

ROBUST SPECTRAL RISK OPTIMIZATION WHEN INFORMATION ON RISK SPECTRUM IS INCOMPLETE*

WEI WANG[†] AND HUIFU XU[‡]

Abstract. A spectral risk measure (SRM) is a weighted average of value at risk where the weighting function (also known as risk spectrum or distortion function) characterizes a decision maker’s risk attitude. In this paper, we consider the case where the decision maker’s risk spectrum is ambiguous and introduce a robust SRM model based on the worst risk spectrum from a ball of risk spectra centered at a nominal risk spectrum. When the ball consists of step-like risk spectra, we show that the robust SRM can be computed by solving a linear programming problem. For the general case, we propose a step-like approximation scheme and derive an error bound for the approximation. As an application, we apply the proposed robust SRM to one-stage stochastic optimization with the objective of minimizing the robust SRM and propose an alternating iterative algorithm for solving the resulting minimax optimization problem. Moreover, to examine stability of the robust spectral risk optimization model with respect to perturbation of observed data from the underlying exogenous uncertainty in data-driven environments, we investigate statistical robustness of the model and derive sufficient conditions for the required stability.

Key words. robust spectral risk measure, robust spectral risk optimization, step-like approximation, alternating iterative algorithm, statistical robustness

AMS subject classifications. 90C15, 60B05, 62P05

DOI. 10.1137/19M1284270

1. Introduction. A spectral risk measure (SRM) is a weighted average of value at risk (VaR) where the weighting function captures an investor’s risk preference. The weighting function is formally called a risk spectrum as it determines the nature of SRM. Since its introduction by Acerbi [1], SRM has received wide attention in that it encompasses all of the important properties that the well-known conditional value at risk (CVaR) enjoys and it provides a more flexible structure than CVaR for practical use. Like CVaR, SRM can also be used as a basis for representing any law invariant coherent risk measure (Pichler and Shapiro [26]) and this ensures a wide scope of applicability of SRM given that the latter is well accepted in risk management. Moreover, SRM is closely related to the distortion risk measure (DRM) in actuarial science [7, 31] because each risk spectrum corresponds to a distortion function (a probability modification function which overstates the probabilities of extreme events, both of adverse and of favorable ones); see [17, 35, 30]. An important common characteristic of SRM and DRM is that they both can be represented as a risk functional on the prospect space of probability distributions or quantile functions (also known as lotteries) instead of random variables (acts). Research of risk measures in these spaces is well documented; we refer readers to Frittelli, Maggis, and Peri [12] for a class of quasi-convex risk measures defined on the space of probability measures and Dentcheva and Ruszczyński [8] for a generic class of law invariant risk measures

*Received by the editors August 28, 2019; accepted for publication (in revised form) August 30, 2020; published electronically November 24, 2020.

<https://doi.org/10.1137/19M1284270>

Funding: The first author is a Ph.D. candidate funded by SCDTP. The second author was supported by a CUHK startup grant and RGC grant 14500620.

[†]School of Business, University of Southampton, Southampton, SO17 1BJ, UK (ww1e17@soton.ac.uk).

[‡]Corresponding author. Department of Systems Engineering & Engineering Management, The Chinese University of Hong Kong, Hong Kong (hfxu@se.cuhk.edu.hk).

defined on the space of quantile functions. In a more recent development, Haskell, Huang, and Xu [16] extend the research to a broad class of multiattribute choice functions defined over the space of survival functions.

In the current research of behavioral economics and risk measurement, the risk spectrum or the probability weighting function is assumed to be known or can be elicited through empirical data. In practice, however, an investor may not have a clear-cut risk spectrum either because there is inadequate information to elicit a unique risk spectrum which captures investor's risk preference precisely or the investment problem involves several stakeholders who fail to reach a consensus. Likewise, an investor may lack overwhelming experimental evidence to identify an appropriate probability weighting function to describe people's attitudes toward gains and losses in a lottery. Under these circumstances, it might be sensible to use partially available information to construct a set of risk spectra which reflect the investor's risk attitude and consider a robust spectral risk measure (RSRM) which is based on the worst risk spectrum from the set to mitigate the risk arising from ambiguity of the investor's risk attitude. For instance, a banker may find a borrower's risk attitude roughly through some questionnaires and/or his historic investment data and make a decision based on the borrower's worst credit scenario. Likewise, an insurer may use some available information to construct a risk profile of a customer and determine the premium on the basis of the worst probability weighting function.

This kind of robust approach was first considered by Armbruster and Delage [2], who propose a preference robust optimization (PRO) model for decision-making problems where information on a decision maker's utility preference is incomplete. Moreover, Armbruster and Delage propose various approaches including pairwise comparison, exponential fit, certainty equivalent, and stochastic dominance to elicit an ambiguity set of utility functions and propose a maximin model where the optimal decision is based on the worst expected utility value calculated from the ambiguity set. Under some moderate conditions, they demonstrate that the resulting PRO problems can be solved efficiently. Delage and Li [6] extend the research to the risk management problem where there is ambiguity to choose a risk measure which represents a decision maker's risk attitude. One of the distinct features of their model is that the risk measure is defined in the output space of random losses. For the PRO model, they derive a tractable formulation for solving the resulting robust risk optimization problem when the loss function is concave with respect to the decision variables. Haskell, Huang, and Xu [16] propose a robust choice function model which is defined on the space of probability distributions. Their choice function is quasi-concave and the resulting maximin problem is a mixed integer program. A decomposition method is proposed to solve the latter. Moreover, Hu and Mehrotra [18] propose a novel moment-type approach for constructing the ambiguity set of a decision maker's utility preference which covers a number of important approaches such as certainty equivalent and pairwise comparison. Hu and Stepanyan [19] propose a reference-based almost stochastic dominance rule which balances the general representation of risk aversion and the individualization of the decision maker's risk preference by constructing a neighborhood of the decision maker's individual utility function and representing a preference relation over this neighborhood.

In this paper, we follow up this stream of research by introducing an RSRM to address the case when information on the true risk spectrum is incomplete. Given the relationship between SRM and the law invariant coherent risk measure [26], the proposed RSRM can be applied to all decision makers whose risk preferences are described by law invariant coherent risk measures. Specifically, we propose a minimax

spectral risk optimization model with the ambiguity set being defined by a ball of risk spectra centered at a nominal risk spectrum in a functional space equipped with some pseudometric. The nominal risk spectrum is determined by either empirical data or subjective judgement. Differing from Delage and Li [6], our model is intrinsically law invariant.

We also take a step further to discuss stability of our robust spectral risk optimization model against variation of exogenous uncertainty and this is particularly relevant in a data-driven environment where the true probability distribution has to be approximated with empirical data. Cont, Deguest, and Scandolo [4] emphasize that measuring the risk of a financial portfolio often involves two steps, estimating the loss distribution of the portfolio from available observations and computing a risk measure that summarizes the risk of this loss distribution, and advocate that these two steps should be considered and studied together. A risk measure is said to be robust if small variations in loss distribution which result from either estimation or misspecification errors lead to small variations in the estimator of the risk measure. Cont, Deguest, and Scandolo also find that expected shortfall is much more sensitive than VaR to large observations of data and remark that statistical robustness, and not only “coherence,” should be a concern for regulators and end-users when choosing or designing risk measurement procedures. The definition of stability is based on Hampel’s classical notion of qualitative robustness in statistics [15, 20]. Krätschmer, Schied, and Zähle [22] argue that the notion is not suitable for risk measurement and propose a refined notion of robustness that applies to tail-dependent law invariant convex risk measures in Orlicz spaces. Guo and Xu [14] take a step further to consider quantitative statistical robustness for utility PRO. The research motivates us to study whether our RSRM is statistically robust.

The main contributions of this paper can be summarized as follows.

First, we propose a preference RSRM model where the decision maker’s ambiguity on risk preference is described by a set of risk spectra and provide a new mathematical framework for studying robust law invariant coherent risk measures. It differs from [6], which deals with ambiguity of convex risk measure, in that the latter is defined over the output space of loss functions and requires some specific treatment to deal with law invariant convex risk measures. Like the robust choice function model [16] where the choice function is defined on the space of probability measures/distributions, the RSRM can also be represented as a risk functional on the space of probability distributions (more precisely on the space of quantile functions), but our model is convex w.r.t. random variables (acts) whereas the robust choice function model is intrinsically quasi-convex although the two models have slightly different focuses.

Second, we propose a new way to construct the ambiguity set, that is, a ball of risk spectra under some pseudometric. The approach differs from various existing approaches in the literature of PRO [2, 6, 18, 19]. To compute the resulting robust risk measure and solve the subsequent robust risk optimization problem efficiently, we propose a step-like function approximation scheme and derive error bounds under some moderate conditions. We also show that the approximated RSRM can be computed by solving a linear programming (LP) problem. In comparison with the robust approach in [2, 6, 16, 18, 19], the problem size of the LP is independent of the samples of the underlying random data. Moreover, we propose an alternating iterative algorithm which always solves the inner linear maximization problem and outer convex minimization problem alternatively for the approximated robust spectral risk optimization problem globally.

Third, to address stability of the newly developed robust spectral risk optimization model against perturbation of the probability distribution of the underlying uncertainty data, we derive a quantitative stability result for the optimal value (Theorem 5.1). Moreover, we take a step further to investigate the case when the unknown true probability distribution is approximated by perceived empirical data which may contain noise and derive a quantitative stability result for examining the quality of the estimator of the optimal value (Theorem 5.2).

The rest of the paper is organized as follows. In section 2, we set up the RSRM and the robust spectral risk optimization models. In section 3, we discuss construction of the ambiguity set and develop numerical schemes for calculating the RSRM and solving the robust spectral risk optimization problem. In section 4, we derive error bounds for the step-like approximation of the ambiguity set and its impact on the optimal value. In section 5, we investigate quantitative statistical stability of the proposed robust models. Some numerical results are reported in section 6.

2. Problem setup. Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ be a financial loss, where Ω is a sample space with sigma algebra \mathcal{F} and \mathbb{P} is a probability measure. Let $F_X(x) := \mathbb{P}(X \leq x)$ be the cumulative distribution function (CDF) of X and for $\alpha \in [0, 1]$ its generalized inverse is defined as $F_X^\leftarrow(\alpha) := \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}$. By writing $F_X(x)$ as $\mathbb{P}(-X + x \geq 0)$, we may interpret $F_X^\leftarrow(\alpha)$ as minimal cash to be injected into the financial position $-X$ such that the probability of the new position (after injection of the cash) being nonnegative (a position of profit) is no less than the specified level α . $F_X^\leftarrow(\alpha)$ is also known as VaR or as a quantile function mapping from $[0, 1]$ to \mathbb{R} . Let $\sigma : [0, 1] \rightarrow \mathbb{R}$ be a nonnegative, nondecreasing function with the normalized property $\int_0^1 \sigma(t)dt = 1$ and \mathfrak{S} denote the set of all such σ . The SRM of X with $\sigma \in \mathfrak{S}$ is defined as

$$(2.1) \quad (\text{SRM}) \quad M_\sigma(X) := \int_0^1 F_X^\leftarrow(t)\sigma(t)dt,$$

where σ is called a *risk spectrum*. Obviously σ plays a role of weighting and $M_\sigma(X)$ is the weighted average loss of X . Moreover, if we change the variable in the integral (2.1) by setting $x = F_X^\leftarrow(t)$, then

$$(2.2) \quad M_\sigma(X) = \int_{-\infty}^{\infty} x\sigma(F_X(x))dF_X(x),$$

in which case σ may be interpreted as a distortion function of probability. It is easy to verify that $M_\sigma(X)$ is a law invariant coherent risk measure (satisfying monotonicity, subadditivity, positive homogeneity, and translation invariance). The increasing property of σ is essential to make $M_\sigma(X)$ a coherent risk measure because it assigns bigger weights to worse cases. Note that if we drop the property, then $M_\sigma(X)$ is still a monetary risk measure which is monotonic, positive homogeneous, and translation invariant. Our focus in this paper will be on the case that $M_\sigma(X)$ is coherent although all of the results developed in this paper are applicable to the risk measures without subadditivity. The following example lists a few well-known SRM in the literature.

Example 2.1. (i) CVaR. Let $\sigma_\alpha(\tau) := \frac{1}{1-\alpha}\mathbf{1}_{[\alpha,1]}(\tau)$, where $\mathbf{1}_{[\alpha,1]}$ is the indicator function and $\alpha \in (0, 1)$. Then $M_{\sigma_\alpha}(X) = \frac{1}{1-\alpha} \int_\alpha^1 F_X^\leftarrow(t)dt =: \text{CVaR}_\alpha(X)$.

(ii) Wang's proportional hazards transform or power distortion [33]. Let $\sigma_\nu(\tau) = \nu(1 - \tau)^{\nu-1}$, where $\nu \in (0, 1]$ is a parameter. Then $M_{\sigma_\nu}(X) = \int_0^\infty (1 - F_X(t))^\nu dt$.

(iii) Gini's measure [33]. Let $\sigma_s(\tau) = (1 - s) + 2s\tau$, where $s \in (0, 1)$. Then $M_{\sigma_s}(X) = \mathbb{E}[X] + s\mathbb{E}(|X - X'|) =: \text{Gini}_s(X)$, where X' is an independent copy of X .

(iv) Convex combination of expected value and CVaR. Let $\sigma_\lambda(\tau) = \lambda\mathbf{1}_{[0,1]}(\tau) + (1 - \lambda)\frac{1}{1-\alpha}\mathbf{1}_{[\alpha,1]}(\tau)$, for some $\lambda \in [0, 1]$, where $\mathbf{1}_{[a,b]}$ is the indicator function over interval $[a, b]$. Then $M_{\sigma_\lambda}(X) = \lambda\mathbb{E}[X] + (1 - \lambda)\text{CVaR}_\alpha(X)$.

2.1. Robust spectral risk measure. Throughout the paper, we focus on decision-making problems where the decision maker's risk attitude can be described by a risk spectrum. In practice, however, identifying a unique risk spectrum which precisely reflects a decision maker's risk attitude may turn out to be difficult either because the decision-making problem involves several stakeholders who fail to reach a consensus or the prospect space is too complex and/or too large to elicit such a risk spectrum. Under these circumstances, it might be sensible to use available partial information to construct a set of risk spectra and consider the worst risk spectrum for the SRM to mitigate the risk arising from ambiguity of the true risk spectrum.

Let \mathfrak{A} denote the ambiguity set of risk spectra constructed by an investor with available information. We define the RSRM by

$$(2.3) \quad (\text{RSRM}) \quad \rho_{\mathfrak{A}}(X) := \sup_{\sigma \in \mathfrak{A}} \int_0^1 F_X^\leftarrow(t)\sigma(t)dt,$$

which is the worst weighted average loss computed from the set of weighting functions \mathfrak{A} . Note that if we let $dG_X(x) = \sigma(F_X(x))dF_X(x)$ in (2.2), then

$$(2.4) \quad \rho_{\mathfrak{A}}(X) = \sup_{G_X \in \mathfrak{B}} \int_{-\infty}^{\infty} xdG_X(x),$$

where $\mathfrak{B} = \{G_X(x) = \int_{-\infty}^x \sigma(F_X(t))dF_X(t) : \sigma \in \mathfrak{A}\}$, in which case we may interpret (2.4) as a *distributionally robust model*; see [30]. However, the latter robust representation is not particularly helpful in the sense that it obscures the sources of ambiguity resulting from σ or from F_X .

Since for each fixed σ , $M_\sigma(\cdot)$ is a law invariant coherent risk measure, then $\rho_{\mathfrak{A}}(X)$ is also a law invariant coherent risk measure because the supremum operation preserves law invariance and coherence. Conversely, by the Kusuoka representation theorem [26], any law invariant coherent risk measure ρ can be represented as

$$(2.5) \quad \rho(X) = \sup_{\mu \in \mathcal{M}} \int_0^1 \text{CVaR}_\alpha(X)\mu(d\alpha),$$

where \mathcal{M} is a subset of \mathcal{P} and \mathcal{P} is the set of all probability measures on $[0, 1]$. Consider the linear mapping $\mathbb{T} : \mathcal{P} \rightarrow \mathfrak{S}$ defined by $(\mathbb{T}\mu)(\tau) := \int_0^\tau (1 - \alpha)^{-1}\mu(d\alpha)$, $\tau \in [0, 1]$. It follows by Pichler and Shapiro [26] that \mathbb{T} is onto, one-to-one with the inverse $\mu = \mathbb{T}^{-1}\sigma$ given by $\mu(\alpha) = (\mathbb{T}^{-1}\sigma)(\alpha) = (1 - \alpha)\sigma(\alpha) + \int_0^\alpha \sigma(\tau)d\tau$, $\forall \alpha \in [0, 1]$. Thus for each fixed μ , by setting $\sigma(\tau) = (\mathbb{T}\mu)(\tau)$ and using Fubini's theorem, we have

$$\begin{aligned} \int_0^1 \text{CVaR}_\alpha(X)\mu(d\alpha) &= \int_0^1 \int_\alpha^1 (1 - \alpha)^{-1} F_X^\leftarrow(\tau)d\tau \mu(d\alpha) = \int_0^1 \int_0^\tau (1 - \alpha)^{-1} \mu(d\alpha) F_X^\leftarrow(\tau)d\tau \\ &= \int_0^1 (\mathbb{T}\mu)(\tau) F_X^\leftarrow(\tau)d\tau = \int_0^1 F_X^\leftarrow(\tau) \sigma(\tau)d\tau. \end{aligned}$$

Let $\mathcal{M} = \{\mu = \mathbb{T}^{-1}\sigma : \sigma \in \mathfrak{A}\}$. Then we obtain a Kusuoka representation for RSRM in (2.5). The following proposition summarizes the discussion which is a combination of the Kusuoka representation theorem [23] and [26, Proposition 3.1].

PROPOSITION 2.1. *A real-valued function ρ is a law invariant coherent risk measure if and only if there exists an ambiguity set \mathfrak{A} of risk spectra such that ρ has a representation of (2.3).*

The proposition implies that RSRM can be used as a mathematical framework for us to study ambiguity of any law invariant coherent risk measure. To see this, let \mathcal{R} denote an ambiguity set of law invariant coherent risk measures. By Proposition 2.1, we can find an ambiguity set of risk spectra $\mathfrak{A}_{\mathcal{R}}$ corresponding to \mathcal{R} such that

$$(2.6) \quad \rho_{\mathfrak{A}_{\mathcal{R}}}(X) = \sup_{\sigma \in \mathfrak{A}_{\mathcal{R}}} M_{\sigma}(X) = \sup_{\rho \in \mathcal{R}} \rho(X).$$

In practice, we do not have to elicit \mathcal{R} in the first place and then translate it into $\mathfrak{A}_{\mathcal{R}}$; rather, we may elicit the latter directly (see Remark 3.1(iv)). In other words, our RSRM model provides an alternative robust approach to address ambiguity of decision makers whose risk attitude can be described by law invariant coherent risk measures.

Pichler and Shapiro [26] call a law invariant risk measure represented in the form of (2.3) *regular* and formulation (2.3) a dual representation with generating set \mathfrak{A} . They discuss in detail existence and uniqueness of the minimal generating set \mathfrak{A} for the dual representation of a given law invariant coherent risk measure but are short of saying how such a generating set can be identified. Our focus here is not on how to find the generating set \mathfrak{A} for a specified law invariant coherent risk measure; rather we concentrate on using partial information to construct an ambiguity set which faithfully represents a decision maker's risk preference. Thus in this paper \mathfrak{A} is an ambiguity set rather than a generating set although mathematically they are in the same position. Depending on availability of information on the decision maker's risk attitude, various ways such as the parametric approach [34], the pairwise comparison [2], and the moment type approach [18] may be exploited to construct the ambiguity set. For example, if we consider a risk measure defined in Example 2.1(v) but are short of identifying the particular λ value, then we may consider an ambiguity set $\mathfrak{A} = \{\sigma_{\lambda}(\tau) : \lambda \in [0, 1]\}$. In this case, $\rho_{\mathfrak{A}}(X) = \max(\mathbb{E}[X], \text{CVaR}_{\alpha}(X))$. In this paper we consider a situation where a decision maker has a nominal risk spectrum obtained from partial information such as empirical data and/or subjective judgment but there is some ambiguity about whether the nominal risk spectrum accurately represents the decision maker's risk attitude. Consequently we construct a ball of risk spectra under some pseudometric centered at the nominal risk spectrum. We will discuss the details in the next section.

As a motivation as well as an application, we will apply the proposed RSRM to optimal decision-making problems where the optimal decision is based on the worst SRM from the ambiguity set \mathfrak{A} , that is,

$$(2.7) \quad (\text{RSRM-Opt}) \quad \min_{z \in Z} \max_{\sigma \in \mathfrak{A}} M_{\sigma}(f(z, \xi(\omega))),$$

where $f : Z \times \mathbb{R}^k \rightarrow \mathbb{R}$ is a continuous loss function, z is a decision vector, $\xi(\omega)$ is a mapping from Ω to \mathbb{R}^k , and Z is a compact and convex set. When \mathfrak{A} is compact, the inner maximum in (RSRM-Opt) can be achieved because $M_{\sigma}(\cdot)$ is continuous and the outer minimum can also be attained because $\rho_{\mathfrak{A}}(f(z, \xi))$ is continuous in z and Z is compact. Moreover, if $f(z, \xi)$ is convex in z for every fixed ξ , then $M_{\sigma}(f(z, \xi(\omega)))$ is convex in z given the fact that $M_{\sigma}(\cdot)$ is a coherent risk measure. In this case the optimal value of the inner maximization problem of (RSRM-Opt) is a convex function of z and hence the outer minimization problem of (RSRM-Opt) is a convex

program. An important component of the research in this paper is to develop efficient computational schemes for solving (RSRM-Opt). We will discuss this in detail in section 4. We will also investigate statistical robustness of the optimal value of problem (RSRM-Opt) in a data-driven environment when the distribution of ξ is obtained from empirical data. Before we conclude this subsection, we remind readers that in both RSRM and (RSRM-Opt) we assume the decision maker's risk attitude can be characterized by a risk spectrum. (RSRM-Opt) cannot be applied to the cases if the decision maker's risk preference can only be described by other risk measures;¹ we refer readers to [5] for preference robust utility-based shortfall risk optimization models and [6] for general preference robust convex risk optimization models.

2.2. Robust distortion risk measure. It is well known that SRM is closely related to DRM, which was introduced by Wang [31] based on Yaari's dual theory of choice [35]; see, for instance, [17]. Over the past few decades, DRM has been extensively studied in behavioral economics and actuarial science; see the monograph [28]. This motivates us to extend the notion of RSRM to that of DRM.

Let $w : [0, 1] \rightarrow [0, 1]$ be a nondecreasing function with $w(0) = 0$ and $w(1) = 1$. The *DRM* induced by w is defined by $D_w(X) := \int_0^\infty w(1 - F_X(x))dx + \int_{-\infty}^0 [w(1 - F_X(x)) - 1]dx$, where the effect of w is to modify ("distort") the decumulative probability distribution $1 - F_X(x)$ and consequently w is called a *distortion function*. Note that DRM can be derived as an insurance premium based on the principal of dual theory of choice in which case X is interpreted as a liability. In particular, when X is strictly positive, the DRM ends up with the first term in the above equation, a DRM introduced in the context of insurance pricing by Denneberg [7] and Wang [31]; see also [30]. Let \mathfrak{D} denote the ambiguity set of distortion functions. We define the *robust distortion risk measure* (RDRM) by

$$(2.8) \quad (\text{RDRM}) \quad \rho_{\mathfrak{D}}(X) := \sup_{w \in \mathfrak{D}} D_w(X).$$

Let us now set w with $w'(t) = \sigma(1 - t)$ and $w(0) = 0$. Then w is a distortion function and

$$\begin{aligned} D_w(X) &= \int_{1-F_X(0)}^0 w(s)dF_X^{\rightarrow}(1-s) + \int_1^{1-F_X(0)} [w(s)-1]dF_X^{\rightarrow}(1-s) \\ &= \int_1^0 w(s)dF_X^{\leftarrow}(1-s) - \int_1^{1-F_X(0)} dF_X^{\leftarrow}(1-s) \\ &= - \int_1^0 F_X^{\leftarrow}(1-s)dw(s) - F_X^{\leftarrow}(F_X(0)) = \int_0^1 F_X^{\leftarrow}(\tau)\sigma(\tau)d\tau = M_{\sigma}(X). \end{aligned}$$

Let $\mathfrak{D} := \{w(t) = \int_0^t \sigma(1 - \tau)d\tau, \text{ for } t \in [0, 1] \text{ with } w(0) = 0 : \sigma \in \mathfrak{A}\}$. Then $\rho_{\mathfrak{D}}(X) = \rho_{\mathfrak{A}}(X)$, which means RDRM coincides with DSRM. In this paper, our focus will be on RSRM and (RSRM-Opt) although the computational schemes and underlying theories can be easily applied to RDRM.

2.3. Domain of the robust spectral risk measure. For a given set of risk spectra \mathfrak{A} , it is natural to ask what domain the resulting robust risk measure has,

¹There could be other risk measures which are not law invariant coherent, in which case our robust framework does not apply. For example, standard deviation (SD) is not a coherent risk measure and therefore if a decision maker's risk preference can only be described by SD, then we cannot characterize it with SRM.

that is, what is the set of random variables X such that $\rho_{\mathfrak{A}}(X) < \infty$? To address this issue, we may recall some well-known results for the case when \mathfrak{A} is a singleton. Let \mathcal{L}^0 denote the set of all random variables mapping from Ω to \mathbb{R} . For any positive number $p \in [1, \infty)$, let $\mathcal{L}^p := \{X \in \mathcal{L}^0 : \mathbb{E}[|X|^p] < \infty\}$, which defines a subset of \mathcal{L}^0 with finite p th order moments. For a fixed risk spectrum σ , let $\mathcal{L}^\sigma := \{X \in \mathcal{L}^0 : M_\sigma(|X|) < \infty\}$. By slightly abusing the notation, we use $\mathcal{L}^q[0, 1]$ to denote the class of functions $h : [0, 1] \rightarrow \mathbb{R}$ such that $\int_0^1 |h(t)|^q dt < \infty$. Pichler [25] calls \mathcal{L}^σ the natural domain of M_σ and shows that if $\sigma \in \mathcal{L}^q[0, 1]$ for some $q \in [1, \infty)$ with conjugate exponent p , i.e., $\frac{1}{p} + \frac{1}{q} = 1$, then $\mathcal{L}^p \subset \mathcal{L}^\sigma \subset \mathcal{L}^1$; see [25, Theorem 11]. Let $\mathcal{L}^{\mathfrak{A}} := \{X \in \mathcal{L}^0 : \rho_{\mathfrak{A}}(|X|) < \infty\}$. Then $\mathcal{L}^{\mathfrak{A}}$ is the natural domain of $\rho_{\mathfrak{A}}$. By definition,

$$\mathcal{L}^{\mathfrak{A}} \subset \bigcap_{\sigma \in \mathfrak{A}} \mathcal{L}^\sigma \quad \text{and} \quad \mathcal{L}^p \subset \mathcal{L}^{\mathfrak{A}} \subset \mathcal{L}^1,$$

when $\sigma \in \mathcal{L}^q[0, 1] \forall \sigma \in \mathfrak{A}$. Moreover equality holds in the first inclusion above when \mathfrak{A} is compact under some weak topology. To see this, let $X \in \bigcap_{\sigma \in \mathfrak{A}} \mathcal{L}^\sigma$. Then $M_\sigma(|X|) < \infty \forall \sigma \in \mathfrak{A}$, which implies $\rho_{\mathfrak{A}}(X) = \sup_{\sigma \in \mathfrak{A}} M_\sigma(|X|) < \infty$ if the set \mathfrak{A} is compact under some weak topology; this shows $X \in \mathcal{L}^{\mathfrak{A}}$ and hence $\bigcap_{\sigma \in \mathfrak{A}} \mathcal{L}^\sigma \subset \mathcal{L}^{\mathfrak{A}}$.

For unbounded risk spectrum σ , since σ is nonnegative and nondecreasing and $\int_0^1 \sigma(t) dt = 1$, then the only unbounded point of it is $t = 1$. In comparison with $\sigma \in \mathcal{L}^1[0, 1]$, the restriction $\sigma \in \mathcal{L}^q[0, 1]$ for $q > 1$ means that $\sigma(t)$ may go to infinity more rapidly than it could be as $t \rightarrow 1$. This means more weights are assigned to tail losses and hence the decision maker is more risk averse. Consider, for example, $\sigma_1(t) = \frac{1}{2}(1-t)^{-\frac{1}{2}}$ and $\sigma_2(t) = \frac{1}{3}(1-t)^{-\frac{2}{3}}$. It is easy to see that $\sigma_1, \sigma_2 \in \mathcal{L}^1[0, 1]$, but $\sigma_1 \notin \mathcal{L}^2[0, 1]$, whereas $\sigma_2 \in \mathcal{L}^2[0, 1]$. Obviously $\sigma_2(t)$ goes to infinity more rapidly than $\sigma_1(t)$ when t approaches to 1. This motivates us to introduce a definition of the so-called tail risk index which captures the behavior of $\sigma(t)$ as t goes to 1.

DEFINITION 2.1 (tail risk index). *Let $\sigma \in \mathfrak{S}$ be unbounded. The tail risk index of σ is a single value r_0 such that $\sigma(t)(1-t)^{\frac{1}{r}}$ goes to either zero or infinity according to whether r is either below or above r_0 as $t \uparrow 1$.*

According to the definition, the tail risk index of Wang's proportional hazards transform defined in Example 2.1(iii) is $r_0 = \frac{1}{1-r}$. Obviously the smaller the tail risk index, the more risk averse a decision maker is. Moreover, if $\sigma \in \mathcal{L}^q[0, 1]$, then the tail risk index of σ is upper bounded by q . Conversely, if σ has a tail risk index of q , then $\sigma \in \mathcal{L}^{q+\epsilon}[0, 1]$ for any $\epsilon > 0$.

3. Construction of the ambiguity set and step-like approximation. In this section, we give details as to how to construct the ambiguity set of risk spectra. There are two main factors that determine construction of the ambiguity set. One is availability of information and the other is tractability of computing the resulting RSRM and solving (RSRM-Opt). We may use various approaches in the literature of PRO to construct \mathfrak{A} if it allows us to do so in practice. Here we consider a situation where there is some partial information for identifying a plausible nominal risk spectrum. For instance, the risk spectrum is a class of parameterized functions with prior distribution of the parameters. By exploiting some sample information (empirical data), we may use sample average or maximal likelihood to identify a risk spectrum [9]. We can then construct a ball of risk spectrum centered at the identified nominal risk spectrum under some metric. This kind of approach is known in the literature of distribution robust optimization; see, for instance, Esfahani and Kuhn [10]. Here we use the approach.

3.1. ψ -ball. An important step to construct a ball of risk spectra is to define an appropriate metric/distance in the space of $\mathcal{L}^1[0, 1]$. Instead of using \mathcal{L}^1 -norm, we use some pseudometric which will not only help us to carry out analysis but also maximize the coverage as we will explain shortly in Remark 3.1(i). Let $\psi : [0, 1] \rightarrow \mathbb{R}_+$ be a positive function with $\int_0^1 \psi(t)dt < +\infty$ and \mathcal{G} be a set of measurable functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $|g(t)| \leq \psi(t)$ for $t \in [0, 1]$. For any $u, v \in \mathfrak{S}$, let

$$(3.1) \quad d_\psi(u, v) := \sup_{g \in \mathcal{G}} |\langle g, u \rangle - \langle g, v \rangle|,$$

where $\langle g, u \rangle = \int_0^1 g(t)u(t)dt$. A sufficient condition for $d_\psi(u, v)$ to be well defined is

$$(3.2) \quad \int_0^1 \psi(t)|u(t) - v(t)|dt < \infty.$$

$d_\psi(u, v)$ is a pseudometric in the sense that $d_\psi(u, v) = 0$ implies $\langle g, u \rangle - \langle g, v \rangle = 0 \forall g \in \mathcal{G}$ but not necessarily $u = v$ unless the set \mathcal{G} is sufficiently large. From a practical point of view, \mathcal{G} may be regarded as a set of test functions such as the quantile functions of some prospects and $d_\psi(u, v) = 0$ means the investor does not tell the difference between u and v for this particular set of investment prospects. Since $u, v \in \mathcal{L}^1[0, 1]$, (3.2) is satisfied when $\psi \in \mathcal{L}^\infty[0, 1]$. Our interest here is in the case that $\psi \in \mathcal{L}^p[0, 1]$ for some $p > 1$, and we may choose $u, v \in \mathcal{L}^q[0, 1]$ with $\frac{1}{p} + \frac{1}{q} = 1$ to guarantee (3.2).

Remark 3.1. We make a few comments to motivate and justify the use of the pseudometric.

(i) Depending on the choice of \mathcal{G} , the pseudometric may cover a number of metrics of interest. For instance, if \mathcal{G} consists of all measurable functions on $[0, 1]$ bounded by 1, then $d_\psi(u, v)$ reduces to \mathcal{L}^1 -distance. If \mathcal{G} consists of all Lipschitz continuous functions with modulus bounded by 1, then $d_\psi(u, v)$ resembles the well-known Kantorovich metric in probability theory [29]. In general, $d_\psi(u, v)$ is not necessarily a full metric.

(ii) The function ψ is usually unbounded. To see this, we may use (RSRM-Opt) as an example in which case \mathcal{G} may contain $\{F_{f(z), \xi}^\leftarrow(t) : z \in Z\}$. If ψ is bounded, then there is a constant C such that $|F_{f(x), \xi}^\leftarrow(t)| \leq C \forall t$, which means the quantile function is bounded uniformly $\forall z \in Z$; this could be too restrictive.

(iii) The pseudometric provides a tighter measurement for discrepancy of two risk spectra. Consider, for example, two risk spectra $u(t) = \frac{1}{2}(1-t)^{-\frac{1}{2}}$ and $v(t) = \frac{1}{3}(1-t)^{-\frac{2}{3}}$. Let $\mathcal{G} = \{u\}$. Then $\psi(t) = u(t)$ for $t \in [0, 1]$, and $d_\psi(u, v) = +\infty$, whereas $\|u - v\|_{\mathcal{L}^1} \leq \|u\|_{\mathcal{L}^1} + \|v\|_{\mathcal{L}^1} = 2$. Moreover, in the case when there is information which indicates that the risk spectrum of a decision maker is in $\mathcal{L}^p([0, 1])$ for some constant $p > 1$, it is inappropriate to use \mathcal{L}^1 -norm to construct an ambiguity ball of risk spectra because otherwise the ambiguity set will contain some elements which are not in $\mathcal{L}^p([0, 1])$.

(iv) Use of the pseudometric will facilitate eliciting a decision maker's risk preference. Consider, for example, two financial losses X_1, X_2 and the decision maker is found to prefer X_1 to X_2 in terms of SRM. We can then describe the risk preference mathematically as $\int_0^1 F_{X_1}^\leftarrow(t)\sigma(t)dt \leq \int_0^1 F_{X_2}^\leftarrow(t)\sigma(t)dt$, which can be equivalently written as $\langle F_{X_1}^\leftarrow(t), \sigma \rangle \leq \langle F_{X_2}^\leftarrow(t), \sigma \rangle$ in a product form in our setting. If we let \mathcal{G} contain all quantile functions in the prospect space, then the pseudometric can be used to measure the differences of all prospects in the space and hence narrow down the

scope of the unknown true risk spectrum. Thus, by adopting the pseudometric, we open a door for including the pairwise comparison approach for eliciting risk spectrum in our proposed model in future research.

DEFINITION 3.1 (ψ -ball of risk spectra). *Let \mathfrak{S} denote the set of all risk spectra $\sigma(\cdot) \in \mathcal{L}^1[0, 1]$ such that σ is nonnegative, nondecreasing, and normalized. Let $\sigma \in \mathfrak{S}$ be the nominal risk spectrum. The ψ -ball of risk spectra in the space of \mathfrak{S} centered at σ with radius r is defined as*

$$(3.3) \quad \mathbb{B}(\sigma, r) := \{\sigma' \in \mathfrak{S} : d_\psi(\sigma', \sigma) \leq r\}.$$

If $\psi \in \mathcal{L}^p[0, 1]$, then $\mathbb{B}(\sigma, r) \supset \{\sigma' \in \mathcal{L}^q[0, 1] : d_\psi(\sigma', \sigma) \leq r\} \neq \emptyset$. Moreover when r is driven to zero, $\mathbb{B}(\sigma, r)$ shrinks to a singleton $\{\sigma\}$ in the case that d_ψ is a full metric. From a practical point of view, σ may be constructed through training data and the radius depends on how much confidence the investor has about σ . In a recent work on the distortion principle for insurance pricing, Escobar and Pflug [9] propose an effective approach to construct a step-like distortion function with empirical data and use it to approximate the true unknown distortion function. Their approach can be used to estimate the nominal risk spectrum σ in our setting.

Note that in the case when \mathfrak{S} is restricted to a specific class of risk spectra, we can derive the ψ -ball explicitly by choosing an appropriate set \mathcal{G} . The next example explains this.

Example 3.1. Consider the three classes of risk spectra defined in Example 2.1.

(i) CVaR decision makers.² Let $\sigma_{\alpha_0} = \frac{1}{1-\alpha_0} \mathbf{1}_{[\alpha_0, 1]}$ denote the nominal spectrum with some $\alpha_0 \in (0, 1)$ and $\mathfrak{S} = \{\frac{1}{1-\alpha} \mathbf{1}_{[\alpha, 1]} : \alpha \in (0, 1)\}$. Let \mathcal{G} be the set of all measurable functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $|g(t)| \leq 1$. For $0 < \alpha < \alpha_0 < 1$, we have

$$\begin{aligned} d_\psi(\sigma_\alpha, \sigma_{\alpha_0}) &= \sup_{g \in \mathcal{G}} |\langle g, \sigma_\alpha \rangle - \langle g, \sigma_{\alpha_0} \rangle| \\ &= \sup_{g \in \mathcal{G}} \left| \int_0^1 g(t) \left(\frac{1}{1-\alpha} \mathbf{1}_{[\alpha, 1]}(t) - \frac{1}{1-\alpha_0} \mathbf{1}_{[\alpha_0, 1]}(t) \right) dt \right| \\ &= \sup_{g \in \mathcal{G}} \left| \int_\alpha^{\alpha_0} \frac{1}{1-\alpha} g(t) dt + \int_{\alpha_0}^1 \left(\frac{1}{1-\alpha} - \frac{1}{1-\alpha_0} \right) g(t) dt \right| \\ &= \sup_{g \in \mathcal{G}} \left| \int_\alpha^{\alpha_0} \frac{1}{1-\alpha} g(t) dt \right| + \sup_{g \in \mathcal{G}} \left| \int_{\alpha_0}^1 \left(\frac{1}{1-\alpha} - \frac{1}{1-\alpha_0} \right) g(t) dt \right| \\ &= \frac{1}{1-\alpha} (\alpha_0 - \alpha) + \left(\frac{1}{1-\alpha_0} - \frac{1}{1-\alpha} \right) (1 - \alpha_0) = \frac{2(\alpha_0 - \alpha)}{1-\alpha}. \end{aligned}$$

Likewise, for $0 < \alpha_0 < \alpha < 1$, we have $d_\psi(\sigma_\alpha, \sigma_{\alpha_0}) = 2(\alpha - \alpha_0)/(1 - \alpha_0)$. Thus, the ψ -ball with radius $r < 2$ can be written as $\mathbb{B}(\sigma_{\alpha_0}, r) = \{\sigma_\alpha : \alpha \in [\frac{2\alpha_0-r}{2-r}, \alpha_0 + \frac{r(1-\alpha_0)}{2}] \cap (0, 1)\}$. Note that in this case, the same ball can be derived by using \mathcal{L}^1 -norm in that

$$\begin{aligned} \|\sigma_\alpha - \sigma_{\alpha_0}\|_{\mathcal{L}^1} &= \int_0^1 |\sigma_\alpha(t) - \sigma_{\alpha_0}(t)| dt = \int_0^1 \left| \frac{1}{1-\alpha} \mathbf{1}_{[\alpha, 1]}(t) - \frac{1}{1-\alpha_0} \mathbf{1}_{[\alpha_0, 1]}(t) \right| dt \\ &= \int_\alpha^{\alpha_0} \frac{1}{1-\alpha} dt + \int_{\alpha_0}^1 \left(\frac{1}{1-\alpha_0} - \frac{1}{1-\alpha} \right) dt = 2 \frac{\alpha_0 - \alpha}{1-\alpha} \end{aligned}$$

for $0 < \alpha < \alpha_0 < 1$, and a similar calculation can be made for the other case.

²Instead of calling decision makers whose risk attitude can be described by CVaR, we take the convention by calling them CVaR decision makers for brevity. A similar convention applies to other risk measures.

(ii) Wang's decision makers. Let \mathfrak{S} be the class of all unbounded risk spectra with the form $\sigma_\nu(t) = \nu(1-t)^{\nu-1}$, where $\nu \in (0, 1]$ and the nominal risk spectrum is σ_{ν_0} . If we choose \mathcal{G} to be the set of all the measurable functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $|g(t)| \leq 1$, then for $0 < \nu < \nu_0 < 1$, we have

$$\begin{aligned} d_\psi(\sigma_\nu, \sigma_{\nu_0}) &= \sup_{g \in \mathcal{G}} |\langle g, \sigma_\nu \rangle - \langle g, \sigma_{\nu_0} \rangle| = \sup_{g \in \mathcal{G}} \left| \int_0^1 g(t) [\nu(1-t)^{\nu-1} - \nu_0(1-t)^{\nu_0-1}] dt \right| \\ &= 2 \left[\left(\frac{\nu}{\nu_0} \right)^{\frac{\nu_0}{\nu_0-\nu}} - \left(\frac{\nu}{\nu_0} \right)^{\frac{\nu}{\nu_0-\nu}} \right]. \end{aligned}$$

Likewise, for $0 < \nu_0 < \nu < 1$, we have $d_\psi(\sigma_\nu, \sigma_{\nu_0}) = 2[(\frac{\nu}{\nu_0})^{\frac{\nu}{\nu_0-\nu}} - (\frac{\nu}{\nu_0})^{\frac{\nu_0}{\nu_0-\nu}}]$. Thus, the ψ -ball with radius r is $\mathbb{B}(\sigma_{\nu_0}, r) = \{\sigma_\nu : 2|(\frac{\nu}{\nu_0})^{\frac{\nu_0}{\nu_0-\nu}} - (\frac{\nu}{\nu_0})^{\frac{\nu}{\nu_0-\nu}}| \leq r \text{ and } 0 < \nu \leq 1\}$. Note that we can derive the same ball using \mathcal{L}^1 -norm and omit the details.

(iii) Gini's decision makers. Let \mathfrak{S} denote the class of all linear risk spectra and σ_{s_0} the nominal risk spectrum (see Example 2.1(iv) for the definition of σ_s). If we choose \mathcal{G} to be the set of all the measurable functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $|g(t)| \leq 1$, then for $0 < s < s_0 < 1$, we have

$$\begin{aligned} d_\psi(\sigma_s, \sigma_{s_0}) &= \sup_{g \in \mathcal{G}} |\langle g, \sigma_s \rangle - \langle g, \sigma_{s_0} \rangle| \\ &= \sup_{g \in \mathcal{G}} \left| \int_0^1 g(t)[((1-s)+2st) - ((1-s_0)+2s_0t)] dt \right| \\ &= \sup_{g \in \mathcal{G}} \left| \int_0^1 g(t)[(2t-1)(s-s_0)] dt \right| \\ &= \sup_{g \in \mathcal{G}} \left| \int_0^1 g(t)(2t-1) dt \right| (s_0-s) = \frac{1}{2}(s_0-s). \end{aligned}$$

Likewise, for $0 < s_0 < s < 1$, we have $d_\psi(\sigma_s, \sigma_{s_0}) = \frac{1}{2}(s-s_0)$. Thus, the ψ -ball with radius $r < 1/2$ can be written as $\mathbb{B}(\sigma_{s_0}, r) = \{\sigma_s : s \in [s_0-2r, s_0+2r] \cap (0, 1)\}$. Note that we can derive the same ball using \mathcal{L}^1 -norm and omit the details.

The example also confirms our claim in Remark 3.1(i) that when \mathcal{G} consists of all measurable functions on $[0, 1]$ bounded by 1, the proposed pseudometric coincides with \mathcal{L}^1 -distance. Note that if we choose \mathcal{G} differently, i.e., a set of all measurable functions $g : [0, 1] \rightarrow \mathbb{R}$ such that $|g(t)| \leq (1-t)^{-\frac{1}{2}}$, then we can obtain a smaller ψ -ball.

3.2. Step-like approximation. With $\mathbb{B}(\sigma, r)$, we can write (2.3) and (2.7) specifically as

$$(3.4) \quad (\text{RSRM}') \quad \rho_{\mathbb{B}(\sigma, r)}(X) := \sup_{\sigma' \in \mathbb{B}(\sigma, r)} \int_0^1 F_X^\leftarrow(t) \sigma'(t) dt$$

and

$$(3.5) \quad (\text{RSRM-Opt}') \quad \min_{z \in Z} \max_{\sigma' \in \mathbb{B}(\sigma, r)} M_{\sigma'}(f(z, \xi(\omega))).$$

Unfortunately both (RSRM') and (RSRM-Opt') are difficult to calculate/solve in general in that the maximization problem w.r.t. σ is infinite dimensional unless $\mathbb{B}(\sigma, r)$

consists of step-like functions with a finite number of breakpoints. This motivates us to approximate the ψ -ball $\mathbb{B}(\sigma, r)$ with a ball of step-like risk spectra. From a practical point of view, a step-like risk spectrum means that the investor applies the same weight to each of the losses between $F_X^\leftarrow(t_i)$ and $F_X^\leftarrow(t)$ for any $t \in [t_i, t_{i+1})$ and increases the weight when the loss goes up to $F_X^\leftarrow(t_{i+1})$.

Let \mathfrak{S}_M denote the set of all nonnegative, nondecreasing, and normalized step-like functions $\sigma_M(\cdot)$ over $[0, 1]$ with breakpoints at points $0 < t_1 < \dots < t_M < 1$. Then any element in \mathfrak{S}_M is a step-like spectrum with at most M breakpoints. For the simplification of notation, let $t_0 = 0$ and $t_{M+1} = 1$. By definition, $\mathfrak{S}_M \subset \mathfrak{S}$. Note that if we impose some additional properties such as Lipschitz continuity on \mathfrak{S} , where \mathfrak{S} is defined in Definition 3.1, then the inclusion no longer holds. This is indeed one of the key technical challenges that we will need to tackle in section 4 when we investigate error bound on the approximation.

DEFINITION 3.2 (projection on space \mathfrak{S}_M). *Let $\sigma \in \mathfrak{S}$. We say $\sigma_M(\cdot)$ is a projection of σ onto the space \mathfrak{S}_M if $\sigma_M(t) := \sigma_i$, for $t \in [t_i, t_{i+1})$, $i = 0, 1, \dots, M$, such that $\sigma_i \in [\sigma(t_i), \sigma(t_{i+1})]$, for $i = 0, \dots, M$, and $\int_0^1 \sigma_M(t) dt = \sum_{i=0}^M \sigma_i(t_{i+1} - t_i) = 1$.*

From Definition 3.2, we can see that $\sigma_M(\cdot)$ is not unique. Note that instead of setting σ_k to the σ value at the left end or right end of the interval $[t_k, t_{k+1})$, we make it flexible between the two extreme values and require the integral of $\sigma_M(t)$ on $[0, 1]$ to be 1.

DEFINITION 3.3 (ψ -ball of step-like risk spectra). *Let $\sigma_M \in \mathfrak{S}_M$. The ψ -ball of step-like risk spectra in the space \mathfrak{S}_M centered at σ_M with radius r is defined as*

$$(3.6) \quad \mathbb{B}_M(\sigma_M, r) := \{\sigma' \in \mathfrak{S}_M : d_\psi(\sigma', \sigma_M) \leq r\}.$$

With $\mathbb{B}_M(\sigma_M, r)$, we propose to solve (RSRM') and (RSRM-Opt') by solving

$$(3.7) \quad (\text{Appr-RSRM}') \quad \rho_{\mathbb{B}_M(\sigma_M, r)}(X) := \sup_{\sigma \in \mathbb{B}_M(\sigma_M, r)} \int_0^1 F_X^\leftarrow(t) \sigma(t) dt$$

and

$$(3.8) \quad (\text{Appr-RSRM-Opt}') \quad \min_{z \in Z} \max_{\sigma \in \mathbb{B}_M(\sigma_M, r)} M_\sigma(f(z, \xi(\omega))),$$

respectively. While the approximation may bring some conveniences in numerical solution of the minimax optimization problem, it also introduces modeling errors. We will investigate significance of the errors in section 4. In the rest of this section, we discuss computational schemes for solving approximated maximization problem (3.7) and approximated minimax optimization problem (3.8).

3.3. Tractable formulation of problem (3.7). We start with problem (3.7) which does not involve decision variables. To this end, we consider the case that X follows a finite discrete distribution with $\mathbb{P}(X = x_k) = p_k$ for $k = 1, \dots, K$, where $x_1 < x_2 < \dots < x_K$. In the case when X is continuously distributed, sample average approximation may be used; see [1] for the case when \mathfrak{A} is a singleton. Consequently we can write (3.7) as

$$(3.9) \quad \rho_{\mathbb{B}_M(\sigma_M, r)}(X) = \sup_{\sigma \in \mathbb{B}_M(\sigma_M, r)} \sum_{k=1}^K x_k \int_{\pi_k}^{\pi_{k+1}} \sigma(t) dt,$$

where $\pi_1 = 0$ and $\pi_k = \sum_{j=1}^{k-1} p_j$ for $k = 2, \dots, K+1$. Let $\sigma, \sigma^0 \in \mathfrak{S}_M$. By definition

$$\begin{aligned}\mathbf{d}_\psi(\sigma, \sigma^0) &= \sup_{g \in \mathcal{G}} \left| \int_0^1 g(t)(\sigma(t) - \sigma^0(t))dt \right| \\ &= \sup_{g \in \mathcal{G}} \left| \sum_{i=0}^M \int_{t_i}^{t_{i+1}} g(t)(\sigma_i - \sigma_i^0)dt \right| = \sup_{g \in \mathcal{G}} \left| \sum_{i=0}^M (\sigma_i - \sigma_i^0) \int_{t_i}^{t_{i+1}} g(t)dt \right|.\end{aligned}$$

Let $y_i = \int_{t_i}^{t_{i+1}} g(t)dt$. If we consider \mathcal{G} which constitutes all measurable functions with $|g(t)| \leq \psi(t)$ for $t \in [0, 1]$, then we have $|y_i| \leq \int_{t_i}^{t_{i+1}} \psi(t)dt$. Consequently,

$$(3.10a) \quad \mathbf{d}_\psi(\sigma, \sigma^0) = \sup_{y_0, \dots, y_M} \sum_{i=0}^M (\sigma_i - \sigma_i^0) y_i$$

$$(3.10b) \quad \text{s.t. } |y_i| \leq \int_{t_i}^{t_{i+1}} \psi(t)dt \quad \text{for } i = 0, \dots, M.$$

Problem (3.10) is a linear program and its Lagrange dual can be written as

$$(3.11a) \quad \inf_{u_i, v_i \geq 0} \sum_{i=0}^M (v_i + u_i) \int_{t_i}^{t_{i+1}} \psi(t)dt$$

$$(3.11b) \quad \text{s.t. } \sigma_i - \sigma_i^0 + v_i - u_i = 0 \text{ for } i = 0, 1, \dots, M,$$

where u_i and v_i are dual variables with $u_i \geq 0$ and $v_i \geq 0$ for $i = 0, \dots, M$. Consequently we can reformulate (3.9) as

$$(3.12a) \quad \sup_{u_i, v_i \geq 0, \sigma_i} \sum_{k=1}^K x_k \sum_{i=0}^M \sigma_i \int_{\pi_k}^{\pi_{k+1}} \mathbf{1}_{[t_i, t_{i+1})}(t)dt$$

$$(3.12b) \quad \text{s.t. } \sigma_i \geq 0 \text{ for } k = 0, 1, \dots, M,$$

$$(3.12c) \quad \sigma_{i+1} \geq \sigma_i \text{ for } i = 0, 1, \dots, M-1,$$

$$(3.12d) \quad \sum_{i=0}^M \sigma_i (t_{i+1} - t_i) = 1,$$

$$(3.12e) \quad \sum_{i=0}^M (v_i + u_i) \int_{t_i}^{t_{i+1}} \psi(t)dt \leq r,$$

$$(3.12f) \quad \sigma_i - \sigma_i^0 + v_i - u_i = 0 \text{ for } i = 0, 1, \dots, M.$$

Problem (3.12) is a linear program. A significant advantage of this formulation is that the problem size is independent of scenarios of X . Note that when $\{t_0, \dots, t_{M+1}\} \subset \{\pi_1, \dots, \pi_{K+1}\}$, $\int_{\pi_k}^{\pi_{k+1}} \mathbf{1}_{[t_i, t_{i+1})}(t)dt = \pi_{k+1} - \pi_k = p_k$. We will exploit this in the numerical tests.

As we commented in Remark 3.1(i), when the set \mathcal{G} constitutes all measurable functions on $[0, 1]$ bounded by 1 the pseudometric reduces to \mathcal{L}^1 -distance in which case $\psi(t) = 1$. Consequently, (3.12e) can be simplified to $\sum_{i=0}^M (v_i + u_i)(t_{i+1} - t_i) \leq r$.

Moreover, when the set \mathcal{G} consists of all Lipschitz continuous functions with modulus bounded by 1, the formulations will be more sophisticated because we will have to incorporate derivative information about $g \in \mathcal{G}$.

3.4. Alternating iterative algorithm for solving (3.8). We now move on to discuss problem (3.8) and assume that ξ is discretely distributed with K scenarios. In the case when ξ is continuously distributed, sample average approximation may be used; see Remark 5.1. By exploiting the dual formulation (3.12) for the inner maximization problem, we can reformulate (3.8) as

$$(3.13a) \quad \min_{z \in Z} \sup_{u_i, v_i \geq 0, \sigma_i} \sum_{k=1}^K f(z, \xi_k) \sum_{i=0}^M \sigma_i \int_{\pi_k}^{\pi_{k+1}} \mathbf{1}_{[t_i, t_{i+1})}(t) dt$$

$$(3.13b) \quad \text{s.t. } \sigma_i \geq 0 \text{ for } i = 0, 1, \dots, M,$$

$$(3.13c) \quad \sigma_{i+1} \geq \sigma_i \text{ for } i = 0, 1, \dots, M-1,$$

$$(3.13d) \quad \sum_{i=0}^M \sigma_i (t_{i+1} - t_i) = 1,$$

$$(3.13e) \quad \sum_{i=0}^M (v_i + u_i) \int_{t_i}^{t_{i+1}} \psi(t) dt \leq r,$$

$$(3.13f) \quad \sigma_i - \sigma_i^0 + v_i - u_i = 0 \text{ for } i = 0, 1, \dots, M,$$

where $\pi_1 = 0$ and $\pi_k = \sum_{j=1}^{k-1} p_j$ for $k = 2, \dots, K+1$ assuming that $f(z, \xi)$ follows a discrete distribution with $p_k = \mathbb{P}(f(z, \xi(\omega)) = f(z, \xi_k))$ for $k = 1, \dots, K$ where $f(z, \xi_1) < f(z, \xi_2) < \dots < f(z, \xi_K)$ ³. To see why strict ordering is needed, consider the case that $f(z, \xi_1) = f(z, \xi_2) < \dots < f(z, \xi_K)$; in this case, we may have $\pi_2 = p_1, \pi_3 = p_1 + p_2$ or alternatively $\pi_2 = p_2, \pi_3 = p_1 + p_2$, and consequently, the formulation (3.13) is not unique. In general, if the strict ordering is not possible (two or more values coincide), we will have to perturb the point z so that a strict order can be found; we will come back to this shortly after the next algorithm. There is one exception, though. If ξ is uniformly distributed over ξ_1, \dots, ξ_K , i.e., $P(\xi = \xi_k) = \frac{1}{K}$ for $k = 1, \dots, K$, then strict ordering is not needed. Consider the case $f(z, \xi_1) \leq f(z, \xi_2) \leq \dots \leq f(z, \xi_K)$, we can still obtain the same formulation as (3.13) with $\pi_k = \frac{k-1}{K}$ for $k = 2, \dots, K+1$.

Problem (3.13) is not a linear program because the objective function is bilinear even when f is a linear function of z . To see this, we note that these terms $\pi_2, \pi_3, \dots, \pi_K$ depend on the order of $f(z, \xi_k)$, $k = 1, \dots, K$, and this order can be sorted out only when z is fixed up. On the other hand, when $f(z, \xi)$ is convex in z , the objective function of problem (3.13) is convex in z and the inner maximization problem is a linear program for each fixed z . This motivates us to consider the following alternating iterative algorithm for solving this saddle point problem.

ALGORITHM 3.1. *Choose some initial $z^0 \in Z$. For $s = 1, \dots$, do*

³Strictly speaking, we should present the inequalities as $f(z, \xi_{i_1}) < f(z, \xi_{i_2}) < \dots < f(z, \xi_{i_K})$ where $\{i_1, \dots, i_K\}$ is a permutation of $\{1, \dots, K\}$. This is because $f(z, \xi_k)$ depends on z and the order of the numbers can be determined only after $f(z, \xi_k)$, $k = 1, \dots, K$, are evaluated.

Step 1. Solve

$$(3.14a) \quad \sup_{u_i, v_i \geq 0, \sigma_i} \sum_{k=1}^K f(z^{s-1}, \xi_k) \sum_{i=0}^M \sigma_i \int_{\pi_k}^{\pi_{k+1}} \mathbf{1}_{[t_i, t_{i+1})}(t) dt$$

$$(3.14b) \quad \text{s.t. } \sigma_i \geq 0 \text{ for } i = 0, 1, \dots, M,$$

$$(3.14c) \quad \sigma_{i+1} \geq \sigma_i \text{ for } i = 0, 1, \dots, M-1,$$

$$(3.14d) \quad \sum_{i=0}^M \sigma_i (t_{i+1} - t_i) = 1,$$

$$(3.14e) \quad \sum_{i=0}^M (u_i + v_i) \int_{t_i}^{t_{i+1}} \psi(t) dt \leq r,$$

$$(3.14f) \quad \sigma_i - \sigma_i^0 + v_i - u_i = 0 \text{ for } i = 0, 1, \dots, M,$$

to obtain σ^s from the optimal solution and then find z^s with

$$(3.15) \quad z^s \in \arg \min_{z \in Z} v(z, \sigma^s)$$

provided that $f(z, \xi_1) < f(z, \xi_2) < \dots < f(z, \xi_K)$, where

$$(3.16) \quad v(z, \sigma) := \sum_{k=1}^K f(z, \xi_k) \sum_{i=1}^M \sigma_i \int_{\pi_k}^{\pi_{k+1}} \mathbf{1}_{[t_i, t_{i+1})}(t) dt.$$

Step 2. Stop when $z^{s+1} = z^s$ and $\sigma^s = \sigma^{s-1}$.

The maximization problem (3.14) is a linear program. The minimization problem (3.15) is a convex program when f is convex in z but it is not necessarily a linear program even when f is linear in z . However, it cannot be plugged into an existing solver for convex optimization directly in that the quantities $\pi_2, \pi_3, \dots, \pi_K$ implicitly depend on the ordering of quantities $f(z, \xi_k)$, $k = 1, \dots, K$. To get around the issue, we suggest to use the classical cutting plane method [21] which only requires calculating the function value and a subgradient at each specified point $z \in Z$. Let $h(z) = v(z, \sigma^s)$ and assume without loss of generality that z_{s-1} as a starting point for the cutting plane method. We consider two cases: (a) there is a strict order of the quantities $f(z, \xi_k)$, $k = 1, \dots, K$, and (b) at least two of the quantities coincide. In case (a), h is continuously differentiable at z_{s-1} because a small perturbation does not affect the order and consequently we can calculate the gradient ∇h at the point and then construct a cutting plane. In case (b), we will have to perturb z a bit so that a strict order can be found at the perturbed point, and then use the function value and the gradient at the perturbed point to construct a cutting plane. For practical applications, ξ_k , $k = 1, \dots, K$, are obtained from empirical data so we assume that $\mathbb{P}(\xi = \xi_k) = \frac{1}{K}$ for $k = 1, \dots, K$. Consequently h is continuously differentiable at z_{s-1} even in case (b). This is what we do in the numerical tests to be reported in section 6.

Note that Guo and Xu [13] propose a similar alternating iterative scheme for solving a class of utility PRO problem. Differing from [13], the size of both inner and outer problems here is independent of sample size K . The following proposition states convergence of the algorithm.

PROPOSITION 3.1. *Suppose that for each fixed ξ , f is Lipschitz continuous in $z \in Z$ with modulus $\kappa(\xi)$. Algorithm 3.1 either terminates in a finite number of steps*

with a solution to the problem (3.13) or generates a sequence $\{(z^s, \sigma^s)\}$ whose cluster points are optimal solutions to the problem (3.13).

Proof. Let (z^*, σ^*) be a cluster point generated by Algorithm 3.1. Then

$$(3.17) \quad v(z^*, \sigma) \leq v(z^*, \sigma^*) \leq v(z, \sigma^*),$$

where $v(z, \sigma)$ is defined in (3.16). (3.17) means that (z^*, σ^*) is a saddle point of $v(z, \sigma)$ and hence an optimal solution of the problem (3.13). For $s = 1, 2, \dots$,

$$v(z^{s-1}, \sigma^s) \geq v(z^{s-1}, \sigma) \quad \forall \sigma \in \mathfrak{S}_M \text{ and } v(z^s, \sigma^s) \leq v(z, \sigma^s) \quad \forall z \in Z.$$

In the case when the algorithm terminates in finite steps, $z^{s+1} = z^s$ and $\sigma^{s+1} = \sigma^s$ for some s and consequently (z^s, σ^s) satisfies (3.17). In what follows, we consider the case that the algorithm generates an infinite sequence $\{(z^s, \sigma^s)\}$. Let $(\hat{z}, \hat{\sigma})$ be a cluster point of $\{(z^s, \sigma^s)\}$ and for simplicity of notation assume that $(z^s, \sigma^s) \rightarrow (\hat{z}, \hat{\sigma})$. Assume for the sake of a contradiction that $(\hat{z}, \hat{\sigma})$ is not a solution of the problem (3.13). Then it violates one of the inequalities in (3.17). Consider the case that there exists z_0 such that $v(z_0, \hat{\sigma}) > v(\hat{z}, \hat{\sigma})$. Then for s sufficiently large, $v(z_0, \sigma^s) > v(z^s, \sigma^s)$. But this contradicts $v(z^s, \sigma^s) \leq v(z, \sigma^s) \forall z \in Z$. In the same manner, we can show that $(\hat{z}, \hat{\sigma})$ satisfies the first inequality in (3.17). \square

4. Error bound. In the previous section, we discussed computational schemes for solving approximated maximization problem (3.7) and approximated minimax optimization problem (3.8). In this section, we investigate errors arising from the approximations and its propagation to the optimal value.

4.1. Error bound on a single step-like approximation. We start by quantifying the difference between risk spectrum $\sigma \in \mathfrak{S}$ and its step-like approximation σ_M on \mathfrak{S}_M . For this purpose, we make the following assumption on the step-like approximation scheme.

Assumption 4.1. Let $C : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a continuous function such that $C(\beta) \rightarrow 0$ as $\beta \downarrow 0$. For each $\sigma \in \mathfrak{S}$, there exists $\sigma_M \in \mathfrak{S}_M$ such that $\mathfrak{d}_\psi(\sigma, \sigma_M) \leq C(\beta_M)$, where $\beta_M := \max_{i=0,1,\dots,M} (t_{i+1} - t_i)$.

To see how the assumption could be possibly satisfied, we consider the case when \mathfrak{S} is restricted to the set of all Lipschitz continuous function defined over $[0, 1]$ with modulus being bounded by L . The following proposition addresses this.

PROPOSITION 4.1. *Let $\sigma \in \mathfrak{S}$ be Lipschitz continuous with modulus being bounded by L and let σ_M be a projection of σ on \mathfrak{S}_M . Then $\sup_{t \in [0, 1]} |\sigma_M(t) - \sigma(t)| \leq L\beta_M$, where β_M is defined as in Assumption 4.1. Moreover, $\mathfrak{d}_\psi(\sigma, \sigma_M) \leq L\beta_M \int_0^1 \psi(t) dt$.*

Proof. By the Lipschitz continuity and nondecreasing property of σ over $[0, 1]$, we have for any $t \in [t_i, t_{i+1}]$, $|\sigma_M(t) - \sigma(t)| = |\sigma_i - \sigma(t)| = \max \{\sigma_{i+1} - \sigma(t), \sigma(t) - \sigma_i\} \leq L(t_{i+1} - t_i) \leq L\beta_M$, which yields the first inequality. Moreover, by definition

$$\mathfrak{d}_\psi(\sigma, \sigma_M) = \sup_{g \in \mathcal{G}} \int_0^1 g(t)(\sigma(t) - \sigma_M(t)) dt \leq L\beta_M \sup_{g \in \mathcal{G}} \int_0^1 g(t) dt = L\beta_M \int_0^1 \psi(t) dt.$$

The proof is complete. \square

The error bounds in Proposition 4.1 depend heavily on the assumption that σ is Lipschitz continuous, which excludes many unbounded risk spectra such as Wang's risk spectrum defined as in Example 2.1(ii). To address this issue, notice first that

since σ is a nonnegative, nondecreasing function with the normalized property, the only unbounded point of an unbounded risk spectrum is $t = 1$. Thus, we may consider a class of functions $\sigma \in \mathfrak{S}$ such that

$$(4.1) \quad \int_{t_{M-1}}^1 |\sigma(t) - \sigma_{M-1}| dt \leq O(1 - t_{M-1}),$$

where $O : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a function independent of σ such that $O(t) \rightarrow 0$ as $t \downarrow 0$. Moreover, we require the functions to be Lipschitz continuous over interval $[0, 1-t_{M-1}]$ with modulus bounded by $\tilde{L}(t_{M-1})$. Following an analysis similar to the proof of Proposition 4.1, we can establish

$$(4.2) \quad d_\psi(\sigma, \sigma_M) \leq \tilde{L}(t_{M-1})\beta_{M-1} \int_0^1 \psi(t) dt + O(1 - t_{M-1}),$$

where $\beta_{M-1} = \max_{i=1,\dots,M-1} (t_i - t_{i-1})$. Let $C(\beta_M) := \tilde{L}(t_{M-1})\beta_{M-1} \int_0^1 \psi(t) dt + O(1 - t_{M-1})$. Then we might not have $C(\beta_M) \rightarrow 0$ as $\beta_M \rightarrow 0$ because $\tilde{L}(t_{M-1})$ depends on t_{M-1} and it goes to infinity as $t_{M-1} \uparrow 1$. To tackle this issue, we may fix up t_{M-1} in advance, which gives us an error bound as specified in (4.1) as well as $\tilde{L}(t_{M-1})$. For the fixed t_{M-1} , we may reduce the error in the first term of $C(\beta_M)$ by increasing the number of grid points for step-like approximation over subinterval $[0, t_{M-1}]$ and subsequently reducing β_{M-1} . The following example explains how this works for both bounded and unbounded risk spectra.

Example 4.1. We reconsider Example 3.1 to verify Assumption 4.1.

(i) For CVaR decision makers, let $\mathfrak{S} = \{\sigma_\alpha : \alpha \in [\alpha_1, \alpha_2] \cap (0, 1)\}$ where the DM's confidence level is restricted to range $[\alpha_1, \alpha_2]$. For any $\sigma_\alpha \in \mathfrak{S}$, we consider its step-like approximation σ_M . Note that we do not know precisely where α is located but are sure that there is i such that $\alpha \in [t_{i-1}, t_i]$. Let $\sigma_M(t) = \sigma_\alpha(t)$ on $[0, 1] \setminus [t_{i-1}, t_i]$ and $\sigma_M(t) = (t_i - \alpha)/[(1 - \alpha)(t_i - t_{i-1})]$ on $[t_{i-1}, t_i]$. Then

$$\begin{aligned} d_\psi(\sigma_\alpha, \sigma_M) &= \sup_{g \in \mathcal{G}} \int_0^1 g(t)[\sigma_\alpha(t) - \sigma_M(t)] dt = \sup_{g \in \mathcal{G}} \int_{t_{i-1}}^{t_i} g(t)[\sigma_\alpha(t) - \sigma_M(t)] dt \\ &= \sup_{g \in \mathcal{G}} \int_{t_{i-1}}^\alpha g(t)[\sigma_\alpha(t) - \sigma_M(t)] dt + \sup_{g \in \mathcal{G}} \int_\alpha^{t_i} g(t)[\sigma_\alpha(t) - \sigma_M(t)] dt \\ &= \sup_{g \in \mathcal{G}} \int_{t_{i-1}}^\alpha g(t) \left(0 - \frac{t_i - \alpha}{(1 - \alpha)(t_i - t_{i-1})} \right) dt \\ &\quad + \sup_{g \in \mathcal{G}} \int_\alpha^{t_i} g(t) \left(\frac{1}{1 - \alpha} - \frac{t_i - \alpha}{(1 - \alpha)(t_i - t_{i-1})} \right) dt \\ &= 2 \frac{(t_i - \alpha)(\alpha - t_{i-1})}{(1 - \alpha)(t_i - t_{i-1})} \leq \frac{1}{2(1 - \alpha)}(t_i - t_{i-1}) \leq \frac{1}{2(1 - \alpha)}\beta_M, \end{aligned}$$

where the first inequality is due to $xy \leq (\frac{x+y}{2})^2$ for any $x, y \in \mathbb{R}$. In this case, the error bound $C(\beta_M)$ is linear in β_M .

(ii) For Wang's decision makers, let \mathfrak{S} denote the class of all unbounded risk spectra of form $\sigma_r(t) = r(1 - t)^{r-1}$, where $r \in [r_1, r_2] \cap (0, 1)$. Then for any $\sigma_r \in \mathfrak{S}$, we may set its step-like approximation σ_M at point t_M satisfying $\int_{t_{M-1}}^1 \sigma_r(t) dt = \sigma_M(t_M)(1 - t_{M-1})$. Consequently we have

$$\begin{aligned}
d_\psi(\sigma_r, \sigma_M) &= \sup_{g \in \mathcal{G}} \int_0^{t_{M-1}} g(t)[\sigma_r(t) - \sigma_M(t)]dt + \sup_{g \in \mathcal{G}} \int_{t_{M-1}}^1 g(t)[\sigma_r(t) - \sigma_M(t)]dt \\
&\leq \tilde{L}(t_{M-1})\beta_{M-1} + \sup_{g \in \mathcal{G}} \int_{t_{M-1}}^1 g(t)[\sigma_r(t) - \sigma_M(t)]dt \\
&= \tilde{L}(t_{M-1})\beta_{M-1} + \int_{t_{M-1}}^{\hat{t}} (\sigma_M(t) - \sigma_r(t))dt + \int_{\hat{t}}^1 (\sigma_r(t) - \sigma_M(t))dt \\
&= \tilde{L}(t_{M-1})\beta_{M-1} + \sigma_r(\hat{t})(2\hat{t} - t_{M-1} - 1) - 2(1 - \hat{t})^r + (1 - t_{M-1})^r \\
&\leq \tilde{L}(t_{M-1})\beta_{M-1} + \sigma_r(\hat{t})(1 - t_{M-1}) + (1 - t_{M-1})^r,
\end{aligned}$$

where $\tilde{L}(t_{M-1}) = r_1(1 - r_1)(1 - t_{M-1})^{r_1-2}$ and \hat{t} is chosen such that $\sigma_r(\hat{t}) = \frac{1}{1-t_{M-1}} \int_{t_{M-1}}^1 \sigma_r(t)dt$. In this case, the error bound $C(\beta_M)$ is presented in the form (4.2) because of the unboundedness of σ_r .

(iii) For Gini's decision makers, let $\mathfrak{S} = \{\sigma_s : s \in [s_1, s_2] \cap (0, 1)\}$. Then for any $\sigma_s \in \mathfrak{S}$, by Proposition 4.1, its step-like approximation with M breakpoints satisfies Assumption 4.1 with $C(\beta_M) = 2s_2\beta_M$ because all $\sigma_s \in \mathfrak{S}$ are Lipschitz continuous with modulus bounded by $2s_2$.

From the example, we can see that if a nominal risk spectrum is found to be in a particular class of risk spectra and we construct the ambiguity set within this class, then Assumption 4.1 will be satisfied. For more general cases, we can also use the above approach to verify Assumption 4.1 because any risk spectrum can be viewed as a convex combination of the three cases essentially.

4.2. Error bound on approximation of the ψ -balls. We now proceed to derive the difference between two ψ -balls in space \mathfrak{S}_M and in space \mathfrak{S} . Let U, V be two sets of risk spectra, $d_\psi(u, V) := \inf_{v \in V} d_\psi(u, v)$ be the distance between u and V , $\mathbb{D}_\psi(U, V) := \sup_{u \in U} d_\psi(u, V)$ be the deviation distance of U from V , and $\mathbb{H}_\psi(U, V) := \max\{\mathbb{D}_\psi(U, V), \mathbb{D}_\psi(V, U)\}$ be the Hausdorff distance between the two sets.

LEMMA 4.1. *Let $\sigma_M \in \mathfrak{S}_M$, $\sigma \in \mathfrak{S}$, and δ, r be any positive numbers. Then the following hold:*

- (i) $\mathbb{D}_\psi(\mathbb{B}_M(\sigma_M, r + \delta), \mathbb{B}_M(\sigma_M, r)) \leq \delta$ and $\mathbb{D}_\psi(\mathbb{B}(\sigma, r + \delta), \mathbb{B}(\sigma, r)) \leq \delta$.
- (ii) If σ_M is a projection of σ on \mathfrak{S}_M , then under Assumption 4.1, $\mathbb{H}_\psi(\mathbb{B}(\sigma_M, r), \mathbb{B}(\sigma, r)) \leq C(\beta_M)$.
- (iii) Under Assumption 4.1, $\mathbb{D}_\psi(\mathbb{B}(\sigma_M, r + \delta), \mathbb{B}(\sigma_M, r)) \leq \delta + 2C(\beta_M)$.

Part (i) says that when two balls are constructed in the same space and with the same center each in the space of the respective ball, their difference under metric d_ψ is bounded by the difference of their radii. Part (ii) derives an error bound for the difference of two balls in the space of \mathfrak{S} with identical radius but different centers where σ_M is a projection of σ on \mathfrak{S}_M . Part (iii) says that when the two balls are in the same space \mathfrak{S} and have the same center σ_M , their difference is bounded by the sum of the difference of their radii and twice the error bound on a single step-like approximation. The second term in the error bound appears because σ_M is not necessarily in \mathfrak{S} if we impose some additional properties such as Lipschitz continuity on \mathfrak{S} .

Proof. Part (i). We only prove the first inequality as the second one can be proved analogously. Let $h_M \in \mathbb{B}_M(\sigma_M, r + \delta) \setminus \mathbb{B}_M(\sigma_M, r)$ and $h_M^\lambda := \lambda h_M + (1 - \lambda)\sigma_M \in \mathfrak{S}_M$, where $\lambda := \frac{r}{d_\psi(h_M, \sigma_M)} \in (0, 1)$. By the definition of d_ψ (see (3.1)),

we have $d_\psi(h_M^\lambda, \sigma_M) := \sup_{g \in \mathcal{G}} \langle g, h_M^\lambda - \sigma_M \rangle = \lambda d_\psi(h_M, \sigma_M) = r$, which implies $h_M^\lambda \in \mathbb{B}_M(\sigma_M, r)$. Thus

$$\begin{aligned} d_\psi(h_M, \mathbb{B}_M(\sigma_M, r)) &\leq d_\psi(h_M, h_M^\lambda) = (1 - \lambda)d_\psi(h_M, \sigma_M) \\ &= d_\psi(\sigma_M, h_M) - r \leq r + \delta - r = \delta. \end{aligned}$$

Since $d_\psi(h_M, \mathbb{B}(\sigma_M, r)) = 0$ for $h_M \in \mathbb{B}_M(\sigma_M, r)$, we obtain $d_\psi(h_M, \mathbb{B}(\sigma_M, r)) \leq \delta \forall h_M \in \mathbb{B}_M(\sigma_M, r + \delta)$. By the definition of \mathbb{D}_ψ , we have

$$\mathbb{D}_\psi(\mathbb{B}_M(\sigma_M, r + \delta), \mathbb{B}_M(\sigma_M, r)) = \sup_{h_M \in \mathbb{B}_M(\sigma_M, r + \delta)} d_\psi(h_M, \mathbb{B}_M(\sigma_M, r)) \leq \delta.$$

The proof is complete.

Part (ii). Let $\sigma' \in \mathbb{B}(\sigma_M, r)$. Then under Assumption 4.1

$$d_\psi(\sigma', \sigma) \leq d_\psi(\sigma', \sigma_M) + d_\psi(\sigma_M, \sigma) \leq r + C(\beta_M),$$

which implies $\mathbb{B}(\sigma_M, r) \subset \mathbb{B}(\sigma, r + C(\beta_M))$. By Part (i),

$$\mathbb{D}_\psi(\mathbb{B}(\sigma_M, r), \mathbb{B}(\sigma, r)) \leq \mathbb{D}_\psi(\mathbb{B}(\sigma, r + C(\beta_M)), \mathbb{B}(\sigma, r)) \leq C(\beta_M).$$

Similarly, we have $\mathbb{D}_\psi(\mathbb{B}(\sigma, r), \mathbb{B}(\sigma_M, r)) \leq C(\beta_M)$. Finally, by the definition of Hausdorff distance under ψ -metric, the proof is complete.

Part (iii). Since $\sigma_M \in \mathfrak{S}_M$, then we can find a $\sigma \in \mathfrak{S}$ such that σ_M is the projection of σ . So for any $u \in \mathbb{B}(\sigma_M, r + \delta)$, we have $d_\psi(u, \sigma) \leq d_\psi(u, \sigma_M) + d_\psi(\sigma_M, \sigma) \leq r + \delta + d_\psi(\sigma_M, \sigma)$, subsequently, $\mathbb{B}(\sigma_M, r + \delta) \subset \mathbb{B}(\sigma, r + \delta + d_\psi(\sigma, \sigma_M))$. On the other hand, for any $u \in \mathbb{B}(\sigma, r - d_\psi(\sigma, \sigma_M))$, we have

$$d_\psi(u, \sigma_M) \leq d_\psi(u, \sigma) + d_\psi(\sigma, \sigma_M) \leq r - d_\psi(\sigma, \sigma_M) + d_\psi(\sigma, \sigma_M) = r$$

and hence $\mathbb{B}(\sigma_M, r) \supset \mathbb{B}(\sigma, r - d_\psi(\sigma, \sigma_M))$. Therefore, according to the definition of \mathbb{D}_ψ and Part (i),

$$\begin{aligned} \mathbb{D}_\psi(\mathbb{B}(\sigma_M, r + \delta), \mathbb{B}(\sigma_M, r)) &\leq \mathbb{D}_\psi(\mathbb{B}(\sigma, r + \delta + d_\psi(\sigma, \sigma_M)), \mathbb{B}(\sigma, r - d_\psi(\sigma, \sigma_M))) \\ &\leq \delta + 2d_\psi(\sigma, \sigma_M) \leq \delta + 2C(\beta_M). \end{aligned}$$

The proof is complete. \square

We are now ready to quantify the difference between $\mathbb{B}(\sigma, r)$ and $\mathbb{B}_M(\sigma_M, r)$ under the ψ -metric.

THEOREM 4.1. *Let $\sigma \in \mathfrak{S}$ and σ_M be a projection of σ on \mathfrak{S}_M . Under Assumption 4.1,*

$$\mathbb{H}_\psi(\mathbb{B}(\sigma, r), \mathbb{B}_M(\sigma_M, r)) \leq 5C(\beta_M).$$

In the case when $\sigma = \sigma_M$, the error bound is $4C(\beta_M)$.

We can use the error bound to identify how small β_M should be for a specified precision ϵ by setting $\beta_M := \inf\{\beta : 5C(\beta) \leq \epsilon\}$. In the case when the breakpoints t_0, \dots, t_M are evenly spread over $[0, 1]$, the formula above enables us to estimate the minimal number of points needed to achieve the specified precision. Note also that in practice, one may construct a step-like nominal risk spectrum σ_M directly using empirical data [9] and σ_M approximates the true as M increases. In that case σ should be understood as the true risk spectrum rather than a nominal one.

Proof. By the triangle inequality property of the Hausdorff distance in the space of \mathfrak{S} , we have

$$\mathbb{H}_\psi(\mathbb{B}(\sigma, r), \mathbb{B}_M(\sigma_M, r)) \leq \mathbb{H}_\psi(\mathbb{B}(\sigma, r), \mathbb{B}(\sigma_M, r)) + \mathbb{H}_\psi(\mathbb{B}(\sigma_M, r), \mathbb{B}_M(\sigma_M, r)).$$

From Lemma 4.1(ii), $\mathbb{H}_\psi(\mathbb{B}(\sigma, r), \mathbb{B}(\sigma_M, r)) \leq C(\beta_M)$. So it suffices to show $\mathbb{H}_\psi(\mathbb{B}(\sigma_M, r), \mathbb{B}_M(\sigma_M, r)) \leq 4C(\beta_M)$. By the definition of \mathbb{D}_ψ ,

$$\begin{aligned} \mathbb{D}_\psi(\mathbb{B}(\sigma_M, r), \mathbb{B}_M(\sigma_M, r)) &= \sup_{u \in \mathbb{B}(\sigma_M, r)} \mathbb{d}_\psi(u, \mathbb{B}_M(\sigma_M, r)) \\ &\leq \sup_{u \in \mathbb{B}(\sigma_M, r)} [\mathbb{d}_\psi(u_M, \mathbb{B}_M(\sigma_M, r)) + \mathbb{d}_\psi(u, u_M)] \\ &\leq \sup_{u \in \mathbb{B}(\sigma_M, r)} \mathbb{d}_\psi(u_M, \mathbb{B}_M(\sigma_M, r)) + C(\beta_M) \\ &\leq \mathbb{D}_\psi(\mathbb{B}_M(\sigma_M, C(\beta_M) + r), \mathbb{B}_M(\sigma_M, r)) + C(\beta_M) \leq 2C(\beta_M), \end{aligned}$$

where u_M is a projection of u . The second inequality follows from Assumption 4.1, and the third inequality is from the fact that for any $u \in \mathbb{B}(\sigma_M, r)$, we can find its projection u_M such that $\mathbb{d}_\psi(u_M, \sigma_M) \leq \mathbb{d}_\psi(u_M, u) + \mathbb{d}_\psi(u, \sigma_M) \leq C(\beta_M) + r$, that is, $u_M \in \mathbb{B}_M(\sigma_M, C(\beta_M) + r)$. The last inequality follows from Lemma 4.1(i). Likewise, we have

$$\begin{aligned} \mathbb{D}_\psi(\mathbb{B}_M(\sigma_M, r), \mathbb{B}(\sigma_M, r)) &= \sup_{u_M \in \mathbb{B}_M(\sigma_M, r)} \mathbb{d}_\psi(u_M, \mathbb{B}(\sigma_M, r)) \\ &\leq \sup_{u_M \in \mathbb{B}_M(\sigma_M, r)} [\mathbb{d}_\psi(u, \mathbb{B}(\sigma_M, r)) + \mathbb{d}_\psi(u, u_M)] \\ &\leq \sup_{u_M \in \mathbb{B}_M(\sigma_M, r)} \mathbb{d}_\psi(u, \mathbb{B}(\sigma_M, r)) + C(\beta_M) \\ &\leq \mathbb{D}_\psi(\mathbb{B}(\sigma_M, r + C(\beta_M)), \mathbb{B}(\sigma_M, r)) + C(\beta_M) \leq 4C(\beta_M), \end{aligned}$$

where the third inequality is derived from the fact that for any $u_M \in \mathbb{B}_M(\sigma_M, r)$, that is, $\mathbb{d}_\psi(u_M, \sigma_M) \leq r$, we have $\mathbb{d}_\psi(u, \sigma_M) \leq \mathbb{d}_\psi(u, u_M) + \mathbb{d}_\psi(u_M, \sigma_M) \leq \mathbb{d}_\psi(u, u_M) + r \leq C(\beta_M) + r$, that is, $u \in \mathbb{B}(\sigma_M, r + C(\beta_M))$. The last inequality follows from Lemma 4.1(iii). Finally, by the definition of Hausdorff distance under metric \mathbb{d}_ψ , the proof is complete. \square

4.3. Error bound on the optimal value. We now move on to investigate the difference between the optimal values of (RSRM-Opt') and (Appr-RSRM-Opt'). Let

$$\vartheta^* := \min_{z \in Z} \max_{\sigma \in \mathbb{B}(\sigma^0, r)} M_\sigma(f(z, \xi(\omega))), \text{ and } \vartheta_M^* := \min_{z \in Z} \max_{\sigma \in \mathbb{B}_M(\sigma_M^0, r)} M_\sigma(f(z, \xi(\omega))).$$

Since the two optimization problems above are identical except the feasible sets of the inner maximization problems, we can use standard stability analysis in parametric programming to derive an error bound for the difference of their optimal values based on the difference of $\mathbb{B}(\sigma^0, r)$ and $\mathbb{B}_M(\sigma_M^0, r)$. The question is how to choose a proper pseudometric \mathbb{d}_ψ which facilitates our analysis. For this purpose, we choose a particular \mathcal{G} associated with f and make the following technical assumption.

Assumption 4.2. Let $\mathcal{G}_f := \{F_{f(z, \xi)}^\leftarrow(\cdot) : z \in Z\}$. There exist a constant $p \in [1, \infty)$ and a positive function $\psi_f \in \mathcal{L}^p[0, 1]$ such that the following growth condition in the space of quantile functions holds:

$$(4.3) \quad |F_{f(z, \xi)}^\leftarrow(t)| \leq \psi_f(t) \quad \forall t \in [0, 1].$$

To see how ψ_f could be possibly identified, we may identify functions $a(\xi)$ and $b(\xi)$ such that $a(\xi) \leq f(z, \xi) \leq b(\xi) \forall \xi \in \mathbb{R}^k, z \in Z$. Let $F_{a(\xi)}^\leftarrow$ and $F_{b(\xi)}^\leftarrow$ denote the quantile functions of $a(\xi)$ and $b(\xi)$ and $\psi_f(t) := \max(F_{a(\xi)}^\leftarrow(t), F_{b(\xi)}^\leftarrow(t)) \forall t \in [0, 1]$. Then ψ_f satisfies (4.3).

THEOREM 4.2 (error bound on the optimal value). *Let Assumption 4.2 hold. Let $\sigma^0 \in \mathfrak{S}$ and σ_M^0 be a projection of σ^0 on the space of \mathfrak{S}_M . Assume Assumption 4.1 holds with $\mathcal{G}_f \subset \mathcal{G}$ and $\psi = \psi_f$. Then*

$$|\vartheta^* - \vartheta_M^*| \leq 5C(\beta_M).$$

Proof. Let $v(z) = \max_{\sigma \in \mathbb{B}(\sigma^0, r)} M_\sigma(f(z, \xi(\omega)))$ and $v_M(z) = \max_{\sigma \in \mathbb{B}_M(\sigma_M^0, r)} M_\sigma(f(z, \xi(\omega)))$. Then

$$\begin{aligned} (4.4) \quad |v(z) - v_M(z)| &= \left| \max_{\sigma \in \mathbb{B}(\sigma^0, r)} M_\sigma(f(z, \xi(\omega))) - \max_{\tilde{\sigma} \in \mathbb{B}_M(\sigma_M^0, r)} M_{\tilde{\sigma}}(f(z, \xi(\omega))) \right| \\ &= \left| \max_{\sigma \in \mathbb{B}(\sigma^0, r)} \min_{\tilde{\sigma} \in \mathbb{B}_M(\sigma_M^0, r)} M_\sigma(f(z, \xi(\omega))) - M_{\tilde{\sigma}}(f(z, \xi(\omega))) \right| \\ &\leq \max_{\sigma \in \mathbb{B}(\sigma^0, r)} \min_{\tilde{\sigma} \in \mathbb{B}_M(\sigma_M^0, r)} |M_\sigma(f(z, \xi(\omega))) - M_{\tilde{\sigma}}(f(z, \xi(\omega)))| \\ &= \max_{\sigma \in \mathbb{B}(\sigma^0, r)} \min_{\tilde{\sigma} \in \mathbb{B}_M(\sigma_M^0, r)} \left| \int_0^1 F_{f(z, \xi(\omega))}^\rightarrow(t)(\sigma(t) - \tilde{\sigma}(t)) dt \right| \\ &\leq \max_{\sigma \in \mathbb{B}(\sigma^0, r)} \min_{\tilde{\sigma} \in \mathbb{B}_M(\sigma_M^0, r)} \int_0^1 |\sigma(t) - \tilde{\sigma}(t)| \psi_f(t) dt \\ &= \mathbb{D}_{\psi_f}(\mathbb{B}(\sigma^0, r), \mathbb{B}_M(\sigma_M^0, r)). \end{aligned}$$

By exchanging the positions of $\mathbb{B}_M(\sigma_M^0, r)$ and $\mathbb{B}(\sigma^0, r)$, we have $|v_M(z) - v(z)| \leq \mathbb{D}_{\psi_f}(\mathbb{B}_M(\sigma_M^0, r), \mathbb{B}(\sigma^0, r))$. Summarizing the discussions above, we have

$$(4.5) \quad |\vartheta^* - \vartheta_M^*| \leq \max_{z \in Z} |v(z) - v_M(z)| \leq \mathbb{H}_{\psi_f}(\mathbb{B}_M(\sigma_M^0, r), \mathbb{B}(\sigma^0, r)),$$

where \mathbb{H} denotes the Hausdorff distance under metric \mathbb{d}_{ψ_f} . The rest follows from Theorem 4.1. \square

It is important to note that the optimal values $\vartheta^*, \vartheta_M^*$ are independent on \mathbb{d}_{ψ_f} but the error bound is dependent on \mathbb{d}_{ψ_f} . From the proof of the theorem, we can see that the error bound (4.5) is indeed tight, because equality might hold in the first inequality of formulae (4.4) under some special circumstances, and likewise, equality might hold in the second inequality when $F_{f(z, \xi(\omega))}^\rightarrow(t)$ and $\sigma(t) - \tilde{\sigma}(t)$ are co-monotonic.

5. Stability analysis and statistical robustness. In this section, we move on to investigate the stability of our robust spectral risk optimization model (RSRM-Opt). We proceed with the investigation in two steps. First, we discuss how a small perturbation of the probability distribution P of ξ affects the optimal value. This is in line with classical stability analysis in stochastic programming. A particularly interesting case is that the true probability distribution is approximated by a discrete distribution such as sample average approximation, and the resulting stability result will serve as theoretical grounding for the numerical schemes developed in section 3 to be extended from discrete distribution to continuous distribution of P . Second, we consider the case that P has to be recovered by empirical data and the data may contain some noise. We investigate the quality of the resulting optimal value of (RSRM-Opt) based on the contaminated data.

5.1. Quantitative stability analysis. Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ be a financial loss and $P_X = \mathbb{P} \circ X^{-1}$ be a push-forward probability measure on \mathbb{R} induced by X . By definition, $F_X(x) = P_X((-\infty, x])$. Let $T_{F_X}^\leftarrow(U) := \inf\{t \in \mathbb{R} : F_X(t) \geq U\}$, where U is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ that is uniformly distributed over $(0, 1)$. By [3, Lemma 2.15], $T_{F_X}^\leftarrow(U)$ is a random variable with distribution F_X and it is known as the quantile transformation. Thus $F_X^\leftarrow(t) = T_{F_X}^\leftarrow(t) \forall t \in [0, 1]$. This means we can write the SRM $M_\sigma(X)$ as $M_\sigma(X) = \int_0^1 T_{F_X}^\leftarrow(t) \sigma(t) dt := \rho_\sigma(F_X)$. From the discussions above, we can see that the SRM may be perceived not only as a classical risk measure defined on the prospect space of losses but also as a risk measure defined on the probability space $\mathcal{P}(\mathbb{R})$ or on the space of quantile functions.

Let us now consider the case that $X = f(z, \xi)$ where f is defined as in (2.7). Let $\mu_z := P \circ f^{-1}(z, \cdot)$ be a probability measure on \mathbb{R} induced by f , where $P = \mathbb{P} \circ \xi^{-1}$ is a probability measure on \mathbb{R}^k induced by ξ . Let $F_{\mu_z}(x) := \mu_z((-\infty, x])$ for any $x \in \mathbb{R}$. Then

$$(5.1) \quad M_\sigma(f(z, \xi)) = \int_0^1 T_{\mu_z}^\leftarrow(t) \sigma(t) dt =: \rho_\sigma(\mu_z),$$

where we write $T_{\mu_z}^\leftarrow(t)$ for $T_{F_{\mu_z}}^\leftarrow(t)$. Let $\mathcal{R}_{\mathfrak{A}}(\mu_z) := \sup_{\sigma \in \mathfrak{A}} \rho_\sigma(\mu_z)$. Then problem (RSRM-Opt) can be equivalently written as

$$(5.2) \quad \vartheta(P) := \min_{z \in Z} \mathcal{R}_{\mathfrak{A}}(\mu_z).$$

Here by writing the optimal value $\vartheta(P)$ as a function of P , we indicate explicitly its dependence on P . Next, we investigate continuity of ϑ in P . To this end, we need some intermediate technical results.

DEFINITION 5.1 (Wasserstein distance/metric). *For probability measures P and \tilde{P} , the Wasserstein distance/metric of order $r \geq 1$ is $d_r(P, \tilde{P}) = (\inf_{\pi} \iint d(\xi, \tilde{\xi})^r \pi(d\xi, d\tilde{\xi}))^{\frac{1}{r}}$, where π is among all probability measures over $\mathbb{R}^k \times \mathbb{R}^k$ with marginals P and \tilde{P} , i.e.,*

$$P(A) = \pi(A \times \mathbb{R}^k) \quad \forall A \in \mathcal{B}(\mathbb{R}^k) \text{ and } \tilde{P}(B) = \pi(\mathbb{R}^k \times B) \quad \forall B \in \mathcal{B}(\mathbb{R}^k).$$

It is well known that convergence under the Wasserstein metric implies weak convergence; see, e.g., [24]. In the case when $r = 1$, it reduces to the Kantorovich metric.

LEMMA 5.1 (see [24, Theorem 2.15]). *The Wasserstein distance of order $r \geq 1$ for measures P and \tilde{P} on the real line \mathbb{R} is*

$$(5.3) \quad d_r(P, \tilde{P}) = \left(\int_0^1 |F_P^\leftarrow(t) - F_{\tilde{P}}^\leftarrow(t)|^r dt \right)^{\frac{1}{r}},$$

where $F_P(y) := P((-\infty, y])$ is the associated CDF and $F_P^\leftarrow(t) := \inf_y \{F_P(y) \geq t\}$ its generalized inverse.

Let $\phi : \mathbb{R}^k \rightarrow [0, \infty)$ be a continuous function and $\mathcal{M}_k^\phi := \{P' \in \mathcal{P}(\mathbb{R}^k) : \int_{\mathbb{R}^k} \phi(t) P'(dt) < \infty\}$. In the particular case when $\phi := \|\cdot\|^p$, where $\|\cdot\|$ denotes the Euclidean norm in \mathbb{R}^k and p is a positive number, we write \mathcal{M}_k^p for $\mathcal{M}_k^{\|\cdot\|^p}$.

DEFINITION 5.2 (ϕ -weak topology). *Let $\phi : \mathbb{R}^k \rightarrow [0, \infty)$ be a gauge function, that is, $\phi \geq 1$ holds outside a compact set. Define \mathcal{C}_k^ϕ to be the linear space of all*

continuous functions $h : \mathbb{R}^k \rightarrow \mathbb{R}$ for which there exists a positive constant c such that $|h(t)| \leq c(\phi(t) + 1) \forall t \in \mathbb{R}^k$. The ϕ -weak topology, denoted by τ_ϕ , is the coarsest topology on \mathcal{M}_k^ϕ for which the mapping $g_h : \mathcal{M}_k^\phi \rightarrow \mathbb{R}$ defined by $g_h(P') := \int_{\mathbb{R}^k} h(t) P'(dt)$, $h \in \mathcal{C}_k^\phi$ is continuous. A sequence $\{P_l\} \subset \mathcal{M}_k^\phi$ is said to converge ϕ -weakly to $P \in \mathcal{M}_k^\phi$ written $P_l \xrightarrow{\phi} P$ if it converges w.r.t. τ_ϕ .

We need the following assumption, which is known as the growth condition in the literature; see [3]. In order to derive our qualitative and quantitative stability results of $\vartheta(\cdot)$ near P , we require the following technical assumption.

Assumption 5.1. There exist $\kappa, \nu \in \mathbb{R}_+$ such that $|f(z, \xi)| \leq \kappa(\|\xi\|^\nu + 1) \forall (z, \xi) \in Z \times \mathbb{R}^k$.

PROPOSITION 5.1. Assume (a) Assumption 5.1 holds, (b) there is a positive number $q > 1$ such that $\mathfrak{A} \subset \mathscr{L}^q[0, 1]$, (c) there is a positive constant $p \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$ such that $P \in \mathcal{M}_k^{\nu p}$, (d) \mathfrak{A} is a compact set in $\mathscr{L}^q[0, 1]$. Then, $\mathcal{R}_{\mathfrak{A}} : \mathcal{M}_1^p \rightarrow \mathbb{R}$ is well defined and is Lipschitz continuous with modulus being bounded by $\sup_{\sigma \in \mathfrak{A}} [\int_0^1 \sigma(t)^q dt]^{\frac{1}{q}}$.

Proof. Under conditions (a) and (c), it follows by [3, Corollary 2.11] that $\mu_z = P \circ f^{-1}(z, \cdot) \in \mathcal{M}_1^p$, i.e., $\int_{\mathbb{R}} |x|^p \mu_z(dx) < \infty \forall P \in \mathcal{M}_k^{\nu p}$. Since $\int_0^1 |T_{\mu_z}^\leftarrow(t)|^p dt = \int_{\mathbb{R}} |x|^p \mu_z(dx) < \infty$, it implies the quantile transformation $T_{\mu_z}^\leftarrow(\cdot) \in \mathscr{L}^p$. The latter ensures, under condition (b), that $\rho_\sigma(\mu_z) := \int_0^1 T_{\mu_z}^\leftarrow(t) \sigma(t) dt < \infty$ for every $\sigma \in \mathfrak{A}$ and consequently $\mathcal{R}_{\mathfrak{A}}(\mu_z) < \infty$ under condition (d). Next, we show the Lipschitz continuity of $\mathcal{R}_{\mathfrak{A}}$. Let $\mu', \mu \in \mathcal{M}_1^p$. By Hölder's inequality and Lemma 5.1,

$$\begin{aligned} |\mathcal{R}_{\mathfrak{A}}(\mu') - \mathcal{R}_{\mathfrak{A}}(\mu)| &= \left| \sup_{\sigma \in \mathfrak{A}} \int_0^1 T_{\mu'}^\leftarrow(t) \sigma(t) dt - \sup_{\sigma \in \mathfrak{A}} \int_0^1 T_\mu^\leftarrow(t) \sigma(t) dt \right| \\ &\leq \sup_{\sigma \in \mathfrak{A}} \left| \int_0^1 (T_{\mu'}^\leftarrow(t) - T_\mu^\leftarrow(t)) \sigma(t) dt \right| \\ &\leq \sup_{\sigma \in \mathfrak{A}} \left\{ \left[\int_0^1 |T_{\mu'}^\leftarrow(t) - T_\mu^\leftarrow(t)|^p dt \right]^{\frac{1}{p}} \left[\int_0^1 \sigma(t)^q dt \right]^{\frac{1}{q}} \right\} \\ &= d_p(\mu', \mu) \sup_{\sigma \in \mathfrak{A}} \left[\int_0^1 \sigma(t)^q dt \right]^{\frac{1}{q}}, \end{aligned}$$

where we use (5.3) to establish the last equality. The conclusion follows. \square

The result is a generalization of Pichler [25, Colloary 13] for SRM. In what follows, we investigate continuity of $\vartheta(\cdot)$ near P .

THEOREM 5.1. Let the assumptions of Proposition 5.1 be satisfied. Then

$$(5.4) \quad \lim_{\substack{P' \xrightarrow{\|\cdot\|^\nu p} P}} \vartheta(P') = \vartheta(P).$$

Moreover, if f is uniformly Hölder continuous in ξ , that is, $|f(z, \xi') - f(z, \xi)| \leq \kappa \|\xi' - \xi\|^\nu \forall \xi', \xi \in \mathbb{R}^k, z \in Z$, where κ and ν are defined as in Assumption 5.1, then

$$(5.5) \quad |\vartheta(P') - \vartheta(P)| \leq \kappa d_{p\nu}(P', P)^\nu \sup_{\sigma \in \mathfrak{A}} \left[\int_0^1 \sigma(t)^q dt \right]^{\frac{1}{q}},$$

where $d_{p\nu}$ is the Wasserstein metric with order $p\nu$.

Before providing a proof of the theorem, we make some comments on the business insights that the technical results may provide. Recall that by Definition 2.1, the exponent q is the tail risk index of a decision maker, i.e., the larger q is, the more risk averse the decision maker will be on the tail loss. On the other hand, for positive numbers $p_1 < p_2$, $\mathcal{M}_k^{\nu p_2} \subset \mathcal{M}_k^{\nu p_1}$. From (5.4) and (5.5), we can see that when q increases, p is forced to decrease, which means that $\mathcal{M}_k^{\nu p}$ becomes larger in order to maintain the required stability. In other words, when the decision maker is more risk averse at tail losses, the optimal value ϑ is more likely to be stable against perturbation of probability P .

Proof. Part (i). Observe first that $\mathscr{P}(\mathbb{R}^k)$ is separable (see [20]). Thus it suffices to show that

$$(5.6) \quad \lim_{\substack{\| \cdot \| \nu p \\ P_l \xrightarrow{\| \cdot \| \nu p} P}} \vartheta(P_l) = \vartheta(P)$$

for any sequence $\{P_l\} \subset \mathcal{M}_k^{\nu p}$ converging to P under topology $\tau_{\| \cdot \| \nu p}$. Let $\mu_z^l = P_l \circ f(z, \cdot)^{-1}$. Then from Proposition 5.1, $\vartheta(\mu_z^l)$ is well defined. The estimation

$$\begin{aligned} |\vartheta(P_l) - \vartheta(P)| &\leq \sup_{z \in Z} |\mathcal{R}_{\mathfrak{A}}(\mu_z^l) - \mathcal{R}_{\mathfrak{A}}(\mu_z)| \\ &\leq \sup_{z \in Z} \left[\int_0^1 |T_{\mu_z^l}^{\leftarrow}(t) - T_{\mu_z}^{\leftarrow}(t)|^p dt \right]^{\frac{1}{p}} \sup_{\sigma \in \mathfrak{A}} \left[\int_0^1 \sigma(t)^q dt \right]^{\frac{1}{q}} \\ (5.7) \quad &= \sup_{z \in Z} d_p(\mu_z^l, \mu_z) \sup_{\sigma \in \mathfrak{A}} \left[\int_0^1 \sigma(t)^q dt \right]^{\frac{1}{q}} \end{aligned}$$

holds, where $p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} = 1$. The first inequality follows from standard stability analysis in parametric programming and the second inequality follows from (5.1), (5.2), and the Hölder inequality; the last equality follows from Lemma 5.1. It remains to show that $\sup_{z \in Z} d_p(\mu_l, \mu) \rightarrow 0$, but this is guaranteed by our setup that Z is a compact set and [3, Lemma 2.74] which states that $\mu = P \circ f^{-1}$ is continuous w.r.t. (P, z) .

Part (ii). The result is essentially covered by [27, Theorem 6]; here we include a simple proof to be self-contained. By transformation of variables, we have

$$d_p(\mu_z^l, \mu_z) = \left(\inf_{\tilde{\pi}} \int \int_{\mathbb{R}^k \times \mathbb{R}^k} |f(z, \xi') - f(z, \xi)|^p \tilde{\pi}(d\xi, \xi') \right)^{\frac{1}{p}},$$

where $\tilde{\pi}$ is a joint probability distribution over $\mathbb{R}^k \times \mathbb{R}^k$ with

$$\tilde{\pi}(A, \mathbb{R}^k) = P_l(A) \quad \forall A \in \mathscr{B}(\mathbb{R}^k) \quad \text{and} \quad \tilde{\pi}(\mathbb{R}^k, B) = P(B) \quad \forall B \in \mathscr{B}(\mathbb{R}^k).$$

By exploiting the Hölder continuity of f in ξ , we have

$$d_p(\mu_z^l, \mu_z) \leq \kappa \left(\inf_{\tilde{\pi}} \int \int_{\mathbb{R}^k \times \mathbb{R}^k} \|\xi' - \xi\|^{p\nu} \tilde{\pi}(d\xi, \xi') \right)^{\frac{1}{p}} = \kappa d_{p\nu}(P_l, P)^{\nu}.$$

The proof is complete. \square

Remark 5.1. In a particular case that $P' = P_N$ where P_N is defined in (5.8), $d_{p\nu}(P_N, P)$ converges to 0 at an exponential rate against an increase of sample size N ; see [11, Theorem 2] for details. This means $\vartheta(P_N)$ converges to $\vartheta(P)$ at an exponential rate.

5.2. Statistical robustness. We now move on to the second step. Let $\mathcal{P}(\mathbb{R}^k)$ denote the set of all probability measures in \mathbb{R}^k and $P, Q \in \mathcal{P}(\mathbb{R}^k)$. Let ξ^1, \dots, ξ^N and $\tilde{\xi}^1, \dots, \tilde{\xi}^N$ be independent and identically distributed (iid) samples generated by ξ with distributions P and Q respectively.⁴ Let

$$P_N(\cdot) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\xi^i}(\cdot) \quad \text{and} \quad Q_N(\cdot) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\tilde{\xi}^i}(\cdot)$$

be the respective empirical distributions, where $\mathbf{1}_\xi(\cdot)$ denotes the Dirac measure at ξ . We refer to P as the true probability distribution and Q its perturbation. In other words, we regard ξ^1, \dots, ξ^N as samples generated by random variable ξ when it follows distribution P although we do not know what it is.

Let $\vartheta(P_N)$ and $\vartheta(Q_N)$ denote the optimal values of problem (5.2) with $P = P_N$ and $P = Q_N$, respectively. Based on the discussions above, $\vartheta(Q_N)$ is obtained from solving (RSRM-Opt) with real data, whereas $\vartheta(P_N)$ is the optimal value based on ideal uncontaminated data. In practice, we can only obtain $\vartheta(Q_N)$. However, we hope that $\vartheta(Q_N)$ is close to $\vartheta(P_N)$ so that the former is a useful estimate of the true robust optimal value $\vartheta(P)$ when the sample size is sufficiently large.

In what follows, we investigate the discrepancy between $\vartheta(P_N)$ and $\vartheta(Q_N)$. Following the convention in the literature of statistical robustness [22], we do so by examining the discrepancy between the push-forward distributions induced by the two statistical estimators, i.e., $P^{\otimes N} \circ \vartheta(P_N)^{-1}$ and $Q^{\otimes N} \circ \vartheta(Q_N)^{-1}$ in the probability space $\mathcal{P}(\mathbb{R})$ under some metrics. Here we use the Kantorovich metric instead of the Prokhorov metric or Levy metric as in [22, 4], and we will show that this helps us to derive a quantitative result which is a significant step forward from a qualitative result. To this end, we need to introduce some intermediate technical notions and results.

Let $\mathcal{F}_p(\mathbb{R}^k) := \{\psi : \mathbb{R}^k \rightarrow \mathbb{R} : |\psi(\xi) - \psi(\tilde{\xi})| \leq c_p(\xi, \tilde{\xi})\|\xi - \tilde{\xi}\| \forall \xi, \tilde{\xi} \in \mathbb{R}^k\}$, where $\|\cdot\|$ denotes some norm on \mathbb{R}^k and $c_p(\xi, \tilde{\xi}) := \max\{1, \|\xi\|, \|\tilde{\xi}\|\}^{p-1} \forall \xi, \tilde{\xi} \in \mathbb{R}^k$ and exponent $p \geq 1$ characterizes the growth of the local Lipschitz constant. Recall that the p th order Fortet–Mourier metric for any probability measures $P, Q \in \mathcal{P}(\mathbb{R}^k)$ is defined by

$$\zeta_p(P, Q) := \sup_{\psi \in \mathcal{F}_p(\mathbb{R}^k)} \left| \int_{\mathbb{R}^k} \psi(\xi) P(d\xi) - \int_{\mathbb{R}^k} \psi(\xi) Q(d\xi) \right|;$$

see [29]. In the case when $p = 1$, the class of functions in $\mathcal{F}_p(\mathbb{R}^k)$ are globally Lipschitz continuous and consequently $\zeta_p(P, Q)$ reduces to the Kantorovich metric, to be denoted by $\mathfrak{d}_{K,k}(P, Q)$, where the first subscript indicates Kontorovich and the second subscript indicates the dimension of \mathbb{R}^k .

LEMMA 5.2. *Let $\xi := (\xi^1, \dots, \xi^N) \in (\mathbb{R}^k)^{\otimes N}$ and*

$$\Psi := \left\{ \psi : (\mathbb{R}^k)^{\otimes N} \rightarrow \mathbb{R} : |\psi(\tilde{\xi}) - \psi(\hat{\xi})| \leq \frac{1}{N} \sum_{k=1}^N c_p(\tilde{\xi}^k, \hat{\xi}^k) \|\tilde{\xi}^k - \hat{\xi}^k\| \forall \tilde{\xi}, \hat{\xi} \in (\mathbb{R}^k)^{\otimes N} \right\}.$$

Let $\mathfrak{d}_\Psi(P^{\otimes N}, Q^{\otimes N}) := \sup_{\psi \in \Psi} \left| \int_{(\mathbb{R}^k)^{\otimes N}} \psi(\tilde{\xi}) P^{\otimes N}(d\tilde{\xi}) - \int_{(\mathbb{R}^k)^{\otimes N}} \psi(\hat{\xi}) Q^{\otimes N}(d\hat{\xi}) \right|$. Then $\mathfrak{d}_\Psi(P^{\otimes N}, Q^{\otimes N}) \leq \zeta_p(P, Q)$.

⁴Here we use the terminology probability measure and probability distribution interchangeably. Likewise, we are slightly abusing the notation ξ^i for samples and random variables. The correct reading should be samples generated by independent random variables ξ^1, \dots, ξ^N whose distribution is identical to that of ξ .

Proof. Let $\xi^{-j} := \{\xi^1, \dots, \xi^{j-1}, \xi^{j+1}, \dots, \xi^N\}$, $\xi^j := \{\xi^1, \dots, \xi^j\}$ and $\xi^{-j} := \{\xi^{j+1}, \dots, \xi^N\}$. For any $P_1, \dots, P_N \in \mathcal{P}(\mathbb{R}^k)$ and any $j \in \{1, \dots, N\}$, denote

$$P_{-j}(d\xi^{-j}) := P_1(d\xi^1) \cdots P_{j-1}(d\xi^{j-1}) P_{j+1}(d\xi^{j+1}) \cdots P_N(d\xi^N)$$

and $h_{\xi^{-j}}(\xi^j) := \int_{(\mathbb{R}^k)^{\otimes(N-1)}} \psi(\xi^{-j}, \xi^j) P_{-j}(d\xi^{-j})$. Then

$$\begin{aligned} |h_{\xi^{-j}}(\tilde{\xi}^j) - h_{\xi^{-j}}(\hat{\xi}^j)| &\leq \int_{(\mathbb{R}^k)^{\otimes(N-1)}} |\psi(\xi^{-j}, \tilde{\xi}^j) - \psi(\xi^{-j}, \hat{\xi}^j)| P_{-j}(d\xi^{-j}) \\ &\leq \int_{(\mathbb{R}^k)^{\otimes(N-1)}} \frac{1}{N} c_p(\tilde{\xi}^j, \hat{\xi}^j) \|\tilde{\xi}^j - \hat{\xi}^j\| P_{-j}(d\xi^{-j}) \\ &\leq \frac{1}{N} c_p(\tilde{\xi}^j, \hat{\xi}^j) \|\tilde{\xi}^j - \hat{\xi}^j\|. \end{aligned}$$

Let \mathcal{H} denote the set of functions $h_{\xi^{-j}}(\xi^j)$ generated by $\psi \in \Psi$. By the definition of d_Ψ and the p th order Fortet–Mourier metric,

$$\begin{aligned} d_\Psi(P_{-j} \times \tilde{P}_j, P_{-j} \times \hat{P}_j) &= \sup_{\psi \in \Psi} \left| \int_{\mathbb{R}^k} \int_{(\mathbb{R}^k)^{\otimes(N-1)}} \psi(\xi^{-j}, \xi^j) P_{-j}(d\xi^{-j}) \tilde{P}_j(d\xi^j) \right. \\ &\quad \left. - \int_{\mathbb{R}^k} \int_{(\mathbb{R}^k)^{\otimes(N-1)}} \psi(\xi^{-j}, \xi^j) P_{-j}(d\xi^{-j}) \hat{P}_j(d\xi^j) \right| \\ &= \sup_{h_{\xi^{-j}} \in \mathcal{H}} \left| \int_{\mathbb{R}^k} h_{\xi^{-j}}(\xi^j) \tilde{P}_j(d\xi^j) \right. \\ &\quad \left. - \int_{\mathbb{R}^k} h_{\xi^{-j}}(\xi^j) \hat{P}_j(d\xi^j) \right| \leq \frac{1}{N} \zeta_p(\tilde{P}_j, \hat{P}_j), \end{aligned}$$

where the inequality is due to $N h_{\xi^{-j}}(\xi^j) \in \mathcal{F}_p(\mathbb{R}^k)$ and the definition of $\zeta_p(P, Q)$. Finally, by the triangle inequality of the pseudometric, we have

$$\begin{aligned} d_\Psi(P^{\otimes N}, Q^{\otimes N}) &\leq d_\Psi(P^{\otimes N}, P^{\otimes(N-1)} \times Q) + d_\Psi(P^{\otimes(N-1)} \times Q, P^{\otimes(N-2)} \times Q^{\otimes 2}) \\ &\quad + \cdots + d_\Psi(P \times Q^{\otimes(N-1)}, Q^{\otimes N}) \leq \frac{1}{N} \zeta_p(P, Q) \times N = \zeta_p(P, Q). \end{aligned}$$

The proof is complete. \square

With the lemma, we are able to derive the desired quantitative statistical robustness for ϑ_N .

THEOREM 5.2 (quantitative statistical robustness). *Let the assumptions of Proposition 5.1 be satisfied. If in addition f is a locally Lipschitz continuous function that increases with p in ξ , i.e.,*

$$|f(z, \xi) - f(z, \xi')| \leq c_p(\xi, \xi') \|\xi - \xi'\| \quad \forall \xi, \xi' \in \mathbb{R}^k, z \in Z,$$

where $c_p(\xi, \xi') := (\max\{1, \|\xi\|, \|\xi'\|\})^{p-1}$. Then for any $N \in \mathbb{N}$ and any $P, Q \in \mathcal{M}_k^{\nu p}$

$$(5.8) \quad d_{K,1}(P^{\otimes N} \circ \vartheta(P_N)^{-1}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}) \leq L_N \zeta_p(P, Q),$$

where $L_N := N \sup_{\sigma \in \mathfrak{A}} \int_{\frac{N-1}{N}}^1 \sigma(t) dt$.

Theorem 5.2 says that if Q is perturbed from P within set $\mathcal{M}_k^{\nu p}$, then $\vartheta(Q_N)$ is close to $\vartheta(P_N)$ (in the sense of their respective push-forward distributions under the Kantorovich metric) $\forall Q$ close to P (under the Fortet-Mourier metric), and (5.8) gives a bound on the propagation of the error to the optimal values. In the case when all risk spectra in \mathfrak{A} are uniformly bounded by a positive constant C , we have $L_N \leq C$, which means the error bound holds uniformly $\forall N$. We refer interested readers to our follow-up work [32] for general risk measures.

In practice, since P is unknown, it is difficult to identify the distance between P and Q . The usefulness of the theorem may therefore be understood as that it provides some theoretical guidance: if the perceived empirical data is generated by some probability distribution Q (this may be predicted using standard statistical methods such as histograms) which is close to P (this is the true probability distribution which the decision maker has in mind or under conjecture) and Q satisfies the moment condition $\int_{t \in \mathbb{R}^k} |t|^{\nu p} Q(dt) < \infty \forall Q \in \mathcal{M}_k^{\nu p}$ (which can be verified through empirical data), then we may believe that the optimal value computed with perceived data is close to the one with ideal data (without noise).

Since $\mathcal{M}^{\nu p}$ is affected by p as we remarked right after Theorem 5.1, a smaller p leads to a larger $\mathcal{M}^{\nu p}$. So we can roughly say that if a decision maker puts more weights on the tail losses (q is larger), then $\vartheta(Q_N)$ is more likely to be statistically robust.

Proof. Let $\xi := (\xi^1, \dots, \xi^N) \in (\mathbb{R}^k)^{\otimes N}$. Notice that both $\vartheta(P_N)$ and $\vartheta(Q_N)$ are essentially functions of iid samples ξ and $\tilde{\xi}$ that P_N and Q_N are based on, respectively. Let \mathcal{G} be the set of all Lipschitz continuous functions on \mathbb{R} with modulus bounded by 1. By definition

$$\begin{aligned} (5.9) \quad & \mathbb{d}_{K,1}(P^{\otimes N} \circ \vartheta(P_N)^{-1}, Q^{\otimes N} \circ \vartheta(Q_N)^{-1}) \\ &= \sup_{g \in \mathcal{G}} \left| \int_{\mathbb{R}} g(t) P^{\otimes N} \circ \vartheta(P_N^{-1})(dt) - \int_{\mathbb{R}} g(t) Q^{\otimes N} \circ \vartheta(Q_N^{-1})(dt) \right| \\ &= \sup_{g \in \mathcal{G}} \left| \int_{(\mathbb{R}^k)^{\otimes N}} g(\vartheta(\xi)) P^{\otimes N}(d\xi) - \int_{(\mathbb{R}^k)^{\otimes N}} g(\vartheta(\tilde{\xi})) Q^{\otimes N}(d\tilde{\xi}) \right|. \end{aligned}$$

Next, we use Lemma 5.2 to estimate the right-hand side of (5.9). Let $\mu_z^N := P_N \circ f^{-1}(z, \cdot)$ and $\tilde{\mu}_z^N := Q_N \circ f^{-1}(z, \cdot)$. By Proposition 5.1, $\vartheta(\mu_z^N)$ is well defined. Moreover, by (5.2)

$$\begin{aligned} \left| g(\vartheta(\xi)) - g(\vartheta(\tilde{\xi})) \right| &\leq \left| \vartheta(\xi) - \vartheta(\tilde{\xi}) \right| = \left| \min_{z \in Z} \sup_{\sigma \in \mathfrak{A}} \rho_{\sigma}(\mu_z^N) - \min_{z \in Z} \sup_{\sigma \in \mathfrak{A}} \rho_{\sigma}(\tilde{\mu}_z^N) \right| \\ &\leq \sup_{z \in Z} \sup_{\sigma \in \mathfrak{A}} \left| \int_0^1 T_{\mu_z^N}^{\leftarrow}(t) \sigma(t) dt - \int_0^1 T_{\tilde{\mu}_z^N}^{\leftarrow}(t) \sigma(t) dt \right| \\ &\leq \sup_{z \in Z} \sup_{\sigma \in \mathfrak{A}} \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \sigma(t) \left| T_{\mu_z^N}^{\leftarrow}(t) - T_{\tilde{\mu}_z^N}^{\leftarrow}(t) \right| dt \\ &\leq \sup_{\sigma \in \mathfrak{A}} \int_{\frac{N-1}{N}}^1 \sigma(t) dt \sup_{z \in Z} \sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| T_{\mu_z^N}^{\leftarrow}(t) - T_{\tilde{\mu}_z^N}^{\leftarrow}(t) \right| dt. \end{aligned}$$

For any fixed $z \in Z$, let $f(z, \xi^{i_1}) \leq \dots \leq f(z, \xi^{i_N})$ and $f(z, \tilde{\xi}^{j_1}) \leq \dots \leq f(z, \tilde{\xi}^{j_N})$ denote the order of sequences $\{f(z, \xi^k)\}_{k=1}^N$ and $\{f(z, \tilde{\xi}^k)\}_{k=1}^N$, respectively. Then

$$\sum_{k=1}^N \int_{\frac{k-1}{N}}^{\frac{k}{N}} \left| T_{\mu_z^N}^{\leftarrow}(t) - T_{\tilde{\mu}_z^N}^{\leftarrow}(t) \right| dt = \sum_{k=1}^N \left| f(z, \xi^{i_k}) - f(z, \tilde{\xi}^{j_k}) \right| \leq \sum_{k=1}^N \left| f(z, \xi^k) - f(z, \tilde{\xi}^k) \right|.$$

The last inequality holds due to the fact that if $\{a_i\}_{i=1}^N$ and $\{b_i\}_{i=1}^N$ are two non-decreasing sequences, then for any permutation $\{k_1, k_2, \dots, k_N\}$ of $\{1, 2, \dots, N\}$, we have $\sum_{i=1}^N |a_i - b_i| \leq \sum_{i=1}^N |a_i - b_{k_i}|$. This result can be easily proved by induction; we omit the details. Therefore, we have

$$\begin{aligned} \left| g(\vartheta(\boldsymbol{\xi})) - g(\vartheta(\tilde{\boldsymbol{\xi}})) \right| &\leq \sup_{\sigma \in \mathfrak{A}} \int_{\frac{N-1}{N}}^1 \sigma(t) dt \sup_{z \in Z} \sum_{k=1}^N \left| f(z, \xi^k) - f(z, \tilde{\xi}^k) \right| \\ &\leq \sup_{\sigma \in \mathfrak{A}} \int_{\frac{N-1}{N}}^1 \sigma(t) dt \sum_{k=1}^N c_p(\xi^k, \tilde{\xi}^k) \|\xi^k - \tilde{\xi}^k\| \\ &= N \sup_{\sigma \in \mathfrak{A}} \int_{\frac{N-1}{N}}^1 \sigma(t) dt \times \frac{1}{N} \sum_{k=1}^N c_p(\xi^k, \tilde{\xi}^k) \|\xi^k - \tilde{\xi}^k\|, \end{aligned}$$

where we exploit the nonincreasing property of σ to establish the second inequality. By setting $L_N := N \sup_{\sigma \in \mathfrak{A}} \int_{\frac{N-1}{N}}^1 \sigma(t) dt$, we arrive at

$$\left| g(\vartheta(\boldsymbol{\xi})) - g(\vartheta(\tilde{\boldsymbol{\xi}})) \right| \leq \frac{L_N}{N} \sum_{k=1}^N c_p(\xi^k, \tilde{\xi}^k) \|\xi^k - \tilde{\xi}^k\| \quad \forall g \in \mathcal{G}, \boldsymbol{\xi}, \tilde{\boldsymbol{\xi}} \in (\mathbb{R}^k)^{\otimes N}.$$

Let $\Psi := \frac{1}{L_N} \{g(\vartheta(\cdot)) : (\mathbb{R}^k)^{\otimes N} \rightarrow \mathbb{R} : g \in \mathcal{G}\}$. By applying Lemma 5.2 to this context, we obtain

$$\sup_{g \in \mathcal{G}} \left| \int_{(\mathbb{R}^k)^{\otimes N}} g(\vartheta(\boldsymbol{\xi})) P^{\otimes N}(d\boldsymbol{\xi}) - \int_{(\mathbb{R}^k)^{\otimes N}} g(\vartheta(\tilde{\boldsymbol{\xi}})) Q^{\otimes N}(d\tilde{\boldsymbol{\xi}}) \right| \leq L_N \zeta_p(P, Q).$$

The proof is complete. \square

6. Numerical tests. We have carried out some numerical tests to validate the theoretical results of this paper in the context of a stylized portfolio selection problem. The subsequent simulation experiments are designed to provide additional insights into the performance of the proposed RSRM based optimization scheme (RSRM-Opt). Specifically, the portfolio selection problem under our optimization scheme is as follows.

Consider a capital market consisting of m assets whose yearly returns are captured by the random vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^T$. If short-selling is forbidden, a portfolio is encoded by a vector of percentage weights $z = (z_1, \dots, z_m)^T$ ranging over the simplex $Z = \{z \in \mathbb{R}_+^m : \sum_{i=1}^m z_i = 1\}$. As portfolio z invests a percentage z_i of the available capital in asset i for each $i = 1, \dots, m$, its return amounts to $z^T \boldsymbol{\xi}$. In the remainder we aim to solve the single-stage stochastic programming problem

$$(6.1) \quad \min_{z \in Z} \max_{\sigma \in \mathbb{B}(\sigma^0, r)} M_\sigma(-z^T \boldsymbol{\xi}),$$

which minimizes the RSRM of the portfolio loss $-z^T \boldsymbol{\xi}$. We use Algorithm 3.1 to solve the problem via the step-like approximation scheme (Appr-RSRM-Opt') in section 3.2.

Our experiments are based on a market with $m = 10$ assets considered by Esfahani and Kuhn in [10]. In view of the capital asset pricing model we may assume that the return ξ_i is decomposable into a systematic risk factor $\varphi \sim \mathcal{N}(0, 2\%)$ common to all assets and an unsystematic or idiosyncratic risk factor $\zeta_i \sim \mathcal{N}(i \times 3\%, i \times 2.5\%)$ specific to asset i . Thus, we set $\xi_i = \varphi + \zeta_i$, where φ and the idiosyncratic risk factors ζ_i , $i = 1, \dots, m$, constitute independent normal random variables. By construction, assets with higher indices promise higher mean returns at a higher risk. We use $\xi := (\xi_1, \dots, \xi_{10})^T$ to generate K iid samples and select the breakpoints t_1, \dots, t_M from $\{\frac{1}{K}, \frac{2}{K}, \dots, \frac{K-1}{K}\}$ so that the integral $\int_{\pi_k}^{\pi_{k+1}} \mathbf{1}_{[t_i, t_{i+1}]}(t) dt$ in (3.14a) is equal to $\pi_{k+1} - \pi_k = p_k$. In the following two experiments, we set $\psi(t) = t$ in the definition of the pseudometric d_ψ and the nominal spectral function $\sigma^0(t) = \frac{1}{2}(1-t)^{-\frac{1}{2}}$. For simplicity, we choose that the projection of σ^0 onto the space \mathfrak{S}_M is

$$(6.2) \quad \sigma_M^0(\tau) := \begin{cases} \sigma^0(t_i) & \text{for } \tau \in [t_i, t_{i+1}), i = 0, 1, \dots, M-1, \\ M+1 - \sum_{i=0}^{M-1} \sigma^0(t_i) & \text{for } \tau \in [t_M, t_{M+1}]. \end{cases}$$

We implement our experiments in MATLAB 2018a installed on a desktop with Windows 10 operating system (i5-6600 CPU, 3.30 GHz, 8 GB memory (RAM)). We use the proposed alternating iterative algorithm, i.e., Algorithm 3.1, to solve (Appr-RSRM-Opt'). Specifically, we use the add-on *Optimization Toolbox* in MATLAB to solve program (3.14) directly and use the classical cutting plane method to solve program (3.15), where for each step, we also use the add-on *Optimization Toolbox* to solve the corresponding program.

6.1. Impact of the radius of ambiguity set. In the first experiment we investigate the impact of the ambiguity set radius r on the optimal robust portfolios and the optimal value of robust portfolios. We solve problem (6.1) via Algorithm 3.1 using training datasets of cardinality $K \in \{30, 300, 3000\}$, the number of breakpoints $M = 5$ with $t_i = \frac{i}{M+1}$ for $i = 0, 1, \dots, M$, and the radius of ambiguity set $r \in [0.01, 100]$. Figures 6.1 and 6.2 visualize the corresponding optimal portfolio weights $z(r)$ and the corresponding optimal value as a function of r , averaged over 100 independent simulation runs separately. Our numerical results show that as the radius of the ambiguity set increases, the optimal portfolio weights shift to the first seven assets and they are pretty evenly spread. Differing from Esfahani and Kuhn in [10], the spread is not even over the 10 assets. We are unable to explain the phenomenon exactly but our conjecture is that it has to do with the difference of spaces that the balls of ambiguity

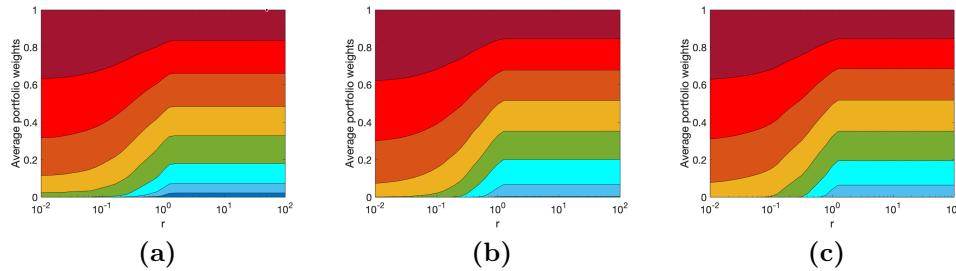


FIG. 6.1. Optimal portfolio composition as a function of the ambiguity set radius r averaged over 100 simulations for $M = 5$ breakpoints; the portfolio weights are depicted in ascending order, i.e., the weight of asset 10 at the top (brick red area) and that of asset 9 second from top (red area), and so on. (a) $K = 30$ training samples. (b) $K = 300$ training samples. (c) $K = 3000$ training samples.

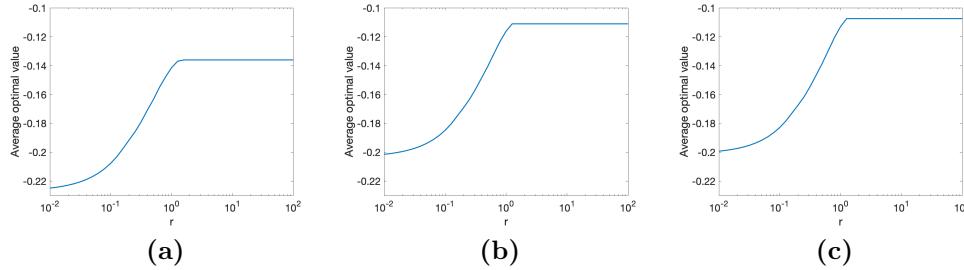


FIG. 6.2. Optimal value of robust portfolios as a function of the ambiguity set radius r averaged over 100 simulations for $M = 5$ breakpoints. (a) $K = 30$ training samples. (b) $K = 300$ training samples. (c) $K = 3000$ training samples.

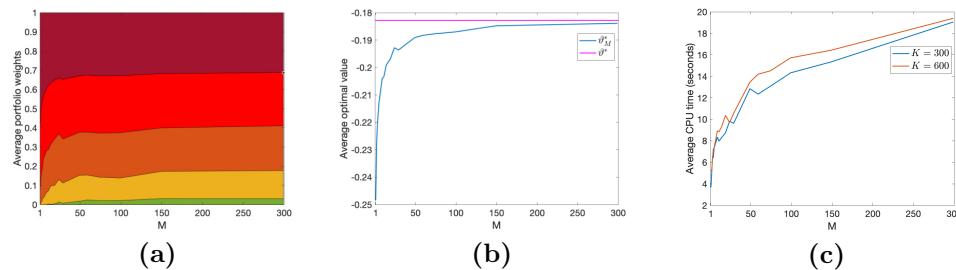


FIG. 6.3. (a) Optimal solution of robust portfolios as M increases averaged over 100 simulations. (b) Optimal value of robust portfolios as M increases averaged over 100 simulations. (c) CPU time as M increases averaged over 100 simulations.

are defined, albeit there is some connection via (2.4). In other words, expansion of the ball of risk spectra does not necessarily match expansion of the Wasserstein ball in the same structure.

6.2. Impact of the number of breakpoints M and CPU time. In the second experiment we investigate the impact of step-like approximation $\mathbb{B}_M(\sigma_M^0, r)$ to $\mathbb{B}(\sigma^0, r)$ on the optimal robust portfolios and the corresponding optimal values. We solve problem (6.1) through Algorithm 3.1 using training datasets of cardinality $K = 300$, radius of ambiguity set $r = 0.01$, and breakpoints $M = 1, 2, 3, 4, 5, 9, 11, 14, 19, 24, 29, 49, 59, 74, 99, 149, 299$ such that $M + 1$ is the divisor of K and then $t_i = \frac{i}{M+1}$ for $i = 1, \dots, M$. It is difficult to calculate the precise optimal value ϑ^* of (RSRM-Opt') since the inner maximization problem is infinite dimensional. In this experiment, we use $M = 10000$ to solve the problem (Appr-RSRM-Opt') and regard its optimal value $\vartheta_{10000}^* = -0.1828$ as the true optimal value ϑ^* of the problem (RSRM-Opt'). Figures 6.3(a) and 6.3(b) visualize the change of the optimal portfolio weights z and the optimal value as the number of breakpoints M increases. The results are averaged over 100 independent simulation runs. From Figure 6.3(b), we can see that the optimal value ϑ_M^* of (Appr-RSRM-Opt') converges to -0.1828 as M increases. This is because the projection of the nominal risk spectrum function, σ_M^0 , converges to σ^0 as M increases and hence the ambiguity set $\mathbb{B}_M(\sigma_M^0, r)$ converges to $\mathbb{B}(\sigma^0, r)$. In this experiment, we also consider the CPU time of Algorithm 3.1 to solve problem (6.1). Figure 6.3(c) visualizes the CPU time as the number of breakpoints M increases, averaged over 100 independent simulation runs, for $K = 300$ and $K = 600$. From Figure 6.3(c), we can roughly state that the average CPU time grows slowly as M increases and the sample size has small impact on the average CPU time.

Acknowledgments. The authors thank the two anonymous referees for their valuable comments, which helped them improve the presentation of the paper. They would also like to thank the associate editor, Professor Andrzej Ruszczyński, for effective handling of the review.

REFERENCES

- [1] C. ACERBI, *Spectral measures of risk: A coherent representation of subjective risk aversion*, J. Banking Finance, 26 (2002), pp. 1505–1518.
- [2] B. ARMBRUSTER AND E. DELAGE, *Decision making under uncertainty when preference information is incomplete*, Management Sci., 61 (2015), pp. 111–128.
- [3] M. CLAUS, *Advancing Stability Analysis of Mean-Risk Stochastic Programs: Bilevel and Two-Stage Models*, Ph.D. thesis, University of Duisburg-Essen, 2016.
- [4] R. CONT, R. DEGUEST, AND G. SCANDOLO, *Robustness and sensitivity analysis of risk measurement procedures*, Quant. Finance, 10 (2010), pp. 593–606.
- [5] E. DELAGE, S. GUO, AND H. XU, *Shortfall Risk Models When Information of Loss Function Is Incomplete*, GERAD HEC Montréal, 2017.
- [6] E. DELAGE AND J. Y.-M. LI, *Minimizing risk exposure when the choice of a risk measure is ambiguous*, Management Sci., 64 (2017), pp. 327–344.
- [7] D. DENNEBERG, *Distorted probabilities and insurance premiums*, Methods Oper. Res., 63 (1990), pp. 3–5.
- [8] D. DENTCHEVA AND A. RUSZCZYŃSKI, *Risk preferences on the space of quantile functions*, Math. Program., 148 (2014), pp. 181–200.
- [9] D. ESCOBAR AND G. CH. PFLUG, *The distortion principle for insurance pricing: Properties, identification and robustness*, Ann. Oper. Res., 292 (2020), pp. 771–794.
- [10] P. M. ESFAHANI AND D. KUHN, *Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations*, Math. Program., 171 (2018), pp. 115–166.
- [11] N. FOURNIER AND A. GUILLIN, *On the rate of convergence in wasserstein distance of the empirical measure*, Probab. Theory Related Fields, 162 (2015), pp. 707–738.
- [12] M. FRITTELLI, M. MAGGIS, AND I. PERI, *Risk measures on and value at risk with probability/loss function*, Math. Finance, 24 (2014), pp. 442–463.
- [13] S. GUO AND H. XU, *Utility preference robust optimization: Piecewise linear approximation, error bounds and stability*, in New Directions in Stochastic Optimisation, J. De Loera, D. Dentcheva, G. Ch. Pflug, and R. Schultz, eds., Mathematisches Forschungsinstitut Oberwolfach, Oberwolfach, Germany, 2018, pp. 2373–2376.
- [14] S. GUO AND H. XU, *Statistical robustness in utility preference robust optimization models*, Math. Program. (2020), <https://doi.org/10.1007/s10107-020-01555-5>.
- [15] F. R. HAMPEL, *A general qualitative definition of robustness*, Ann. Math. Statist., 42 (1971), pp. 1887–1896.
- [16] W. B. HASKELL, W. HUANG, AND H. XU, *Preference Robust Optimization for Lotteries*, submitted, 2020.
- [17] G. HENRYK AND M. SILVIA, *On a relationship between distorted and spectral risk measures*, (2006).
- [18] J. HU AND S. MEHROTRA, *Robust decision making over a set of random targets or risk-averse utilities with an application to portfolio optimization*, IIE Trans., 47 (2015), pp. 358–372.
- [19] J. HU AND G. STEPANYAN, *Optimization with reference-based robust preference constraints*, SIAM J. Optim., 27 (2017), pp. 2230–2257.
- [20] P. J. HUBER AND E. M. RONCHETTI, *Robust Statistics*, Springer, New York, 2011.
- [21] J. E. KELLEY, JR., *The cutting-plane method for solving convex programs*, J. SIAM, 8 (1960), pp. 703–712.
- [22] V. KRÄTSCHMER, A. SCHIED, AND H. ZÄHLE, *Comparative and qualitative robustness for law-invariant risk measures*, Finance Stoch., 18 (2014), pp. 271–295.
- [23] S. KUSUOKA, *On law invariant coherent risk measures*, in Advances in Mathematical Economics, Springer, New York, 2001, pp. 83–95.
- [24] G. CH. PFLUG AND A. PICHLER, *Multistage Stochastic Optimization*, Springer, New York, 2014.
- [25] A. PICHLER, *The natural Banach space for version independent risk measures*, Insurance Math. Econom., 53 (2013), pp. 405–415.
- [26] A. PICHLER AND A. SHAPIRO, *Minimal representation of insurance prices*, Insurance Math. Econom., 62 (2015), pp. 184–193.
- [27] A. PICHLER AND H. XU, *Quantitative stability analysis for minimax distributionally robust risk optimization*, Math. Program. (2018).

- [28] C. PUPPE, *Distorted Probabilities and Choice Under Risk*, Lecture Notes in Econom. and Math. Systems 363, Springer, New York, 2012.
- [29] W. RÖMISCH, *Stability of stochastic programming problems*, in Stochastic Programming, Handbooks Oper. Res. Management Sci., 10, Elsevier, Amsterdam, pp. 483–554.
- [30] A. TSANAKAS AND E. DESLI, *Risk measures and theories of choice*, British Actuarial J., 9 (2003), pp. 959–991.
- [31] S. WANG, *Premium calculation by transforming the layer premium density*, ASTIN Bull., 26 (1996), pp. 71–92.
- [32] W. WANG, H. XU, AND T. MA, *Quantitative Statistical Robustness for Tail-Dependent Law Invariant Risk Measures*, preprint, arXiv:2006.15491, 2020.
- [33] D. WOZABAL, *Robustifying convex risk measures for linear portfolios: A nonparametric approach*, Oper. Res., 62 (2014), pp. 1302–1315.
- [34] G. WU AND R. GONZALEZ, *Curvature of the probability weighting function*, Management Sci., 42 (1996), pp. 1676–1690.
- [35] M. E. YAARI, *The dual theory of choice under risk*, Econometrica, 55 (1987), pp. 95–115.