Adv. Appl. Probab. 1–20 (2024)
*doi:*10.1017/apr.2024.10

AN EXTREME WORST-CASE RISK MEASURE BY EXPECTILE

YANLIN HU,*
YU CHEN ,* ** AND
TIANTIAN MAO,* *** *University of Science and Technology of China*

Abstract

Expectiles have received increasing attention as a risk measure in risk management because of their coherency and elicitability at the level $\alpha \geq 1/2$. With a view to practical risk assessments, this paper delves into the worst-case expectile, where only partial information on the underlying distribution is available and there is no closed-form representation. We explore the asymptotic behavior of the worst-case expectile on two specified ambiguity sets: one is through the Wasserstein distance from a reference distribution and transforms this problem into a convex optimization problem via the well-known Kusuoka representation, and the other is induced by higher moment constraints. We obtain precise results in some special cases; nevertheless, there are no unified closed-form solutions. We aim to fully characterize the extreme behaviors; that is, we pursue an approximate solution as the level α tends to 1, which is aesthetically pleasing. As an application of our technique, we investigate the ambiguity set induced by higher moment conditions. Finally, we compare our worst-case expectile approach with a more conservative method based on stochastic order, which is referred to as 'model aggregation'.

Keywords: Expectile; extreme value theory; ambiguity; Wasserstein distance; moment constraints

2020 Mathematics Subject Classification: Primary 60G70
Secondary 91B05

1. Introduction

In recent years there has been increasing interest in using expectiles as an alternative risk measure to value at risk (VaR) in risk theory. Expectiles, a new least-squares analogue of quantiles introduced by [21], have recently become very popular in applications such as quantitative finance and risk management; see [4] and [32]. Indeed, the papers [3] and [32] have shown that expectiles with $\alpha \geq 1/2$ are the only risk measures that are coherent and elicitable. For a given level $\alpha \in (0, 1)$ and an integrable random variable Y, the αth expectile is defined as the minimizer of an asymmetrically weighted mean-squared deviation,

$$e_\alpha(Y) = \operatorname*{argmin}_\theta \mathbb{E}\left[\eta_\alpha(Y-\theta) - \eta_\alpha(Y)\right], \tag{1.1}$$

Received 9 December 2022; accepted 16 January 2024.
* Postal address: Department of Statistics and Finance, IIF, School of Management, University of Science and Technology of China, Hefei 230026, Anhui, P. R. China.
** Email address: cyu@ustc.edu.cn
*** Email address: tmao@ustc.edu.cn

where $\eta_\alpha(y) = |\alpha - \mathbb{1}_{\{y \le 0\}}|y^2$ is the expectile check function. Subtracting $\eta_\alpha(Y)$ in the expectation makes the integrand well-defined and finite without the need to assume that $\mathbb{E}(Y^2) < \infty$, while for Y we only require a finite first-order moment.

Expectiles have emerged as an important risk measure, especially in the study of extreme risks characterized by their low-probability, high-consequence nature. A paramount challenge in this area is data uncertainty, as the loss distribution is often only partially known or ambiguous. Typically, we only have access to limited information about the underlying distribution. This leads to the pivotal question in risk management: how can we robustly measure risk amidst such distributional uncertainties? In this paper, we investigate a worst-case expectile problem for datasets with ambiguous data, proposing a robust risk measure constrained by certain parameters. Recently, we have been confronted with a variety of extreme risks in many fields, such as finance, insurance, and environmental science. In many of these fields, the partial availability of distribution information can result in decisions that are overly sensitive to shifts in the underlying loss distribution, as pointed out in studies such as [10] and [23]. Hence, the lack of complete information can render traditional risk measures, such as VaR and expectiles, less effective when based solely on stochastic data.

We aim to investigate the corresponding optimization problem based on a set of uncertain distributions, i.e., ambiguity sets. To address this problem, many researchers have proposed robust distribution optimization methods, the most widely used of which is the worst-case risk measure. Specifically, the worst-case risk measure is the worst risk measure calculated for all distributions under the ambiguity set. To be precise, for a random variable Y, given the level $\alpha \ge 1/2$, which corresponds to the case where e_α is law-invariant and coherent [4], we define the *worst-case value of the expectile* (WCE) under a set $\mathcal{M}$ as follows:

$$\mathrm{WCE}_\alpha^{\mathcal{M}} = \sup_{F_Y \in \mathcal{M}} e_\alpha(Y), \tag{1.2}$$

where F_Y denotes the distribution of Y and $\mathcal{M}$ is a set of plausible distributions for Y.

This paper investigates the worst-case expectiles on two specified ambiguity sets, namely, moment-based and Wasserstein-metric-based ambiguity sets. These robust measures play a pivotal role in risk management, enabling market risk assessment under stressed scenarios. They offer insights into data modeling during financial crises through the worst-case approach. The worst-case expectile measure we introduce aligns with what has recently been termed the distributionally robust stochastic optimization problem. As shown in [26], there are two natural methods of constructing the ambiguity set $\mathcal{M}$. On the one hand, an ambiguity set of distributions can be defined as the set of all distributions that are in a neighborhood of a given reference distribution with respect to a transportation distance, such as the Wasserstein distance; this is discussed by [19], [7], and others. On the other hand, an alternative ambiguity set can be described by specifying the moments of the distributions, as discussed by [8] and [30], among others. These two kinds of ambiguity sets are fundamentally different. In this paper we provide the worst-case expectiles under both scenarios.

The contributions of this paper are summarized as follows. First, we formally define the worst-case expectiles on the two types of ambiguity sets defined. Second, we transform the problem into a convex optimization problem for the ambiguity set specified by Wasserstein distance and a reference distribution G. Moreover, we obtain precise solutions to this problem in several special cases. Third, we focus on a more general ambiguity set defined by the moments of the distribution, i.e., where only the expectation and the pth-order central moment with $p > 1$ are acknowledged; this encompasses the scenario in which only the initial two

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

moments are recognized. In addition, we derive some asymptotic behaviors of the worst-case expectiles for these two ambiguity sets.

The paper is organized as follows. In Section 2, we start with an overview of distortion risk measures and then transform the infinite-dimensional problem (1.2) into a finite one, focusing on the maximum of a series of worst-case distortion risk measures for two specified ambiguity sets induced by Wasserstein distance and moment constraints. In Section 3, we first give some precise results on the worst-case value of the expectile over the p-Wasserstein ball and analyze the asymptotic behavior of the worst-case expectile as α tends to 1, with most circumstances taken into account. Then we compare our worst-case approach with the recently proposed 'model aggregation' approach of [18] and offer an alternative way of reformulating the problem (1.2), with full discussion. Finally, in Section 4 we discuss our findings and research perspectives. The proofs of all theoretical results and the details of extensive simulation studies are given in our supplementary material.

2. Reformulation of worst-case expectile through two important ambiguity sets

We first introduce some notation. Consider an atomless probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{L}^p$ be the set of all random variables in $(\Omega, \mathcal{F}, \mathbb{P})$ with finite pth moment, $p \in (0, \infty)$. We denote by $\mathcal{L}^0$ the set of all random variables and by $\mathcal{L}^\infty$ the set of essentially bounded random variables. In addition, we denote by $\mathcal{M}^p$ the set of the distribution functions of random variables in $\mathcal{L}^p$, i.e., $\mathcal{M}^p = \{F_Y(y) : Y \in \mathcal{L}^p\}$. For a random variable Y, $F_Y(y) = \mathbb{P}(Y \le y)$ represents its distribution function, and the corresponding left-continuous inverse is defined as

$$F_Y^{-1}(u) = \inf\{y \in \mathbb{R} : F_Y(y) \ge u\}, \ u \in (0, 1].$$

The mappings ess inf$(\cdot)$ and ess sup$(\cdot)$ on $\mathcal{L}^0$ give the essential infimum and the essential supremum, respectively, of a random variable. For a distribution $G \in \mathcal{M}^1$, we denote its essential supremum and first moment by ess $\sup_G :=$ ess sup(G) and $\mu_G := \int x dG(x)$, respectively. Throughout this paper, we denote the right and left derivatives of the function g at the point x_0 (if they exist) by $g'_+(x_0)$ and $g'_-(x_0)$. Similarly, we denote the left and right limits of the function g at x_0 by $g(x_0^-)$ and $g(x_0^+)$, respectively. For a real x, we denote its positive and negative parts by x^+ and x^-, respectively; that is, $x^+ = \max(x, 0)$ and $x^- = \max(0, -x)$.

2.1. Distortion risk measure

We now give the definition and some examples of *distortion risk measures*, which are commonly used in risk management. Recall that a non-decreasing function H:$[0, 1] \to [0, 1]$ satisfying the boundary condition $H(0) = 0$, $H(1) = 1$ is called a *distortion function*. A precise definition and detailed discussion of the distortion risk measure associated with H can be found in [28] and [29]. In particular, when H is continuous, the corresponding distortion risk measure ρ_H for the random variable Y has the following representation:

$$\rho_H(Y) = \int_0^1 F_Y^{-1}(u) dH(u),$$

if the integral is finite. Obviously, for $Y \in \mathcal{L}^1$, if we take $H(u) = u$, then $\rho_H(Y) = \int_0^1 F_Y^{-1}(u) du = \mathbb{E}[Y]$.

Another important risk measure that is highly relevant to expectiles is the *expected shortfall* (ES). Given $\alpha \in (0, 1)$, the expected shortfall at level α [25] is a distortion risk measure with

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

$H(u) = (1-\alpha)^{-1}(u-\alpha)^+$. Specifically, for $Y \in \mathcal{L}^1$,

$$\mathrm{ES}_\alpha(Y) = \frac{1}{1-\alpha}\int_\alpha^1 F_Y^{-1}(u)du. \tag{2.1}$$

2.2. Reformulation

In addition to the original definition introduced in Equation (1.1), there are a number of other representations of the expectile, such as those mentioned in [4]. For our problem, a useful tool is the well-known Kusuoka representation, which expresses the expectile as the maximum of a convex combination of the expected shortfall at a certain level and the expectation.

Proposition 1. ([4, Proposition 9].) *Let $Y \in \mathcal{L}^1$, $\alpha \geq 1/2$, and $\beta = \alpha/(1-\alpha)$; then*

$$e_\alpha(Y) = \max_{\gamma\in[\frac{1}{\beta},1]} \{(1-\gamma)\mathrm{ES}_\tau(Y) + \gamma\mathbb{E}[Y]\}, \tag{2.2}$$

where $\tau = (\beta - 1/\gamma)\big/(\beta-1)$.

Although there is no closed-form representation for the expectile, Proposition 1 allows us to reformulate the problem (1.2). Since a convex combination of two distortion risk measures is again a distortion risk measure, the expectile is actually the maximum of a one-parameter family of distortion risk measures indexed by $\gamma \in [1/\beta, 1]$. Later, we will use τ to represent $(\beta - 1/\gamma)\big/(\beta-1)$; we also use $\tau(\gamma)$ to emphasize the interdependence between γ and τ. Specifically, combining Proposition 1 and (2.1), we can rewrite $\mathrm{WCE}_\alpha^{\mathcal{M}}$ as follows:

$$\begin{aligned}
\mathrm{WCE}_\alpha^{\mathcal{M}} &= \sup_{F_Y\in\mathcal{M}} e_\alpha(Y) \\
&= \sup_{F_Y\in\mathcal{M}} \sup_{\gamma\in[\frac{1}{\beta},1]} \{(1-\gamma)\mathrm{ES}_\tau(Y) + \gamma\mathbb{E}[Y]\} \\
&= \sup_{\gamma\in[\frac{1}{\beta},1]} \sup_{F_Y\in\mathcal{M}} \{(1-\gamma)\mathrm{ES}_\tau(Y) + \gamma\mathbb{E}[Y]\} \\
&= \sup_{\gamma\in[\frac{1}{\beta},1]} \sup_{F_Y\in\mathcal{M}} \left\{\frac{1-\gamma}{1-\tau}\int_0^1 F_Y^{-1}(u)\mathbb{1}_{(\tau,1)}(u)du + \gamma\int_0^1 F_Y^{-1}(u)du\right\} \\
&=: \sup_{\gamma\in[\frac{1}{\beta},1]} \sup_{F_Y\in\mathcal{M}} \rho_{H_\gamma}(Y),
\end{aligned} \tag{2.3}$$

where the distortion function $H_\gamma(u) = \frac{1-\gamma}{1-\tau}(u-\tau)^+ + \gamma u$. The inner supremum has been well studied for a wide variety of ambiguity sets [5, 17, 24]. If the inner problem has an explicit representation for γ, then the infinite-dimensional problem (1.2) is reduced to an optimization problem for some function of γ. For this reason, we start by introducing two ambiguity sets, defined in terms of the Wasserstein distance and moment conditions, respectively, and illustrate some existing results that we will use later.

2.3. Ambiguity set induced by the Wasserstein distance

The Wasserstein distance of order $p \in [1, \infty)$ between two univariate distribution functions $F_1, F_2 \in \mathcal{M}^p$ is defined as the smallest $\mathcal{L}^p$ *distance* between a pair of random variables with

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

marginal distributions F_1, F_2, i.e.,

$$W_p(F_1, F_2) = \inf\left\{\left[\mathbb{E}\left|Y_1 - Y_2\right|^p\right]^{1/p} : Y_1 \sim F_1,\, Y_2 \sim F_2\right\}.$$

For univariate distributions, the Wasserstein distance has the following closed-form expression:

$$W_p(F_1, F_2) = \left(\int_0^1 \left|F_1^{-1}(u) - F_2^{-1}(u)\right|^p du\right)^{1/p},\ p \in [1, \infty).$$

For details on the Wasserstein distance, see e.g. [27]. The ambiguity set induced by the Wasserstein distance is then defined as a ball centered on the reference distribution G with radius $\varepsilon > 0$,

$$\mathcal{W}_{p,G,\varepsilon} = \left\{F \in \mathcal{M}^p : W_p(F, G) \le \varepsilon\right\}. \tag{2.4}$$

First we consider the reformulation (2.3) in the context of $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$. Note that the expectile reduces to expectation at the level of $\alpha = 1/2$ in Equation (1.1); hence the problem (1.2) is reduced to the following:

$$\begin{aligned}
\mathrm{WCE}_{1/2}^{\mathcal{W}_{p,G,\varepsilon}} &:= \sup_{F_Y \in \mathcal{W}_{p,G,\varepsilon}} e_{1/2}(Y) = \sup_{F_Y \in \mathcal{W}_{p,G,\varepsilon}} \int_0^1 F_Y^{-1}(u)du \\
&= \sup_{F \in \mathcal{W}_{p,G,\varepsilon}} \int_0^1 \left(F^{-1}(u) - G^{-1}(u)\right)du + \int_0^1 G^{-1}(u)du \\
&\le \sup_{F \in \mathcal{W}_{p,G,\varepsilon}} \left(\int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right|^p du\right)^{1/p} + \mu_G \\
&\le \varepsilon + \mu_G,
\end{aligned} \tag{2.5}$$

where the first inequality is due to the Hölder inequality. Then, letting $F_0^{-1}(x) := G^{-1}(x) + \varepsilon$, $x \in \mathbb{R}$, we obtain the opposite inequality to (2.5); thus we have

$$\mathrm{WCE}_{1/2}^{\mathcal{W}_{p,G,\varepsilon}} = \varepsilon + \mu_G.$$

As a result, we only need to concentrate on the case where $\alpha > 1/2$ and $\beta > 1$. The following proposition is a special version of Proposition 4 in [17], giving the precise value for $\sup_{F_Y \in \mathcal{W}_{p,G,\varepsilon}} \rho_{H_\gamma}(Y)$ and an explicit expression for the distribution function at which the supremum is attained.

Proposition 2. *For any fixed $\gamma \in [1/\beta, 1]$, let $H_\gamma(u) = \frac{1-\gamma}{1-\tau}(u-\tau)^+ + \gamma u$. Let h_γ represent the left derivative of the distortion function H_γ, and let $\|\cdot\|_q$ represent the standard q-norm. Then*

$$\sup_{F_Y \in \mathcal{W}_{p,G,\varepsilon}} \rho_{H_\gamma}(Y) = \varepsilon \|h_\gamma\|_q + \int_0^1 G^{-1}(u) h_\gamma(u) du, \tag{2.6}$$

where q is the conjugate index of p, i.e., $q = p/(p-1)$ with $p > 1$, and $q = \infty$ with $p = 1$. Moreover, the supremum in Equation (2.6) is attained by a certain distribution function, whose quantile function is given by

$$F_{p,G,\varepsilon,\gamma}^{-1}(u) = G^{-1}(u) + C_{p,\gamma,\varepsilon}\left(\mathbb{1}_{\{0<u\le\tau\}} + \beta^{q-1}\mathbb{1}_{\{\tau<u\le 1\}}\right), \qquad p > 1, \tag{2.7}$$

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

where $C_{p,\gamma,\varepsilon} = \varepsilon/[\tau + \beta^q(1-\tau)]^{1/p}$. *For* $p = 1$, *we have*

$$F^{-1}_{1,G,\varepsilon,\gamma}(u) = \begin{cases} G^{-1}(u) + \frac{\varepsilon}{1-\tau}\mathbb{1}_{\{\tau<u\le 1\}} & \text{if } \gamma < 1, \\ G^{-1}(u) + \varepsilon & \text{if } \gamma = 1. \end{cases} \tag{2.8}$$

The above proposition can easily be verified using Proposition 4 in [17], where the result for the worst-case value of a general distortion risk measure is provided. For convenience, let us write

$$z_{p,G,\varepsilon}(\gamma) := \varepsilon\|h_\gamma\|_q + \int_0^1 G^{-1}(u)h_\gamma(u)du; \tag{2.9}$$

then, using the reformulation (2.3) and the setting $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$, we obtain that the problem (1.2) essentially becomes a convex optimization problem (the concavity of $z_{p,G,\varepsilon}(\gamma)$ will be clarified in Theorem 1), i.e.,

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}} = \sup_{\gamma\in[\frac{1}{\beta},1]} z_{p,G,\varepsilon}(\gamma). \tag{2.10}$$

Therefore, we devote a great deal of space to describing the behavior of $z_{p,G,\varepsilon}(\gamma)$ in the next section.

2.4. Ambiguity set induced by moment conditions

We now focus on another important ambiguity set, this one induced by moment conditions; for details see [6] and [24]. Assume that $p > 1$, $\mu \in \mathbb{R}$, and σ is a positive real number. Specifically, we concentrate on the ambiguity set defined as follows:

$$\mathcal{M}_{p,\mu,\sigma} = \left\{ F \in \mathcal{M}^p : \int xdF(x) = \mu, \int |x-\mu|^p dF(x) \le \sigma^p \right\}. \tag{2.11}$$

Let $Y \in \mathcal{L}^p$ be a random variable such that $F_Y \in \mathcal{M}_{p,\mu,\sigma}$; then its expectation is clarified by μ, while its pth center moment is confirmed to have an upper bound σ^p. We would like to emphasize here that there are many other comparable versions of $\mathcal{M}_{p,\mu,\sigma}$, in one of which the restriction on the expectation is given by a pre-specified range rather than an equality; this indicates that the information about the expectation of an unknown distribution is not fully determined, which is often the case in parameter estimation problems. Many risk measures, including those explored in this article, share a similar structure in relation to the worst-case problem; thus we do not elaborate further here.

The case $p \neq 2$ is less commonly discussed in literature than $\mathcal{M}_{2,\mu,\sigma}$. This is not just because the expectation and variance are the most fundamental attributes of random variables, but also because of computational convenience. For example, [18] establishes an explicit solution for $\mathrm{WCE}_\alpha^{\mathcal{M}_{2,\mu,\sigma}}$. However, it is still necessary to consider the general order condition when we have information on higher or lower moments than $p = 2$. To the best of our knowledge, there is no literature concentrating on the worst-case value of the expectile through $\mathcal{M}_{p,\mu,\sigma}$ when $p \neq 2$. Note that if $\alpha = 1/2$, then obviously $\mathrm{WCE}_{1/2}^{\mathcal{M}_{p,\mu,\sigma}} = \mu$; therefore, we focus on the situation $\mathcal{M} = \mathcal{M}_{p,\mu,\sigma}$ at level $\alpha > 1/2$, similarly to Section 2.3.

In parallel to Proposition 2, we can derive the following, which covers the case of a moment-constrained ambiguity set.

Proposition 3. *Following the notation and definitions in Proposition 2, we have*

$$\sup_{F_Y\in\mathcal{M}_{p,\mu,\sigma}} \rho_{H_\gamma}(Y) = \mu + \min_{x\in\mathbb{R}} \sigma\|h_\gamma - x\|_q = \mu + \sigma[h_\gamma]_q,$$

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

where $[h_\gamma]_q = \min_{x\in\mathbb{R}} \|h_\gamma - x\|_q$. *Moreover, the above supremum can be attained by a two-point distribution, whose quantile function is defined by*

$$F^{-1}_{p,\mu,\sigma,\gamma}(u) = \mu + \frac{\sigma}{\left[\tau^{p-1} + (1-\tau)^{p-1}\right]^{1/p}} \left(-\frac{(1-\tau)^{1/q}}{\tau^{1/p}} \mathbb{1}_{\{0<u\le\tau\}} + \frac{\tau^{1/q}}{(1-\tau)^{1/p}} \mathbb{1}_{\{\tau<u\le 1\}} \right). \tag{2.12}$$

The specific form of the above quantile function requires careful computational verification and is essentially contained in the proof of Theorem 7; hence we omit its proof. According to Proposition 3, Equation (2.3) is reduced to

$$\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}} = \mu + \sigma \sup_{\gamma\in[\frac{1}{\beta},1]} [h_\gamma]_q. \tag{2.13}$$

Therefore, in the next section, we need to solve Equation (2.13), where the optimized function is not concave in general.

3. Main results

In this section, we first provide some exact results on the function $z_{p,G,\varepsilon}(\gamma)$, discussing both scenarios $p = 1$ and $p > 1$ in detail (see Theorems 1 and 2). Moreover, we find either a single distribution function $F_{\alpha,\mathcal{M}}$ or a sequence of distribution functions $F_{n,\alpha,\mathcal{M}}$ to attain or approximate the supremum in Equations (2.10) and (2.13), i.e., to satisfy $e_\alpha(F_{\alpha,\mathcal{M}}) = \mathrm{WCE}_\alpha^{\mathcal{M}}$ or $\lim_{n\to\infty} e_\alpha(F_{n,\alpha,\mathcal{M}}) = \mathrm{WCE}_\alpha^{\mathcal{M}}$, for $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$ (Theorems 1 and 2) and $\mathcal{M} = \mathcal{M}_{p,\mu,\sigma}$ (Theorem 7). As expected, the construction of these functions is extremely dependent on the behavior of the optimized functions. Unlike in most of the existing literature, which focuses on the exact worst-case value of certain risk measures, there is no explicit solution in most cases for the worst-case value of the expectile; we will explore this further below. However, when we consider extreme situations, i.e., $\alpha \to 1$, or equivalently $\beta \to \infty$, we can obtain asymptotic representations under extra conditions (see Theorems 3–7). In addition to studying $\mathrm{WCE}_\alpha^{\mathcal{M}}$, in Section 3.1.3 we give a concise introduction to an alternative approach called 'model aggregation' for prudent risk evaluation based on stochastic dominance, which was recently proposed in [18]. We then present a comprehensive comparison between model aggregation and our approach, simultaneously giving an upper bound to our problem for $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$.

3.1. Results on $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$

3.1.1. *Some precise results.* As shown in Equation (2.10), $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$ is the supremum of a function $z_{p,G,\varepsilon}(\gamma)$ on $\left[1/\beta, 1\right]$. Unfortunately, there is no closed-form solution for this problem in general, because of extreme dependence on both the reference distribution G and the order p (even in some simple cases, such as when $p = 2$ and G is the standard normal distribution function). Fortunately, the function $z_{p,G,\varepsilon}(\gamma)$ possesses good properties, in particular continuity and concavity, which implies that we can transform the infinite-dimensional problem (1.2) into a tractable one. Before we give the main results, it is worth mentioning that when $p = 1$, $l_p(\gamma) := \|h_\gamma\|_q$ is neither continuous at $\gamma = 1$ nor strictly concave on $(1/\beta, 1)$, which is different from the case $p > 1$. Consequently, we discuss the two cases of $p > 1$ and $p = 1$ separately.

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

Theorem 1. *When $p > 1$, $z_{p,G,\varepsilon}(\gamma)$ is continuous and strictly concave on $\left[1/\beta, 1\right]$ for any G; thus, its maximum value point γ^* exists and $\gamma^* \in (1/\beta, 1)$, that is to say,*

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}} = z_{p,G,\varepsilon}\left(\gamma^*\right).$$

Moreover, we have

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}} = e_\alpha(F_{p,G,\varepsilon,\gamma^*}),$$

where $F_{p,G,\varepsilon,\gamma^*}$ is defined in (2.7).

Based on the fact that τ is a strictly increasing function of γ that maps $\left[1/\beta, 1\right]$ into [0,1], we may seek the corresponding τ^* instead of γ^*, i.e.,

$$\tau^* = \sup_{\tau \in (0,1)} \left\{ z'_{p,G,\varepsilon-}(\gamma(\tau)) \geq 0 \right\}. \tag{3.1}$$

In the following, we focus on the situation when $p = 1$. Noting that $z_{1,G,\varepsilon}(1) = z_{1,G,\varepsilon}(1/\beta) = \varepsilon + \mu_G$ and $\lim_{\gamma \to 1^-} z_{1,G,\varepsilon}(\gamma) = \varepsilon\beta + \mu_G > \varepsilon + \mu_G$, we consider only $\sup_{\gamma \in \left(\frac{1}{\beta}, 1\right)} z_{1,G,\varepsilon}(\gamma)$.

Theorem 2. *When $p = 1$, we have the following:*

(i) If $\operatorname{ess\,sup}_G \leq \mu_G + \varepsilon\beta$, then $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \mu_G + \varepsilon\beta$, and there exists a sequence of distribution functions $F_{n,G,\varepsilon}$ such that $\lim_{n\to\infty} e_\alpha(F_{n,G,\varepsilon}) = \mu_G + \varepsilon\beta$, whose quantile function is given by

$$F^{-1}_{n,G,\varepsilon}(u) = G^{-1}(u) + \frac{\varepsilon}{1-\tau_n} \mathbb{1}_{\{\tau_n < u \leq 1\}},$$

where $\tau_n \in (0, 1)$ is an arbitrary sequence tending to 1 as $n \to \infty$.

(ii) If $\operatorname{ess\,sup}_G > \mu_G + \varepsilon\beta$, then $z_{1,G,\varepsilon}(\gamma)$ attains its maximum on $\gamma^* \in (1/\beta, 1)$ and $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = z_{1,G,\varepsilon}(\gamma^*)$. Analogously to Theorem 1, we have $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = e_\alpha(F_{1,G,\varepsilon,\gamma^*})$, where $F_{1,G,\varepsilon,\gamma^*}$ is as defined in (2.8).

Recalling Equations (2.7) and (2.8), we know that the quantile function that attains the supremum is in fact equal to the sum of G^{-1} and the quantile function corresponding to some non-negative two-point distribution indexed by γ^*, p, and ε. See Figure 3.1 for a numerical illustration.

Theorem 10 in [4] reveals that expectiles are Lipschitz with respect to the 1-Wasserstein metric, with Lipschitz constant β, i.e., for any distribution functions $F_1, F_2 \in \mathcal{M}^1$, it holds that

$$|e_\alpha(F_1) - e_\alpha(F_2)| \leq \beta W_1(F_1, F_2). \tag{3.2}$$

In conjunction with our Theorem 2, there are two counterintuitive aspects worth mentioning. First, there are circumstances in which the supremum in (1.2) may not be attained, which seems to contradict the fact that the expectile is a continuous functional under the 1-Wasserstein distance. However, this should not be surprising, because the Wasserstein ball is not compact under the induced Wasserstein topology (see Proposition 2.2.9 in [22].) Second, one may conjecture that

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = e_\alpha(G) + \beta\varepsilon \tag{3.3}$$

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

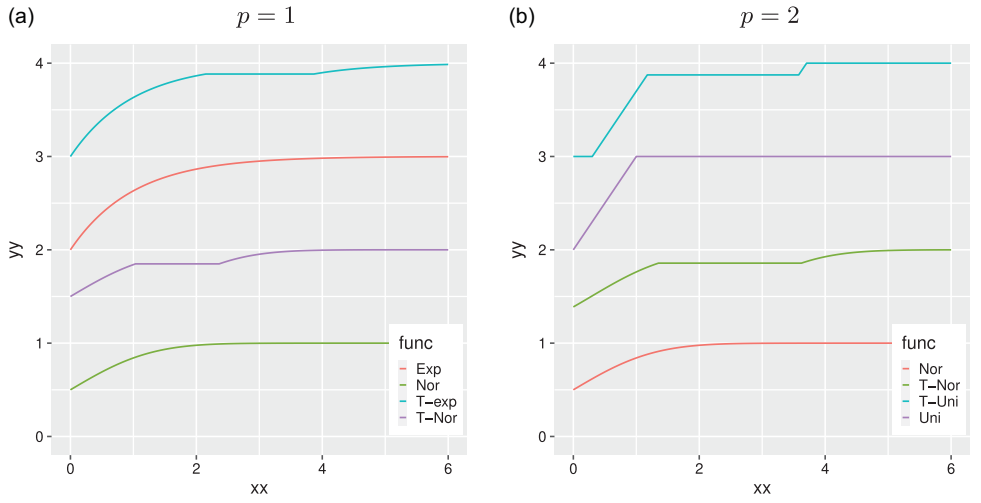

FIGURE 3.1: Illustration of the distribution function at which the supremum is attained in Theorems 1 and 2: (a) Here $p = 1$, $\varepsilon = 0.2$, and $\alpha = 0.8$; the standard normal distribution (green line) and exponential distribution (red line) and the corresponding transformed distributions $F_{1,G,\varepsilon,\gamma^*}$ (purple and blue lines) are shown. (b) Here $p = 2$, $\varepsilon = 1$, and $\alpha = 0.9$; the standard normal distribution (red line) and uniform distribution (purple line) and the corresponding transformed distributions $F_{2,G,\varepsilon,\gamma^*}$ (green and blue lines) are shown.

from Equation (3.2). However, interestingly, this is clearly not the case, since from Theorem 2(i), it follows that under the condition ess $\sup_G \leq \mu_G + \varepsilon\beta$, we have

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \mu_G + \beta\varepsilon,$$

which is strictly less than $e_\alpha(G) + \beta\varepsilon$, for any non-degenerate distribution function G. A natural question is under which assumptions on G and ε the Lipschitz bound (3.3) is optimal, and when it can be improved. The following proposition provides a detailed answer.

Proposition 4. *If G is a non-degenerate distribution function, then* $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} < e_\alpha(G) + \beta\varepsilon$. *By contrast, if $G = \mathbb{1}_{\{x \geq x_0\}}$, then* $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = e_\alpha(G) + \beta\varepsilon$.

Proposition 4 implies that the Lipschitz bound (3.3) is optimal if and only if G is degenerate, in which case the Wasserstein constraint is reduced to a restriction on the absolute first moment, i.e., $\mathcal{M} = \{F_Y, \mathbb{E}|Y - x_0| \leq \varepsilon\}$ for some real number x_0. The proof of this proposition is deferred to Appendix B, as a continuation of that of Theorem 2. In addition, we have the following corollary, which is a direct consequence of Theorem 2 and Proposition 4.

Corollary 1. *When p=1, we have the following:*

(i) If G is bounded from above, i.e. ess $\sup_G < \infty$, *then there exists some level α close enough to 1 so that* $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \varepsilon\beta + \mu_G$.

(ii) If G is not bounded from above, then $z_{1,G,\varepsilon}(\gamma)$ attains its maximum on $(1/\beta, 1)$, and

$$\varepsilon\beta + \mu_G < \mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} < \varepsilon\beta + e_\alpha(G). \tag{3.4}$$

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

3.1.2. *Some asymptotic results.* So far we have generally analyzed the properties of the function $z_{p,G,\varepsilon}$ and obtained some precise results for $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$, along with a rough range (3.4) for $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}}$. The worst-case value of the expectile has a closed-form solution only in a few special circumstances, namely $p = 1$ and ess $\sup_G \le \mu_G + \varepsilon\beta$, which excludes a large class of distribution functions. It is clearly unsatisfactory simply to know that (2.10) is a convex optimization problem and can be calculated with respect to the specific reference distribution, since this does not provide meaningful understanding of the characteristics of the worst-case value of the expectile. Therefore, we take an alternative approach and explore its extreme behavior, i.e., we consider $\alpha \to 1$, which is a case often relevant in risk management. For this, we need to make some additional assumptions on the distribution function G. We start by collecting some notation.

Let $\mathbb{E}^G[\cdot]$ represent taking the expectation under the distribution function G. Throughout this paper, the notation $f(\alpha) = O(g(\alpha))$ means that there exists a constant c such that $f(\alpha)/g(\alpha) \to c$ as $\alpha \to 1$. All limits are at $\alpha \to 1$, or equivalently $\beta \to \infty$, unless otherwise specified.

Theorem 3. *When $p = 1$, if $\mathbb{E}^G[(Y^+)^2] < \infty$, G^{-1} is continuous, and* ess $\sup_G = \infty$, *then*

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \mu_G + \varepsilon\beta + o(1) \quad \text{as } \alpha \to 1.$$

Remark 1. From the proof in Appendix B, it can easily be seen that the remainder in Theorem 3 is actually $O\left(\beta\mathbb{E}^G\left[Y\mathbb{1}_{\{Y>\varepsilon\beta/2\}}\right]\right)$, which is of lower order than $\varepsilon\beta$ because $G \in \mathcal{M}^1$. A trivial application of Chebyshev's inequality implies that the higher the moments of G, the faster the residuals converge. In particular, when G is a normally distributed distribution function, the residuals are of exponential order infinitesimal. However, when G is a heavy-tailed distribution with an infinite moment of order $p \in (1, 2)$, the residual term is large, which motivates us to relax the assumption that $\mathbb{E}^G[(Y^+)^2] < \infty$ (see Theorem 4).

The following example shows that even if $G \in \bigcap_{p\in(1,2)} \mathcal{M}^{2-\delta}$ for all $\delta > 0$, but $G \notin \mathcal{M}^2$, the residual term is just $O(1)$ rather than $o(1)$.

Example 1. *Let G be a Pareto distribution function with density $2x^{-3}\mathbb{1}_{(1,\infty)}(x)$; then we derive the following:*

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \varepsilon\beta + 2 + 1/\varepsilon + \mathrm{o}(1).$$

Now we see that the approximation in Theorem 3 is valid only for $\mathbb{E}^G[(Y^+)^2] < \infty$. Therefore, we generally consider a class of heavy-tailed distributions whose survival function satisfies

$$\overline{G}(x) := 1 - G(x) = x^{-\theta}L(x), \tag{3.5}$$

where $L(x)$ is a slowly varying function at infinity, i.e., $\lim_{x\to\infty} L(cx)/L(x) = 1$ for any $c > 0$. Intuitively, $\overline{G}(x)$ is 'almost' $x^{-\theta}$ as $x \to \infty$, and the 'tail' becomes heavier as $\theta \to 0$. When $\theta > 2$, the condition in Theorem 3 is fulfilled. Moreover, $\theta > 1$ is required to guarantee that $G \in \mathcal{M}^1$. Consequently, we assume that $\theta \in (1, 2]$. However, when θ approaches 1, it seems difficult to give uniform and simple bounds connected to θ. Worse still, when L is a general slowly varying function, the analysis is beyond the scope of this paper. Therefore, we impose additional restrictions on both θ and L.

Theorem 4. *Suppose that when x is large enough, the survival function of G has the representation* (3.5), *and assume $3/2 < \theta \le 2$, $\lim_{x\to\infty} L(x) =: \zeta > 0$. In addition, assume L is*

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

differentiable with $\lim_{x\to\infty} x^{3-\theta}L'(x) = 0$. *Then*

$$\mathrm{WCE}_{\alpha}^{\mathcal{W}_{1,G,\varepsilon}} = \varepsilon\beta + \mu_G + \frac{\zeta\varepsilon^{1-\theta}}{\theta-1}\beta^{2-\theta} + o(1), \qquad as \quad \alpha \to 1.$$

Remark 2. Through a careful check of the proof in Appendix B, we can verify that our approximation for τ^* only requires the assumption $\lim_{x\to\infty} L(x) =: \zeta > 0$, which is not restrictive because there are large numbers of common heavy-tailed distributions satisfying these conditions. This is evidently the central condition to be checked as part of Theorem 4, excluding some irregular slowly varying functions such as $\log(x)$, which are difficult to analyze using our techniques. The other assumptions we made are utilized to guarantee that the residual term is $o(1)$ rather than $o(\beta^{2-\theta})$. Even so, a large category of distributions is still included; see Table 1. A similar table can be found in [1].

It is easily confirmed from the above table that our assumption on $L'(x)$ is automatically satisfied whenever the parameter $\theta > 1$, except for the Burr distribution, which requires $2 - \alpha\beta - \alpha < 0$. Therefore, the assumptions in Theorem 4 should not be viewed as restrictive in practice.

We now turn to the case $p > 1$. We can establish a more accurate approximation for $\mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}}$ when $\operatorname{ess\,sup}_G < \infty$.

Theorem 5. *If* $p > 1$ *and* G *has bounded support, i.e.* $\operatorname{ess\,sup}_G < \infty$, *then*

$$\mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}} = p^{-1}\left[\varepsilon\,(p-1)^{1/q}\,\beta^{1/p} + \mu_G\right] + q^{-1}\operatorname{ess\,sup}_G + o\,(1)\,, \qquad as \quad \alpha \to 1. \quad (3.6)$$

In fact, as noted in Remark 1, the remainder $o(1)$ in Equation (3.6) may also be replaced by a precise order of particular relevance to p and G, through a slight modification of the proof in Appendix B. However, tedious arguments are required to analyze the relationship between terms related to p and those related to G, or even both of them. The form of Theorem 5 is consistent with Theorem 2(i), in the sense that $\mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}} \to \mathrm{WCE}_{\alpha}^{\mathcal{W}_{1,G,\varepsilon}}$ as $p \to 1$, regardless of the residual term. In detail, $(p-1)^{1/q} \to 1$ as $p \to 1$. We must inevitably make the strong assumption $\operatorname{ess\,sup}_G < \infty$, since there are essential difficulties in tackling a general distribution G, on account of the complexity of $z_{p,G,\varepsilon}$ and the entanglement of l_p and k_G, which are defined in the proof of Theorem 1. However, a broad class of distributions is still covered, including the empirical distribution, which has attracted considerable research interest recently. In the following supporting example, G is chosen to be a degenerate distribution function, and the restrictions related to the Wasserstein distance are effectively translated into moment constraints.

Example 2. *Let* G *be the distribution function corresponding to a point measure* δ_{x_0}. *Then we can provide a specific expression for* $\mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}}$ *as follows:*

$$\mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}} = \begin{cases} \varepsilon\beta + x_0, & p = 1, \\ \varepsilon p^{-1}(p-1)^{1/q}\beta^{1/p}\left(1 + \frac{\beta-1}{\beta^q-\beta}\right)\left(1 + \frac{1-\beta^{2-q}}{\beta-1}\right)^{1/q} + x_0, & p > 1. \end{cases}$$

From Theorems 3–5, it appears that the worst-case value of the expectile through $\mathcal{W}_{p,G,\varepsilon}$ is mainly influenced by the level α when extreme risk is involved. On the one hand, it is intuitively clear that for a distribution G bounded from above, there is a bounded expectile at any level,

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

TABLE 1. A list of heavy-tailed distributions satisfying the assumptions in Theorem 4 with corresponding θ, ζ, and $L'(x)$. Here $\Gamma(\cdot)$ and $B(\cdot,\cdot)$ denote the gamma and beta functions, respectively.

Distribution (parameters)	Density function	θ	ζ	$L'(x)$
Generalized Pareto (σ, $\xi > 0$)	$\frac{1}{\sigma}\left(1+\frac{\xi x}{\sigma}\right)^{-1/\xi-1}\mathbb{1}_{\{x>0\}}$	$1/\xi$	$\left(\frac{\sigma}{\xi}\right)^{1/\xi}$	$O(x^{-2})$
Burr (α, $\beta > 0$)	$\alpha\beta x^{\alpha-1}(1+x^{\alpha})^{-\beta-1}\mathbb{1}_{\{x>0\}}$	$\alpha\beta$	1	$O(x^{-\alpha-1})$
Fréchet ($\alpha > 0$)	$\alpha x^{-\alpha-1}e^{-x^{-\alpha}}\mathbb{1}_{\{x>0\}}$	α	1	$O(x^{-\alpha-1})$
Fisher (ν_1, $\nu_2 > 0$)	$\frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{B(\frac{\nu_1}{2},\frac{\nu_2}{2})}x^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1 x}{\nu_2}\right)^{-\frac{\nu_1+\nu_2}{2}}\mathbb{1}_{\{x>0\}}$	$\nu_2/2$	$\frac{2\left(\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_2}{2}}}{\nu_2 B(\frac{\nu_1}{2},\frac{\nu_2}{2})}$	$O(x^{-2})$
Inverse gamma (α, $\beta > 0$)	$\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{-\alpha-1}e^{-\frac{\beta}{x}}\mathbb{1}_{\{x>0\}}$	α	$\frac{\beta^{\alpha}}{\Gamma(\alpha)}$	$O(x^{-2})$
Student ($\nu > 0$)	$\frac{1}{\sqrt{\nu\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	ν	$\frac{\nu^{\frac{\nu}{2}-1}}{\sqrt{\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$	$O(x^{-3})$

and so it is reasonable to expect that at extreme levels, the effect of G on $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$ is limited. In detail, Theorems 3 and 5 quantitatively characterize this effect in terms of a weighted average of expectation and essential supremum, i.e., $\mu_G/p + \operatorname{ess\,sup}_G/q$ (μ_G by continuity when $p = 1$). On the other hand, as the tail of the reference distribution becomes heavier and heavier, the 'invasion' of G becomes more and more prominent, because more and more terms related to G are included in the asymptotic expression for $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}}$ (see Theorems 3 and 4). An analogous phenomenon is found for $\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}}$, which will be discussed in Section 3.2. Even if ε is small and G is a degenerate distribution, we find that the worst-case value of the expectile explodes as $\alpha \to 1$. This is because the shape of the distribution may be greatly altered, regardless of how slight the perturbation induced by the Wasserstein distance is (see Figure 3.1 for a simple illustration). Specifically, we observe that the transformed distribution function $F_{p,G,\varepsilon,\gamma^*}$ elongates the tail of G dramatically when α is sufficiently close to 1; this can be theoretically verified in many cases. Indeed, in the several scenarios contemplated in Theorems 3–5, $\tau^* \to 1$ and $\beta^{q-1}/[\tau^* + \beta^q(1-\tau^*)]^{1/p} \to \infty$ always hold when $p > 1$, indicating that $F^{-1}_{p,G,\varepsilon,\gamma^*}$ substantially elevates G^{-1} in an extremely narrow interval $[\tau^*, 1]$. These results should be significant as a feature of model ambiguity at extreme levels.

3.1.3. *Comparison with the model aggregation approach.* As suggested by [18], instead of directly calculating the maximum (or supremum) of $e_\alpha(F)$ over $F \in \mathcal{M}$, it may be better to calibrate a robust (conservative) distribution F^* from $\mathcal{M}$ and calculate $e_\alpha(F^*)$; the latter is called the 'model aggregation' approach. As will be seen later, this element F^* does not necessarily belong to $\mathcal{M}$, although it 'dominates' $\mathcal{M}$ in some sense. In the presence of model uncertainty, a natural problem in risk management is how to generate such a robust model F^* from the collection of models generated by various scenarios. Indeed, F^* is typically referred to as a maximal element (or supremum) of the ambiguity set $\mathcal{M}$ in the sense of stochastic dominance; it has been explicitly derived in many situations. Below, we will use the abbreviation 'MA' to refer to the model aggregation approach. It can be argued that both the worst-case approach and the MA approach are reasonable ways to assess the risk posed by model uncertainty, although they may yield different values over the same ambiguity set $\mathcal{M}$. Figure 1 in [18] illustrates the two methods.

Now we are in a position to properly formulate the MA approach by giving a partial order $\preceq$ on $\mathcal{M}^1$. At this point, $(\mathcal{M}^1, \preceq)$ is called an order set. The most commonly used partial orders in finance and economics are the usual stochastic order $\preceq_{st}$ and the increasing convex order $\preceq_{icx}$, defined as follows.

Definition 1. *For $F_1, F_2 \in \mathcal{M}^1$,*

(i) *$F_1 \preceq_{st} F_2$ if $\int u dF_1 \le \int u dF_2$ for increasing functions u;*

(ii) *$F_1 \preceq_{icx} F_2$ if $\int u dF_1 \le \int u dF_2$ for increasing convex functions u.*

A natural interpretation of a partial order is that F_2 is riskier than F_1 if $F_1 \preceq F_2$. This is the case when F_1 and F_2 are viewed as loss distributions rather than wealth distributions: a larger element with respect to $\preceq_{st}$ or $\preceq_{icx}$ corresponds to higher risk. Thus, we expect the risk measure ρ to maintain the order, i.e., $\rho(F_1) \le \rho(F_2)$ if $F_1 \preceq F_2$, which is formally defined being as $\preceq$-consistent. The expectile is consistent with both $\preceq_{st}$ and $\preceq_{icx}$. For an ambiguity set $\mathcal{M} \in \mathcal{M}^1$, the supremum of $\mathcal{M}$ in $\mathcal{M}^1$ is defined by $\bigvee \mathcal{M} \in \mathcal{M}^1$ such that $F_1 \preceq \bigvee \mathcal{M} \preceq F_2$ for all $F_1 \in \mathcal{M}$ and for any F_2 that dominates every element in $\mathcal{M}$. If such an F_2 exists, we say that $\mathcal{M}$ is bounded from above. In what follows, we denote the supremum of $\mathcal{M}$ with respect to $\preceq_{st}$ and $\preceq_{icx}$ on $\mathcal{M}^1$ by $\bigvee_{st} \mathcal{M}$ and $\bigvee_{icx} \mathcal{M}$, respectively.

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

For an ambiguity set $\mathcal{M}$ whose supremum $\bigvee \mathcal{M}$ exists, we formally formulate the MA approach for an expectile, in parallel with (1.2), as

$$\mathrm{MAE}_{\alpha,\preceq}^{\mathcal{M}} = e_\alpha\left(\bigvee \mathcal{M}\right), \tag{3.7}$$

and $\mathrm{MAE}_{\alpha,\preceq}^{\mathcal{M}} = \infty$ if $\mathcal{M}$ is not bounded from above. From the definition and the $\preceq_{icx}$-consistency of the expectile, it is easily verified that $e_\alpha\left(\bigvee_{icx} \mathcal{M}\right) \geq e_\alpha(F)$ for any $F \in \mathcal{M}$. Moreover, since the expectile is consistent with the two partial orders $\preceq_{st}$ and $\preceq_{icx}$, the MA approach with the stronger partial order $\preceq_{st}$ leads to a higher risk evaluation, i.e., $\mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{M}} \leq \mathrm{MAE}_{\alpha,\preceq_{st}}^{\mathcal{M}}$. Combining these two statements, we have the chain of inequalities

$$\mathrm{WCE}_{\alpha}^{\mathcal{M}} \leq \mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{M}} \leq \mathrm{MAE}_{\alpha,\preceq_{st}}^{\mathcal{M}}, \tag{3.8}$$

which indicates that the MA approach may produce a more conservative risk evaluation. Although the supremum of Wasserstein balls with respect to $\preceq_{st}$ and $\preceq_{icx}$ is explicitly derived in [18, Theorem 6], there is no closed-form expression for the supremum under the partial order $\preceq_{st}$, and the $\mathcal{W}_{p,G,\varepsilon}$ is not bounded from above under $\preceq_{icx}$ when $p = 1$. Therefore, in the following we will restrict our attention to $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$ with $p > 1$ and the partial order $\preceq_{icx}$. From [18, Theorem 6], the quantile function of $\bigvee_{icx} \mathcal{W}_{p,G,\varepsilon}$ with $p > 1$ is given by

$$(F_{p,G,\varepsilon}^{icx})^{-1}(u) = G^{-1}(u) + \left(1 - \frac{1}{p}\right)(1-u)^{-1/p}\varepsilon, \tag{3.9}$$

which is the sum of the quantile function G^{-1} and a Pareto quantile with a tail index $p > 1$. It is surprising that $F_{p,G,\varepsilon}^{icx}$ is a heavy-tailed distribution even if the reference distribution G is light-tailed, while $F_{p,G,\varepsilon,\gamma^*}$ is not so. Another significant difference between the worst-case approach and the MA approach is that the construction for $F_{p,G,\varepsilon,\gamma^*}$ is highly dependent on the specific choice of risk measure, rather than simply arising from the ambiguity set $\mathcal{W}_{p,G,\varepsilon}$. Also, one may wonder whether the inequality in Equation (3.8) could be refined to an equality. The theorem below provides a negative answer, as well as an asymptotic representation for Equation (3.7), where $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$.

Theorem 6. *Let $p > 1$ and let $F_{p,G,\varepsilon}^{icx}$ be as defined in (3.9). Furthermore, let $z_{p,G,\varepsilon}^{icx}(\gamma) = \int_0^1 (F_{p,G,\varepsilon}^{icx})^{-1}(u)dH_\gamma(u)$. Then the following hold:*

(i) We have that $z_{p,G,\varepsilon}^{icx}(\gamma)$ is continuous and strictly concave on $[1/\beta, 1]$, and

$$\mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{W}_{p,G,\varepsilon}} = \max_{\gamma\in[1/\beta,1]} z_{p,G,\varepsilon}^{icx}(\gamma).$$

(ii) For any G, ε, and level $\alpha > 1/2$, $\mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}}$ is strictly less than $\mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{W}_{p,G,\varepsilon}}$.

(iii) If we further assume $\operatorname{ess\,sup}_G < \infty$, then

$$\mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{W}_{p,G,\varepsilon}} = \mathrm{WCE}_{\alpha}^{\mathcal{W}_{p,G,\varepsilon}} + \frac{\varepsilon}{p} + \mathrm{o}(1), \qquad \text{as} \quad \alpha \to 1.$$

Parts (i) and (iii) are parallel to Theorems 1 and 5 respectively, in which the worst-case approach yields similar results. Moreover, they are proved via identical steps and using similar techniques; thus we only give the proof of Part (ii), which is deferred to Appendix B.

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

Additionally, because of technical constraints, we can only obtain an asymptotic solution for $\mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{W}_{p,G,\varepsilon}}$ when ess $\sup_G$ is finite. Comparing these two methods, the MA approach yields a more cautious risk assessment, whereas our worst-case approach takes full advantage of the properties of expectiles. However, when extreme levels are considered, they are separated by almost a constant ε/p, which is negligible relative to the diverging term $\mathrm{MAE}_{\alpha,\preceq_{icx}}^{\mathcal{W}_{p,G,\varepsilon}}$.

3.1.4. *Alternative formulation of* $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$. We present an alternative way of formulating $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$, based on the general results of [2] on robust shortfall risk measures. As a particular category of risk measure, the utility-based *shortfall risk measure* (abbreviated as SR hereafter) was proposed by [11] and has attracted rapidly increasing interest in recent years [9, 13, 16]. We first give a brief definition of the shortfall risk measure, which is adopted from [15] with some slight modifications for simplicity. Let $l:\mathbb{R}\to\mathbb{R}$ be a convex, increasing, and non-constant loss function, and let λ be a pre-specified constant in the interior of the range of l, indicating the risk level. Then the SR of $X\in\mathcal{L}^1$ is defined as

$$\mathrm{SR}_{\lambda,l}(X)=\inf\{t\in\mathbb{R}:\mathbb{E}(l(X-t))\le\lambda\}, \tag{3.10}$$

whenever $\mathbb{E}(l(X-t))$ is finite for some $t\in\mathbb{R}$. Viewing $-X$ as a financial position, we see from the definition that the SR is the smallest amount of cash that must be added to the position $-X$ to make it below a certain risk level, i.e., $\mathbb{E}(l(-(-X+t)))\le\lambda$. Theorem 4.9 in [3] shows that the expectile is indeed a shortfall risk measure with $l(t)=\alpha t^+-(1-\alpha)t^-$ and $\lambda=0$.

To hedge the risk arising from the ambiguity of the true probability distribution, much of the extant literature considers the *distributionally robust shortfall risk measure* (abbreviated as DRSR in the sequel); see [15, 2, 30]. Specifically, for an ambiguity set $\mathcal{M}$, we concentrate on the following problem:

$$\mathrm{DRSR}_{\lambda,l}^{\mathcal{M}}=\inf\left\{t\in\mathbb{R}:\sup_{G\in\mathcal{M}}\mathbb{E}^G(l(X-t))\le\lambda\right\}. \tag{3.11}$$

In [2], Equation (3.11) is reformulated as a finite-dimensional problem with $\mathcal{M}=\mathcal{W}_{p,G,\varepsilon}$ and a different loss function l. In fact, we have

$$\mathrm{DRSR}_{0,l}^{\mathcal{W}_{p,G,\varepsilon}}=\inf_{\mu\ge0}\mathrm{SR}_{0,l^{\mu p}+\mu\varepsilon^p}(X_G), \tag{3.12}$$

where X_G is an arbitrary random variable with distribution function G and $l^{\mu p}$ is the μc-transform of function l with cost function $c(x,y)=|x-y|^p$, which is formally defined as

$$l^{\mu p}(x)=\sup\left\{l(y)-\mu|x-y|^p,\ y\in\mathbb{R}\text{ such that }l(y)<\infty\right\}. \tag{3.13}$$

Meanwhile, [15] reveals that the infimum and supremum in Equation (3.11) can be exchanged in many circumstances, and in particular for $l(t)=\alpha t^+-(1-\alpha)t^-$, which gives rise to an expectile. Indeed, we have

$$\mathrm{DRSR}_{\lambda,l}^{\mathcal{M}}=\sup_{G\in\mathcal{M}}\mathrm{SR}_{\lambda,l}(X_G),$$

where the right-hand side is exactly the worst-case shortfall risk measure over $\mathcal{M}$. Now, by taking $\mathcal{M}=\mathcal{W}_{p,G,\varepsilon}$ and $\lambda=0$ and combining this with Equation (3.12), we obtain the following reformulation:

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}=\inf_{\mu\ge0}\mathrm{SR}_{-\mu\varepsilon^p,l_\alpha^{\mu p}}(X_G), \tag{3.14}$$

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

where $l_\alpha^{\mu p}$ can be calculated explicitly from Equation (3.13) with $l(y) = \alpha y^+ - (1-\alpha)y^-$. Specifically, when $p > 1$ we have

$$l_\alpha^{\mu p} = \begin{cases} \alpha x + \frac{\alpha^q}{q(\mu p)^{q-1}} & \text{if } x > \frac{1}{q(\mu p)^{q-1}} \left(\frac{\alpha^q - (1-\alpha)^q}{1-2\alpha} \right), \\ (1-\alpha)x + \frac{(1-\alpha)^q}{q(\mu p)^{q-1}} & \text{if } x \le \frac{1}{q(\mu p)^{q-1}} \left(\frac{\alpha^q - (1-\alpha)^q}{1-2\alpha} \right). \end{cases}$$

We believe it would be difficult to simplify Equation (3.14) further in this case, not only because of the complexity of $l_\alpha^{\mu p}$, but also because of the generality of G. However, when $p = 1$, it is easy to obtain $l_\alpha^{\mu 1} = l$ if $\mu \ge \alpha$ and $l_\alpha^{\mu p} = \infty$ otherwise, which further transforms Equation (3.14) into

$$\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \inf_{\mu \ge \alpha} \inf\{t \in \mathbb{R} : \alpha \mathbb{E}^G (X-t)^+ - (1-\alpha)\mathbb{E}^G (X-t)^- \le -\mu\varepsilon\}. \tag{3.15}$$

Notice that the function

$$g(t) = \alpha \mathbb{E}^G (X-t)^+ - (1-\alpha)\mathbb{E}^G (X-t)^- = \alpha \int_t^{+\infty} (1 - G(s))ds - (1-\alpha) \int_{-\infty}^t G(s)ds$$

is a strictly decreasing continuous function that tends to $-\infty$ as $t \to +\infty$ and tends to $+\infty$ as $t \to -\infty$. Equation (3.15) can therefore be translated into a more concise form; that is, $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}}$ is the unique solution to the following equation:

$$\alpha \int_t^{+\infty} (1 - G(s))ds - (1-\alpha) \int_{-\infty}^t G(s)ds = -\alpha\varepsilon. \tag{3.16}$$

The above equation is quite succinct, although it is still difficult to use it to acquire general results for $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}}$. However, if we assume $\operatorname{ess\,sup}_G \le \beta\varepsilon + \mu_G$ and substitute $\varepsilon\beta + \mu_G$ into Equation (3.16), we obtain

$$\begin{aligned} &\alpha \int_{\varepsilon\beta+\mu_G}^{+\infty} (1 - G(s))ds - (1-\alpha) \int_{-\infty}^{\varepsilon\beta+\mu_G} G(s)ds \\ &= -(1-\alpha) \left(\varepsilon\beta + \mu_G - G^{-1}(1) + \int_{-\infty}^{G^{-1}(1)} G(s)ds \right) \\ &= -(1-\alpha) \left(\varepsilon\beta + \mu_G - G^{-1}(1) + \int_0^1 \left(G^{-1}(1) - G^{-1}(s) \right) ds \right) \\ &= -(1-\alpha)\varepsilon\beta = -\alpha\varepsilon, \end{aligned}$$

which is consistent with Theorem 2(i), which shows that $\mathrm{WCE}_\alpha^{\mathcal{W}_{1,G,\varepsilon}} = \varepsilon\beta + \mu_G$.

In summary, Equations (2.10) and (3.14) are both tractable optimization problems that can be solved easily by many popular algorithms, once the distribution function G is provided. Nevertheless, it is clear that Equation (2.10) is easier to deal with, and it is possible to obtain a transformed distribution function whose expectile is exactly $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$ by solving Equation (2.10). Moreover, when considering extreme levels, it is more difficult to obtain asymptotic results analogous to those of Theorems 3–7 by analyzing Equation (3.14).

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

3.2. Results on $\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}}$

We consider the problem (1.2) where $\mathcal{M} = \mathcal{M}_{p,\mu,\sigma}$. As in Section 3.1, we observe that there is no closed-form solution in most cases, apart from $p = 2$. Another difficulty is that the function being optimized is not concave in general, so it is more fruitful to study the asymptotic nature of the solution than simply to discuss the solvability of the problem (2.13). In other words, we can obtain an asymptotic solution when $\alpha \to 1$, or equivalently $\beta \to \infty$. We show this in the following theorem.

Theorem 7. *For any $p > 1$, there exists only one $\gamma^* \in (1/\beta, 1)$ maximizing $[h_\gamma]_q$, and $\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}} = e_\alpha(F_{p,\mu,\sigma,\gamma^*})$, where $F_{p,\mu,\sigma,\gamma^*}$ is defined in (2.12). Moreover, we have the following:*

- *If $p = 2$, then $\gamma^* = (1/\beta + 1)/2$ and*

$$\mathrm{WCE}_\alpha^{\mathcal{M}_{2,\mu,\sigma}} = \mu + \sigma \frac{\beta - 1}{2\sqrt{\beta}}.$$

- *If $p >> 2$, then $\gamma^* = 1/p + \mathrm{o}(1)$ and*

$$\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}} = \mu + \sigma \frac{(\beta - 1)^{1/p}}{p^{1/p} q^{1/q}} + \mathrm{o}\left(\beta^{2-p-1/p}\right), \qquad as \quad \alpha \to 1.$$

- *If $p < 2$, then $\gamma^* = 1/p + \mathrm{o}(1)$ and*

$$\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}} = \mu + \sigma \frac{(\beta - 1)^{1/p}}{p^{1/p} q^{1/q}} \left(1 - \frac{\eta_1}{\beta^{p-1}} + \frac{\eta_2}{\beta^{2p-2}}\right) + \mathrm{o}\left(\beta^{1/p-2p+2}\right), \quad as \quad \alpha \to 1,$$

where $\eta_1 = p^{-1}(p-1)^{p-1}$ and

$$\eta_2 = p^{-2}(p-1)^{2p-1} + p^{-1}(p-1)^{2p} + (2p^2)^{-1}(p-1)^{p-1}(p+1).$$

The strategy of the proof is basically to use the same technique as in the proofs of Theorems 4 and 5, with some tiny modifications when dealing with the case $p < 2$. The calculations are so tedious that we do not elaborate further here, deferring the proof to Appendix B; instead, we now discuss the results of the theorem in detail. Surprisingly, Theorems 5 and 7 indicate that the main terms of the asymptotic representations of $\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}}$ and $\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}$ are of the same order, and the corresponding coefficients are also equal, when $p > 1$, $\operatorname{ess\,sup}_G < \infty$, and $\sigma = \varepsilon$ (noting that $p^{-1/p} q^{-1/q} = (p-1)^{1/q}/p$). Alternatively, if we ignore the constants and the terms of lower order than $\beta^{1/p}$, the asymptotic behavior of $\mathrm{WCE}_\alpha^{\mathcal{M}}$ is totally captured by $C p^{-1/p} q^{-1/q} \beta^{1/p}$, where $C = \varepsilon$ for $\mathcal{M} = \mathcal{W}_{p,G,\varepsilon}$ and $C = \sigma$ for $\mathcal{M}_{p,\mu,\sigma}$. We formally record this observation as follows.

Corollary 2. *If $p > 1$ and $\operatorname{ess\,sup}_G < \infty$, then*

$$\lim_{\alpha \to 1} \frac{\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}}}{p^{-1/p} q^{-1/q} \beta^{1/p}} = \sigma, \qquad \lim_{\alpha \to 1} \frac{\mathrm{WCE}_\alpha^{\mathcal{W}_{p,G,\varepsilon}}}{p^{-1/p} q^{-1/q} \beta^{1/p}} = \varepsilon.$$

It can be seen from Equation (2.11) that for any given μ and σ, $\mathrm{WCE}_\alpha^{\mathcal{M}_{p_1,\mu,\sigma}} < \mathrm{WCE}_\alpha^{\mathcal{M}_{p_2,\mu,\sigma}}$ for any $p_1 > p_2 > 1$. This roughly characterizes the effect of the reduction in ambiguity on

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press

the worst-case value of the expectile, while our results describe in detail the influence of p on $\mathrm{WCE}_\alpha^{\mathcal{M}_{p,\mu,\sigma}}$, at extreme levels. Regarding Theorem 7 alone, the first residual term $\mathrm{O}(\beta^{1/p-p+1})$ is $o(1)$ if and only if $p > (\sqrt{5}+1)/2 \approx 1.618$, and the second residual term $\mathrm{O}(\beta^{1/p-2p+2})$ is $o(1)$ if and only if $p > (\sqrt{3}+1)/2 \approx 1.366$. It seems that when $p \to 1$, the error increases greatly and there are essential difficulties in finding uniform bounds to control the error.

4. Concluding remarks and future work

In this paper, we have presented the worst-case value of the expectile over two common ambiguity sets, one specified as a ball in the p-Wasserstein metric and the other specified by a moment constraint. We initially focus on the p-Wasserstein ball, for which we reformulate the problem as a convex optimization problem and provide a method of constructing elements that attain or approach the worst-case expectile. In this case, it is established that the precise value of the worst-case expectile can be derived only under specific conditions. However, the asymptotic behavior of the worst-case expectile at extreme levels is thoroughly understood for a wide range of distribution functions. The technical results and simulation studies described in the supplementary material show that our approximation is remarkably accurate, and the general trends of worst-case expectiles at extreme levels are independent of the reference distribution. Furthermore, we observe significant alterations in the shape of the distribution function within the p-Wasserstein ball for which the extreme worst-case expectile is achieved; this highlights a feature of distributional ambiguity.

We also compare our worst-case approach with the model aggregation approach in the case of the expectile; the latter relies not on a specific risk measure but on the choice of stochastic order. We observe that the model aggregation approach yields a more conservative risk evaluation at both fixed and extreme levels. Additionally, we attempt to reformulate our problem using the general results of [2] on robust shortfall risk measures, and we re-derive Theorem 2(i) from this new perspective. In parallel, we apply our techniques to investigate the worst-case expectile over an ambiguity set with constraints on higher-order moments.

Although we have systematically studied the worst-case value of the expectile, there are still some restrictions on our results, which reveal a few promising directions for future study. First, we would like to remove the strong assumption $\operatorname{ess\,sup}_G < \infty$ in Theorems 5 and 6, which is essential for our proof. Second, one may want to offer more accurate and even unified approximations for the worst-case expectiles over both the p-Wasserstein ball and the moment-constrained ambiguity set, to describe their asymptotic behavior in a more general setting. Finally, a potential future problem would be to adapt the worst-case expectile to robust portfolio selection, a topic that has already attracted great interest in recent years [12, 14, 20, 31].

Acknowledgements

The authors thank the editor and the anonymous referees for their helpful comments, which greatly improved the quality of this paper. Y. Chen and T. Mao are co-corresponding authors.

Funding information

This work is supported by the National Natural Science Foundation of China (Nos. 12371279, 12371476) and NSF of Anhui Province (Nos. 2208085MA05, 2208085MA07).

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

Supplementary material

The supplementary material for this article can be found at https://dx.doi.org/10.1017/apr.2024.10.

References

[1] AHMAD, A. A. *et al.* (2020). Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model. *Electron. J. Statist.* **14**, 4421–4456.

[2] BARTL, D., DRAPEAU, S. AND TANGPI, L. (2020). Computational aspects of robust optimized certainty equivalents and option pricing. *Math. Finance* **30**, 287–309.

[3] BELLINI, F. AND BIGNOZZI, V. (2015). On elicitable risk measures. *Quant. Finance* **15**, 725–733.

[4] BELLINI, F., KLAR, B., MÜLLER, A. AND ROSAZZA-GIANIN, E. (2014). Generalized quantiles as risk measures. *Insurance Math. Econom.* **54**, 41–48.

[5] BERNARD, C., PESENTI, S. M. AND VANDUFFEL, S. (2022). *Robust distortion risk measures*. Preprint. Available at https://arxiv.org/abs/2205.08850.

[6] CHEN, L., HE, S. AND ZHANG, S. (2011). Tight bounds for some risk measures, with applications to robust portfolio selection. *Operat. Res.* **59**, 847–865.

[7] CHEN, Z., YU, P. AND HASKELL, W. B. (2019). Distributionally robust optimization for sequential decision-making. *Optimization* **68**, 2397–2426.

[8] DELAGE, E. AND YE, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operat. Res.* **58**, 595–612.

[9] DUNKEL, J. AND WEBER, S. (2010). Stochastic root finding and efficient estimation of convex risk measures. *Operat. Res.* **58**, 1505–1521.

[10] EMBRECHTS, P., WANG, B. AND WANG, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance Stoch.* **19**, 763–790.

[11] FÖLLMER, H. AND SCHIED, A. (2002). Convex measures of risk and trading constraints. *Finance Stoch.* **6**, 429–447.

[12] GHAOUI, L. E., OKS, M. AND OUSTRY, F. (2003). Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. *Operat. Res.* **51**, 543–556.

[13] GIESECKE, K., SCHMIDT, T. AND WEBER, S. (2008). Measuring the risk of large losses. *J. Investment Manag.* **6**, 1–15.

[14] GLASSERMAN, P. AND XU, X. (2014). Robust risk measurement and model risk. *Quant. Finance* **14**, 29–58.

[15] GUO, S. AND XU, H. (2019). Distributionally robust shortfall risk optimization model and its approximation. *Math. Program.* **174**, 473–498.

[16] HU, Z. AND DALI, Z. (2016). Convex risk measures: efficient computations via Monte Carlo. Available at https://ssrn.com/abstract=2758713.

[17] LIU, F., MAO, T., WANG, R. AND WEI, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Math. Operat. Res.* **47**, 2494–2519.

[18] MAO, T., WANG, R. AND WU, Q. (2022). *Model aggregation for risk evaluation and robust optimization*. Preprint. Available at https://arxiv.org/abs/2201.06370.

[19] MOHAJERIN-ESFAHANI, P. AND KUHN, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Math. Program.* **171**, 115–166.

[20] NATARAJAN, K., PACHAMANOVA, D. AND SIM, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Manag. Sci.* **54**, 573–585.

[21] NEWEY, W. K. AND POWELL, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**, 819–847.

[22] PANARETOS, V. M. AND ZEMEL, Y. (2020). *An Invitation to Statistics in Wasserstein Space*. Springer, Cham.

[23] PESENTI, S. M., MILLOSSOVICH, P. AND TSANAKAS, A. (2016). Robustness regions for measures of risk aggregation. *Dependence Modeling* **4**, 348–367.

[24] PESENTI, S. M., WANG, Q. AND WANG, R. (2020). *Optimizing distortion risk metrics with distributional uncertainty*. Preprint. Available at https://arxiv.org/abs/2011.04889.

[25] ROCKAFELLAR, R. T. AND URYASEV, S. (2000). Optimization of conditional value-of-risk. *J. Risk* **2**, 21–41.

[26] SHAPIRO, A. (2017). Distributionally robust stochastic programming. *SIAM J. Optimization* **27**, 2258–2275.

[27] VILLANI, C. (2009). *Optimal Transport: Old and New*. Springer, Berlin.
[28] WANG, Q., WANG, R. AND WEI, Y. (2020a). Distortion risk metrics on general spaces. *ASTIN Bull.* **50**, 827–851.
[29] WANG, R., WEI, Y. AND WILLMOT, G. E. (2020b). Characterization, robustness and aggregation of signed Choquet integrals. *Math. Operat. Res.* **45**, 993–1015.
[30] WIESEMANN, W., KUHN, D. AND SIM, M. (2014). Distributionally robust convex optimization. *Operat. Res.* **62**, 1203–1466.
[31] ZHU, M. S. AND FUKUSHIMA, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operat. Res.* **57**, 1155–1168.
[32] ZIEGEL, J. F. (2016). Coherence and elicitability. *Math. Finance* **26**, 901–918.

https://doi.org/10.1017/apr.2024.10 Published online by Cambridge University Press