tractability of DRRO based on $\mathcal{F}_{\vec{m}, \mathcal{G}}$. We proceed by focusing first on its inner maximization problem, i.e. the evaluation of the worst case distortion risk measure:

$$\sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}} \rho_h^F(f(\vec{w}, \vec{X})). \quad (3.3)$$

The above problem appears particularly difficult to solve when the distortion function h is non-convex, in which case the problem is a non-convex optimization problem. We tackle this difficulty by taking a common strategy for non-convex optimization problems. Namely, we first approximate the problem by a more conservative convex optimization problem, and then prove the tightness of the approximation whenever possible. In particular, we use the fact that given any two distortion

functions satisfying $h_1 \leq h_2$, their corresponding distortion risk measures must satisfy $\rho_{h_1}(X) \geq \rho_{h_2}(X)$ for any X . Thus, by finding a convex distortion function h_{cx} that satisfies $h_{\text{cx}} \leq h$, we can bound from above the objective function in (3.3), i.e. $\rho_h^F(f(\vec{w}, \vec{X}))$ by a convex objective function $\rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X}))$. To tighten the bound, we look for the “largest” convex distortion function that is dominated by the original distortion function h . This largest convex function is also known as the convex envelope of h .

Definition 3.1. (Convex envelope (e.g. Brighi and Chipot (1994))) For any distortion function h , the largest convex distortion function dominated by h is denoted by h_{cx} , which is called convex envelope of h and is defined as

$$h_{\text{cx}}(t) = \sup \{g(t) \mid g : [0, 1] \rightarrow [0, 1], g \leq h, g \text{ is increasing and convex on } [0, 1]\}, \quad t \in [0, 1].$$

We can thus write down the following problem that bounds from above the original problem

$$\sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}} \rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X})). \quad (3.4)$$

To show how tightly the above problem approximates the original problem (3.3), we will need some intermediate steps. We should mention here that as our goal is to present the steps from a high level point of view, we defer many parts of the proof, particularly those that are technically heavy, to the appendix. The key step of our analysis is to narrow down a subset of distributions in the ambiguity set $\mathcal{F}_{\vec{m}, \mathcal{G}}$ that maximize (3.4). In particular, we rely on the following structure of a convex envelope h_{cx} to characterize the subset of distributions.

Lemma 3.1. *Let h be a distortion function and h_{cx} be the convex envelope of h .*

(i) *If h is right-continuous, then the set*

$$\{t \in [0, 1] : h(t) \neq h_{\text{cx}}(t)\} = \cup_{k \in I_1} [a_k, b_k) \cup (\cup_{\ell \in I_2} (a_\ell, b_\ell)) \quad (3.5)$$

is the union of some disjoint left-closed and right-open intervals and open intervals. Moreover, h is continuous at b_k for $k \in I_1 \cup I_2$ and at a_ℓ for $\ell \in I_2$; h_{cx} is linear on each open interval (a_k, b_k) for $k \in I_1 \cup I_2$; and

$$h(a_k-) = h_{\text{cx}}(a_k) \text{ for } k \in I_1 \text{ and } h(x) = h_{\text{cx}}(x) \text{ for } x \in \{b_k : k \in I_1 \cup I_2\} \cup \{a_\ell : \ell \in I_2\}. \quad (3.6)$$

(ii) *If h is left-continuous, then the set*

$$\{t \in [0, 1] : h(t) \neq h_{\text{cx}}(t)\} = \cup_{k \in I} (a_k, b_k) \quad (3.7)$$

is the union of some disjoint open intervals. Moreover, h is continuous at a_k and b_k for $k \in I$; h_{cx} is linear on each open interval (a_k, b_k) for $k \in I$; and

$$h(a_k) = h_{\text{cx}}(a_k) \text{ and } h(b_k) = h_{\text{cx}}(b_k), \quad k \in I. \quad (3.8)$$

The above lemma, simply put, states that any convex envelope would either coincide with the original distortion function or take the shape of a linear function over intervals where the two do not coincide. Note that a similar result was first proved in Brighi and Chipot (1994) for the case of continuous distortion function. Here we generalise it to cover the case of right and left continuous distortion functions. It turns out that in many cases the distributions that maximize (3.4) have a simpler structure for the parts that correspond to the linear pieces of the convex envelope. We summarize this observation below.

Proposition 3.2. *In the case where $f(\vec{w}, \vec{x})$ is concave in \vec{x} , we have*

$$\rho_{h_{\text{cx}}, f, \vec{w}}^{\uparrow}(\vec{X}) \triangleq \sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}} \rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}} \rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X})), \quad (3.9)$$

where

$$\mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}} = \left\{ F \in \mathcal{F}_{\vec{m}, \mathcal{G}} \mid \begin{array}{l} \text{VaR}_{\alpha}^F(f(\vec{w}, \vec{X})) \text{ is constant} \\ \text{on each open interval on which } h_{\text{cx}}(\alpha) \text{ is linear.} \end{array} \right\}.$$

Moreover, if the worst case value $\rho_{h_{\text{cx}}, f, \vec{w}}^{\uparrow}(\vec{X})$ is attainable, then a worst-case distribution F can be obtained in the set $\mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$, namely, there exists a distribution $F^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$ such that $\rho_{h_{\text{cx}}, f, \vec{w}}^{\uparrow}(\vec{X}) = \rho_{h_{\text{cx}}}^{F^*}(f(\vec{w}, \vec{X}))$.

While we leave the technical details of the proof of Proposition 3.2 in the appendix, we point out here the main idea behind the proof. Namely, our key observation is that in the case where the cost function $f(\vec{w}, \vec{x})$ is concave in \vec{x} , it is always possible to construct a worst-case random vector $\vec{X}_{F_{\text{wc}}}$ that takes one point on each subset $\{a_k \leq U \leq b_k\}$ for $k \in I$, where U and $f(\vec{w}, \vec{X}_{F_{\text{wc}}})$ are comonotonic, and $U \sim \text{U}(0, 1)$. In other words, whenever the distortion function is not convex, i.e. there are parts where the function differs from its convex envelope, the distributions that maximize (3.4) would always assign some point masses according to where the envelope is linear. Each point mass corresponds to quantiles that are constant over the interval where the envelope is linear, i.e. (a_k, b_k) , and thus the probability of the point mass is determined by the range of the interval (a_k, b_k) .

Remark 3.3. *When $h(x) = x$ for $x \in [0, 1]$, i.e. the distortion risk measure reduces to the expectation, we have from Proposition 3.2 that the worst-case distribution to the optimization problem (3.4) is simply a point mass at \vec{m} . This recovers one of the key results in Delage et al. (2014), which states that the solution optimized based on expected value is in fact robust with respect to worst-case expected cost.*

We are now ready to present the main result of this section regarding the tightness of the problem (3.4). For readability we present only the proof for the case where the distortion function is left continuous and leave other more technical details to the appendix. The general idea behind the proof is fairly straightforward. Namely, it can be shown that the worst case distribution that maximizes (3.4) also attains the worst case value for the original problem (3.3).

Theorem 3.4. *Given any distortion function h and a cost function $f(\vec{w}, \vec{x})$ that is concave in \vec{x} , we have*

$$\sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}} \rho_h^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}} \rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X})) \quad (3.10)$$

for any \vec{w} . In addition, if the problem (3.9) is attainable and h is left-continuous, then the problem (3.3) is also attainable at the worst-case distribution $F_{\text{wc}} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$.

Proof. First let h be a left-continuous distortion function. Note that $\rho_h(X) \leq \rho_{h_{\text{cx}}}(X)$ for each random variable X . Then by (3.9), it suffices to show for each $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$ and $\vec{X} \sim F$, we have

$$\rho_h(f(\vec{w}, \vec{X})) = \rho_{h_{\text{cx}}}(f(\vec{w}, \vec{X})). \quad (3.11)$$

By Lemma 3.1 (ii), we have that the set $\Delta =: \{t \in [0, 1] : h_-(t) \neq h_{\text{cx}}(t)\}$ has the expression (3.7), that is, $\Delta = \cup_{k \in I} (a_k, b_k)$, and h_{cx} is linear in each interval of (a_k, b_k) , $k \in I$. Thus, $\text{VaR}_\alpha(f(\vec{w}, \vec{X}))$ is a constant on each interval of (a_k, b_k) , $k \in I$. We denote $x_k = \text{VaR}_\alpha(f(\vec{w}, \vec{X}))$ when $\alpha \in (a_k, b_k)$ for $k \in I$. Note that for an increasing function g and $a < b$, we have $\int_{(a,b)} dg(\alpha) = \int_{(a,b)} dg^+(\alpha) = g^+(b-) - g^+(a) = g(b-) - g(a+)$. Hence, by (2.8), we have

$$\begin{aligned} \rho_h(f(\vec{w}, \vec{X})) &= \int_0^1 \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh(\alpha) \\ &= \int_{(0,1) \setminus \Delta} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh(\alpha) + \sum_{k \in I} \int_{(a_k, b_k)} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh(\alpha) \\ &= \int_{(0,1) \setminus \Delta} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh_{\text{cx}}(\alpha) + \sum_{k \in I} x_k (h(b_k-) - h(a_k+)) \\ &= \int_{(0,1) \setminus \Delta} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh_{\text{cx}}(\alpha) + \sum_{k \in I} x_k (h(b_k) - h(a_k)) \\ &= \int_{(0,1) \setminus \Delta} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh_{\text{cx}}(\alpha) + \sum_{k \in I} x_k (h_{\text{cx}}(b_k) - h_{\text{cx}}(a_k)) \\ &= \int_{(0,1) \setminus \Delta} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh_{\text{cx}}(\alpha) + \sum_{k \in I} \int_{(a_k, b_k)} \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) dh_{\text{cx}}(\alpha) \\ &= \rho_{h_{\text{cx}}}(f(\vec{w}, \vec{X})), \end{aligned}$$

where the forth equality follows that h is continuous at a_k and b_k (Theorem 3.1(ii)) and the fifth equality follows from (3.8). Thus, we showed (3.10) for left-continuous h . If $\rho_{h_{\text{cx}}, f, \vec{w}}^\uparrow(\vec{X})$ is attainable, then a worst-case distribution can be chosen in the set $\mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$. Then by (3.11), we have $\rho_{h, f, \vec{w}}^\uparrow(\vec{X})$ is also attainable at the worst-case distribution $F_{\text{wc}} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$. \square

Hence, we arrive at the conclusion that in many cases, the convex relaxation of a DRRO problem can in fact be tight. What is perhaps surprising is the generality of the result. In the literature of DRO, such kind of tightness result, while highly desirable, usually can only be established for

fairly specialized cases (see e.g. Ben-Tal et al. (2015), Mak et al. (2015)). The above result, on the other hand, reveals that such "hidden convexity" can actually be systematically found in DRRO problems through the use of convex envelope. The problem now boils down to identifying the exact functional form of the convex envelope of a distortion function. As shown below, this is usually straightforward to do.

Example 3.1. (Inverse S-shaped distortion risk measures)

(i) It is easy to see that the convex envelope of h defined by (2.10) is

$$h_{\text{cx}}(t) = \begin{cases} h'(t_{\text{cx}}) t, & 0 \leq t \leq t_{\text{cx}}, \\ h(t), & t_{\text{cx}} < t \leq 1, \end{cases} \quad (3.12)$$

where t_{cx} is the unique solution to the following equation

$$(2 - \alpha)(t^\alpha + (1 - t)^\alpha) = (1 - t)^{\alpha-1}$$

and

$$h'(t) = h(t) \left(\frac{\alpha}{t} - \frac{t^{\alpha-1} - (1 - t)^{\alpha-1}}{t^\alpha + (1 - t)^\alpha} \right). \quad (3.13)$$

(ii) For the two-parameter inverse S-shaped distortion function defined by (2.11), its convex envelope is $h_{\text{cx}}(t) = t$, $t \in [0, 1]$, if $0 < \delta \leq 2/3$ and

$$h_{\text{cx}}(t) = \begin{cases} a \left(\frac{\delta^2}{8} + \beta \right) t, & 0 < t \leq \frac{3}{2}\delta, \\ a \left(\frac{t^3}{6} - \frac{\delta}{2} t^2 + \left(\frac{\delta^2}{2} + \beta \right) t \right), & \frac{3}{2}\delta < t < 1, \end{cases} \quad (3.14)$$

if $2/3 < \delta < 1$.

(iii) The convex envelope of the inverse S-shaped distortion function defined by (2.12) is form:

$$h(t) = \begin{cases} 2(\sqrt{2} - 1)t, & 0 \leq t \leq \frac{\sqrt{2}}{2}, \\ 2t^2 - 2t + 1, & \frac{\sqrt{2}}{2} < t \leq 1. \end{cases}$$

Example 3.2. (GlueVaR and RVaR) For the GlueVaR distortion function $g_{\alpha, \beta}^{h_1, h_2}$ defined in (2.13), it is easy to see that if $\frac{1-h_1}{\beta-\alpha} \leq \frac{h_1}{1-\beta}$, the convex envelope of $g_{\alpha, \beta}^{h_1, h_2}$ is

$$g_{\text{cx}}(t) = \begin{cases} 0, & 0 \leq t \leq \alpha, \\ \frac{(1-h_1)(t-\alpha)}{\beta-\alpha}, & \alpha \leq t \leq \beta, \\ 1 + \frac{h_1(t-1)}{1-\beta}, & \beta \leq t \leq 1, \end{cases} \quad (3.15)$$

and if $\frac{1-h_1}{\beta-\alpha} \geq \frac{h_1}{1-\beta}$, the convex envelope of $g_{\alpha, \beta}^{h_1, h_2}$ is

$$g_{\text{cx}}(t) = \begin{cases} 0, & 0 \leq t \leq \alpha, \\ \frac{t-\alpha}{1-\alpha}, & \alpha < t \leq 1, \end{cases} \quad (3.16)$$

which is equal to the distortion function of CVaR_α . It thus follows immediately that the convex envelope of the distortion function $h_{\alpha,\beta}(t)$ of $\text{RVaR}_{\alpha,\beta}$ is the distortion function of CVaR_α given by (3.16).

Example 3.3. (Beta-distortion function) For the Beta distortion function h defined by (2.17), we discuss its convex envelope h_{cx} for the following cases.

- (i) If $\alpha \geq 1, \beta \leq 1$, then h is convex and thus its convex envelope $h_{\text{cx}} = h$.
- (ii) If $\alpha \leq 1, \beta \geq 1$, then h is concave and thus its convex envelope is identity function $h_{\text{cx}}(t) = t$ for $t \in [0, 1]$.
- (iii) If $\alpha, \beta \geq 1$, then h is an S-shaped function and its convex envelope is

$$h_{\text{cx}}(t) = \begin{cases} h(t), & 0 \leq t \leq t_{\text{cx}}, \\ 1 + \frac{1}{B(\alpha,\beta)} t_{\text{cx}}^{\alpha-1} (1 - t_{\text{cx}})^{\beta-1} (t - 1), & t_{\text{cx}} < t \leq 1. \end{cases}$$

where $t_{\text{cx}} \in [0, 1)$ is the unique solution to the following equation

$$t^{\alpha-1} (1 - t)^\beta = \int_t^1 u^{\alpha-1} (1 - u)^{\beta-1} du.$$

- (iv) If $\alpha, \beta \leq 1$, then h is an inverse S-shaped function and its convex envelope is

$$h_{\text{cx}}(t) = \begin{cases} \frac{1}{B(\alpha,\beta)} t_{\text{cx}}^{\alpha-1} (1 - t_{\text{cx}})^{\beta-1} t, & 0 \leq t \leq t_{\text{cx}}, \\ h(t), & t_{\text{cx}} < t \leq 1. \end{cases}$$

where $t_{\text{cx}} \in (0, 1]$ is the unique solution to the following equation

$$t^\alpha (1 - t)^{\beta-1} = \int_0^t u^{\alpha-1} (1 - u)^{\beta-1} du.$$

Up to this point, we have shown that any DRRO problem can be solved by replacing its inner maximization problem by its convex counterpart (3.4). We next discuss how to solve numerically the problem (3.4) and the corresponding DRRO problem. In particular, our goal is to show how the problem can be generally solved as conic programs (Ben-Tal and Nemirovski (2001)). There are several conic programs that can be solved efficiently by off-the-shelf solvers, and we refer to the cones used in these programs as tractable cones, including non-negative cone, second-order cone, positive semidefinite cone, and their Cartesian product. We say a set $\mathcal{X} \subset \mathbb{R}^m$ is conic-representable if it can be represented by the projection of a tractable cone \mathcal{K} . We make the following assumptions **(C1)**–**(C3)** about the ambiguity set $\mathcal{F}_{\vec{m}, \mathcal{G}}$. Recall that the conic inequality $\vec{x} \preceq_{\mathcal{K}} \vec{y}$ refers to the set constraint $\vec{y} - \vec{x} \in \mathcal{K}$ and the dual cone \mathcal{K}^* refers to the set $\mathcal{K}^* = \{\vec{y} \mid \vec{y}^\top \vec{x} \geq 0, \forall \vec{x} \in \mathcal{K}\}$.

(C1) The set of support G is conic-representable and the set of convex functions \mathcal{G} admits the representation

$$\mathcal{G} = \left\{ g : \mathbb{R}^d \rightarrow \mathbb{R} \mid \exists \vec{z} \in \mathcal{K}^*, g(\vec{x}) = \vec{z}^\top (\bar{g}(\vec{x}) - \bar{b}) \right\}$$

for some function $\bar{g} : \mathbb{R}^d \rightarrow \mathbb{R}^p$, constant vector $\bar{b} \in \mathbb{R}^p$ and proper cone $\mathcal{K} \subset \mathbb{R}^p$, where $\vec{z}^\top \bar{g}(\vec{x})$ is convex for any $\vec{z} \in \mathcal{K}^*$ and the \mathcal{K} -epigraph of \bar{g} , i.e. $\{(\vec{x}, \vec{y}) \mid \bar{g}(\vec{x}) \preceq_{\mathcal{K}} \vec{y}\}$ is conic-representable.

(C2) $\bar{g}(\vec{m}) \preceq_{\mathcal{K}} \bar{b}$ and $\vec{m} \in \text{int}(G)$ hold.

With the condition (C1), we can equivalently express any set $\mathcal{F}_{\vec{m}, \mathcal{G}}$ defined in (3.1) as

$$\mathcal{F}_{\vec{m}, \mathcal{K}} = \left\{ F \in \mathcal{F}_0(\mathbb{R}^d) \mid \begin{array}{l} \int \vec{X} dF = \vec{m} = (m_1, \dots, m_d)^\top, \\ \int \bar{g}(\vec{X}) dF \preceq_{\mathcal{K}} \bar{b}, \\ \text{where } \mathcal{K} \subset \mathbb{R}^p \text{ is a proper cone,} \\ \int_{\{\vec{X} \in G\}} dF = 1, \text{ where } G \subseteq \mathbb{R}^d \text{ is conic-representable.} \end{array} \right\}, \quad (3.17)$$

which recovers the representation of ambiguity set applied in the work of Wiesemann et al. (2014). The condition (C2) allows us to verify certain Slater condition required for deriving an equivalent formulation of the problem (3.4). Note that the condition is equivalent to checking if the distribution $F = \mathbf{1}_{\vec{m}}$, i.e. putting all the mass at the mean, is strictly feasible in (3.17). Next, we assume that the convex distortion function h_{cx} is piecewise linear.

(C3) There exist $0 = a_0 < a_1 < \dots < a_m = 1$, $m \in \mathbb{N}$, such that $(0, 1] = \cup_{j=1}^m (a_{j-1}, a_j]$ and h_{cx} is linear on $(a_{j-1}, a_j]$, $j = 1, \dots, m$.

Such kind piecewise linear assumption is common in the literature of DRO. For example, in the context of utility-based DRO, the utility function is often assumed to take a piecewise linear form (see e.g. Wiesemann et al. (2014)), and the idea is that such a functional form can always be applied to well approximate a general function over a bounded domain. The condition (C3) obviously holds for VaR, VaR⁺, CVaR, RVaR, and GlueVaR, since their convex envelopes are piecewise linear. Note that for any convex distortion function h_{cx} that has bounded derivatives, one can always find a piecewise linear function h_m that satisfies $\rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X})) \leq \rho_{h_m}^F(f(\vec{w}, \vec{X})), \forall \vec{X}$. That is, the problem $\sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}} \rho_{h_m}^F(f(\vec{w}, \vec{X}))$ provides a conservative approximation to the problem (3.4). We show in the following result that with the above conditions, the problem of worst-case distortion risk measure (3.4) can be solved by a convex optimization problem with a finite number of variables and constraints.

Theorem 3.5. *Assume that the conditions (C1)-(C3) hold. The optimization problem (3.4) can*

be equivalently solved by the following convex optimization problem:

$$\sup_{\{\vec{x}_j\}_{j=1}^m} \sum_{j=1}^m \tilde{\phi}_j p_j f(\vec{w}, \vec{x}_j) \quad (3.18)$$

$$\text{subject to } \sum_{j=1}^m p_j \vec{x}_j = \vec{m}, \quad (3.19)$$

$$\sum_{j=1}^m p_j \bar{g}(\vec{x}_j) \preceq_{\mathcal{K}} \bar{b}, \quad (3.20)$$

$$\vec{x}_j \in G, \quad j = 1, \dots, m, \quad (3.21)$$

where $p_j := a_j - a_{j-1}$ and $\tilde{\phi}_j$ denotes the slope of h_{cx} on $(a_{j-1}, a_j]$. The worst-case distribution F_{wc} , if attainable, takes the form of

$$F_{\text{wc}} = \sum_{j=1}^m p_j \mathbf{1}_{\vec{x}_j^*}, \quad (3.22)$$

where \vec{x}_j^* , $j = 1, \dots, m$, denote the optimal solution to (3.18).

Proof. From Proposition 3.2 and its proof, we know that in the case where h_{cx} is piecewise linear (satisfying **(C3)**), it suffices to search for a worst-case distribution in the set $\mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$ that is supported by at most m points. In particular, given any distribution $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{\text{cx}}}$, we seek m points \vec{x}_j , $j = 1, \dots, m$ that satisfy

$$f(\vec{w}, \vec{x}_j) = \text{VaR}_{\alpha}(f(\vec{w}, \vec{X})), \quad \alpha \in (a_{j-1}, a_j], \quad j = 1, \dots, m, \quad \text{with } \vec{X} \sim F.$$

This condition can be equivalently stated as seeking \vec{x}_j , $j = 1, \dots, m$ that satisfy the ordering

$$f(\vec{w}, \vec{x}_1) \leq f(\vec{w}, \vec{x}_2) \leq \dots \leq f(\vec{w}, \vec{x}_m) \quad (3.23)$$

and the probability p_j of each point \vec{x}_j is equal to $p_j = a_j - a_{j-1}$. In this case, the objective function $\rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X}))$ reduces to (3.18) and the constraints in (3.17) reduce to (3.19)-(3.21).

The ordering constraint (3.23), however, is usually difficult to deal with, since it is not convex in general, i.e. the function $f(\vec{w}, \vec{x}_j) - f(\vec{w}, \vec{x}_{j+1})$ is nonconvex in \vec{x} (and \vec{w}). Fortunately, we now show that there is no loss of optimality by relaxing the ordering constraint. We proceed by deriving the dual of the problem (3.18). Letting $\vec{\lambda}_m \in \mathbb{R}^d$, $\vec{\lambda}_g \in \mathcal{K}^* \subset \mathbb{R}^p$ denote the Lagrange multipliers for the constraint (3.19) and (3.20), we can write down the following dual problem

$$\inf_{\vec{\lambda}_m, \vec{\lambda}_g \in \mathcal{K}^*} \sup_{\vec{x}_j \in G} \sum_{j=1}^m \tilde{\phi}_j p_j f(\vec{w}, \vec{x}_j) + \vec{\lambda}_m^\top (\vec{m} - \sum_{j=1}^m p_j \vec{x}_j) + \vec{\lambda}_g^\top (\bar{b} - \sum_{j=1}^m p_j \bar{g}(\vec{x}_j)). \quad (3.24)$$

Given the condition **(C2)**, we can confirm that the problem (3.18) satisfies the Slater condition and hence strong duality holds. We know from strong duality that any solution that is optimal to the

problem (3.18) must also be optimal to the inner maximization in (3.24). Observe now the following

$$(3.24) = \inf_{\vec{\lambda}_m, \vec{\lambda}_g \in \mathcal{K}^*} \sup_{\vec{x}_j \in G} \vec{\lambda}_m^\top \vec{m} + \vec{\lambda}_g^\top \vec{b} + \sum_{j=1}^m p_j (\tilde{\phi}_j f(\vec{w}, \vec{x}_j) - \vec{\lambda}_m^\top \vec{x}_j - \vec{\lambda}_g^\top g(\vec{x}_j)) \quad (3.25)$$

$$= \inf_{\vec{\lambda}_m, \vec{\lambda}_g \in \mathcal{K}^*} \vec{\lambda}_m^\top \vec{m} + \vec{\lambda}_g^\top \vec{b} + \sum_{j=1}^m p_j (\sup_{\vec{x}_j \in G} \tilde{\phi}_j f(\vec{w}, \vec{x}_j) - \vec{\lambda}_m^\top \vec{x}_j - \vec{\lambda}_g^\top g(\vec{x}_j)). \quad (3.26)$$

Now two cases arise.

- (i) The optimal solution to the problem (3.18) subject to the constraints (3.19)-(3.21) (that is, without the ordering constraint) is attainable. We aim to show any solution that is optimal to (3.18) will necessarily satisfy the ordering constraint (3.23). Note that any optimal solution $\vec{x}_j, j = 1, \dots, m$ to the inner maximization in (3.24) must satisfy that each \vec{x}_j can separately maximize the function $g_{\tilde{\phi}}(\vec{x}) := \tilde{\phi} f(\vec{w}, \vec{x}) - \vec{\lambda}_m^\top \vec{x} - \vec{\lambda}_g^\top g(\vec{x})$ with $\tilde{\phi} = \tilde{\phi}_j$. Since ϕ_j is increasing, we can easily conclude that any optimal solution $\vec{x}_j, j = 1, \dots, m$ to (3.24) must satisfy the ordering constraint (3.23), which in turn implies that the optimal solution to the problem (3.18) also has to satisfy the ordering constraint.
- (ii) The optimal solution to the problem (3.18) subject to the constraints (3.19)-(3.21) is not attainable. We aim to show that the problem (3.18) subject to the constraint (3.19)-(3.21) and the ordering constraint (3.23) is also not attainable; and the optimal values of the two optimization problems are equal. It suffices to show that for any feasible solution without the ordering constraint, there exists a feasible solution with the ordering constraint such that its value of objective function is no less than that of the original solution. Note that for any feasible solution $\{\vec{x}_j\}_{j=1}^m$ to the problem (3.18) subject to the constraints (3.19)-(3.21), let $G^* \subseteq G$ be a compact convex set such that $\{\vec{x}_j\}_{j=1}^m \subseteq G^*$. Now we consider a new optimization problem (3.18) subject to (3.19)-(3.20) and the following constraint

$$\vec{x}_j \in G^*, j = 1, \dots, m. \quad (3.27)$$

Then the new problem is attainable. Applying Case (i) to this new problem, we have its optimal solution has to satisfy the ordering constraint. Denote by $\{\vec{x}_j^*\}_{j=1}^m$ this optimal solution, and then $\sum_{j=1}^m \tilde{\phi}_j p_j f(\vec{w}, \vec{x}_j^*) \geq \sum_{j=1}^m \tilde{\phi}_j p_j f(\vec{w}, \vec{x}_j)$. Hence, we can find a feasible solution $\{\vec{x}_j^*\}_{j=1}^m \subseteq G^* \subseteq G$ satisfying the ordering constraint such that its objective function is no less than that of $\{\vec{x}_j\}_{j=1}^m$.

Combining the above two cases, we have that the ordering constraint (3.23) can be safely removed, and thus, we complete the proof. \square

From the above optimization problem, one can also easily confirm that the worst-case risk must be attainable if the set of convex support G is bounded, i.e. compact. As the final step, we

discuss now the tractability of solving the DRRO problem, i.e. the minmax problem (1.3). One can recast the whole problem into a single minimization problem by employing the dual problem of (3.18), i.e. (3.26), which leads to the following reformulation of the DRRO problem

$$\begin{aligned} \inf_{\vec{w} \in \mathcal{W}, \vec{\lambda}_m, \vec{\lambda}_g \in \mathcal{K}^*, \vec{s}} \quad & \vec{\lambda}_m^\top \vec{m} + \vec{\lambda}_g^\top \vec{b} + \sum_{j=1}^m p_j s_j \\ \text{subject to } \quad & s_j \geq \sup_{\vec{x} \in G} \{ \tilde{\phi}_j f(\vec{w}, \vec{x}) - \vec{\lambda}_m^\top \vec{x} - \vec{\lambda}_g^\top \vec{g}(\vec{x}) \}, \quad j = 1, \dots, m, \end{aligned} \quad (3.28)$$

where additional variables $s_j, j = 1, \dots, m$ are introduced to simplify the objective function. This optimization problem may not appear fully tractable (i.e. solvable by off-the-shelf solvers) at its current form, as it involves infinitely many constraints due to the maximization term in the right-hand-side of constraints. For readers who are familiar with the literature of robust optimization however, this type of constraint is common in robust optimization and can be handled by well-established techniques. Namely, based on Assumption 3.1 and convex (conic) duality theory, the maximization problem can often be recast as a minimization problem

$$\inf_{\vec{w}, \vec{\lambda}_m, \vec{\lambda}_g, \vec{y}} \left\{ h(\vec{w}, \vec{\lambda}_m, \vec{\lambda}_g, \vec{y}) \mid \vec{y} \in \Psi(\vec{w}, \vec{\lambda}_m, \vec{\lambda}_g) \right\},$$

where \vec{y} denotes the dual variable and the set $\Psi(\vec{w}, \vec{\lambda}_m, \vec{\lambda}_g)$ is specified through a finite number of constraints on the variables $\vec{w}, \vec{\lambda}_m, \vec{\lambda}_g, \vec{y}$. Replacing the maximization problem in the constraints of (3.28) by the above minimization problem, we end up at a formulation with only a finite number of constraints. What we outline here is a general principle established in the literature of robust optimization (see e.g. Ben-Tal et al. (2015)) to derive a tractable reformulation for problems involving infinitely many constraints, i.e. a robust constraint. For sake of space, we refer readers to Ben-Tal et al. (2015) and Wiesemann et al. (2014) for more detailed discussions about how different functional forms of $f(\vec{w}, \vec{x})$ and the (conic-representable) set \mathcal{W} would lead to different exact reformulations of convex or conic programs.

3.2 The Case of Fixed Mean and Covariance

As an important exception, the classical mean-covariance ambiguity set proposed in the earlier works of DRO and now widely referred to (see e.g. Popescu (2007), Bertsimas et al. (2010), Natarajan et al. (2010)) cannot be expressed using the ambiguity set studied in the previous section. Namely it is defined by

$$\mathcal{F}_{\vec{m}, \Sigma} = \left\{ F \in \mathcal{F}_0(\mathbb{R}^d) \mid \begin{array}{l} \mathbb{E}_F[\vec{X}] = \vec{m} \\ \mathbb{E}_F[(\vec{X} - \vec{m})(\vec{X} - \vec{m})^\top] = \Sigma \end{array} \right\}, \quad (3.29)$$

where $\vec{m} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d \times d}$ stand for the mean and covariance. Unlike the version of mean-covariance ambiguity set mentioned in the previous section, i.e. (3.2), which takes into account only

the upper bound of the covariance, the above set imposes a fixed level of covariance across all the distributions. In this section, we investigate the tractability of DRRO based on the set $\mathcal{F}_{\vec{m}, \Sigma}$.

We can follow similar steps presented in the previous section to tackle first the inner maximization problem of DRRO, i.e. the worst-case distortion risk measure $\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_h^F(f(\vec{w}, \vec{X}))$. While it is still possible to bound first the worst-case risk measure by its convex counterpart, i.e. $\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_{h_{\text{cx}}}^F(f(\vec{w}, \vec{X}))$, additional investigation is required to prove the tightness of the convex counterpart. In particular, the hard requirement that the distributions must now have certain degree of dispersion prohibits us from applying directly the analysis about worst-case distributions developed in the previous section. The key finding that enables us to speak about the tightness of the convex counterpart is that the hard constraint on covariance can actually be relaxed without the loss of optimality for a large class of distortion function h . To present this finding, we define first the following set of distributions that relaxes the set $\mathcal{F}_{\vec{m}, \Sigma}$

$$\mathcal{F}_{\vec{m}, \Sigma}^{\leq} = \left\{ F \in \mathcal{F}_0(\mathbb{R}^d) \mid \begin{array}{l} \mathbb{E}_F[\vec{X}] = \vec{m} \\ \mathbb{E}_F[(\vec{x}^\top (\vec{X} - \vec{m}))^2 - \vec{x}^\top \Sigma \vec{x}] \leq 0, \quad \forall \vec{x} \in \mathbb{R}^d \end{array} \right\}. \quad (3.30)$$

We show that relaxing the set $\mathcal{F}_{\vec{m}, \Sigma}$ to the set $\mathcal{F}_{\vec{m}, \Sigma}^{\leq}$ actually does not affect the worst-case risk if the distortion function h is insensitive to "very good" outcomes.

Proposition 3.6. *Given any right-continuous distortion function h that satisfies $h(x) = 0$, $x \in (0, \varepsilon)$ for some arbitrary small $\varepsilon > 0$, and a cost function $f(\vec{w}, \vec{x})$ that is concave in \vec{x} , we have*

$$\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_h^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{\leq}} \rho_h^F(f(\vec{w}, \vec{X})). \quad (3.31)$$

Proof. Denote by ρ_1^\uparrow and ρ_2^\uparrow the left side and right side of (3.31), respectively. Note that $\mathcal{F}_{\vec{m}, \Sigma} \subseteq \mathcal{F}_{\vec{m}, \Sigma}^{\leq}$, and thus, $\rho_2^\uparrow \geq \rho_1^\uparrow$. We only need to show $\rho_2^\uparrow \leq \rho_1^\uparrow$.

We first show the result for the case that $\rho_2^\uparrow < \infty$. We assert that for any $\delta > 0$, there exists a distribution $G_\delta \in \mathcal{F}_{\vec{m}, \Sigma}^{\leq}$ and $\vec{X}_{G_\delta} \sim G_\delta$ such that $f(\vec{w}, \vec{X}_{G_\delta})$ has a lower bound and

$$\rho_h(f(\vec{w}, \vec{X}_{G_\delta})) > \rho_2^\uparrow - \delta. \quad (3.32)$$

To see it, by the definition of ρ_2^\uparrow , we know that for any $\delta > 0$, there exists a distribution $G_\delta \in \mathcal{F}_{\vec{m}, \Sigma}^{\leq}$ and $\vec{X}_{G_\delta} \sim G_\delta$ such that (3.32) holds. If $f(\vec{w}, \vec{X}_{G_\delta})$ does not have a lower bound, then we can replace \vec{X}_{G_δ} by $\vec{X}_{G_\delta}^*$ which is defined as follows. Let $U \sim \text{U}[0, 1]$ be a random variable such that $f(\vec{w}, \vec{X}_{G_\delta}) = F_{\vec{w}, f}^{-1}(U)$ a.s., where $F_{\vec{w}, f}$ is the distribution function of $f(\vec{w}, \vec{X}_{G_\delta})$, and we define

$$\vec{X}_G^* = \mathbb{E}[\vec{X}_{G_\delta} | U \leq \varepsilon] I_{\{U \leq \varepsilon\}} + \vec{X}_{G_\delta} I_{\{U > \varepsilon\}} =: \vec{x}_G^* I_{\{U \leq \varepsilon\}} + \vec{X}_{G_\delta} I_{\{U > \varepsilon\}}.$$

Then we have

$$\begin{aligned} f(\vec{w}, \vec{X}_{G_\delta}^*) &= f(\vec{w}, \vec{x}_G^*) I_{\{U \leq \varepsilon\}} + f(\vec{w}, \vec{X}_{G_\delta}) I_{\{U > \varepsilon\}} \\ &\geq \min\{f(\vec{w}, \vec{x}_G^*), \text{VaR}_\varepsilon(f(\vec{w}, \vec{X}_{G_\delta}))\} I_{\{U \leq \varepsilon\}} + f(\vec{w}, \vec{X}_{G_\delta}) I_{\{U > \varepsilon\}} \\ &=: W_G. \end{aligned}$$

Hence, it holds that

$$\rho_h(f(\vec{w}, \vec{X}_{G_\delta}^*)) \geq \rho_h(W_G) = \rho_h(f(\vec{w}, \vec{X}_{G_\delta})) > \rho_2^\uparrow - \delta, \quad (3.33)$$

where the equality follows from Lemma 6.7 and the fact that U , $f(\vec{w}, \vec{X}_{G_\delta})$ and W_G are comonotonic, and $[f(\vec{w}, \vec{X}_{G_\delta})|U > \varepsilon] = [W_G|U > \varepsilon]$ a.s. In addition, the second (strict) inequality in (3.33) is due to (3.32). In addition, by Lemma 6.3, we have the distribution of $\vec{X}_{G_\delta}^*$ is in $\mathcal{F}_{\vec{m}, \Sigma}^\leq$. Therefore, we can find $G_\delta \in \mathcal{F}_{\vec{m}, \Sigma}^\leq$ and $\vec{X}_{G_\delta} \sim G_\delta$ such that $f(\vec{w}, \vec{X}_{G_\delta})$ has a lower bound and (3.32) holds.

Next, we aim to construct a distribution in $\mathcal{F}_{\vec{m}, \Sigma}$ satisfying the corresponding distortion risk measure is larger than $\rho_2^\uparrow - 2\delta$. To this end, denote by Σ_δ the covariance matrix of \vec{X}_{G_δ} which satisfies $\Sigma_\delta \preceq \Sigma$ as $G_\delta \in \mathcal{F}_{\vec{m}, \Sigma}^\leq$. Let $\{\beta_k, k \in \mathbb{N}\}$ be a strictly increasing sequence of positive constants such that $\lim_{k \rightarrow \infty} \beta_k = 1$, we define

$$\Sigma_k = \frac{\Sigma - \beta_k \Sigma_\delta}{1 - \beta_k}, \quad k \in \mathbb{N}.$$

By $\Sigma_\delta \preceq \Sigma$, we have for any $\vec{x} \in \mathbb{R}^d$

$$\vec{x}^\top \Sigma_k \vec{x} = \frac{\vec{x}^\top \Sigma \vec{x} - \beta_k \vec{x}^\top \Sigma_\delta \vec{x}}{1 - \beta_k} \geq \frac{\vec{x}^\top \Sigma \vec{x} - \vec{x}^\top \Sigma_\delta \vec{x}}{1 - \beta_k} \geq 0,$$

that is, $\{\Sigma_k, k \in \mathbb{N}\}$ is a sequence of positive semi-definite matrices.

For each $k \in \mathbb{N}$, let H_k be a distribution of random vector in \mathbb{R}^d such that it has the expectation \vec{m} and covariance matrix Σ_k , that is,

$$\int \vec{X} dH_k = \vec{m} \quad \text{and} \quad \int (\vec{X} - \vec{m})(\vec{X} - \vec{m})^\top dH_k = \Sigma_k.$$

For each $k \in \mathbb{N}$, let \vec{Y}_k be a random vector having the distribution H_k . Then,

$$\mathbb{E}[\vec{Y}_k] = \vec{m} \quad \text{and} \quad \text{Cov}(\vec{Y}_k) = \mathbb{E}[(\vec{Y}_k - \vec{m})(\vec{Y}_k - \vec{m})^\top] = \Sigma_k.$$

Define \vec{Z}_k as a random vector with distribution W_k given by

$$W_k = \beta_k G_\delta + (1 - \beta_k) H_k, \quad k \in \mathbb{N}.$$

Then we have $\mathbb{E}[\vec{Z}_k] = \vec{m}$ and

$$\text{Cov}(\vec{Z}_k) = \beta_k \text{Cov}(\vec{X}_{G_\delta}) + (1 - \beta_k) \text{Cov}(\vec{Y}_k) = \Sigma, \quad k \in \mathbb{N},$$

that is, $W_k \in \mathcal{F}_{\vec{m}, \Sigma}$ for $k \in \mathbb{N}$. In addition, we can show for k large enough,

$$\rho_h(f(\vec{w}, \vec{Z}_k)) > \rho_h(f(\vec{w}, \vec{X}_{G_\delta})) - \delta. \quad (3.34)$$

To see it, let Z_k^* be a random variable such that its distribution G_k^* on \mathbb{R} is a mixture of $f(\vec{w}, \vec{X}_{G_\delta})$ and $f_0 < \text{ess-inf} f(\vec{w}, \vec{X}_{G_\delta})$ with respective weights β_k and $1 - \beta_k$, that is,

$$G_k^*(x) = \beta_k \mathbb{P}(f(\vec{w}, \vec{X}_{G_\delta}) \leq x) + (1 - \beta_k) I_{\{f_0 \leq x\}}, \quad x \in \mathbb{R}.$$

We have the following two facts.

(i) $\lim_{k \rightarrow \infty} \rho_h(Z_k^*) = \rho_h(f(\vec{w}, \vec{X}_{G_\delta}))$. To see it, note that

$$G_k^*(x) = \begin{cases} \beta_k \mathbb{P}(f(\vec{w}, \vec{X}_{G_\delta}) \leq x) = 0, & x < f_0 \\ \beta_k \mathbb{P}(f(\vec{w}, \vec{X}_{G_\delta}) \leq x) + (1 - \beta_k), & x \geq f_0. \end{cases}$$

This implies that for $\alpha \geq 1 - \beta_k$,

$$\text{VaR}_\alpha(Z_k^*) = \text{VaR}_{\frac{\alpha-1+\beta_k}{\beta_k}}(f(\vec{w}, \vec{X}_{G_\delta})). \quad (3.35)$$

Note that there exists $k_0 \in \mathbb{N}$ such that $1 - \beta_k < \varepsilon/2 < \varepsilon$ for $k \geq k_0$. Thus, (3.35) holds for $\alpha \geq \varepsilon$ and $k \geq k_0$. Then since $\frac{\alpha-1+\beta_k}{\beta_k} = 1 - (1 - \alpha)/\beta_k$ increasingly converges to α as $k \rightarrow \infty$, VaR_α is left-continuous and increasing in α , we have the sequence $\{\text{VaR}_\alpha(Z_k^*), k \geq k_0\}$ increasingly converges to $\text{VaR}_\alpha(f(\vec{w}, \vec{X}_{G_\delta}))$ as $k \rightarrow \infty$ for $\alpha \geq \varepsilon$. Also, note that $\beta_k > 1 - \varepsilon/2$ for $k \geq k_0$, and thus, for $\alpha \geq \varepsilon$ and $k \geq k_0$,

$$\frac{\alpha - 1 + \beta_k}{\beta_k} = 1 - \frac{1 - \alpha}{\beta_k} > 1 - \frac{1 - \alpha}{1 - \varepsilon/2} = \frac{\alpha - \varepsilon/2}{1 - \varepsilon/2} > \alpha - \frac{\varepsilon}{2} > \frac{\varepsilon}{2}.$$

By (3.35), this implies that $\text{VaR}_\alpha(Z_k^*)$ for $\alpha \geq \varepsilon$ has a lower bound $\text{VaR}_{\varepsilon/2}(f(\vec{w}, \vec{X}_{G_\delta}))$. Then by monotone convergence theorem, we have

$$\lim_{k \rightarrow \infty} \rho_h(Z_k^*) = \lim_{k \rightarrow \infty} \int_{[\varepsilon, 1]} \text{VaR}_\alpha(Z_k^*) dh(\alpha) = \int_{[\varepsilon, 1]} \text{VaR}_\alpha(f(\vec{w}, \vec{X}_{G_\delta})) dh(\alpha) = \rho_h(f(\vec{w}, \vec{X}_{G_\delta})),$$

where the first equality is due to (6.56) of Lemma 6.7.

(ii) There exists $k_0 \in \mathbb{N}$ such that for any $k \geq k_0$,

$$\text{VaR}_\alpha(Z_k^*) \leq \text{VaR}_\alpha(f(\vec{w}, \vec{Z}_k)), \quad \alpha \geq \varepsilon. \quad (3.36)$$

To see it, note that for $x \geq f_0$, we have

$$\begin{aligned} \mathbb{P}(f(\vec{w}, \vec{Z}_k) \leq x) &= \beta_k \mathbb{P}(f(\vec{w}, \vec{X}_{G_\delta}) \leq x) + (1 - \beta_k) \mathbb{P}(f(\vec{w}, \vec{Y}_k) \leq x) \\ &\leq \beta_k \mathbb{P}(f(\vec{w}, \vec{X}_{G_\delta}) \leq x) + (1 - \beta_k) = G_k^*(x). \end{aligned}$$

Also, note that $f_0 = \text{VaR}_{1-\beta_k}(Z_k^*)$. Hence, for each $k \in \mathbb{N}$, we have

$$\text{VaR}_\alpha(f(\vec{w}, \vec{Z}_k)) \geq \text{VaR}_\alpha(Z_k^*), \quad \alpha \geq 1 - \beta_k.$$

As $\lim_{k \rightarrow \infty} \beta_k = 1$, there exists k_0 such that $1 - \beta_k < \varepsilon$ for $k \geq k_0$. Hence, we have (3.36) holds.

By fact (i), we have for the given $\delta > 0$, there exists k_1 such that $\rho_h(Z_k^*) > \rho_h(f(\vec{w}, \vec{X}_{G_\delta})) - \delta$, and by fact (ii), we have

$$\rho_h(f(\vec{w}, \vec{Z}_k)) \geq \rho_h(Z_k^*) > \rho_h(f(\vec{w}, \vec{X}_{G_\delta})) - \delta, \quad k \geq \max\{k_0, k_1\}. \quad (3.37)$$

That is, (3.34) holds. This combined with (3.32) implies that

$$\rho_h(f(\vec{w}, \vec{Z}_k)) > \rho_2^\uparrow - 2\delta, \quad k \geq \max\{k_0, k_1\}. \quad (3.38)$$

Then by that the distribution W_k of \vec{Z}_k belongs to $\mathcal{F}_{\vec{m}, \Sigma}$ for $k \in \mathbb{N}$, we have

$$\rho_1^\uparrow \geq \sup_{k \geq \max\{k_0, k_1\}} \rho_h(f(\vec{w}, \vec{Z}_k)) > \rho_2^\uparrow - 2\delta. \quad (3.39)$$

As δ can be chosen arbitrarily small, we have $\rho_1^\uparrow \geq \rho_2^\uparrow$ and thus, $\rho_1^\uparrow = \rho_2^\uparrow$. That is, we complete the proof for the case that $\rho_2^\uparrow < \infty$.

If $\rho_2^\uparrow = \infty$, we can exactly employ the above arguments to show $\rho_1^\uparrow = \rho_2^\uparrow$, except for replacing $\rho_2^\uparrow - \delta$ of (3.32) and (3.33) by $1/\delta$, replacing $\rho_h(f(\vec{w}, \vec{X}_{G_\delta})) - \delta$ of (3.34) and (3.37) by $1/(2\delta)$, and replacing $\rho_2^\uparrow - 2\delta$ of (3.38) and (3.39) by $1/(2\delta)$. Then we obtain

$$\rho_1^\uparrow \geq \frac{1}{2\delta}$$

holds for any $\delta > 0$. Then letting $\delta \downarrow 0$, we have $\rho_1^\uparrow = \infty$, and thus, $\rho_1^\uparrow = \rho_2^\uparrow$. Thus, we complete the proof. \square

We should point out here that in the case where the distortion function h is convex, the tightness of the relaxation can be established for any distortion function if the cost function is Lipschitz continuous. That is, the insensitivity condition is basically only required in the case of nonconvex distortion function h . Note also that since the set $\mathcal{F}_{\vec{m}, \Sigma}^\leq$ is a special case of $\mathcal{F}_{\vec{m}, \mathcal{G}}$, we know from Theorem 3.2 that

$$\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^\leq} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X})), \quad (3.40)$$

where

$$\mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq} = \left\{ F \in \mathcal{F}_{\vec{m}, \Sigma}^\leq \left| \begin{array}{l} \text{VaR}_\alpha(f(\vec{w}, \vec{X}_F)) \text{ is a constant on each interval on} \\ \text{which } h_{cx} \text{ is linear, where } \vec{X}_F \text{ has distribution } F. \end{array} \right. \right\}.$$

Combining Proposition 3.6, the observation (3.40), and the results established in the previous section, we arrive at the following result about hidden convexity, which is analogous to the main result in the case of bounded moments.

Theorem 3.7. *Given any distortion function h that satisfies $h(x) = 0$, $x \in (0, \varepsilon)$ for some arbitrary small $\varepsilon > 0$, and a cost function $f(\vec{w}, \vec{x})$ that is concave in \vec{x} , we have*

$$\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_h^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X})). \quad (3.41)$$

Proof. Denote by ρ_h^\uparrow and $\rho_{h_{cx}}^\uparrow$ the left side and the right side of (3.41), respectively. Let h_+ be the right-continuous copy of h , that is, $h_+(x) = \lim_{y \downarrow x} h(y)$, $x \in (0, 1)$, and denote

$$\rho_{h_+}^\uparrow := \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_{h_+}^F(f(\vec{w}, \vec{X})).$$

Since $\rho_{h_+} \leq \rho_h \leq \rho_{h_{cx}}$, we have $\rho_{h_+}^\uparrow \leq \rho_h^\uparrow \leq \rho_{h_{cx}}^\uparrow$. To show (3.41), it suffices to show $\rho_{h_+}^\uparrow \geq \rho_{h_{cx}}^\uparrow$. By (3.31) of Theorem 3.6, we have

$$\rho_{h_+}^\uparrow = \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{\leq}} \rho_{h_+}^F(f(\vec{w}, \vec{X})) \geq \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}} \rho_{h_+}^F(f(\vec{w}, \vec{X})),$$

and by (3.40), we have

$$\rho_{h_{cx}}^\uparrow = \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X})).$$

Hence, to show $\rho_{h_+}^\uparrow \geq \rho_{h_{cx}}^\uparrow$, it suffices to show

$$\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}} \rho_h^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X})),$$

that is, for any $F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}$ and $\vec{X} \sim F$,

$$\rho_{h_+}(f(\vec{w}, \vec{X})) = \rho_{h_{cx}}(f(\vec{w}, \vec{X})). \quad (3.42)$$

Let h_- be the left-continuous version of h . It is obvious that $h_{cx} \leq h_- \leq h_+$ and thus, $\rho_{h_+} \leq \rho_{h_-} \leq \rho_{h_{cx}}$. Hence, to show (3.42), we first show for any $F \in \mathcal{F}_{\vec{m}, \Sigma}^{h_{cx}, \leq}$ and $\vec{X} \sim F$

$$\rho_{h_-}(f(\vec{w}, \vec{X})) = \rho_{h_{cx}}(f(\vec{w}, \vec{X})) \quad (3.43)$$

and then show

$$\rho_{h_+}^\uparrow \geq \rho_{h_-}(f(\vec{w}, \vec{X})). \quad (3.44)$$

The proof of (3.43) is similar to that of (3.11) in the proof of Theorem 3.4 and the proof of (3.44) is similar to that of (6.45). Thus, we omit the remaining proof. \square

To see how the problem $\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X}))$ (in (3.41)) formulated based on the convex envelope h_{cx} can be solved numerically, one can invoke the observation (3.31) that the problem can be equivalently solved by the problem $\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}^{\leq}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X}))$. The latter can be solved by the optimization problem presented in Theorem 3.5. We should remark here that to the best of our knowledge, prior to our work the problem $\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_{h_{cx}}^F(f(\vec{w}, \vec{X}))$ is only known to be solvable in few special cases of the cost function $f(\vec{w}, \vec{x})$. The observation (3.31) implies also that one can follow the discussion at the end of Section 3.1 to solve the DRRO problem (1.3) formulated based on $\mathcal{F}_{\vec{m}, \Sigma}$. With the insensitivity condition provided in Theorem 3.7, one can easily confirm that DRRO problems formulated based on a S-shaped distortion function can be solved exactly by its convex counterpart, but this might not be the case for DRRO problems formulated based on an inverse S-shaped distortion function. This provides some general sense of when the hidden convexity might exist for DRRO problems with an ambiguity set defined based on fixed mean and covariance.

3.3 The Case of Linear Cost Function

In this section, we continue the discussion of the case of fixed mean and covariance but pay a particular attention to the case of linear cost function. We show that in this case we can obtain tightness results that are applicable to any distortion function h (i.e. the insensitivity condition in Theorem 3.7 can be dropped). We start by simplifying the problem defined over multiple random variables to a problem over univariate random variable X . For convenience, from here on we denote by $\mathcal{F}(\mu, \sigma)$ the distribution set $\mathcal{F}_{\vec{m}, \Sigma}$ in the case of $d = 1$ with mean μ and standard deviation σ . The result below is a direct application of the projection theorem first developed in Popescu (2007).

Theorem 3.8. *In the case where $f(\vec{w}, \vec{x}) = \vec{w}^\top \vec{x}$, we have*

$$\sup_{F \in \mathcal{F}_{\vec{m}, \Sigma}} \rho_h^F(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}(\mu, \sigma)} \rho_h^F(X)$$

where $\mu = \vec{w}^\top \vec{m}$ and $\sigma = \sqrt{\vec{w}^\top \Sigma \vec{w}}$.

It was found recently by Li (2018) that the worst-case distortion risk measure $\rho_h^\uparrow(X) := \sup_{F \in \mathcal{F}(\mu, \sigma)} \rho_h^F(X)$ can actually be solved in closed-form in the case where the distortion function h is convex (see also Liu et al. (2020)). This result, which is summarized in the following lemma, will be particularly useful in proving our main result.

Lemma 3.9. *Let h be a convex distortion function. The worst-case distortion risk measure in this case can be evaluated by*

$$\rho_h^\uparrow(X) = \mu + \sigma \sqrt{\|h'\|_2^2 - 1},$$

where $\|h'\|_2 = \sqrt{\int_0^1 (h'(t))^2 dt}$ is the L^2 -norm of h' and h' is the left (or right) derivative function of h . Moreover, $\rho_h^\uparrow(X)$ is attainable at the worst-case distribution F_{wc} satisfying

$$F_{wc}^{-1}(\alpha) = h'(\alpha)c + d, \quad \alpha \in (0, 1), \quad (3.45)$$

where c and d are two constants such that the mean and variance of the distribution F_{wc} are equal to μ and σ^2 , respectively.

Recall from the previous sections that we rely heavily on the characterization of worst-case distributions to reveal the hidden convexity of DRRO problems. The fact that the above result actually provides us the exact functional form of the worst-case distributions (in the case of convex distortion function) facilitate us greatly to discover hidden convexity. In particular, this makes it possible to derive the following tightness result that holds for any distortion function h .

Theorem 3.10. *Given any distortion function h , we have*

$$\rho_h^\uparrow(X) = \mu + \sigma \sqrt{\|h'_{cx}\|_2^2 - 1}. \quad (3.46)$$

Moreover, if h is left-continuous, then $\rho_h^\uparrow(X)$ is attainable at a worst-case distribution F_{wc} satisfying

$$F_{\text{wc}}^{-1}(\alpha) = h'_{\text{cx}}(\alpha)c + d, \quad \alpha \in (0, 1), \quad (3.47)$$

where c and d are two constants such that the mean and variance of the distribution F_{wc} are equal to μ and σ^2 , respectively.

Proof. For a distortion function h , denote by h_- and h_+ the left continuous and right continuous copies of h , respectively. It is obvious that the convex envelopes of h_- and h_+ are the same as the convex envelop of h , namely $(h_-)_{\text{cx}} = (h_+)_{\text{cx}} = h_{\text{cx}}$. Then by Lemma 2.1(i), it holds that $\rho_{h_+}^\uparrow(X) \leq \rho_h^\uparrow(X) \leq \rho_{h_-}^\uparrow(X)$. Hence, to show (3.46), it suffices to show that $\rho_{h_-}^\uparrow(X) = \rho_{h_+}^\uparrow(X) = \mu + \sigma \sqrt{\|h'_{\text{cx}}\|_2^2 - 1}$.

We first show that (3.46) holds for h_- . Note that $h_{\text{cx}} = (h_-)_{\text{cx}} \leq h_-$ and the distortion risk measure ρ_h is decreasing in h , thus we have $\rho_{h_-} \leq \rho_{h_{\text{cx}}}$ and hence

$$\rho_{h_-}^\uparrow(X) = \sup_{F \in \mathcal{F}(\mu, \sigma)} \rho_{h_-}(X_F) \leq \sup_{F \in \mathcal{F}(\mu, \sigma)} \rho_{h_{\text{cx}}}(X_F) = \rho_{h_{\text{cx}}}^\uparrow(X).$$

Therefore, by Lemma 3.9, we know that $\rho_{h_{\text{cx}}}^\uparrow(X) = \rho_{h_{\text{cx}}}(X_{F_{\text{wc}}}) = \mu + \sigma \sqrt{\|h'_{\text{cx}}\|_2^2 - 1}$, where $X_{F_{\text{wc}}}$ is a random variable such that its distribution is the worst-case distribution F_{wc} given by (3.47). Hence, it suffices to show that the distribution given by (3.47) is also the worst-case distribution of ρ_{h_-} , that is, to show that $\rho_{h_-}(X_{F_{\text{wc}}}) = \rho_{h_{\text{cx}}}(X_{F_{\text{wc}}})$.

By Theorem 3.1(ii), we have that the set $I =: \{t \in [0, 1] : h_-(t) \neq h_{\text{cx}}(t)\}$ has the expression (3.7), h_{cx} is linear in each interval of $\cup_{k \in I} (a_k, b_k)$. Thus, the worst case distribution given by $F_{\text{wc}}^{-1}(\alpha)$ in (3.47) is a constant on each interval of $\cup_{k \in I} (a_k, b_k)$. We denote $x_k = \text{VaR}_\alpha^+(X_{F_{\text{wc}}})$ when $\alpha \in (a_k, b_k)$ for $k \in I$. Note that for an increasing function g and $a < b$, we have $\int_{(a,b)} dg(\alpha) = \int_{(a,b)} dg^+(\alpha) = g^+(b-) - g^+(a) = g(b-) - g(a+)$. Hence, by (2.8), we have

$$\begin{aligned} \rho_{h_-}(X_{F_{\text{wc}}}) &= \int_0^1 \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh(\alpha) \\ &= \int_{(0,1) \setminus I} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh(\alpha) + \sum_{k \in I} \int_{(a_k, b_k)} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh(\alpha) \\ &= \int_{(0,1) \setminus I} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh_{\text{cx}}(\alpha) + \sum_{k \in I} x_k (h(b_k-) - h(a_k+)) \\ &= \int_{(0,1) \setminus I} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh_{\text{cx}}(\alpha) + \sum_{k \in I} x_k (h(b_k) - h(a_k)) \\ &= \int_{(0,1) \setminus I} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh_{\text{cx}}(\alpha) + \sum_{k \in I} x_k (h_{\text{cx}}(b_k) - h_{\text{cx}}(a_k)) \\ &= \int_{(0,1) \setminus I} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh_{\text{cx}}(\alpha) + \sum_{k \in I} \int_{[a_k, b_k)} \text{VaR}_\alpha^+(X_{F_{\text{wc}}}) dh_{\text{cx}}(\alpha) \\ &= \rho_{h_{\text{cx}}}(X_{F_{\text{wc}}}), \end{aligned}$$

where the forth equality follows that h is continuous at a_k and b_k (Theorem 3.1(ii)) and the fifth equality follows from (3.8). Thus, we have proved that $\rho_{h_-}^\uparrow(X) = \mu + \sigma\sqrt{\|h'_{\text{cx}}\|_2^2 - 1}$ and $\rho_{h_-}^\uparrow(X)$ is attainable at the worst case distribution F_{wc} .

In the following, we show that (3.46) holds for h_+ . It suffices to show $\rho_{h_+}^\uparrow = \rho_{h_-}^\uparrow$ as we have proved that $\rho_{h_-}^\uparrow = \rho_{h_{\text{cx}}}^\uparrow$. Let F_{wc} be the worst-case distribution given by (3.47) and let F be a distribution function such that $F(x) < F_{\text{wc}}(x)$ when $0 < F_{\text{wc}}(x) < 1$, $F(x) = 1$ when $F_{\text{wc}}(x) = 1$, and F has finite mean and variance. For any $\varepsilon \in (0, 1)$, define a mixed distribution function F_ε by

$$F_\varepsilon = (1 - \varepsilon)F_{\text{wc}} + \varepsilon F, \quad \varepsilon \in (0, 1),$$

and let X_ε be a random variable with distribution function F_ε , whose mean and variance are denoted by μ_ε and σ_ε^2 , respectively. Note that for any $x \in \mathbb{R}$, $F_\varepsilon(x) \uparrow F_{\text{wc}}(x)$ as $\varepsilon \downarrow 0$ and $F_\varepsilon(x) < F_{\text{wc}}(x)$ when $0 < F_{\text{wc}}(x) < 1$. Hence for any $x \in \mathbb{R}$, $h_+(F_\varepsilon(x)) \uparrow h_-(F_{\text{wc}}(x))$ as $\varepsilon \downarrow 0$. Then we have

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \rho_{h_+}(X_\varepsilon) &= \lim_{\varepsilon \downarrow 0} \left(\int_0^\infty (1 - h_+(F_\varepsilon(x))) dx - \int_{-\infty}^0 h_+(F_\varepsilon(x)) dx \right) \\ &= \int_0^\infty (1 - h_-(F_{\text{wc}}(x))) dx - \int_{-\infty}^0 h_-(F_{\text{wc}}(x)) dx \\ &= \rho_{h_-}(X_{F_{\text{wc}}}), \end{aligned} \quad (3.48)$$

where the second equality follows from the monotone convergence theorem and that $F_\varepsilon \uparrow F_{\text{wc}}$ and $h_+(F_\varepsilon) \uparrow h_-(F_{\text{wc}})$ as $\varepsilon \downarrow 0$. Let $X_\varepsilon^* = \mu + \frac{\sigma}{\sigma_\varepsilon}(X_\varepsilon - \mu_\varepsilon)$. Then we have $X_\varepsilon^* \in \mathcal{F}(\mu, \sigma)$ and

$$\lim_{\varepsilon \downarrow 0} \rho_{h_+}(X_\varepsilon^*) = \lim_{\varepsilon \downarrow 0} \left(\mu + \frac{\sigma}{\sigma_\varepsilon}(\rho_{h_+}(X_\varepsilon) - \mu_\varepsilon) \right) = \rho_{h_-}(X_{F_{\text{wc}}}),$$

where the first equality follows from that ρ_{h_+} is positively homogeneous and translation-invariant and the second equality follows from (3.48), $\lim_{\varepsilon \downarrow 0} \mu_\varepsilon = \mu$ and $\lim_{\varepsilon \downarrow 0} \sigma_\varepsilon^2 = \sigma^2$. Hence, we have $\rho_{h_+}^\uparrow(X) \geq \lim_{\varepsilon \downarrow 0} \rho_{h_+}(X_\varepsilon^*) = \rho_{h_-}(X_{F_{\text{wc}}}) = \rho_{h_-}^\uparrow(X)$, which implies that $\rho_{h_+}^\uparrow(X) = \rho_{h_-}^\uparrow(X)$ since it holds that $\rho_{h_+}^\uparrow(X) \leq \rho_{h_-}^\uparrow(X)$. Thus, (3.46) holds. In addition, if h is left continuous, then $\rho_h = \rho_{h_-}$ by Lemma 2.1(iii) and thus $\rho_h^\uparrow(X) = \rho_{h_-}^\uparrow(X)$. Hence, $\rho_h^\uparrow(X)$ is also attainable at the worst-case distribution F_{wc} satisfying (3.47) when h is left continuous. \square

The above result, together with Theorem 3.8, immediately imply that DRRO problems in the case of linear cost function can be reduced to

$$\min_{\vec{w} \in \mathcal{W}} \vec{w}^\top \vec{m} + \sqrt{\vec{w}^\top \Sigma \vec{w}} \sqrt{\|h'_{\text{cx}}\|_2^2 - 1}. \quad (3.49)$$

The objective function of this form is known to be easy to optimize, since its epigraph can be represented via a second-order cone (c.f. Ben-Tal and Nemirovski (2001)). In the case where the feasible set \mathcal{W} can also be represented via a second-order cone, the whole problem (3.49) can be efficiently solved as a second-order cone program (see e.g. Li (2018)). It came to our attention

that a special case of the above result was also obtained in a recent work of Zhu and Shao (2018). In particular, they take a fairly different approach to derive the closed-form result for distortion risk measures defined by (2.7), i.e. the case of right-continuous distortion functions. They however falsely claim the existence of worst-case distributions. Our approach, on the other hand, gains its power from exploiting specifically the characterization of worst-case distributions in terms of convex envelope, which makes it possible to handle worst-case distortion risk measures in its full generality.

Remark 3.11. (The worst-case Value-at-Risk of $\text{VaR}_\alpha(X)$ is not attainable) At the end of this section, we point out that the well-known worst-case Value-at-Risk derived in El Ghaoui et al. (2003) is based on the Value-at-Risk's definition of $\inf\{x \in \mathbb{R} : \mathbb{P}(X \geq x) \leq 1 - \alpha\}$, which is equivalently to defining Value-at-Risk at level α as $\text{VaR}_\alpha^+(X)$. It is shown in El Ghaoui et al. (2003) that

$$\sup_{F \in \mathcal{F}(\mu, \sigma)} \text{VaR}_\alpha^+(X_F) = \mu + \sigma \sqrt{\frac{\alpha}{1 - \alpha}} \quad (3.50)$$

and the worst-case Value-at-Risk is attainable at a two-point distribution

$$\mathbb{P}\left(X^* = \mu - \sigma \sqrt{\frac{1 - \alpha}{\alpha}}\right) = \alpha, \quad \mathbb{P}\left(X^* = \mu + \sigma \sqrt{\frac{\alpha}{1 - \alpha}}\right) = 1 - \alpha,$$

where X_F represents a random variable with distribution F . This result of El Ghaoui et al. (2003) is now a direct corollary of Theorem 3.10 by noticing that the distortion function of $\text{VaR}_\alpha^+(X)$ is a left-continuous function $h_{\text{VaR}^+}(x) = I_{\{\alpha < x \leq 1\}}$.

What appears less noticed however is that if one employs the standard definition of Value-at-Risk, i.e. $\text{VaR}_\alpha(X)$, the worst-case Value-at-Risk of $\text{VaR}_\alpha(X)$ is not attainable although the worst-case value is the same, namely

$$\sup_{F \in \mathcal{F}(\mu, \sigma)} \text{VaR}_\alpha(X_F) = \mu + \sigma \sqrt{\frac{\alpha}{1 - \alpha}}. \quad (3.51)$$

Notice that the distortion function of $\text{VaR}_\alpha(X)$ is a right-continuous function $h_{\text{VaR}}(x) = I_{\{\alpha \leq x \leq 1\}}$, which shares the same convex envelope $h_{\text{CVaR}}(x) = (x - \alpha)_+ / (1 - \alpha)$ as $h_{\text{VaR}^+}(x)$. To see why (3.51) is not attainable, one can first verify that the worst case distribution for $\sup_{F \in \mathcal{F}(\mu, \sigma)} \text{CVaR}_\alpha(X_F)$, which is equal to (3.51) and is attainable, has to be a two-point distribution belonging to the following class of two-point distributions (see, e.g. Li (2018)):

$$\mathcal{F}_{\text{bi}} = \left\{ \text{df } F_\beta \text{ of } X_\beta : \mathbb{P}\left(X_\beta = \mu - \sigma \sqrt{\frac{1 - \beta}{\beta}}\right) = \beta, \mathbb{P}\left(X_\beta = \mu + \sigma \sqrt{\frac{\beta}{1 - \beta}}\right) = 1 - \beta, \beta \in (0, 1) \right\}.$$

Then for any distribution $F \in \mathcal{F}(\mu, \sigma) \setminus \mathcal{F}_{\text{bi}}$, $\text{CVaR}_\alpha(X_F) < \mu + \sigma \sqrt{\frac{\alpha}{1 - \alpha}}$, and thus,

$$\text{VaR}_\alpha(X_F) \leq \text{CVaR}_\alpha(X_F) < \mu + \sigma \sqrt{\frac{\alpha}{1 - \alpha}} \quad \text{for any } F \in \mathcal{F}(\mu, \sigma) \setminus \mathcal{F}_{\text{bi}}. \quad (3.52)$$

On the other hand, for any random variable X_β with distribution $F_\beta \in \mathcal{F}_{\text{bi}}$, we have

$$\text{VaR}_\alpha(X_\beta) = \begin{cases} \mu - \sigma \sqrt{\frac{1-\beta}{\beta}}, & \alpha \leq \beta < 1, \\ \mu + \sigma \sqrt{\frac{\beta}{1-\beta}}, & 0 < \beta < \alpha. \end{cases} \quad (3.53)$$

It is easy to see

$$\text{VaR}_\alpha(X_\beta) < \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}} = \lim_{\beta \uparrow \alpha} \text{VaR}_\alpha(X_\beta) \quad \text{for each } \beta \in (0, 1). \quad (3.54)$$

Combining (3.52) and (3.54), we know that for any random variable X_F with a distribution $F \in \mathcal{F}(\mu, \sigma)$, we must have $\text{VaR}_\alpha(X_F) < \mu + \sigma \sqrt{\frac{\alpha}{1-\alpha}}$. Hence, $\sup_{F \in \mathcal{F}(\mu, \sigma)} \text{VaR}_\alpha(X_F)$ is not attainable. \square

4 Numerical Example

The development of DRRO offers an opportunity to examine numerically how an individual who overly weights "very good" and "overly bad" outcomes, as depicted by the Cumulative Prospect Theory (CPT), makes optimal decisions when facing distributional ambiguity. In particular, in this section we apply the inverse S-shaped distortion function (2.10) in Example 2.2(i) to examine the impact of the parameter α on the solutions of DRRO. Recall that the smaller the value of α , the more weights are put on extreme (low and high) quantiles, whereas as α approaches to 1, the distorted expectation converges to risk-neutral expectation. Existing literature suggests that a reasonable choice of α would be somewhere between 0.56 and 0.71, which are supported by the empirical survey results in Camerer and Ho (1994), Tversky and Kahneman (1992), and Wu and Gonzalez (1996).

As a testbed for implementing DRRO, we consider the two-stage problem of production and transportation planning, a classical example that has been considered in Bertsimas et al. (2010) and Yang and Xu (2016). In this problem, given m facilities and n customer locations, a company needs to first determine how many products to produce in each facility, and then decide how many products to transport from each facility to each customer location. The demand from each customer location j is known beforehand as d_j units, and the production cost at each facility i is also known with certainty as c_i per unit. The transportation cost t_{ij} between each facility i and customer location j however is uncertain. The company needs to make a "here-and-now" production decision but can make "wait-and-see" transportation decision after the realization of the random cost t_{ij} . This two-stage problem can be formulated as

$$\begin{aligned} \min_{x_i} \quad & \sum_{i=1}^m c_i x_i + \rho_h^F(Q(\vec{T}, \vec{x})) \\ \text{subject to} \quad & 0 \leq x_i \leq 1, \forall i, \end{aligned} \quad (4.1)$$

where x_i denotes the amount to produce at each facility i , the second stage cost $Q(\vec{t}, \vec{x})$, which captures the total transportation cost, is defined by

$$\begin{aligned} Q(\vec{t}, \vec{x}) := \min_{w_{ij}} \quad & \sum_{i=1}^m \sum_{j=1}^n t_{ij} w_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^m w_{ij} = d_j, \quad \forall j \\ & \sum_{j=1}^n w_{ij} = x_i, \quad \forall i \\ & w_{ij} \geq 0, \quad \forall i, j, \end{aligned}$$

and ρ_h^F in (4.1) denotes a distorted expectation based on the inverse S-shaped distortion function h defined in (2.10).

We can solve the following DRRO problem that captures aversion towards ambiguity

$$\begin{aligned} \min_{x_i} \quad & \sum_{i=1}^m c_i x_i + \sup_{F \in \mathcal{F}} \rho_h^F(Q(\vec{T}, \vec{x})) \\ \text{subject to} \quad & 0 \leq x_i \leq 1, \quad \forall i. \end{aligned} \tag{4.2}$$

It is straightforward to confirm that the function $Q(\vec{t}, \vec{x})$ is convex in \vec{x} and concave in \vec{t} . One classical setup in DRO to address the issue of distribution ambiguity is to assume that the distribution F of transportation cost can only be characterized by the fixed mean, bounded second order moments, and an ellipsoidal set of support, i.e. the following uncertainty set

$$\mathcal{F} := \left\{ F \in \mathcal{F}_0(\mathbb{R}^d) \left| \begin{array}{l} \mathbb{E}_F[\vec{T}] = \vec{m} \\ \mathbb{E}_F[(\vec{T} - \vec{m})(\vec{T} - \vec{m})^\top] \preceq \Sigma \\ \mathbb{P}_F((\vec{T} - \vec{t}_0)^\top \Theta (\vec{T} - \vec{t}_0) \leq 1) = 1 \end{array} \right. \right\}, \tag{4.3}$$

for some $\vec{t}_0 \in \mathbb{R}^d$ and $\Theta \in \mathbb{R}^{d \times d}$. Following the discussion throughout Section 3.1, we can solve the above DRRO problem (approximately) by identifying first the convex envelop h_{cx} of the inverse S-shaped distortion function (2.10) (see Example 3.1), approximating the envelope h_{cx} by a piecewise

linear function h_m , and then solving the following semi-definite program (SDP)

$$\begin{aligned}
& \min_{x_i, \vec{\lambda}_m, \Lambda \in \mathbb{R}^{m \times n}, s_k, \tau_k, w_{ijk}} && \sum_{i=1}^m c_i x_i + \vec{\lambda}_m^\top \vec{m} + \Lambda \cdot (\Sigma + \vec{m} \vec{m}^\top) + \sum_{k=1}^K p_k s_k \\
& \text{subject to} && \begin{bmatrix} \Lambda & \frac{1}{2}(\vec{\lambda}_m - \tilde{\phi}_k \vec{w}_k) \\ \frac{1}{2}(\vec{\lambda}_m - \tilde{\phi}_k \vec{w}_k)^\top & s_k \end{bmatrix} \succeq -\tau_k \begin{bmatrix} \Theta & -\Theta \vec{t}_0 \\ -\vec{t}_0^\top \Theta & \vec{t}_0^\top \Theta \vec{t}_0 - 1 \end{bmatrix}, \quad \forall k \in \{1, \dots, K\} \\
& && \tau_k \geq 0, \quad \forall k \in \{1, \dots, K\} \\
& && \sum_{i=1}^m w_{ijk} = d_j, \quad \forall j, \quad \forall k \in \{1, \dots, K\} \\
& && \sum_{j=1}^n w_{ijk} = x_i, \quad \forall i, \quad \forall k \in \{1, \dots, K\} \\
& && w_{ijk} \geq 0, \quad \forall i, j, \forall k \in \{1, \dots, K\} \\
& && 0 \leq x_i \leq 1, \quad \forall i,
\end{aligned}$$

where $\vec{w}_{k'} := \text{vec}(\{w_{ijk}\}_{k=k'})$, $p_k := \alpha_k - \alpha_{k-1}$, and $\tilde{\phi}_k$ denotes the slope of each linear piece of h_m . Throughout our experiments, we implement the above optimization model in Matlab using YALMIP as the modelling language and Sedumi as the optimization solver.

To set up parameters for the model, we follow closely the setup applied in Bertsimas et al. (2010) and Yang and Xu (2016), and we consider 10 suppliers and 3 customers. In particular, we randomly generate the demand d_j from the uniform distribution defined over the interval $[0.5m/n, m/n]$, which ensures $\sum_j d_j < m$, i.e. total demand is less than total supply. The mean and variance of the transportation costs q_{ij} are set to be $100 + 0.1\sqrt{n(i-1) + j}$ and $5/\sqrt{n(i-1) + j}$ respectively. That is, the average cost is lower (but the variability of the cost is higher) for the transportation with lower serial number. The support set is calibrated by setting it to cover 95% of a multivariate normal distribution with the set mean and variance.

In Figure 4, we present the optimal solutions to the DRRO model (4.2), i.e. the production decisions x_i , in terms of how productions are allocated across different suppliers. As seen, as α decreases there are more allocations to suppliers whose transportation costs are higher in average but lower in variance. Moreover, the allocation in the case of lower α is more "diversified", i.e. more suppliers are involved. We can interpret that optimal production decisions tend to be more conservative as α decreases. This might not necessarily be intuitive since as α decreases, not only more weights would be put on "very bad" outcomes but also on "very good" outcomes in the inverse S-shaped distortion function. It appears that one's sensitivity towards "very good" outcomes does not "materialize" when the individual is averse towards distribution ambiguity. To facilitate explaining the results, we provide in Figure 5 the convex envelope (3.12) of the inverse S-shaped distortion function (2.10) for various distortion parameter α . Taking a closer look at Figure 5, we see that the convex envelope simply takes a linear form over lower-quantile region, i.e. from probability zero to around 0.7, regardless of α -value. In other words, an individual characterized by an inverse

S-shaped distortion function would "act" as if he/she is risk neutral towards lower-quantile events when the individual is ambiguity-averse. In the meanwhile, the individual would remain sensitive to "very bad" outcomes since the envelope coincides with the distortion function over high-quantile region. This explains why overall the decisions become more risk-averse as α decreases.

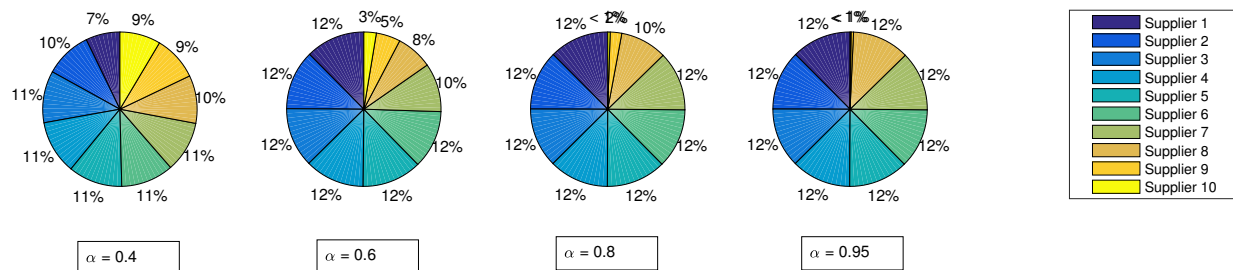


Figure 4: Optimal allocations for various distortion parameters α

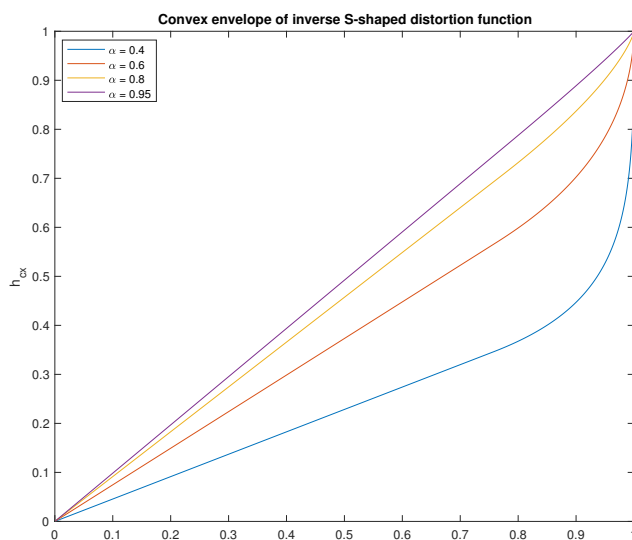


Figure 5: Convex envelopes of inverse S-shaped distortion functions (2.10)

In addition to the explanation based on the envelope, we can also establish some intuition from the perspective of worst-case distribution. Since an inverse S-shaped distortion function is

sensitive to both lower quantiles and higher quantiles, a distribution that has larger (resp. smaller) spread over lower quantiles and smaller (resp. larger) spread over higher quantiles is likely to be considered favourable (resp. unfavourable). In Figure 6, we see that this indeed is the case for the worst-case distributions, which we compute based on the result in Theorem 3.5. We see that the worst-case distributions always put a point mass (with $> 70\%$ weight) at the lowest value, i.e. it reaches the minimum spread over lower quantiles, and then assign weights in a monotonically decreasing fashion to larger values. We see in Figure 5 that the range of probabilities where the concave form of distortion takes place is always between 0 and some value less than 0.7, and hence this part of distortion essentially applies to the same point mass (the left-end point) of the worst-case distribution and has no real effect. This indicates that the most conservative view an "inverse S-shaped" individual can hold over "very good" outcomes is that their corresponding values concentrate on a single point, which makes the individual feel indifferent.

Unlike the worst-case distributions in classical DRO framework however, where the number of supports is always less or equal to the number of moment constraints, we see in Figure 6 that the worst-case distribution spreads over to the right and the degree of the spread depends on the shape of the convex envelope. In particular, we see that the worst-case distribution has a "fatter" tail that corresponds to the "steepness" of the envelope (at different quantile levels) as α decreases. The worst-case distribution converges to a point mass at the mean μ , as α increases to 1, which corresponds to the constant slope of the envelope (in the case $\alpha = 1$). As mentioned earlier, this also recovers the result in Delage et al. (2014).

5 Conclusion

In this paper, we advocate the use of distorted expectation to model a decision maker's risk attitude in distributionally robust optimization. The scheme can be attractive from both the theoretical point of view, given its roots in Yaari's dual theory of choice under risk (Yaari (1987)), and from the practical point of view, given the interpretability of a distortion function and the easiness to specify the function. In particular, we highlight in this paper the computational tractability of solving distributionally robust optimization problems formulated based on distorted expectations, and show that in many cases the problem can be solved by an alternative distributionally robust optimization problem formulated based on some convex distortion function. The latter can often be solved (exactly or approximately) by convex optimization techniques. The key of our approach lies in the use of convex envelope to discover hidden convexity in the proposed distributionally robust risk optimization problems. As an important example, our results can be applied to discover how a decision maker whose risk attitude is captured by a non-convex distortion function, e.g. an inverse S-shaped function, would act when facing distributional ambiguity.

References

- Acerbi, C. (2002). A coherent representation of subjective risk aversion *Journal of Banking and Finance*, **26**(7), 1505–1518.
- Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.
- Benati, S. and Rizzi, R. (2007) A mixed integer linear programming formulation of the optimal mean/Value-at-Risk problem. *European journal of Operational Research*, **176**(1), 423–434.
- Ben-Tal, A., den Hertog, D., Vial, J. (2015) Deriving robust counterparts of nonlinear uncertain inequalities *Mathematical Programming*, **149**, 265–299.
- Ben-Tal, A., den Hertog, D., De Waegenaere, A., Melenberg, B., Rennen, G. (2013) Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, **59**(2), 341–357.
- Ben-Tal A, Nemirovski, A (2001) *Lectures on modern convex optimization: analysis, algorithms, and engineering applications* (SIAM, Philadelphia).
- Bertsimas, D., Doan, X.V., Natarajan, K., Teo, C.P. (2010) Models for minimax stochastic linear optimization problems with risk aversion *Mathematics of Operations Research* **35**(3), 580–602.
- Belles-Sampera J, Guillen M, Santolino M (2014) Beyond Value-at-Risk: GlueVaR distortion risk measure. *Risk Analysis*, **34**(1), 121–134.
- Belles-Sampera J, Guillen M, Santolino M (2016) What attitudes to risk underlie distortion risk measure choices? *Insurance: Mathematics and Economics*, **68**, 101–109.
- Bleichrodt, H. and Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management science*, **46**(11), 1485–1496.
- Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Brighi, B. and Chipot, M. (1994). Approximated convex envelope of a function. *SIAM journal on numerical analysis*, **31**(1), 128–148.
- Cai J., Liu H., Wang R. (2018) Asymptotic equivalence of risk measures under dependence uncertainty. *Mathematical Finance* **28**, 29–49.
- Cai, J., Wang, Y. and Mao, T. (2017). Tail subadditivity of distortion risk measures and multivariate tail distortion risk measures. *Insurance: Mathematics and Economics*, **75**, 105–116.

- Camerer, C. F., Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, **8**, 167–196.
- Chen L, He S, Zhang S (2011) Tight bounds for some risk measures, with applications to robust portfolio selection *Operations Research*, **59**(4), 847–865.
- Cont R, Deguest R, Scandolo G (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, **10**, 593–606.
- Cornilly D. and Vanduffel S. (2019) Equivalent distortion risk measures on moment spaces. *Statistics & Probability Letters*, **146**, 187–192.
- Delage, E., Arroyo, S., and Ye, Y. (2014). The value of stochastic modeling in two-stage stochastic programs with cost uncertainty. *Operations Research*, **62**(6): 1377–1393.
- Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* **58**(3): 595–612.
- Denneberg D (1994a). *Non-additive measure and integral*. Kluwer Academic Publishers, Dordrecht.
- Denneberg D (1994b). Conditioning (updating) non-additive measures. *Annals of Operations Research*, **52**(1994), 21–42.
- El Ghaoui L, Oks M, Oustry F (2003) Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, **51**(4): 543–556.
- Esfahani P M, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, **171**(1-2): 115–166.
- Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and Stochastics*, **6**(4), 429–447.
- Guegan, D. and Hassani, B. (2015). Distortion risk measure or the transformation of unimodal distributions into multimodal functions. In *Future perspectives in risk models and finance*, (pp. 71-88). Springer, Cham.
- Hanasusanto, G., Roitch, V., Kuhn, D., Wiesemann, W. (2017). Ambiguous joint chance constraints under mean and dispersion information *Operations Research*, **65**(3), 751–767.
- Jiang, R. and Guan Y. (2016). Data-driven chance constrained stochastic program. *Mathematical Programming*, **158**(1-2), 291–327.
- Jin H, Zhou, X.Y. (2008). Behavioral portfolio selection in continuous time. *Mathematical Finance*, **18**, 385–426.

- Li, J. Y. M. (2018). Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research*, **66**(6), 1533–1541.
- Li L, Shao H, Wang R, Yang J (2018) Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics*, **9**(1), 190–218.
- Liu F., Cai J., Lemieux C., Wang R. (2020) Convex risk functionals: representation and applications. *Insurance: Mathematics and Economics*, **90**, 66–79.
- Mak, H.Y., Rong, Y., Zhang, J. (2015). Appointment scheduling with limited distributional information. *Management Science*, **61**(2), 316–334.
- Mao T., Cai J. (2018) Risk measures based on the behavioural economics theory. *Finance and Stochastics*, **22**, 367–393.
- Merkle, M., Marinescu, D., Merkle, M.M.R., Monea M., Stroe M. (2014). Lebesgue-Stieljes integral and Young’s inequality. *Applicable Analysis and Discrete Mathematics*, **8**, 60–72.
- Natarajan K, Sim M, Uichanco J (2010) Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, **20**(4), 695–731.
- Popescu I (2007) Robust mean-covariance solutions for stochastic optimization. *Operations Research*, **51**(1), 98–112.
- Postek, K., Ben-Tal, A., den Hertog, D., Melenberg B. (2018) Robust optimization with ambiguous stochastic constraints under mean and dispersion information *Operations Research*, **66**(3), 597–892.
- Prelec, D. (1998) The probability weighting function. *Econometrica*, **66**, 497–527.
- Tversky A, Fox C.R. (1995) Weighing risk and uncertainty. *Psychological Review*, **102**, 269–283.
- Tversky A, Kahneman D (1992) Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**, 297–323.
- Wang R., Bignozzi, V., Tsakanas, A. (2015). How superadditive can a risk measure be? *SIAM Journal on Financial Mathematics*, **6**, 776–803.
- Wang, S. S. (2000). A class of distortion operators for pricing financial and insurance risks. *Journal of risk and insurance*, 15–36.
- Wiesemann W., Kuhn, D., Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, **62**(6), 1203–1466.

- Wirch, J. L. (1999). Coherent beta risk measures for capital requirements. *Phd thesis*, University of Waterloo.
- Wozabal, D (2014) Robustifying convex risk measures for linear portfolios: a nonparametric approach. *Operations Research*, **62**, 1302–1315.
- Wu, G. and R. Gonzalez. (1996). Curvature of the probability weighting function. *Management Science*, **42**, 1676–1690.
- Xu Z.Q. and Zhou X.Y. (2013) Optimal stopping under probability distortion. *Annals of Applied Probability*, **23**(1), 251–282.
- Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, 95–115.
- Yang W. and Xu H. (2016) Distributionally robust chance constraints for non-linear uncertainties. *Mathematical Programming*, **155**(1-2): 231-265.
- Zhu, W., Shao H. (2018). Closed-form solutions for extreme-case distortion risk measures and applications to robust portfolio management. Available at SSRN: <https://ssrn.com/abstract=3103458>.
- Zymler S., Kuhn D., Rustem B. (2013a) Distributionally robust joint chance constraints with second-order moment information. *Mathematical Programming*, **137**(1-2): 167–198.
- Zymler S., Kuhn D., Rustem B. (2013b) Worst-case value-at-risk of non-linear portfolios. *Management Science*, **59**(1), 172–188.

6 Appendix

In the appendix, we give the proofs of the preliminary lemmas and most of the main results in the paper.

6.1 Proof of Lemma 2.1

Lemma 2.1 is used in the proof of Theorem 3.10. This lemma gives the relationships among the existing expressions of distortion functions and will be helpful in the study related to distortion functions.

Proof of Lemma 2.1: Since ρ_h is decreasing in h and $h(x)$ is an increasing function, (i) holds obviously. In addition, (iii) follows immediately from (ii) since $h = h_+$ and $h = h_-$ when h is right continuous and left continuous, respectively. Hence, we only need to show (ii). Let F be the distribution function of X . As h_+ is a right-continuous distortion function, h_+ can be viewed as a distribution function of a random variable defined on $(0, 1)$. Moreover, $x \mapsto h_+(F(x))$ is also right-continuous and is also a distribution function. Let Y be a random variable having distribution h_+ . Then we have $F^{-1}(Y) = \text{VaR}_Y(X)$ is a random variable having distribution $\mathbb{P}(F^{-1}(Y) \leq x) = h_+(F(x))$ and hence

$$\begin{aligned} \rho_{h_+}(X) &= \int_{-\infty}^{\infty} x dh_+(F(x)) = \mathbb{E}[F^{-1}(Y)] \\ &= \int_0^1 F^{-1}(\alpha) dh_+(\alpha) = \int_0^1 \text{VaR}_\alpha(X) dh_+(\alpha) = \int_0^1 \text{VaR}_\alpha(X) dh(\alpha), \end{aligned}$$

where the first equality follows from the integration by parts. Hence, (2.7) holds. Now, we prove (2.8). Note that the set of discontinuities of a monotonic function is at most countable. Without loss of generality, denote by $\{\alpha_i, i \in M\}$ with $M \subseteq \mathbb{N}$ the set of discontinuities of h . Let $x_i = \text{VaR}_{\alpha_i}(X)$ and $y_i = \text{VaR}_{\alpha_i}^+(X)$, $i \in M$. Then $h_-(\alpha) = h_+(\alpha)$ for $\alpha \in (0, 1) \setminus \{\alpha_i, i \in M\}$, that is, $h_-(F(x)) = h_+(F(x))$ for $x \in \mathbb{R} \setminus \cup_{i \in M} [x_i, y_i]$ as $F(x) = \alpha_i$ for $x \in [x_i, y_i]$, $i \in M$. Then by (2.1), we have

$$\begin{aligned} \rho_{h_-}(X) - \rho_{h_+}(X) &= \sum_{i \in M} \int_{x_i}^{y_i} (h_+(F(x)) - h_-(F(x))) dx \\ &= \sum_{i \in M} (h_+(\alpha_i) - h_-(\alpha_i))(y_i - x_i) \\ &= \sum_{i \in M} (\text{VaR}_{\alpha_i}^+(X) - \text{VaR}_{\alpha_i}(X))(h_+(\alpha_i) - h_-(\alpha_i)) \\ &= \sum_{i \in M} \int_{\{\alpha_i\}} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) dh(\alpha). \end{aligned} \tag{6.1}$$

By (2.7) and $\text{VaR}_\alpha^+(X) = \text{VaR}_\alpha(X)$ for $\alpha \in (0, 1) \setminus \{\alpha_i, i \in M\}$, we have

$$\begin{aligned} \rho_{h-}(X) &= \int_0^1 \text{VaR}_\alpha(X) dh(\alpha) + \sum_{i \in M} \int_{\{\alpha_i\}} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) dh(\alpha) \\ &= \int_{(0,1) \setminus \{\alpha_i, i \in M\}} \text{VaR}_\alpha(X) dh(\alpha) + \sum_{i \in M} \int_{\{\alpha_i\}} \text{VaR}_\alpha^+(X) dh(\alpha) \\ &= \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha). \end{aligned}$$

Thus, we complete the proof. \square

6.2 Proof of Lemma 3.1

Lemma 3.1 plays an important role in the paper, which identifies a particular structure of convex envelopes that can be used to characterize worst-case distributions. Mathematically, it is the generalization of the following well-know result in Brighi and Chipot (1994).

Lemma 6.1. *(Lemma 5.1 of Brighi and Chipot (1994)). Suppose h is a continuous distortion function and h_{cx} is the convex envelope of h . Then, the set $\{t \in [0, 1] : h(t) \neq h_{\text{cx}}(t)\}$ is the union of some disjoint open intervals, and h_{cx} is linear on each of the open intervals.*

Proof of Lemma 3.1: (i) Note that the set of discontinuities of a monotonic function is at most countable. Without loss of generality, denote by $\{x_i, i \in M\}$ with $M = \{1, 2, \dots\}$ the set of discontinuities of h . Since h is nondecreasing, the left limit of h at any discontinuity x_i exists, say $h(x_i-) = \lim_{y \uparrow x_i} h(y)$, and $h(x_i-) < h(x_i+) = h(x_i)$. Note that h_{cx} is continuous and $h_{\text{cx}}(y) \leq h(y)$ for any $y \in [0, 1]$. Hence $h_{\text{cx}}(x_i) = \lim_{y \uparrow x_i} h_{\text{cx}}(y) \leq \lim_{y \uparrow x_i} h(y) = h(x_i-) < h(x_i)$ for any $i \in M$. Note that the right derivative $(h_{\text{cx}})'_+$ of h_{cx} exists at any $x \in [0, 1)$ satisfying that $(h_{\text{cx}})'_+(x)$ is finite and nondecreasing in $x \in [0, 1)$, and

$$\lim_{\delta \downarrow 0} \frac{h(x_i + \delta) - h(x_i-)}{\delta} = \infty, \quad \text{for each } i \in M.$$

Also, note that the set of discontinuities of h is at most countable. Hence, for each $i \in M$, there exists $\delta_i^0 > 0$ such that $h(x)$ is continuous at $x_i + \delta_i^0$, and

$$(h_{\text{cx}})'_+(x) < \Delta_i(\delta_i^0) \quad \text{for all } x \in [x_i, x_i + \delta_i^0] \quad \text{with } \Delta_i(\delta) := \frac{h(x_i + \delta) - h(x_i-)}{\delta}. \quad (6.2)$$

If $h(x)$ does not cross⁷ the function $g(x) := h(x_i-) + \Delta_i(\delta_i^0)(x - x_i)$ in the interval $(x_i, x_i + \delta_i^0)$, then $h(x) \geq g(x)$ for all $x \in (x_i, x_i + \delta_i^0)$ and $g(x_i + \delta_i^0) = h(x_i + \delta_i^0)$. Then in this case, we have with $\delta_i = \delta_i^0$ and $\Delta_i = \Delta_i(\delta_i^0)$, for $x \in (x_i, x_i + \delta_i)$

$$h_{\text{cx}}(x) \leq h_{\text{cx}}(x_i) + (h_{\text{cx}})'_+(x)(x - x_i) < h(x_i-) + \Delta_i(x - x_i) \leq h(x). \quad (6.3)$$

⁷A function $f(x) : E \rightarrow \mathbb{R}$ is said to cross another function $g(x)$ on $x \in E$, if there exists $x_0 \in E$ such that $f(x) - g(x)$ changes its sign from $x_0 - \delta$ to $x_0 + \delta$ for some $\delta > 0$ with $(x_0 - \delta, x_0 + \delta) \subset E$. Specifically, we say f crosses g at x_0 .

Otherwise, if $h(x)$ crosses $g(x)$, let $x_i^* := \inf\{x \in (x_i, x_i + \delta_i^0) : h \text{ crosses } g \text{ at } x\}$, then we have

$$x_i^* \geq x_i + \frac{h(x_i) - h(x_i-)}{\Delta_i(\delta_i^0)},$$

where the inequality follows from h is nondecreasing and $h(x_i) > g(x_i)$. It is obvious that h crosses g at x_i^* and h is continuous at x_i^* . Let $\delta_i = x_i^* - x_i$ and $\Delta_i = \Delta_i(\delta_i)$. Then we have $h(x) \geq g(x)$ for all $x \in [x_i, x_i + \delta_i]$ and $h(x_i + \delta_i) = g(x_i + \delta_i)$, and thus, (6.3) holds. Also, note that for each $i \in M$, $\delta_i^0 > 0$ can be chosen to be arbitrarily small for (6.2) holds and $\delta_i \leq \delta_i^0$. We can choose arbitrarily small $\delta_i > 0$ such that (6.3) holds. Then for $j = 1, 2, \dots$, if there exists $i < j$ such that $x_j < x_i$, we take $\delta_j < \min_{i < j} \{x_i - x_j : x_i > x_j\}$.

Note that for each $i \in M$, we find arbitrarily small $\delta_i > 0$ with $\Delta_i = \Delta_i(\delta_i)$ defined by (6.2) such that the function g_i defined as follows satisfies

$$h_{\text{cx}}(x) < g_i(x) := h(x_i-) + \Delta_i(x - x_i) < h(x), \quad x \in (x_i, x_i + \delta_i), \quad (6.4)$$

$$g_i(x_i) = h(x_i-), \quad g_i(x_i + \delta_i) = h(x_i + \delta_i), \quad (6.5)$$

and h is continuous at $x_i + \delta_i$. Also, (6.4) and (6.5) to hold. For $i, j \in M$ with $i < j$, if $x_j \in (x_i, x_i + \delta_i]$, then $h(x_j) > g_i(x_j)$. Then we remove all such j 's from the set M , that is, let

$$M^* = \{i \in M : x_i \notin [x_j, x_j + \delta_j] \text{ for } j < i\}.$$

Then it is obvious the intervals $\{[x_i, x_i + \delta_i]\}_{i \in M^*}$ are disjoint and $M \subset \cup_{i \in M^*} [x_i, x_i + \delta_i]$. Define

$$h^*(x) = \begin{cases} h(x_i-) + \Delta_i(x - x_i), & x \in [x_i, x_i + \delta_i], i \in M^*, \\ h(x), & \text{otherwise.} \end{cases}$$

It is easy to see that $h_{\text{cx}} \leq h^* \leq h$ and h^* is a continuous distortion function. Hence, h_{cx} is also the largest convex function dominated by h^* . By Lemma 6.1, we have that the set $\{t \in [0, 1] : h^*(t) \neq h_{\text{cx}}(t)\}$ is the union of some disjoint open intervals, denoted by $\cup_{k \in I} (a_k, b_k)$, and h_{cx} is linear on each of the open intervals. Hence, by the definition of h^* and (6.4), we have

$$\begin{aligned} \{t \in [0, 1] : h(t) \neq h_{\text{cx}}(t)\} &= \{t \in [0, 1] : h(t) > h_{\text{cx}}(t)\} \\ &= \{x_i, i \in M^*\} \cup \{t \in [0, 1] : h^*(t) > h_{\text{cx}}(t)\}, \end{aligned}$$

where the second equality follows from $h(x) > h^*(x) > h_{\text{cx}}(x)$ for all $x \in (x_i, x_i + \delta_i)$ and $h(x) = h^*(x)$ for all $x \in [0, 1] \setminus (\cup_{i \in M^*} [x_i, x_i + \delta_i])$. We assert that for each $i \in M^*$, either $x_i \in (a_k, b_k)$ for some $k \in I$ or $x_i = a_k$ for some $k \in I$. If this is not true, then $x_i = b_k$ for some $k \in I$ or $x_i \in [0, 1] \setminus (\cup_{k \in I} [a_k, b_k])$, and then there exists $\delta > 0$ such that $(x_i, x_i + \delta) \not\subset \{t \in [0, 1] : h^*(t) \neq h_{\text{cx}}(t)\}$, that is, $h^*(x) = h_{\text{cx}}(x)$ for $x \in (x_i, x_i + \delta)$. This is a contradiction with (6.4). Now we have the

$$\{t \in [0, 1] : h(t) \neq h_{\text{cx}}(t)\} = \cup_{k \in I_1} [a_k, b_k) \cup (\cup_{\ell \in I_2} (a_\ell, b_\ell)),$$

where $a_k = x_{i_k}$, $i_k \in M^*$ for any $k \in I_1$. Hence, from (6.5), we have

$$h(a_k-) = h^*(a_k) = h_{\text{cx}}(a_k), \quad k \in I_1,$$

and it is obvious that $b_k, a_\ell, b_\ell \in [0, 1] \setminus (\cup_{i \in M^*} [x_i, x_i + \delta_i))$, $k \in I_1, \ell \in I_2$, h and h^* coincides at above points and h is continuous at above points, and thus

$$h(b_k) = h_{\text{cx}}(b_k), \quad k \in I_1; \quad h(a_\ell) = h_{\text{cx}}(a_\ell), \quad h(b_\ell) = h_{\text{cx}}(b_\ell), \quad \ell \in I_2,$$

Thus, we complete the proof of (i).

(ii) For a left-continuous distortion function h , let h^+ be the right-continuous version of h , that is, $h^+(x) = \lim_{y \downarrow x} h(y)$, $x \in [0, 1]$. Then $h_{\text{cx}} = h_{\text{cx}}^+$ and $h(x) \leq h^+(x)$ as h is nondecreasing. We also have $\{t \in [0, 1] : h(t) < h^+(t)\} = \{x_i, i \in M\}$ and thus,

$$\{t \in [0, 1] : h^+(t) > h_{\text{cx}}(t)\} = \{x_i, i \in M\} \cup \{t \in [0, 1] : h(t) > h_{\text{cx}}(t)\}.$$

From the proof of (i), we have $\{t \in [0, 1] : h(t) > h_{\text{cx}}(t)\} = \cup_{k \in I} (a_k, b_k)$ as h is left continuous at a_k , $k \in I$. Thus, we complete the proof. \square

6.3 Proof of Proposition 3.2

Proposition 3.2 is one of main results in the paper. Before proceeding further, we should note first that the proof may appear technical in that it requires a number of additional definition, assumptions, and lemmas. In particular, the proof relies on Lemma 6.2, 6.3, 6.4, 6.5 that will be presented first in sequence. Readers might consider reading first the main proof presented after these lemmas and come back to these lemmas whenever needed.

Now, we start with Lemma 6.2 which requires the following definition.

Definition 6.1. A function $f^* : [a, b] \rightarrow \mathbb{R}$ is called an *increasing rearrangement* of a measurable function $f : [a, b] \rightarrow \mathbb{R}$, if f^* is non-decreasing and its level sets have the same measure as the level sets of f , i.e.

$$\mu(\{x \in [a, b] : f^*(x) > t\}) = \mu(\{x \in [a, b] : f(x) > t\}), \quad t \in \mathbb{R},$$

where μ is the Lebesgue measure in $[a, b] \subset \mathbb{R}$. Specially, if $[a, b] = [0, 1]$, then on the probability space $([0, 1], \mathcal{B}([0, 1]), \mu)$, we have f^* is an increasing rearrangement of f if and only if $f^*(x)$ is non-decreasing in $x \in [0, 1]$ a.s., and $f^* \stackrel{\text{d}}{=} f$, where $\mathcal{B}([0, 1])$ is the Borel field on $[0, 1]$ and $f^* \stackrel{\text{d}}{=} f$ represents for that they have the same distribution.

Lemma 6.2. (*Hardy-Littlewood inequality*) If f and g are measurable real functions on $[a, b] \subset \mathbb{R}$, then

$$\int_a^b f(x)g(x) \, dx \leq \int_a^b f^*(x)g^*(x) \, dx, \quad (6.6)$$

where f^* and g^* are the increasing rearrangements of f and g , respectively. The equality in (6.6) holds if and only if

$$\begin{aligned} & \mu(\{x \in [a, b] : f(x) > s\} \cap \{x \in [a, b] : g(x) > r\}) \\ &= \min \{ \mu(\{x \in [a, b] : f(x) > s\}), \mu(\{x \in [a, b] : g(x) > r\}) \} \text{ for } (s, r) \in \mathbb{R}^2 \text{ a.e.,} \end{aligned} \quad (6.7)$$

where μ is the Lebesgue measure in \mathbb{R} .

To state the following lemma, we recall that a class of sets $\{B_i, i \in I\}$ is called a *partition* of the probability space Ω , if $B_i \cap B_j = \emptyset$ a.s. for any $i \neq j$ and $\cup_{i \in I} B_i = \Omega$ a.s. Without loss of generality, we assume $\mathbb{P}(B_i) > 0$ for each $i \in I$ when $\{B_i, i \in I\}$ is a partition of a probability space Ω .

Lemma 6.3. *For any $\vec{X} = (X_1, \dots, X_d) \in \mathcal{F}_{\vec{m}, G}$ and a partition $\{B_i, i \in I\}$ of Ω , define a random vector $\vec{X}^* = (X_1^*, \dots, X_d^*)$ associated with \vec{X} as*

$$\vec{X}^* = \sum_{k \in I_1} \mathbb{E}[\vec{X}|B_k] I_{B_k} + \sum_{\ell \in I_2} \vec{X} I_{B_\ell} =: \sum_{k \in I_1} \vec{x}_k I_{B_k} + \sum_{\ell \in I_2} \vec{X} I_{B_\ell}, \quad (6.8)$$

where I_A is the indicator function of a set A , $I_1 \cup I_2 = I$ and $I_1 \cap I_2 = \emptyset$. Then, $\vec{X}^* = (X_1^*, \dots, X_d^*) \in \mathcal{F}_{\vec{m}, G}$.

Proof. Denote $p_k = \mathbb{P}(B_k)$, $k \in I$. First, by (6.8), we have that

$$\mathbb{E}[\vec{X}^*] = \sum_{k \in I_1} \vec{x}_k p_k + \sum_{\ell \in I_2} \mathbb{E}[\vec{X} I_{B_\ell}] = \sum_{k \in I} p_k \mathbb{E}[\vec{X}|B_k] = \mathbb{E}[\vec{X}],$$

that is, \vec{X}^* satisfies (3.17). Second, note that

$$\vec{x}_k = [\vec{X}^*|B_k] = \mathbb{E}[\vec{X}|B_k] =: \vec{x}_k \text{ a.s., } k \in I_1. \quad (6.9)$$

Further, for any random vector \vec{X} and any convex set $C \subseteq \mathbb{R}^d$, it holds that $\mathbb{E}[\vec{X}|\vec{X} \in C] \in C$ a.s. (see e.g., Section 2.1.4 of Boyd and Vandenberghe (2004)). For each $k \in I_1$, denote $\vec{Y}_k = [\vec{X}|B_k]$. Since $\mathbb{P}(B_k) > 0$ and $B_k \subset \{\vec{X} \in G\}$ a.s., we have $\mathbb{P}(\vec{Y}_k \in G) = \mathbb{P}(\{\vec{X} \in G\} \cap B_k) / \mathbb{P}(B_k) = 1$. Then

$$\vec{x}_k = \mathbb{E}[\vec{X}|B_k] = \mathbb{E}[\vec{Y}_k] = \mathbb{E}[\vec{Y}_k | \vec{Y}_k \in G] \in G, \text{ a.s., } k \in I_1,$$

which, together with $\vec{X} \in G$ and (6.8), implies

$$\vec{X}^* = \sum_{k \in I_1} \vec{x}_k I_{B_k} + \vec{X} I_{\{\cup_{\ell \in I_2} B_\ell\}} \in G \text{ a.s.}$$

Hence, \vec{X}^* satisfies (3.17). Third, for any convex function $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$, we have

$$\begin{aligned} \mathbb{E}[\phi(\vec{X}^*)] &= \sum_{k \in I_1} p_k \phi(\mathbb{E}[\vec{X}|B_k]) + \sum_{\ell \in I_2} p_\ell \mathbb{E}[\phi(\vec{X})|B_\ell] \\ &\leq \sum_{k \in I_1} p_k \mathbb{E}[\phi(\vec{X})|B_k] + \sum_{\ell \in I_2} p_\ell \mathbb{E}[\phi(\vec{X})|B_\ell] = \mathbb{E}[\phi(\vec{X})], \end{aligned} \quad (6.10)$$

where the inequality follows from Jensen's inequality. The inequality (6.10) implies that \vec{X}^* satisfies the constraint (3.17) as \vec{X} satisfies the constraint (3.17). Thus, $\vec{X}^* = (X_1^*, \dots, X_d^*) \in \mathcal{F}_{\vec{m}, \mathcal{G}}$. \square

Lemma 6.4. *Assume that a convex distortion function h is linear on some intervals of $[0, 1]$. Then, the following two results hold.*

- (a) *There exists a partition of $(0, 1]$ with $(0, 1] = \cup_{k \in I} (a_k, b_k]$, where $I = I_1 \cup I_2 \subset \mathbb{N}$ and $I_1 \cap I_2 = \emptyset$, such that $\{(a_k, b_k]\}_{k \in I}$ are disjoint intervals; h is linear on $(a_k, b_k]$ for $k \in I_1$; h is strictly convex on $(a_k, b_k]$ for $k \in I_2$; and the slopes of h on the intervals $(a_k, b_k]$ for $k \in I_1$ are mutually unequal.*
- (b) *Under the notation of (a), for any $\vec{X} = (X_1, \dots, X_d) \in \mathcal{F}_{\vec{m}, \mathcal{G}}$, define random vector $\vec{X}^* = (X_1^*, \dots, X_d^*)$ associated with \vec{X} by (6.8) with $B_k = \{a_k \leq U < b_k\}$, $k \in I$, where U is a random variable uniformly distributed on $(0, 1)$ such that $F_{\vec{w}}^{-1}(U) = f(\vec{w}, \vec{X})$ a.s., and $F_{\vec{w}}$ is the distribution function of $f(\vec{w}, \vec{X})$. Then, $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$. Furthermore, if $f(\vec{w}, \vec{x}) : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ is concave in $\vec{x} = (x_1, \dots, x_d)$, then*

$$\rho_h(f(\vec{w}, \vec{X})) \leq \int_0^1 v(\alpha) dh(\alpha) \leq \rho_h(f(\vec{w}, \vec{X}^*)), \quad (6.11)$$

where

$$v(\alpha) = \begin{cases} f(\vec{w}, \vec{x}_k), & \alpha \in (a_k, b_k], \quad k \in I_1, \\ \text{VaR}_\alpha(f(\vec{w}, \vec{X})), & \alpha \in (a_k, b_k], \quad k \in I_2. \end{cases} \quad (6.12)$$

- (c) *Under the notation of (b), if the equalities in (6.11) hold, that is $\rho_h(f(\vec{w}, \vec{X})) = \rho_h(f(\vec{w}, \vec{X}^*))$, then $v(\alpha)$ defined by (6.12) is a non-decreasing function of α and $v(\alpha) = \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ for $\alpha \in (0, 1)$.*

Proof. The result (a) follows immediately from the definition and properties of a convex distortion function. We prove that the results (b) and (c) hold.

(b) First, note that $\{B_k, k \in I\}$ is a partition of Ω . We get $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ by Lemma 6.3. Next, we prove (6.11). Since h is convex, it is continuous. Thus, by (2.9), we have

$$\rho_h(f(\vec{w}, \vec{X})) = \int_0^1 \text{VaR}_\alpha(f(\vec{w}, \vec{X})) dh(\alpha) = \int_0^1 \text{VaR}_\alpha(f(\vec{w}, \vec{X})) h'(\alpha) d\alpha, \quad (6.13)$$

where $h'(\alpha)$ is the left-derivative of h at α . Note that from the definition of \vec{X}^* , we have

$$[f(\vec{w}, \vec{X}^*)|B_k] = \begin{cases} f(\vec{w}, \vec{x}_k), & k \in I_1, \\ [f(\vec{w}, \vec{X})|B_k], & k \in I_2, \end{cases} \quad \text{a.s.}, \quad (6.14)$$

with $\mathbb{P}(B_k) = b_k - a_k$, $k \in I$, which, together with (6.12), implies that $\alpha \mapsto \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ is a non-decreasing rearrangement of the function v . Then by the Hardy-Littlewood inequality and the fact that $h'(\alpha)$ is non-decreasing in α , we have

$$\rho_h(f(\vec{w}, \vec{X}^*)) = \int_0^1 \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*)) h'(\alpha) d\alpha \geq \int_0^1 v(\alpha) h'(\alpha) d\alpha \quad (6.15)$$

$$\begin{aligned} &= \sum_{k \in I_1} \int_{a_k}^{b_k} f(\vec{w}, \vec{x}_k) dh(\alpha) + \sum_{k \in I_2} \int_{a_k}^{b_k} \text{VaR}_\alpha(f(\vec{w}, \vec{X})) dh(\alpha) \\ &\geq \sum_{k \in I_1} \int_{a_k}^{b_k} \mathbb{E}[f(\vec{w}, \vec{X}) | B_k] dh(\alpha) + \sum_{k \in I_2} \int_{a_k}^{b_k} \text{VaR}_\alpha(f(\vec{w}, \vec{X})) dh(\alpha) \\ &= \sum_{k \in I_1} \mathbb{E}[F_{\vec{w}}^{-1}(U) I_{\{a_k \leq U < b_k\}}] \ell_k + \sum_{k \in I_2} \int_{a_k}^{b_k} \text{VaR}_\alpha(f(\vec{w}, \vec{X})) dh(\alpha) \\ &= \sum_{k \in I} \int_{a_k}^{b_k} \text{VaR}_\alpha(f(\vec{w}, \vec{X})) dh(\alpha) = \rho_h(f(\vec{w}, \vec{X})), \end{aligned} \quad (6.16)$$

where the second inequality follows from Jensen's inequality and the concavity of $f(\vec{w}, \vec{x})$ in $\vec{x} \in \mathbb{R}^d$. This completes the proof of (6.11).

(c) Note that if the equalities in (6.11) hold, that is $\rho_h(f(\vec{w}, \vec{X})) = \rho_h(f(\vec{w}, \vec{X}^*))$, then the equality in (6.15) holds. Note that $\alpha \mapsto \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ is a non-decreasing rearrangement of the function v defined by (6.12). By Lemma 6.2, we have

$$\mu(\{v(\alpha) > s, h'(\alpha) > r\}) = \min\{\mu(\{v(\alpha) > s\}), \mu(\{h'(\alpha) > r\})\} \text{ for } (s, r) \in \mathbb{R}^2 \text{ a.e.}, \quad (6.17)$$

Thus, let $U \sim \text{U}[0, 1]$, we see that (6.17) is equivalent to

$$\mathbb{P}(v(U) > s, h'(U) > r) = \min\{\mathbb{P}(v(U) > s), \mathbb{P}(h'(U) > r)\} \text{ for } (s, r) \in \mathbb{R}^2 \text{ a.e.},$$

which means that $v(U)$ and $h'(U)$ are comonotonic.⁸ Then by the properties of h' and the result in (a), we know that

$$h'(\alpha) = \begin{cases} \ell_k, & \alpha \in (a_k, b_k], \ k \in I_1, \\ h'(\alpha), & \alpha \in (a_k, b_k], \ k \in I_2. \end{cases}$$

is non-decreasing in $\alpha \in [0, 1]$, strictly increasing in each $(a_k, b_k]$, $k \in I_2$ with mutually different ℓ_k , $k \in I_1$ and (6.12), we have v is non-decreasing. Thus, $v(\alpha) = \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ for $\alpha \in (0, 1)$ almost everywhere. Last, note that both the two functions $v(\alpha)$ and $\text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ are left-continuous in α , we have $v(\alpha) = \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ for $\alpha \in (0, 1)$. This completes the proof. \square

The last lemma, Lemma 6.5, that we need in order to prove Proposition 3.2 is presented below. To state it, we introduce the following assumptions.

⁸Two random variables X, Y are said to be comonotonic if $\mathbb{P}(X \leq x, Y \leq y) = \min\{\mathbb{P}(X \leq x), \mathbb{P}(Y \leq y)\}$ for any $(x, y) \in \mathbb{R}^2$ or equivalently $\mathbb{P}(X > x, Y > y) = \min\{\mathbb{P}(X > x), \mathbb{P}(Y > y)\}$ for any $(x, y) \in \mathbb{R}^2$.

Assumption 6.1. The distortion function h is convex and piecewise linear, that is, there exist $0 = a_0 < a_1 < \dots < a_m = 1$, $m \in \mathbb{N}$, such that h is linear on $(a_{k-1}, a_k]$, $k = 1, \dots, m$, and $\ell_1 < \dots < \ell_m$, where ℓ_k is the slope of h on the interval $(a_{k-1}, a_k]$, $k = 1, \dots, m$.

Assumption 6.2. Under the notations of Assumption 6.1, let B_1, \dots, B_m be a partition of Ω with $\mathbb{P}(B_k) = p_k = a_k - a_{k-1}$, $k = 1, \dots, m$. A random vector $\vec{X} \in \mathbb{R}^d$ is a constant vector on each B_k , $k = 1, \dots, m$, namely,

$$\vec{X} = \vec{x}_k \text{ on } B_k, \quad k = 1, \dots, m,$$

where \vec{x}_k , $k = 1, \dots, m$, are any given constant vectors.

Lemma 6.5. Under Assumptions 6.1 and 6.2, if there exist $i < j$ with $f(\vec{w}, \vec{x}_i) > f(\vec{w}, \vec{x}_j)$, then we have

$$|f(\vec{w}, \vec{x}_k) - f(\vec{w}, \vec{x}_\ell)| < 2L\Delta \quad \text{for all } k, \ell \in \{i, i+1, \dots, j\}, \quad (6.18)$$

where $L = 1/\min_{i < j} |\ell_j - \ell_i|$ and $\Delta = \rho_h(f(\vec{w}, \vec{X})) - \sum_{k=1}^m \ell_k p_k f(\vec{w}, \vec{x}_k) > 0$.

Proof. Note that for $1 \leq i < j \leq m$ with $f(\vec{w}, \vec{x}_i) > f(\vec{w}, \vec{x}_j)$, we have $(\ell_j - \ell_i)f(\vec{w}, \vec{x}_j) < (\ell_j - \ell_i)f(\vec{w}, \vec{x}_i)$, which is equivalent to

$$\ell_i f(\vec{w}, \vec{x}_i) + \ell_j f(\vec{w}, \vec{x}_j) < \ell_i f(\vec{w}, \vec{x}_j) + \ell_j f(\vec{w}, \vec{x}_i). \quad (6.19)$$

Thus, if $p_i = a_i - a_{i-1} \leq p_j = a_j - a_{j-1}$, let

$$V(\alpha) = \begin{cases} f(\vec{w}, \vec{x}_k), & \alpha \in (a_{k-1}, a_k], \quad k \neq i, j, \\ f(\vec{w}, \vec{x}_i), & \alpha \in (a_{j-1}, a_{j-1} + p_i], \\ f(\vec{w}, \vec{x}_j), & \alpha \in (a_{i-1}, a_i] \cup (a_{j-1} + p_i, a_j], \end{cases}$$

and note that there exists a permutation (i_1, \dots, i_m) of $(1, \dots, m)$ such that $f(\vec{w}, \vec{x}_{i_1}) \leq \dots \leq f(\vec{w}, \vec{x}_{i_m})$, then we have

$$\text{VaR}_\alpha(f(\vec{w}, \vec{X})) = f(\vec{w}, \vec{x}_{i_k}), \quad \alpha \in (p_{i_1} + \dots + p_{i_{k-1}}, p_{i_1} + \dots + p_{i_k}], \quad k = 1, \dots, m,$$

where $p_{i_0} = 0$. Hence, the function $\alpha \mapsto \text{VaR}_\alpha(f(\vec{w}, \vec{X}))$ is a non-decreasing rearrangement of $V(\alpha)$.

Thus, by (6.19), we have

$$\begin{aligned}
& \sum_{k=1}^m \ell_k p_k f(\vec{w}, \vec{x}_k) \\
&= \sum_{k \neq i, j} \ell_k p_k f(\vec{w}, \vec{x}_k) + p_i \left(\ell_i f(\vec{w}, \vec{x}_i) + \ell_j f(\vec{w}, \vec{x}_j) \right) + \ell_j (p_j - p_i) f(\vec{w}, \vec{x}_j) \\
&< \sum_{k \neq i, j} \ell_k p_k f(\vec{w}, \vec{x}_k) + \ell_i p_i f(\vec{w}, \vec{x}_i) + \ell_j p_i f(\vec{w}, \vec{x}_i) + \ell_j (p_j - p_i) f(\vec{w}, \vec{x}_j) \quad (6.20) \\
&= \sum_{k \neq i, j} \int_{a_{k-1}}^{a_k} f(\vec{w}, \vec{x}_k) dh(\alpha) + \int_{a_{i-1}}^{a_i} f(\vec{w}, \vec{x}_j) dh(\alpha) \\
&\quad + \int_{a_{j-1}}^{a_{j-1}+p_i} f(\vec{w}, \vec{x}_i) dh(\alpha) + \int_{a_{j-1}+p_i}^{a_j} f(\vec{w}, \vec{x}_j) dh(\alpha) \\
&= \int_0^1 V(\alpha) dh(\alpha) = \int_0^1 V(\alpha) h'(\alpha) d\alpha \\
&\leq \int_0^1 \text{VaR}_\alpha(f(\vec{w}, \vec{X})) h'(\alpha) d\alpha = \rho_h(f(\vec{w}, \vec{X})), \quad (6.21)
\end{aligned}$$

where $h'(\alpha)$ is the left-derivative of h at α and is non-decreasing in α , and the last inequality follows from that the function $\alpha \mapsto \text{VaR}_\alpha(f(\vec{w}, \vec{X}))$ is a non-decreasing rearrangement of $V(\alpha)$ and the Hardy-Littlewood inequality.

Similarly, if $p_i = a_i - a_{i-1} > p_j = a_j - a_{j-1}$, we have

$$\begin{aligned}
& \sum_{k=1}^m \ell_k p_k f(\vec{w}, \vec{x}_k) \\
&= \sum_{k \neq i, j} \ell_k p_k f(\vec{w}, \vec{x}_k) + p_j \left(\ell_i f(\vec{w}, \vec{x}_i) + \ell_j f(\vec{w}, \vec{x}_j) \right) + \ell_i (p_i - p_j) f(\vec{w}, \vec{x}_i) \\
&< \sum_{k \neq i, j} \ell_k p_k f(\vec{w}, \vec{x}_k) + \ell_j p_j f(\vec{w}, \vec{x}_i) + \ell_i p_j f(\vec{w}, \vec{x}_j) + \ell_i (p_i - p_j) f(\vec{w}, \vec{x}_i) \quad (6.22) \\
&= \sum_{k \neq i, j} \int_{a_{k-1}}^{a_k} f(\vec{w}, \vec{x}_k) dh(\alpha) + \int_{a_{j-1}}^{a_j} f(\vec{w}, \vec{x}_i) dh(\alpha) \\
&\quad + \int_{a_{i-1}}^{a_{i-1}+p_j} f(\vec{w}, \vec{x}_j) dh(\alpha) + \int_{a_{i-1}+p_j}^{a_i} f(\vec{w}, \vec{x}_i) dh(\alpha) \\
&= \int_0^1 W(\alpha) dh(\alpha) = \int_0^1 W(\alpha) h'(\alpha) d\alpha \\
&\leq \int_0^1 \text{VaR}_\alpha(f(\vec{w}, \vec{X})) h'(\alpha) d\alpha = \rho_h(f(\vec{w}, \vec{X})), \quad (6.23)
\end{aligned}$$

where

$$W(\alpha) = \begin{cases} f(\vec{w}, \vec{x}_k), & \alpha \in (a_{k-1}, a_k], \quad k \neq i, j, \\ f(\vec{w}, \vec{x}_j), & \alpha \in (a_{i-1}, a_{i-1} + p_j], \\ f(\vec{w}, \vec{x}_i), & \alpha \in (a_{j-1}, a_j] \cup (a_{i-1} + p_j, a_i], \end{cases}$$

and the last inequality follows from that the function $\alpha \mapsto \text{VaR}_\alpha(f(\vec{w}, \vec{X}))$ is a non-decreasing rearrangement of $W(\alpha)$ and the Hardy-Littlewood inequality.

By comparing both sides of the inequality (6.20), we see that the difference between the two sides of the inequality in (6.20) is less than Δ . Thus, if $p_i \leq p_j$, we have

$$0 < \ell_i f(\vec{w}, \vec{x}_j) + \ell_j f(\vec{w}, \vec{x}_i) - (\ell_i f(\vec{w}, \vec{x}_i) + \ell_j f(\vec{w}, \vec{x}_j)) < \Delta,$$

that is,

$$0 < f(\vec{w}, \vec{x}_i) - f(\vec{w}, \vec{x}_j) < \frac{\Delta}{\ell_j - \ell_i}. \quad (6.24)$$

Similarly, if $p_i > p_j$, by comparing both sides of the inequality (6.22), we see that (6.24) holds as well. Thus, for any $i < j$ with $f(\vec{w}, \vec{x}_i) > f(\vec{w}, \vec{x}_j)$, (6.24) holds. Hence,

$$|f(\vec{w}, \vec{x}_j) - f(\vec{w}, \vec{x}_i)| < L\Delta \text{ if } f(\vec{w}, \vec{x}_i) > f(\vec{w}, \vec{x}_j) \text{ and } 1 \leq i < j \leq m. \quad (6.25)$$

Next, we aim to show (6.18) holds for any $k, \ell \in \{i, \dots, j\}$. Note that $f(\vec{w}, \vec{x}_i) > f(\vec{w}, \vec{x}_j)$, we have one of the following three cases holds:

$$(1) f(\vec{w}, \vec{x}_k) < f(\vec{w}, \vec{x}_j); \quad (2) f(\vec{w}, \vec{x}_k) \in [f(\vec{w}, \vec{x}_j), f(\vec{w}, \vec{x}_i)]; \quad \text{and} \quad (3) f(\vec{w}, \vec{x}_k) > f(\vec{w}, \vec{x}_j).$$

In Case (1), we have $f(\vec{w}, \vec{x}_k) < f(\vec{w}, \vec{x}_j) < f(\vec{w}, \vec{x}_i)$. By $i \leq k$ and (6.25), we have $f(\vec{w}, \vec{x}_k) > f(\vec{w}, \vec{x}_i) - L\Delta$. Then we have

$$f(\vec{w}, \vec{x}_k) \in [f(\vec{w}, \vec{x}_i) - L\Delta, f(\vec{w}, \vec{x}_j)].$$

In Case (2), noting that $f(\vec{w}, \vec{x}_i) - L\Delta < f(\vec{w}, \vec{x}_j) < f(\vec{w}, \vec{x}_i) < f(\vec{w}, \vec{x}_j) + L\Delta$, we have

$$f(\vec{w}, \vec{x}_k) \in [f(\vec{w}, \vec{x}_i) - L\Delta, f(\vec{w}, \vec{x}_j) + L\Delta].$$

In Case (3), we have $f(\vec{w}, \vec{x}_k) > f(\vec{w}, \vec{x}_i) > f(\vec{w}, \vec{x}_j)$. By $k \leq j$ and (6.25), we have $f(\vec{w}, \vec{x}_k) < f(\vec{w}, \vec{x}_j) + L\Delta$. Then we have

$$f(\vec{w}, \vec{x}_k) \in [f(\vec{w}, \vec{x}_i), f(\vec{w}, \vec{x}_j) + L\Delta].$$

Combining the above three cases, we have

$$f(\vec{w}, \vec{x}_k) \in [f(\vec{w}, \vec{x}_i) - L\Delta, f(\vec{w}, \vec{x}_j) + L\Delta], \quad k \in \{i, \dots, j\}.$$

Note that $0 < f(\vec{w}, \vec{x}_i) - f(\vec{w}, \vec{x}_j^*) < L\Delta$. Thus, the length of the interval $[f(\vec{w}, \vec{x}_i) - L\Delta, f(\vec{w}, \vec{x}_j) + L\Delta]$ is less than $2L\Delta$, which implies (6.18). We complete the proof. \square

With Lemma 6.2, 6.3, 6.4, 6.5, we are now ready to give the proof of Proposition 3.2.

Proof of Proposition 3.2: We first show that (3.9) holds when h is a piecewise linear convex distortion function and then show that (3.9) holds for any convex distortion function by approximating a convex distortion function by a series of piecewise linear convex distortion functions. To do so, we prove (3.9) in the following four cases.

Case 1. Suppose that $\rho_{h,f,\vec{w}}^\uparrow(\vec{X}) < \infty$ and the distortion function h is a piecewise linear convex function, that is, there exist $0 = a_0 < a_1 < \dots < a_m = 1$, $m \in \mathbb{N}$, such that h is linear on $(a_{k-1}, a_k]$, $k = 1, \dots, m$, and $\ell_1 < \dots < \ell_m$, where ℓ_k is the slope of h on the interval $(a_{k-1}, a_k]$, $k = 1, \dots, m$. For this case, to show (3.9), it suffices to show that for any $\varepsilon > 0$, there exists a random vector $\vec{W} \in \mathcal{F}_{\vec{m},\mathcal{G}}^h$ such that

$$\rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h,f,\vec{w}}^\uparrow(\vec{X}) - \varepsilon C, \quad (6.26)$$

where C is a constant that depends only on h and is independent of ε . In fact, (6.26) implies

$$\sup_{F \in \mathcal{F}_{\vec{m},\mathcal{G}}^h} \rho_h^F(f(\vec{w}, \vec{X})) \geq \rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h,f,\vec{w}}^\uparrow(\vec{X}) - \varepsilon C,$$

which yields (3.9) by letting $\varepsilon \downarrow 0$ and the fact $\rho_{h,f,\vec{w}}^\uparrow(\vec{X}) \geq \sup_{F \in \mathcal{F}_{\vec{m},\mathcal{G}}^h} \rho_h^F(f(\vec{w}, \vec{X}))$.

In the following, we show how to define a random vector $\vec{W} \in \mathcal{F}_{\vec{m},\mathcal{G}}^h$ such that (6.26) holds. First, by the definition (3.3), we know that for any $\varepsilon > 0$, there exists a random vector $\vec{X}_\varepsilon = (X_1, \dots, X_d) \in \mathcal{F}_{\vec{m},\mathcal{G}}$ such that

$$\rho_h(f(\vec{w}, \vec{X}_\varepsilon)) > \rho_{h,f,\vec{w}}^\uparrow(\vec{X}) - \varepsilon. \quad (6.27)$$

Note that when h is a piecewise linear convex function, the set I_2 in Lemma 6.4(a) is empty. Define $\vec{X}^* = (X_1^*, \dots, X_d^*)$ based on \vec{X}_ε by the same form as (6.8) with a partition $B_k = \{a_{k-1} \leq U < a_k\}$, $k = 1, \dots, m$, of Ω , namely, define \vec{X}^* as

$$\vec{X}^* = \sum_{k=1}^m \mathbb{E}[\vec{X}_\varepsilon | B_k] I_{B_k} =: \sum_{k=1}^m \vec{x}_k^* I_{B_k}, \quad (6.28)$$

where $\vec{x}_k^* = \mathbb{E}[\vec{X}_\varepsilon | B_k]$ for $k = 1, \dots, m$; $p_k \triangleq \mathbb{P}(B_k) = a_k - a_{k-1}$ for $k = 1, \dots, m$; U is a random variable uniformly distributed on $(0, 1)$ such that $F_{\vec{w}}^{-1}(U) = f(\vec{w}, \vec{X}_\varepsilon)$ a.s.; and $F_{\vec{w}}$ is the distribution function of $f(\vec{w}, \vec{X}_\varepsilon)$. Then, by Lemma 6.4(b), we have that $\vec{X}^* \in \mathcal{F}_{\vec{m},\mathcal{G}}$ and (6.11) reduces to

$$\rho_h(f(\vec{w}, \vec{X}_\varepsilon)) \leq \int_0^1 v(\alpha) dh(\alpha) = \sum_{i=1}^m f(\vec{w}, \vec{x}_k^*) \ell_k p_k \leq \rho_h(f(\vec{w}, \vec{X}^*)), \quad (6.29)$$

which, together with (6.27), yields

$$\rho_{h,f,\vec{w}}^\uparrow(\vec{X}) - \varepsilon \leq \rho_h(f(\vec{w}, \vec{X}_\varepsilon)) \leq \sum_{k=1}^m \ell_k p_k f(\vec{w}, \vec{x}_k^*) \leq \rho_h(f(\vec{w}, \vec{X}^*)) \leq \rho_{h,f,\vec{w}}^\uparrow(\vec{X}). \quad (6.30)$$

If $m = 1$, that is, the distortion risk measure $\rho_h^F(\cdot) = \mathbb{E}_F[\cdot]$ is expectation, we simply obtain $\vec{W} = \vec{X}^* = \mathbb{E}[\vec{X}_\varepsilon]$. Then $\text{VaR}_\alpha(f(\vec{w}, \vec{X}^*)) = (f(\vec{w}, \mathbb{E}[\vec{X}_\varepsilon]))$ is a constant for any $\alpha \in [0, 1]$, and thus by $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$, we have $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. By (6.30), we have $\rho_h(f(\vec{w}, \vec{X}^*)) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon$, that is, $\vec{W} = \vec{X}^*$ satisfies (6.26) with $C = 1$. Therefore, we get the desired \vec{W} for $m = 1$.

Now, if $m \geq 2$ and $f(\vec{w}, \vec{x}_1^*) \leq f(\vec{w}, \vec{x}_2^*) \leq \dots \leq f(\vec{w}, \vec{x}_m^*)$, define $\vec{W} := \vec{X}^*$. Then we have

$$\text{VaR}_\alpha(f(\vec{w}, \vec{X}^*)) = f(\vec{w}, \vec{x}_k), \quad \alpha \in (a_{k-1}, a_k], \quad k = 1, \dots, m,$$

that is, $\text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ is a constant on each open interval on which $h(\alpha)$ is linear. Then by $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$, we have $\vec{W} = \vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. By (6.30), we have $\rho_h(f(\vec{w}, \vec{X}^*)) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon$, that is, $\vec{W} = \vec{X}^*$ satisfies (6.26) with $C = 1$. Therefore, we get the desired \vec{W} .

Otherwise, if $m \geq 2$ and there exist $1 \leq i < j \leq m$ such that $f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_j^*)$. Then, we define \vec{W} based on \vec{X}^* as follows. If $m = 2$ and $f(\vec{w}, \vec{x}_1^*) > f(\vec{w}, \vec{x}_2^*)$, then $\vec{W} = \mathbb{E}[\vec{X}^*]$ satisfies the requirements. To see it, note that $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$, and then by Lemma 6.3, we have that $\vec{W} = \mathbb{E}[\vec{X}^*] \in \mathcal{F}_{\vec{m}, \mathcal{G}}$. Also, by that $\text{VaR}_\alpha(f(\vec{w}, \vec{W})) = (f(\vec{w}, \mathbb{E}[\vec{X}^*]))$ is a constant for any $\alpha \in [0, 1]$, and thus it is a constant on each open interval on which $h(\alpha)$ is linear. It then follows that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. In addition, note that \vec{X}^* satisfies Assumption 2, and by (6.30), $\Delta := \rho_h(f(\vec{w}, \vec{X}^*)) - \sum_{k=1}^2 \ell_k p_k f(\vec{w}, \vec{x}_k^*) < \varepsilon$. Then by Lemma 6.5, we have

$$|f(\vec{w}, \vec{x}_1^*) - f(\vec{w}, \vec{x}_2^*)| \leq 2L\varepsilon, \quad (6.31)$$

where $L = 1/|\ell_2 - \ell_1|$. Then we have with $L^* = 2\ell_2 L$,

$$\begin{aligned} \rho_h(f(\vec{w}, \vec{W})) &= \rho_h(f(\vec{w}, \mathbb{E}[\vec{X}^*])) = f(\vec{w}, \mathbb{E}[\vec{X}^*]) \geq \mathbb{E}[f(\vec{w}, \vec{X}^*)] \\ &= p_1 f(\vec{w}, \vec{x}_1^*) + p_2 f(\vec{w}, \vec{x}_2^*) \\ &= (\ell_1 p_1 + \ell_2 p_2) (p_1 f(\vec{w}, \vec{x}_1^*) + p_2 f(\vec{w}, \vec{x}_2^*)) \\ &\geq \sum_{j=1}^2 \ell_j p_j (f(\vec{w}, \vec{x}_j^*) - 2L\varepsilon) = \sum_{j=1}^2 \ell_j p_j f(\vec{w}, \vec{x}_j^*) - \max_{1 \leq i \leq 2} \ell_i 2L\varepsilon \\ &\geq \rho_h(f(\vec{w}, \vec{X}_\varepsilon)) - \ell_2 2L\varepsilon > \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - L^* \varepsilon, \end{aligned} \quad (6.32)$$

where the first inequality is due to Jensen's inequality, the fourth equality is due to $\ell_1 p_1 + \ell_2 p_2 = h(1) = 1$, the second inequality follows from (6.31), the fourth inequality follows from (6.29), and the last inequality follows from (6.27). This implies \vec{W} satisfies (6.26) with $C = L^*$.

Furthermore, if $m = 3$ and there exist $1 \leq i < j \leq 3$ such that $f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_j^*)$. we construct \vec{W} by the following steps, in which we consider all possible relationships among $f(\vec{w}, \vec{x}_1^*)$, $f(\vec{w}, \vec{x}_2^*)$, and $f(\vec{w}, \vec{x}_3^*)$.

Step 1. If there does not exist $i > 1$ such that $f(\vec{w}, \vec{x}_i^*) < f(\vec{w}, \vec{x}_1^*)$, we go to Step 3. Otherwise, there exists $i > 1$ such that $f(\vec{w}, \vec{x}_i^*) < f(\vec{w}, \vec{x}_1^*)$, let

$$i_1 = \max\{i : f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_1^*)\}.$$

Note that \vec{X}^* satisfies Assumption 2, and by (6.30), $\Delta := \rho_h(f(\vec{w}, \vec{X}^*)) - \sum_{k=1}^m \ell_k p_k f(\vec{w}, \vec{x}_k^*) < \varepsilon$. Then by Lemma 6.5, we have

$$|f(\vec{w}, \vec{x}_j^*) - f(\vec{w}, \vec{x}_i^*)| \leq 2L\varepsilon \quad \text{for any } 1 \leq i, j \leq i_1, \quad (6.33)$$

where $L = 1/\min_{i < j} |\ell_j - \ell_i|$. We define \vec{W}_1^* as

$$\begin{aligned} \vec{W}_1^* &= \mathbb{E}[\vec{X}^* | U^* < a_{i_1}] I_{\{U < a_{i_1}\}} + \vec{X}^* I_{\{U^* \geq a_{i_1}\}} \\ &=: \vec{y}_{i_1} I_{\{U^* < a_{i_1}\}} + \vec{X}^* I_{\{U^* \geq a_{i_1}\}}, \end{aligned} \quad (6.34)$$

where $U^* \sim U(0, 1)$ such that $\vec{X}^* = \vec{x}_i^*$ in $\{a_{i-1} \leq U^* < a_i\}$, $i = 1, 2, 3$. By Lemma 6.3, we can verify that $\vec{W}_1^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$. Moreover, with $L^* = 2L\ell_3 + 1$, we have

$$\begin{aligned} \rho_h(f(\vec{w}, \vec{W}_1^*)) &\geq \sum_{j=1}^{i_1} \ell_j p_j f(\vec{w}, \vec{y}_{i_1}) + \sum_{j>i_1} \ell_j p_j f(\vec{w}, \vec{x}_j^*) \\ &\geq \sum_{j=1}^{i_1} \ell_j p_j \mathbb{E}[f(\vec{w}, \vec{X}^*) | U < a_{i_1}] + \sum_{j>i_1} \ell_j p_j f(\vec{w}, \vec{x}_j^*) \\ &= \sum_{j=1}^{i_1} \ell_j p_j \left(\frac{\sum_{j=1}^{i_1} p_j f(\vec{w}, \vec{x}_j^*)}{\sum_{j=1}^{i_1} p_j} \right) + \sum_{j>i_1} \ell_j p_j f(\vec{w}, \vec{x}_j^*) \\ &\geq \sum_{j=1}^{i_1} \ell_j p_j (f(\vec{w}, \vec{x}_j^*) - 2L\varepsilon) + \sum_{j>i_1} \ell_j p_j f(\vec{w}, \vec{x}_j^*) \\ &= \sum_{j=1}^3 \ell_j p_j f(\vec{w}, \vec{x}_j^*) - \max_{1 \leq i \leq 3} \ell_i 2L\varepsilon \geq \rho_h(f(\vec{w}, \vec{X}_\varepsilon)) - \max_{1 \leq i \leq 3} \ell_i 2L\varepsilon \\ &> \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - L^* \varepsilon, \end{aligned} \quad (6.35)$$

where the first inequality is due to the Hardy-Littlewood inequality of Lemma 6.2, the second inequality is due to Jensen's inequality, the third inequality follows from (6.33), the fourth inequality follows from (6.29), and the last inequality follows from (6.27).

Step 1(a) If $i_1 = 3$, then by (6.34), we have that $\vec{W}_1^* = \mathbb{E}[\vec{X}^*] = \vec{y}_3$ is a degenerated random vector, and thus, $\text{VaR}_\alpha(f(\vec{w}, \vec{W}_1^*)) = f(\vec{w}, \vec{y}_3)$ for any $\alpha \in (0, 1)$. This implies $\vec{W}_1^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. Hence, by letting $\vec{W} := \vec{W}_1^*$, we have found the desired \vec{W} such that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ and $\rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon C$ with $C = L^*$.

Step 1(b) If $i_1 = 2$, then by (6.34), we have

$$\vec{W}_1^* = \vec{y}_2 I_{\{U < a_2\}} + \vec{X}^* I_{\{U \geq a_2\}} = \vec{y}_2 I_{\{U < a_2\}} + \vec{x}_3^* I_{\{U \geq a_2\}}.$$

If $f(\vec{w}, \vec{y}_2) > f(\vec{w}, \vec{x}_3^*)$, we go to Step 2. Otherwise, if $f(\vec{w}, \vec{y}_2) \leq f(\vec{w}, \vec{x}_3^*)$, then we have

$$\text{VaR}_\alpha(f(\vec{w}, \vec{W}_1^*)) = \begin{cases} f(\vec{w}, \vec{y}_2), & \alpha \in (0, a_2] \\ f(\vec{w}, \vec{x}_3^*), & \alpha \in (a_2, 1), \end{cases}$$

which is constant on each open interval on which $h(\alpha)$ is linear. This implies $\vec{W}_1^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. Hence, define $\vec{W} := \vec{W}_1^*$, then we have found the desired \vec{W} such that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ and $\rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon C$ with $C = L^*$.

Step 2. Now, we have $i_1 = 2$ and $f(\vec{w}, \vec{y}_2) > f(\vec{w}, \vec{x}_3^*)$. In this case, (6.34) reduces to

$$\vec{W}_1^* = \vec{y}_2 I_{\{U < a_2\}} + \vec{x}_3^* I_{\{U \geq a_2\}} =: \sum_{j=1}^3 \vec{w}_j^* I_{\{a_{i-1} < U \leq a_i\}},$$

which satisfies Assumption 6.2, and by (6.35), $\Delta := \rho_h(f(\vec{w}, \vec{W}^*)) - \sum_{k=1}^3 \ell_k p_k f(\vec{w}, \vec{w}_k^*) < L^* \varepsilon$. Then by Lemma 6.5, we have

$$|f(\vec{w}, \vec{y}_2) - f(\vec{w}, \vec{x}_3^*)| \leq 2LL^* \varepsilon. \quad (6.36)$$

We define \vec{W} based on \vec{W}_1^* as

$$\vec{W} = \mathbb{E}[\vec{W}_1^*] =: \vec{w}^*.$$

By the arguments similar to those for Lemma 6.4(b), we can verify that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and thus $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ as \vec{W} is a degenerated random vector. Moreover, we have

$$\begin{aligned} \rho_h(f(\vec{w}, \vec{W})) &= \sum_{j=1}^3 \ell_j p_j f(\vec{w}, \vec{w}_j^*) \geq \sum_{j=1}^3 \ell_j p_j \mathbb{E}[f(\vec{w}, \vec{W}_1^*)] \\ &= \sum_{j=1}^3 \ell_j p_j \left(\sum_{j=1}^3 p_j f(\vec{w}, \vec{w}_j^*) \right) \geq \sum_{j=1}^3 \ell_j p_j (f(\vec{w}, \vec{x}_j^*) - 2LL^* \varepsilon) \\ &= \sum_{j=1}^3 \ell_j p_j f(\vec{w}, \vec{w}_j^*) - 2 \max_{1 \leq i \leq 3} \ell_i LL^* \varepsilon \\ &\geq \rho_h(f(\vec{w}, \vec{X}_\varepsilon)) - 2 \max_{1 \leq i \leq 3} \ell_i LL^* \varepsilon > \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - (L^*)^2 \varepsilon, \end{aligned}$$

where the first inequality is due to Jensen's inequality, the second inequality follows from (6.33), the third inequality follows from (6.29), and the last inequality follows from (6.27). Therefore, we have found the desired \vec{W} such that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ and $\rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon C$ with $C = (L^*)^2$.

Step 3. In this case, we have $f(\vec{w}, \vec{x}_1^*) \leq f(\vec{w}, \vec{x}_i^*)$ for $i = 2, 3$. Then by assumption that there exist $1 \leq i < j \leq m$ such that $f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_j^*)$, we have $f(\vec{w}, \vec{x}_2^*) > f(\vec{w}, \vec{x}_3^*)$. Then we define

$$\vec{W} = \vec{x}_1^* I_{\{U \leq a_1\}} + \mathbb{E}[\vec{X}^* | U > a_1] I_{\{a_1 < U \leq 1\}} =: \vec{x}_1^* I_{\{U \leq a_1\}} + \vec{w}_2^* I_{\{a_1 < U \leq 1\}},$$

where $U \sim U(0, 1)$ such that $\vec{X}^* = \vec{x}_i^*$ in $\{a_{i-1} \leq U < a_i\}$, $i = 1, 2, 3$. By the arguments similar to those for Lemma 6.4(b), we can verify that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}$. Furthermore, by the concavity of $f(\vec{w}, \vec{x})$ in \vec{x} , we have

$$f(\vec{w}, \vec{w}_2^*) \geq \mathbb{E}[f(\vec{w}, \vec{X}^*) | U > a_1] = \frac{p_2 f(\vec{w}, \vec{x}_2^*) + p_3 f(\vec{w}, \vec{x}_3^*)}{p_2 + p_3} > f(\vec{w}, \vec{x}_3^*) \geq f(\vec{w}, \vec{x}_1^*). \quad (6.37)$$

It then follows that

$$\text{VaR}_\alpha(f(\vec{w}, \vec{W})) = \begin{cases} f(\vec{w}, \vec{x}_1^*), & \alpha \in [0, a_1] \\ f(\vec{w}, \vec{w}_2^*), & \alpha \in (a_1, 1], \end{cases}$$

which is constant on each open interval on which $h(\alpha)$ is linear. This implies $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. Moreover, note that \vec{X}^* satisfies Assumption 6.2, and by (6.30),

$$\Delta := \rho_h(f(\vec{w}, \vec{X}^*)) - \sum_{k=1}^m \ell_k p_k f(\vec{w}, \vec{x}_k^*) < \varepsilon.$$

Then by Lemma 6.5, we have

$$f(\vec{w}, \vec{x}_3^*) > f(\vec{w}, \vec{x}_2^*) - 2L\varepsilon. \quad (6.38)$$

Then we have

$$\begin{aligned} \rho_h(f(\vec{w}, \vec{W})) &= \ell_1 p_1 f(\vec{w}, \vec{x}_1^*) + \sum_{j=2}^3 \ell_j p_j f(\vec{w}, \vec{w}_2^*) \\ &\geq \ell_1 p_1 f(\vec{w}, \vec{x}_1^*) + \sum_{j=2}^3 \ell_j p_j \frac{p_2 f(\vec{w}, \vec{x}_2^*) + p_3 f(\vec{w}, \vec{x}_3^*)}{p_2 + p_3} \\ &\geq \ell_1 p_1 f(\vec{w}, \vec{x}_1^*) + \sum_{j=2}^3 \ell_j p_j (f(\vec{w}, \vec{x}_i^*) - 2L\varepsilon) \\ &\geq \sum_{j=1}^3 \ell_j p_j f(\vec{w}, \vec{x}_i^*) - 2 \max_{1 \leq j \leq 3} \ell_j L\varepsilon \\ &\geq \rho_h(f(\vec{w}, \vec{X}_\varepsilon)) - 2\ell_3 L\varepsilon > \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - L^* \varepsilon, \end{aligned}$$

where the first inequality follows from (6.37), the second inequality follows from (6.38), the third inequality follows from (6.29), and the last inequality follows from (6.27). Therefore, we have found the desired \vec{W} such that $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ and $\rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon C$ with $C = L^*$.

Following Step 1 to Step 3 above, we construct the random vector \vec{W} as desired for all possible cases when there exist $1 \leq i < j \leq 3$ such that $f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_j^*)$: Step 1(a) deals with the case that $f(\vec{w}, \vec{x}_3^*) < f(\vec{w}, \vec{x}_1^*)$; Step 1(b) deals with the case of $f(\vec{w}, \vec{x}_2^*) < f(\vec{w}, \vec{x}_1^*) \leq f(\vec{w}, \vec{x}_3^*)$ and $\frac{p_2 f(\vec{w}, \vec{x}_2^*) + p_1 f(\vec{w}, \vec{x}_1^*)}{p_1 + p_2} \leq f(\vec{w}, \vec{x}_3^*)$; Step 1(b) + Step 2 deal with the case of $f(\vec{w}, \vec{x}_2^*) < f(\vec{w}, \vec{x}_1^*) \leq f(\vec{w}, \vec{x}_3^*)$ and $\frac{p_2 f(\vec{w}, \vec{x}_2^*) + p_1 f(\vec{w}, \vec{x}_1^*)}{p_1 + p_2} > f(\vec{w}, \vec{x}_3^*)$; and Step 3 deals with the case that $f(\vec{w}, \vec{x}_1^*) \leq f(\vec{w}, \vec{x}_3^*) < f(\vec{w}, \vec{x}_2^*)$. In addition, it is worth noting that in the case of $m = 3$, the constant C can be taken as $(L^*)^2$, which is a constant independent of ε .

For the general case that $m \geq 3$ and there exist $1 \leq i < j \leq m$ such that $f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_j^*)$, By following the similar steps as those for the case that $m = 3$ and there exist $1 \leq i < j \leq 3$ such

that $f(\vec{w}, \vec{x}_i^*) > f(\vec{w}, \vec{x}_j^*)$, we can show there exists a constant $C > 0$ such that for any $\varepsilon > 0$, there is a random vector $\vec{W} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ satisfies $\rho_h(f(\vec{w}, \vec{W})) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon C$. From this, we have for any $m \geq 1$,

$$\sup_{F \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h} \rho_h^F(f(\vec{w}, \vec{X})) \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon C \quad \text{for any } \varepsilon > 0.$$

Letting $\varepsilon \downarrow 0$, we have (3.9) holds for any piecewise linear convex distortion function h . This completes the proof of Case 1.

Case 2. Suppose that $\rho_{h, f, \vec{w}}^\uparrow(\vec{X}) = \infty$ and h is a piecewise linear convex distortion function. For each $n \in \mathbb{N}$, define a convex set $G_n = G \cap \{\vec{x} : \|\vec{x}\|_2 \leq n\}$, where $\|\vec{x}\|_2 = (\sum_{i=1}^d x_i^2)^{1/2}$. Consider the optimization problem:

$$\sup_F \rho_h^F(f(\vec{w}, \vec{X})) \quad \text{subject to (3.17), (3.17) and } \int_{\{\vec{X} \in G_n\}} dF = 1. \quad (6.39)$$

Denote the worst-case value in the problem (6.39) by $\rho_{h, n}^\uparrow(f(\vec{w}, \vec{X}))$, namely

$$\rho_{h, n}^\uparrow(f(\vec{w}, \vec{X})) = \sup_{F \in \mathcal{F}_n} \rho_h^F(f(\vec{w}, \vec{X})),$$

where

$$\mathcal{F}_n = \left\{ \text{Distributions } F : \mathbb{R}^d \rightarrow [0, 1] \mid F \text{ satisfies (3.17), (3.17), and } \int_{\{\vec{X} \in G_n\}} dF = 1 \right\}.$$

Note that for any random vector \vec{X}_F with distribution F satisfying (3.17)-(3.17) or for any $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}$, we have

$$\rho_h(f(\vec{w}, \vec{X}^*)) \geq \rho_h(f(\vec{w}, \vec{X}_F)),$$

where $\vec{X}^* = (X_1^*, \dots, X_d^*)$ is defined by (6.28). Note that the number of the possible outcomes of \vec{X}^* are at most m . There exists $n_0 \in \mathbb{N}$ such that $\vec{X}^* \in G_{n_0}$ a.s. It has been already verified in Case 1 that the distribution of \vec{X}^* satisfies (3.17) and (3.17). Thus, we have

$$\rho_{h, n_0}^\uparrow(f(\vec{w}, \vec{X})) \geq \rho_h(f(\vec{w}, \vec{X}^*)) \geq \rho_h(f(\vec{w}, \vec{X}_F)).$$

Since $\rho_{h, f, \vec{w}}^\uparrow(\vec{X}) = \infty$, there exists a sequence of random vectors $\{\vec{X}_{F_n}\}_{n \in \mathbb{N}}$ with distribution F_n satisfying (3.17)-(3.17) such that

$$\rho_h(f(\vec{w}, \vec{X}_{F_n})) > n, \quad n \in \mathbb{N}.$$

Hence, by the above arguments, for each \vec{X}_{F_n} , there exists $k_n \in \mathbb{N}$ such that

$$\rho_{h, k_n}^\uparrow(f(\vec{w}, \vec{X})) \geq \rho_h(f(\vec{w}, \vec{X}_{F_n})) > n. \quad (6.40)$$

Note that G_{k_n} is a compact set and $f(\vec{w}, \vec{x})$ has a finite maximum value in $\vec{x} \in G_{k_n}$. We have $\rho_{h, k_n}^\uparrow(f(\vec{w}, \vec{X})) < \infty$ for each $n \in \mathbb{N}$. Then by the proof for Case 1, we have there exists \vec{X}_n

satisfying its distribution is in \mathcal{F}_n and $\text{VaR}(f(\vec{w}, \vec{X}_n))$ is a constant one each interval when h is a linear function, which implies that its distribution is in $\mathcal{F}_{\vec{m}, \mathcal{G}}^h$, and

$$\rho_h(f(\vec{w}, \vec{X}_n)) > \rho_{h, k_n}^\uparrow(f(\vec{w}, \vec{X})) - 1,$$

which combined with (6.40) implies $\lim_{n \rightarrow \infty} \rho_h(f(\vec{w}, \vec{X}_n)) = \infty$. It follows that (3.9).

Case 3. Suppose that the distortion function h is a general convex function and $\rho_{h, f, \vec{w}}^\uparrow(\vec{X}) < \infty$. In this case, there exists a sequence of convex piecewise linear distortion functions $\{h_n\}_{n \in \mathbb{N}}$ such that $h_n \geq h$, $n \in \mathbb{N}$ and $h_n \downarrow h$ as $n \rightarrow \infty$. Without loss of generality, assume that all the break points⁹ of h_n are in the set of $\{a_k, b_k, k \in I\}$ and the break points of h_n are in the set of all the break points of h_{n+1} for each $n \in \mathbb{N}$. Note that for each distribution $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and $\vec{X} \sim F$, by monotone convergence theorem, we have $\lim_{n \rightarrow \infty} \rho_{h_n}(f(\vec{w}, \vec{X})) = \rho_h(f(\vec{w}, \vec{X}))$. Thus, for any $\varepsilon > 0$, there exists $F_\varepsilon \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and $\vec{X}_\varepsilon \sim F_\varepsilon$ such that $\rho_h(f(\vec{w}, \vec{X}_\varepsilon)) > \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon/2$, and then for this distribution F_ε , there exists $n_0 \in \mathbb{N}$,

$$\rho_{h_{n_0}}(f(\vec{w}, \vec{X}_\varepsilon)) > \rho_h(f(\vec{w}, \vec{X}_\varepsilon)) - \frac{\varepsilon}{2} > \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon \geq \rho_{h_{n_0}, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon, \quad (6.41)$$

where the third inequality follows from $\rho_h \geq \rho_{h_{n_0}}$ as $h \leq h_{n_0}$. Since h_{n_0} is a piecewise linear function, by the proof for Case 1, there exists a distribution $F^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and $\vec{X}^* \sim F^*$ such that $\text{VaR}_\alpha(f(\vec{w}, X^*))$ is a constant on each interval when h_{n_0} is a linear function, and

$$\rho_{h_{n_0}}(f(\vec{w}, \vec{X}^*)) > \rho_{h_{n_0}}(f(\vec{w}, \vec{X}_\varepsilon)) - C\varepsilon,$$

where C is a constant independent of ε . Note that $\rho_h(f(\vec{w}, \vec{X}^*)) \geq \rho_{h_{n_0}}(f(\vec{w}, \vec{X}^*))$ as $h \leq h_{n_0}$, and thus,

$$\rho_h(f(\vec{w}, \vec{X}^*)) \geq \rho_{h_{n_0}}(f(\vec{w}, \vec{X})) - C\varepsilon \geq \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) - (C+1)\varepsilon,$$

where the last inequality follows from (6.41). Also, note that all the break points of h_n are subset of $\{a_k, b_k, k \in I\}$ and thus $\text{VaR}_\alpha(f(\vec{w}, X_n^*))$ is a constant on each interval of $\{(a_k, b_k], k \in I_1\}$, that is, the intervals where h is linear.

Case 4. Suppose that the distortion function h is a general convex function and $\rho_{h, f, \vec{w}}^\uparrow(\vec{X}) = \infty$. Let h_n , $n \in \mathbb{N}$ be a sequence of piecewise linear convex distortion functions defined in Case 3. Note that for each distribution $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and $\vec{X} \sim F$, by monotone convergence theorem, we have $\lim_{n \rightarrow \infty} \rho_{h_n}(f(\vec{w}, \vec{X})) = \rho_h(f(\vec{w}, \vec{X}))$. Then for each $n \in \mathbb{N}$, there exist $F_n \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ such that

$$\rho_h(f(\vec{w}, \vec{X}_n)) > n + 1 \quad \text{with } \vec{X}_n \sim F_n.$$

⁹For a piecewise linear function h , a point x is called a break point if the slopes of h at the left of x and right of x are different.

For each chosen X_n , there exists $k_n \in \mathbb{N}$ such that

$$\rho_{h_{k_n}}(f(\vec{w}, \vec{X}_n)) > \rho_h(f(\vec{w}, \vec{X}_n)) - 1 > n.$$

By the above proof of Case 1 and 2 for piecewise linear distortion functions, there exists $F_n^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^{h_{k_n}} \subset \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ such that

$$\rho_h(f(\vec{w}, \vec{X}_n^*)) \geq \rho_{h_{k_n}}(f(\vec{w}, \vec{X}_n^*)) \geq \rho_{h_{k_n}}(f(\vec{w}, \vec{X}_n)) - 1 \geq n - 1,$$

where $\vec{X}_n^* \sim F_n^*$. This implies $\lim_{n \rightarrow \infty} \rho_h(f(\vec{w}, \vec{X}_n^*)) = \infty$, which means (3.9) holds. Hence, combining Cases 1 - 4, we have (3.9) holds.

Next, we show that the worst case distribution can be obtained in the set $\mathcal{F}_{\vec{m}, \mathcal{G}}^h$ when the worst case value $\rho_{h, f, \vec{w}}^\uparrow(\vec{X})$ is attainable. Note that if $\rho_{h, f, \vec{w}}^\uparrow(\vec{X})$ is available, then there exists a distribution $F_{\text{wc}} \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ and a random vector $\vec{X}_{\text{wc}} = (X_1, \dots, X_d) \sim F_{\text{wc}}$ such that

$$\rho_h(f(\vec{w}, \vec{X}_{\text{wc}})) = \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) < \infty. \quad (6.42)$$

We define $\vec{X}^* = (X_1^*, \dots, X_d^*)$ as in Lemma 6.4(b) by replacing \vec{X} by \vec{X}_{wc} . Then by Lemma 6.4(b), $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$ such that $\rho_h(f(\vec{w}, \vec{X}_{\text{wc}})) \leq \rho_h(f(\vec{w}, \vec{X}^*))$, which is no larger than the worst-case value $\rho_{h, f, \vec{w}}^\uparrow(\vec{X})$. Therefore, by (6.42), we have

$$\rho_h(f(\vec{w}, \vec{X}_{\text{wc}})) = \rho_{h, f, \vec{w}}^\uparrow(\vec{X}) = \rho_h(f(\vec{w}, \vec{X}^*)). \quad (6.43)$$

Employing the notation of the proof of Lemma 6.4(b) and noting that \vec{X}^* is defined based on \vec{X}_{wc} , The above equation implies the inequality of (6.15) is in fact an equality. That is,

$$\int_0^1 \text{VaR}_\alpha(f(\vec{w}, \vec{X}^*)) h'(\alpha) d\alpha = \int_0^1 v(\alpha) h'(\alpha) d\alpha, \quad (6.44)$$

with v defined by (6.12). Then by Lemma 6.4(c), we have $\text{VaR}_\alpha(f(\vec{w}, \vec{X}^*)) = v(\alpha)$ for any $\alpha \in [0, 1]$, that is, $\text{VaR}_\alpha(f(\vec{w}, \vec{X}^*))$ is a constant when h is linear. This implies $\vec{X}^* \in \mathcal{F}_{\vec{m}, \mathcal{G}}^h$. Then by (6.43), we complete the proof. \square

6.4 Proof of Theorem 3.4

We have given the proof of Theorem 3.4 in the main body for the case where the distortion function h is left-continuous. To prove Theorem 3.4 for the other cases, we need the following result which is a corollary of Lemma 2.1.

Corollary 6.6. *Let h be a distortion function with the set of discontinuities $\{\alpha_i, i \in M\}$, and let $h_+(x) \triangleq h(x+)$ and $h_-(x) \triangleq h(x-)$ be the right-continuous and left-continuous copies of h , respectively. If $h(x) = h_+(x)$ for $x \in \{\alpha_i, i \in I_1\}$ and $h(x) = h_-(x)$ for $x \in \{\alpha_i, i \in I_2\}$ where*

$I_1 \cup I_2 = M$, then $\rho_h(X)$ has the following Lebesgue-Stieltjes integral representation:

$$\begin{aligned}\rho_h(X) &= \int_0^1 \text{VaR}_\alpha(X) dh(\alpha) + \sum_{i \in I_2} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) \Delta_h(\alpha_i) \\ &= \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha) - \sum_{i \in I_1} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) \Delta_h(\alpha_i).\end{aligned}$$

where VaR_α and VaR_α^+ are the left-continuous and right-continuous Value-at-Risk, respectively, and $\Delta_h(\alpha_i) = h_+(\alpha_i) - h_-(\alpha_i)$, $i \in M$.

Proof. The proof is similar to that of (2.8) of Lemma 2.1. We only give the proof of the first equality in the representation of $\rho_h(X)$. Let $x_i = \text{VaR}_{\alpha_i}(X)$ and $y_i = \text{VaR}_{\alpha_i}^+(X)$, $i \in M$. Then $h(\alpha) = h_+(\alpha)$ for $\alpha \in (0, 1) \setminus \{\alpha_i, i \in I_2\}$, that is, $h(F(x)) = h_+(F(x))$ for $x \in \mathbb{R} \setminus \cup_{i \in I_2} [x_i, y_i]$ as $F(x) = \alpha_i$ for $x \in [x_i, y_i]$, $i \in I_2$. Then by (2.1), we have

$$\begin{aligned}\rho_h(X) - \rho_{h_+}(X) &= \sum_{i \in I_2} \int_{x_i}^{y_i} (h_+(F(x)) - h_-(F(x))) dx \\ &= \sum_{i \in I_2} (h_+(\alpha_i) - h_-(\alpha_i)) (y_i - x_i) \\ &= \sum_{i \in I_2} (\text{VaR}_{\alpha_i}^+(X) - \text{VaR}_{\alpha_i}(X)) (h_+(\alpha_i) - h_-(\alpha_i)) \\ &= \sum_{i \in I_2} \int_{\{\alpha_i\}} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) dh(\alpha).\end{aligned}$$

Then by (2.7), we have

$$\begin{aligned}\rho_h(X) &= \rho_{h_+}(X) + \sum_{i \in I_2} \int_{\{\alpha_i\}} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) dh(\alpha) \\ &= \int_0^1 \text{VaR}_\alpha(X) dh(\alpha) + \sum_{i \in I_2} (\text{VaR}_\alpha^+(X) - \text{VaR}_\alpha(X)) \Delta_h(\alpha_i).\end{aligned}$$

Thus, we complete the proof. \square

Proof of Theorem 3.4 (continued): In the following, we show that (3.10) holds for h_+ which is defined by $h_+(x) = \lim_{y \downarrow x} h(y)$. It suffices to show

$$\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) = \rho_{h_-, f, \vec{w}}^\uparrow(\vec{X}), \quad (6.45)$$

where $h_-(x) = \lim_{y \uparrow x} h(y)$, as we have proved that (3.11) for any left-continuous h in the proof of Theorem 3.4 in the main body. Since h is a monotone function, h has at most countable discontinuous points. We show (6.45) for the following two cases.

- (i) Suppose that h has finite discontinuous points, denoted by $a_1 < \dots < a_s$. Since h is continuous at 0 and 1, we know $0 < a_1 < \dots < a_s < 1$. Denote $\delta_1 := \min\{\min_{2 \leq i \leq s}(a_i - a_{i-1}), a_1, 1 - a_s\} > 0$. For any $n \in \mathbb{N}$, let $0 < \delta < \delta_1$ and

$$B_k^n = \left\{ a_k - \frac{1}{n} < U < a_k + \delta \right\}, \quad k = 1, \dots, s, \quad B_0^n = \Omega - \cup_{k=1}^s B_k^n. \quad (6.46)$$

Then for n large enough, $\{B_k^n, k = 0, \dots, s\}$ are disjoint and thus constructs a partition of Ω . Hence, there exist $\varepsilon \in (0, 1)$ and $n_0 \in \mathbb{N}$ such that for $n \geq n_0$

$$\varepsilon < \min_{k=1, \dots, s} a_k - \frac{1}{n} < \max_{k=1, \dots, s} a_k + \delta < 1 - \varepsilon, \quad (6.47)$$

and $\{B_k^n, k = 0, \dots, s\}$ constructs a partition of Ω . For any $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and $\vec{X} \sim F$, we define

$$\vec{X}_n = \sum_{k=1}^s \mathbb{E}[\vec{X} | B_k^n] I_{B_k^n} + \vec{X} I_{B_0^n}, \quad n \geq n_0. \quad (6.48)$$

By concavity of $f(\vec{w}, \cdot)$, we have

$$\begin{aligned} f(\vec{w}, \vec{X}_n) &= \sum_{k \in I_1} f(\vec{w}, \mathbb{E}[\vec{X} | B_k^n]) I_{B_k^n} + f(\vec{w}, \vec{X}) I_{B_0^n} \\ &\geq \sum_{k \in I_1} \mathbb{E}[f(\vec{w}, \vec{X}) | B_k^n] I_{B_k^n} + f(\vec{w}, \vec{X}) I_{B_0^n} =: f_{B_n}(\vec{X}). \end{aligned} \quad (6.49)$$

Let $U \sim U[0, 1]$ be comonotonic with $f(\vec{w}, \vec{X})$. Then for $n \geq n_0$, by (6.47), we have $f_{B_n}(\vec{X}) = f(\vec{w}, \vec{X})$ a.s. on $\{U \in [0, \varepsilon] \cup [1 - \varepsilon, 1]\}$ and $f_{B_n}(\vec{X})$ takes values that equal to the conditional expectations of $f(\vec{w}, \vec{X})$ on the set $\{\varepsilon < U < 1 - \varepsilon\}$. Thus, for $n \geq n_0$

$$|f_{B_n}(\vec{X}) - f(\vec{w}, \vec{X})| < \text{VaR}_{1-\varepsilon}(f(\vec{w}, \vec{X})) - \text{VaR}_{\varepsilon}(f(\vec{w}, \vec{X})) < \infty, \quad \text{a.s.} \quad (6.50)$$

Then by dominated convergence theorem and noting that

$$\lim_{n \rightarrow \infty} \text{VaR}_{\alpha}(f_{B_n}(\vec{X})) = \begin{cases} \text{VaR}_{\alpha}^+(f(\vec{w}, \vec{X})), & \alpha = a_k, \quad k = 1, \dots, s, \\ \text{VaR}_{\alpha}(f(\vec{w}, \vec{X})), & \text{else,} \end{cases} \quad (6.51)$$

we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \rho_{h+}(f_{B_n}(\vec{X})) &= \lim_{n \rightarrow \infty} \int_0^1 \text{VaR}_{\alpha}(f_{B_n}(\vec{X})) dh_+(\alpha) \\ &= \sum_{k=1}^s \text{VaR}_{\alpha}^+(f(\vec{w}, \vec{X})) \Delta_h(\alpha_k) + \int_C \text{VaR}_{\alpha}(f(\vec{w}, \vec{X})) dh_-(\alpha) \\ &= \int_0^1 \text{VaR}_{\alpha}^+(f(\vec{w}, \vec{X})) dh(\alpha) = \rho_{h-}(f(\vec{w}, \vec{X})), \end{aligned}$$

where $\Delta_h(a_k) := h_+(a_k) - h_-(a_k)$, $k = 1, \dots, s$, $C := [0, 1] \setminus \{a_k, k = 1, \dots, s\}$ and the second equality is due to h is continuous on C . By Lemma 6.3, we have that the distribution of \vec{X}_n is in $\mathcal{F}_{\vec{m}, \mathcal{G}}$ for $n \geq n_0$. Thus, we have

$$\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) \geq \liminf_{n \rightarrow \infty} \rho_{h_+}(f(\vec{w}, \vec{X}_n)) \geq \lim_{n \rightarrow \infty} \rho_{h_+}(f_{B_n}(\vec{X})) = \rho_{h_-}(f(\vec{w}, \vec{X})).$$

From the arbitrariness of the distribution of \vec{X} in $\mathcal{F}_{\vec{m}, \mathcal{G}}$, we have $\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) \geq \rho_{h_-, f, \vec{w}}^\uparrow(\vec{X})$, that is, (6.45) holds.

- (ii) Suppose that h has countable discontinuous points, denoted by $\{a_k, k \in \mathbb{N}\}$. That is, $h_-(a_k) < h_+(a_k)$ for $k \in \mathbb{N}$, and $h_-(x) = h_+(x)$ for $x \in C := [0, 1] \setminus \{a_k, k \in \mathbb{N}\}$. Define

$$h_n(x) = \begin{cases} h_-(x), & x = a_1, \dots, a_n, \\ h_+(x), & \text{else,} \end{cases} \quad n \in \mathbb{N}.$$

Then we have $h_+ \geq h_n \geq h_-$ and h_n decreasingly converges to h_- as $n \rightarrow \infty$. By Corollary 6.6, we have for any random variable X ,

$$\begin{aligned} \rho_{h_n}(X) &= \sum_{k=1}^n \text{VaR}_{a_k}^+(X) \Delta_h(a_k) + \int_{C_n} \text{VaR}_\alpha(X) dh(\alpha) \\ &= \sum_{k=1}^n (\text{VaR}_{a_k}^+(X) - \text{VaR}_{a_k}(X)) \Delta_h(a_k) + \int_0^1 \text{VaR}_\alpha(X) dh(\alpha), \quad n \in \mathbb{N} \end{aligned} \quad (6.52)$$

with $C_n := [0, 1] \setminus \{a_1, \dots, a_n\}$, $n \in \mathbb{N}$. Then by monotone convergence theorem,

$$\begin{aligned} \lim_{n \rightarrow \infty} \rho_{h_n}(X) &= \sum_{k \in \mathbb{N}} (\text{VaR}_{a_k}^+(X) - \text{VaR}_{a_k}(X)) \Delta_h(a_k) + \int_0^1 \text{VaR}_\alpha(X) dh(\alpha) \\ &= \int_0^1 \text{VaR}_\alpha^+(X) dh(\alpha) = \rho_{h_-}(X). \end{aligned}$$

For any given $F \in \mathcal{F}_{\vec{m}, \mathcal{G}}$ and $\vec{X} \sim F$, if $\rho_{h_-}(f(\vec{w}, \vec{X}))$ is finite, then for any $\varepsilon^* > 0$, there exists $s \in \mathbb{N}$ such that

$$\rho_{h_s}(f(\vec{w}, \vec{X})) > \rho_{h_-}(f(\vec{w}, \vec{X})) - \varepsilon^*. \quad (6.53)$$

Otherwise, if $\rho_{h_-}(f(\vec{w}, \vec{X})) = \infty$, then for any $M > 0$, there exists $s \in \mathbb{N}$ such that

$$\rho_{h_s}(f(\vec{w}, \vec{X})) > M. \quad (6.54)$$

Similar to Case (i), we have $0 < a_1 < \dots < a_s < 1$, and we then follow all the arguments from (6.46) to (6.51) in Case (i) which are all true here. For the $f_{B_n}(\vec{X})$, $n \geq n_0$, defined by (6.49),

by dominated convergence theorem, we have

$$\begin{aligned}
\lim_{n \rightarrow \infty} \rho_{h_+}(f_{B_n}(\vec{X})) &= \lim_{n \rightarrow \infty} \int_0^1 \text{VaR}_\alpha(f_{B_n}(\vec{X})) dh_+(\alpha) \\
&= \sum_{k=1}^s \text{VaR}_\alpha^+(f(\vec{w}, \vec{X})) \Delta_h(\alpha_k) + \int_{C_s} \text{VaR}_\alpha(f(\vec{w}, \vec{X})) dh_+(\alpha) \\
&= \rho_{h_s}(f(\vec{w}, \vec{X})),
\end{aligned}$$

where the last equality follows from (6.52). Since the distribution of \vec{X}_n defined by (6.48) is in $\mathcal{F}_{\vec{m}, \mathcal{G}}$ and $f(\vec{w}, \vec{X}_n) \geq f_{B_n}(\vec{X})$ a.s., this implies

$$\begin{aligned}
\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) &\geq \limsup_{n \rightarrow \infty} \rho_{h_+}(f(\vec{w}, \vec{X}_n)) \geq \lim_{n \rightarrow \infty} \rho_{h_+}(f_{B_n}(\vec{X})) \\
&= \rho_{h_s}(f(\vec{w}, \vec{X})) \\
&> \begin{cases} \rho_{h_-}(f(\vec{w}, \vec{X})) - \varepsilon^*, & \text{if } \rho_{h_-}(f(\vec{w}, \vec{X})) < \infty, \\ M, & \text{if } \rho_{h_-}(f(\vec{w}, \vec{X})) = \infty, \end{cases}
\end{aligned}$$

where the last inequality follows from (6.53) and (6.54).

If $\rho_{h_-}(f(\vec{w}, \vec{X})) < \infty$, then from the arbitrariness of the distribution of \vec{X} in $\mathcal{F}_{\vec{m}, \mathcal{G}}$, we have $\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) \geq \rho_{h_-, f, \vec{w}}^\uparrow(\vec{X}) - \varepsilon^*$ for any $\varepsilon^* > 0$. Letting $\varepsilon^* \downarrow 0$ yields $\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) \geq \rho_{h_-, f, \vec{w}}^\uparrow(\vec{X})$. That is, (6.45) holds for $\rho_{h_-}(f(\vec{w}, \vec{X})) < \infty$.

If $\rho_{h_-}(f(\vec{w}, \vec{X})) = \infty$, then by the arbitrariness of M , we have $\rho_{h_+, f, \vec{w}}^\uparrow(\vec{X}) = \infty$, that is, (6.45) holds.

Combining the two cases, we complete the proof of Theorem 3.4. \square

6.5 Lemma 6.7 for the proof of Proposition 3.6

The proof of Proposition 3.6, which is another main result in the paper, is given in the main text. In the proof, the following Lemma 6.7 is employed.

Lemma 6.7. *Let h be a distortion function satisfying $h(x) = 0$, $x \in (0, \varepsilon)$ for some $\varepsilon > 0$. Then for two random variables X and Y such that $[X|U > \varepsilon] \leq [Y|V > \varepsilon]$ a.s., where $U, V \sim \text{U}[0, 1]$ are comonotonic with X and Y , respectively, we have*

$$\rho_h(X) \leq \rho_h(Y). \quad (6.55)$$

If, moreover, h is a right-continuous distortion function, then

$$\rho_h(X) = \int_{[\varepsilon, 1]} \text{VaR}_\alpha(X) dh(\alpha) = \text{VaR}_\varepsilon(X) h(\varepsilon) + \int_{(\varepsilon, 1]} \text{VaR}_\alpha(X) dh(\alpha). \quad (6.56)$$

Proof. Note that (6.56) is an immediate consequence of Lemma 2.1 (iii). We only need to show (6.55). Without loss of generality, assume that $x_0 := \text{VaR}_\varepsilon^+(X) > 0$ and $y_0 := \text{VaR}_\varepsilon^+(Y) > 0$. Otherwise let $X^* = X - \min\{x_0, y_0\} + \varepsilon$ and $Y^* = Y - \min\{x_0, y_0\} + \varepsilon$. Denote by F and G the distribution functions of X and Y , respectively. We first assert that

$$(-\infty, \text{VaR}_\varepsilon^+(X)) \subseteq \{x \in \mathbb{R} : F(x) \leq \varepsilon\} \subseteq (-\infty, \text{VaR}_\varepsilon^+(X)]. \quad (6.57)$$

To show it, note that

$$\text{VaR}_\varepsilon^+(X) = \inf\{x \in \mathbb{R} : F(x) > \varepsilon\} = \sup\{x \in \mathbb{R} : F(x) \leq \varepsilon\}. \quad (6.58)$$

Then from the first equality of (6.58), we have for any $x > \text{VaR}_\varepsilon^+(X)$, it holds that $F(x) > \varepsilon$, and thus, $(\text{VaR}_\varepsilon^+(X), \infty) \subseteq \{x \in \mathbb{R} : F(x) > \varepsilon\}$. That is, the second implication of (6.57) holds. Similarly, from the second equality of (6.58), we have for any $x < \text{VaR}_\varepsilon^+(X)$, $F(x) \leq \varepsilon$. This means the first implication of (6.57) holds. Then we have (6.57) holds.

Next, from $[X|U > \varepsilon] \leq [Y|U > \varepsilon]$, we can show

$$x_0 \leq y_0. \quad (6.59)$$

To see it, for any $x \in \mathbb{R}$, we have

$$\mathbb{P}(X > x, U > \varepsilon) \leq \mathbb{P}(Y > x, U > \varepsilon),$$

and then by the comonotonicity of U and X, Y , the above inequality is equivalent to

$$\min\{1 - F(x), 1 - \varepsilon\} \leq \min\{1 - G(x), 1 - \varepsilon\}. \quad (6.60)$$

On the other hand, for any $x > y_0$, by the definition of $\text{VaR}_\varepsilon^+(Y)$, we have $G(x) > \varepsilon$, that is, $1 - G(x) < 1 - \varepsilon$. Then by (6.60),

$$1 - F(x) \leq 1 - G(x) < 1 - \varepsilon. \quad (6.61)$$

This combined with (6.58) implies $x_0 \leq x$. Thus, we have (6.59) holds.

By (6.57), we have $h(F(x)) = 0$ for $x < x_0$, and thus,

$$\begin{aligned} \rho_h(X) &= \int_0^\infty 1 - h(F(x))dx - \int_{-\infty}^0 h(F(x))dx \\ &= x_0 + \int_{x_0}^\infty 1 - h(F(x))dx. \end{aligned}$$

Similarly, we have

$$\begin{aligned} \rho_h(Y) &= y_0 + \int_{y_0}^\infty 1 - h(G(x))dx \\ &= x_0 + \int_{x_0}^{y_0} 1dx + \int_{y_0}^\infty 1 - h(G(x))dx. \end{aligned}$$

By $x_0 \leq y_0$ and (6.61) holds for $x > y_0$, we have $\rho_h(X) \leq \rho_h(Y)$. This completes the proof. \square

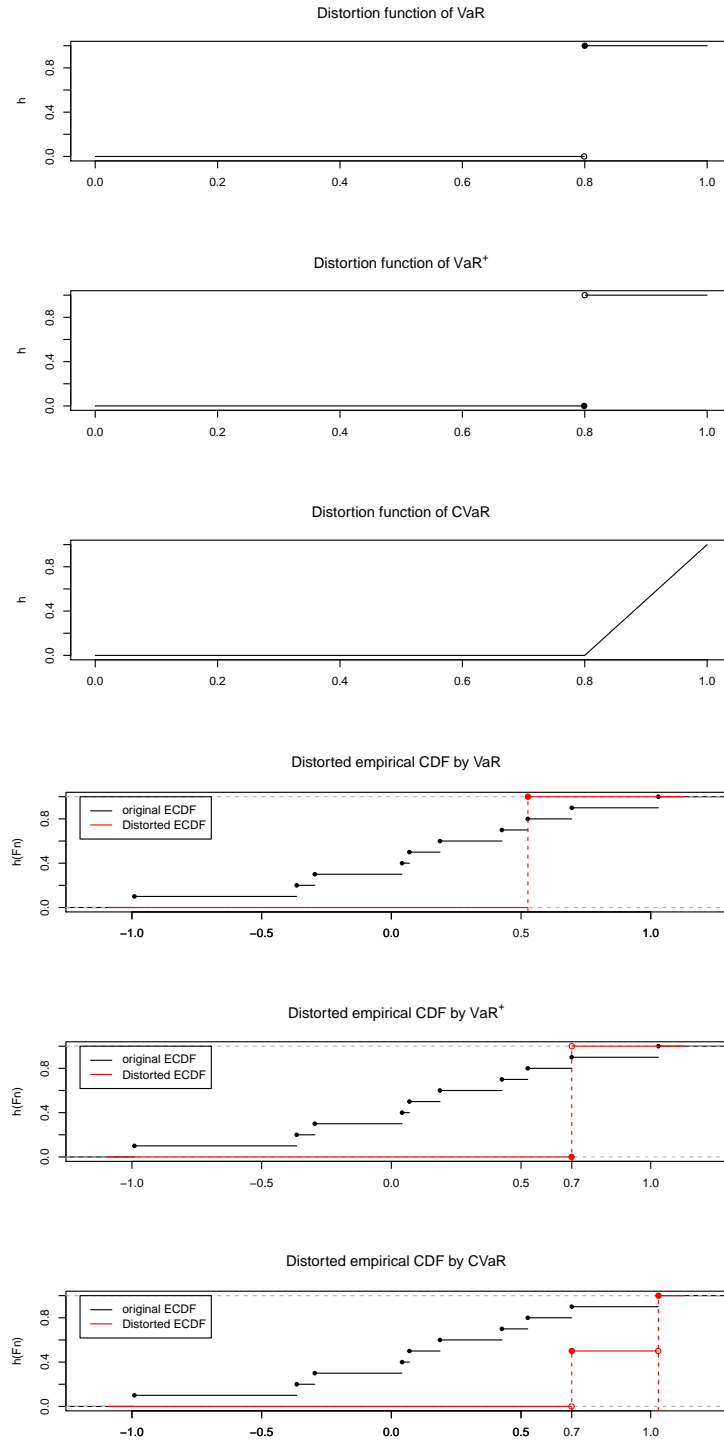


Figure 1: Distortion functions and distorted empirical CDFs by VaR, VaR^+ , CVaR

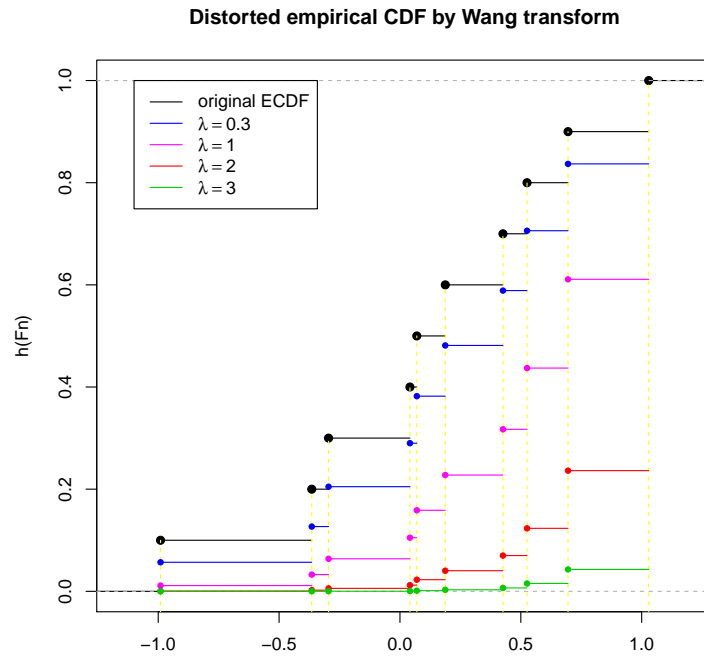
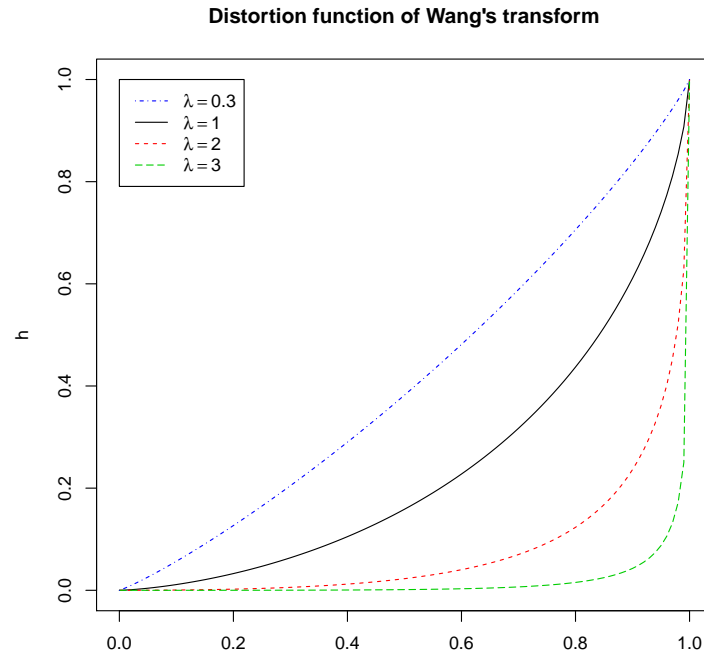


Figure 2: Distortion functions and distorted empirical CDFs by Wang's Transform

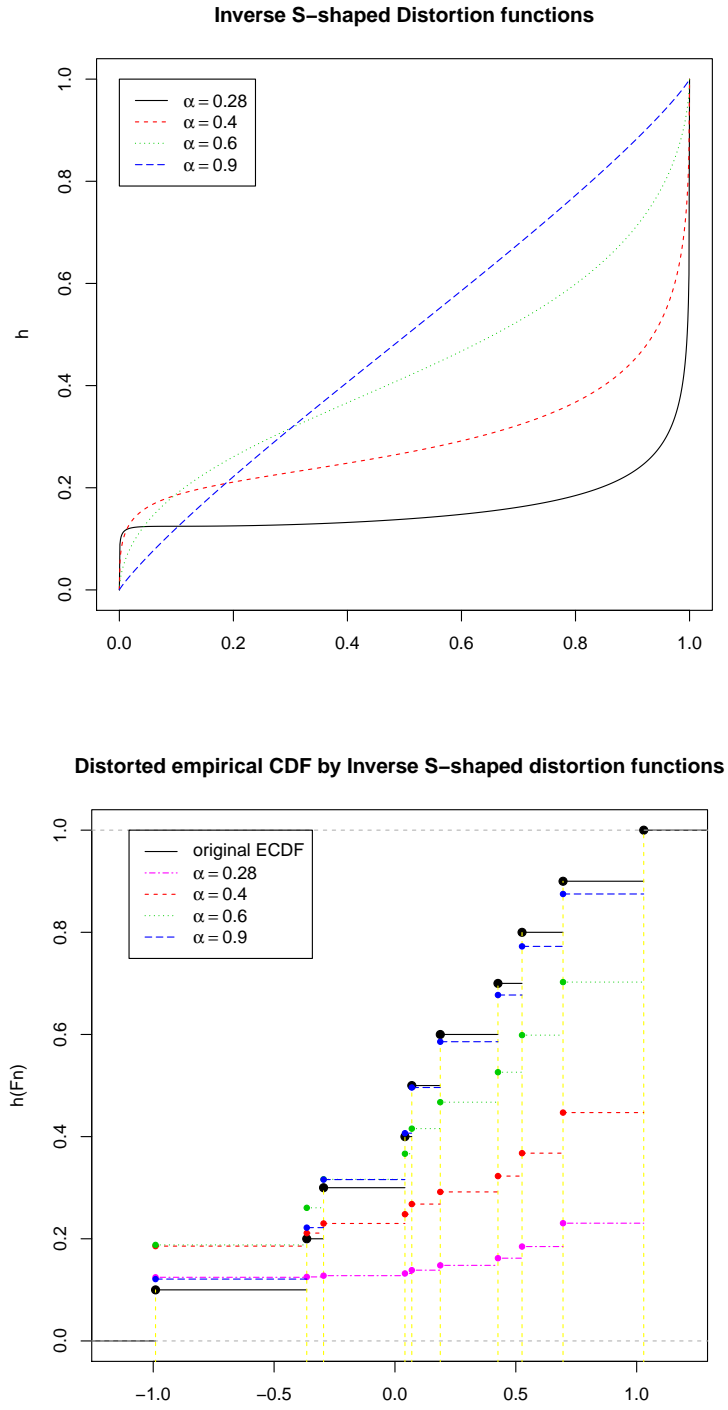


Figure 3: Distortion functions and distorted empirical CDFs by inverse S-shaped distortion functions (2.10)

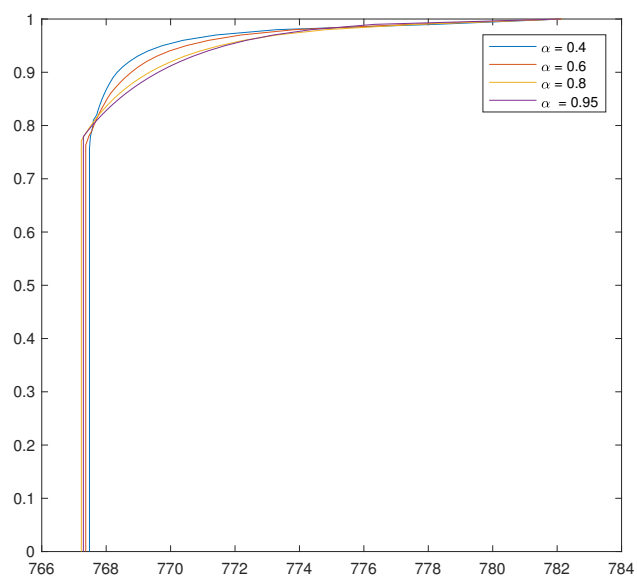


Figure 6: CDFs of worst-case distributions for various distortion parameters α